CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Sidgwick's Dualism of Practical Reason, Evolutionary Debunking, and Moral Psychology

Peter Andes*

University of Alberta
*Corresponding author. Email: andes@ualberta.ca

Abstract

Sidgwick's seminal text *The Methods of Ethics* left off with an unresolved problem that Sidgwick referred to as the dualism of practical reason. The problem is that employing Sidgwick's methodology of rational intuitionism appears to show that there are reasons to favour both egoism and utilitarianism. Katarzyna de Lazari-Radek and Peter Singer offer a solution in the form of an evolutionary debunking argument: the appeal of egoism is explainable in terms of evolutionary theory. I argue that like rational prudence, rational benevolence is subject to debunking arguments and so problematic, but also – and more importantly – that debunking arguments are irrelevant in the debate over the dualism of practical reason on the view of reason and rational intuitionism that Lazari-Radek and Singer embrace. Either both egoism and utilitarianism are debunked, or neither are. If I am right, Sidgwick's dualism is left standing.

Introduction

In *The Point of View of the Universe* Katarzyna de Lazari-Radek and Peter Singer offer an articulation and defence of Henry Sidgwick's ideas about ethics.[1] They seek to demonstrate a number of ways in which his work can inform contemporary debates in the field. Sidgwick's seminal text *The Methods of Ethics* left off with an unresolved problem that Sidgwick referred to as the dualism of practical reason.[2] The problem is that employing Sidgwick's methodology of rational intuitionism appears to show that there are both reasons for egoism and utilitarianism – reasons to adopt both Sidgwick's principle of rational prudence and his principle of rational benevolence – such that reason alone cannot decide what we ought to do when these two principles conflict.[3]

---

[1] Katarzyna de Lazari-Radek and Peter Singer, *The Point of View of the Universe: Sidgwick and Contemporary Ethics* (Oxford, 2014).

[2] H. Sidgwick, *The Methods of Ethics*, 7th edn. (London, 1907).

[3] As was very helpfully pointed out to me by an anonymous reviewer, one might argue that the two principles are not of equal weight, and so that there is not a dualism of practical reason. I simply wish here to consider whether the solution offered to the dualism of practical reason by Lazari-Radek and Singer can succeed, and not whether we might deny the dualism altogether.

Lazari-Radek and Singer offer a solution in the form of an evolutionary debunking argument: the appeal of egoism is explainable in terms of evolutionary theory. We can provide an evolutionary explanation for why it would be beneficial that we care about our own interests, and thus why thoughts like those expressed in the formulation of Sidgwick's principle of rational prudence would appeal to us. If this is the case, then we can dispense with the principle of rational prudence, since we would only find it plausible because of evolutionary pressures that have influenced our judgement. These pressures are not rational, and so should not be considered in attempting to judge rationally to arrive at principles of morality under Sidgwick's method of rational intuitionism. At the same time, an evolutionary debunking argument cannot be offered, Lazari-Radek and Singer allege, that would lead us to dispense with the principle of rational benevolence. If this solution succeeds, it undermines egoism, leaving utilitarianism standing as the correct theory of normative ethics and resolving Sidgwick's dualism of practical reason.

In this article I explore the success of this treatment of Sidgwick's dualism. I begin by outlining Sidgwick's approach before turning to Lazari-Radek and Singer's debunking solution. I then consider whether we can offer similar debunking arguments for rational benevolence. Ultimately, I argue that, like rational prudence, rational benevolence is subject to debunking arguments and so problematic, but also – and more importantly – that debunking arguments are irrelevant in the debate over the dualism of practical reason on the view of reason and rational intuitionism that Lazari-Radek and Singer embrace. I pose a dilemma for their position such that either rational benevolence is exposed to debunking arguments, or it is insulated from them completely, with this latter state of affairs being irreconcilable with their own debunking solution that attempts to eliminate rational prudence. Either both rational prudence and rational benevolence are subject to debunking arguments, or neither are. If I am right, this dilemma presents a serious obstacle for Lazari-Radek and Singer's solution to the dualism of practical reason.

Along the way I bring recent developments in moral psychology to bear on the issue of evolutionary debunking and the principles Sidgwick sees as generated by rational intuitionism. I offer Jonathan Haidt's account of our moral psychology as able to provide a general debunking explanation of rational benevolence, and Joshua Greene's research as an empirical perspective that might be friendlier to Lazari-Radek and Singer's aims.[4] I consider whether Greene's research could be used by Lazari-Radek and Singer to defend their approach. I also explore whether the empirical conclusions of Haidt and Greene generate problematic implications for rational intuitionism, and how the methodology itself might be considered so as to be subject to evolutionary debunking. I conclude with an examination of Lazari-Radek and Singer's conception of reason as a unity, and how this plays a role in their approach to debunking arguments and rational intuitionism, before offering the dilemma that I think stands in the way of their approach.

Sidgwick's dualism of practical reason

Before we can begin outlining the moves Lazari-Radek and Singer want to make in order to offer their solution to Sidgwick's dualism of practical reason, we first need to examine Sidgwick's method and sketch how he ultimately arrives at this view.

---

[4]Jonathan Haidt, 'The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment', *Psychological Review* 108 (2001), pp. 814–34 and *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York, 2012); Joshua Greene, 'The Secret Joke of Kant's Soul', *Moral Psychology*, 5 vols., ed. Walter Sinnott-Armstrong (Cambridge, MA., 2017), vol. 3, pp. 35–80, and *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them* (New York, 2013).

Sidgwick's project in *The Method of Ethics* is to evaluate various approaches to ethics and ultimately argue for a coherent set of basic, self-evident principles that can guide our practical reason in order to determine what to do.[5] In order to accomplish this he employs a method referred to as rational intuitionism.

Rational intuitionism involves the use of reason to find principles that can be considered axiomatic in terms of their self-evidence and direct plausibility, which we can apprehend clearly. Sidgwick offers three principles that he believes are generated by the proper application of rational intuitionism. Two have been named already, those of prudence and benevolence. As Sidgwick describes it, the principle of rational prudence involves the thought that 'One ought to aim at one's own good … on the whole'.[6] The principle of rational benevolence is characterized as stating that 'The good of any one individual is of no more importance, from the point of view of the Universe, than the good of any other', and that 'as a rational being I am bound to aim at good generally – so far as it is attainable by my efforts – not merely at a particular part of it'.[7] The third principle for Sidgwick is the principle of justice, which states that 'whatever action any of us judges to be right for himself, he implicitly judges to be right for all similar persons in similar circumstances'.[8]

The self-evidence of these principles on the basis of our intuitions is meant to be distinct from the kind of self-evidence that we sometimes attribute to the intuitions that make up our common-sense understanding of morality, our intuitional judgements when first confronted with a moral situation. The intuitions involved in rational intuitionism are carefully reflected upon and apply to general principles, seeming to differ from the kind of initial judgements we have in particular cases. For example, when we have an intuition about the morally permissible course of action in trolley problem cases this is said to be different from when we are thinking about the plausibility of a fundamental general moral principle.

In contrast with the typical narrative of ethics, whereby intuitions clash with utilitarian thinking, Sidgwick thus seeks to use intuitions to argue for utilitarianism, albeit employing intuitions that have undergone a process of critical reflection.[9] He contends that common-sense intuitions like the kind that we have about particular cases ultimately require appeals to utility to solve a number of problems. A morality based on particular intuitions in cases is thus not the method of ethics to which we should turn. He argues that his rational intuitionism results in the acceptance of axiomatic moral principles that offer a rational foundation for utilitarianism.[10] However, there is more to the story.

As Sidgwick has framed the debate, common-sense morality is off the table, leaving utilitarianism and egoism as the remaining ethical theories. And these two remaining theories appear to be irreconcilable. In some situations, a conflict can arise between my own interests and the interests of others. The need for a solution to this conflict Sidgwick dubbed 'the profoundest problem of ethics'.[11] There are a number of ways in which the utilitarian can attempt to persuade the egoist to promote more than her own good, such as with the idea that by pursuing the impartial good of all (as

---

[5]Lazari-Radek and Singer, *Point of View*, p. 149.
[6]Sidgwick, *Methods*, p. 381.
[7]Sidgwick, *Methods*, p. 382.
[8]Sidgwick, *Methods*, p. 379.
[9]Sidgwick, *Methods*, p. 496.
[10]Sidgwick, *Methods*, p. 496.
[11]Sidgwick, *Methods*, p. 386 n. 297.

utilitarians would do) she will also achieve her own good. But these strategies do not show utilitarianism to be the preferred theory. Rather, they are merely arguments for utilitarian outcomes on egoist grounds.[12]

The difficulty, Sidgwick emphasizes, is that there is a conflict between his principles of rational prudence and rational benevolence, which lead to egoism and utilitarianism respectively. He writes:

> It would be contrary to Common Sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently 'I' am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals: and this being so, I do not see how it can be proved that this distinction is not to be taken as fundamental in determining the ultimate end of rational action for an individual.[13]

This is the thinking that underlies the principle of prudence, and it opposes the thinking behind the principle of benevolence. How can I simultaneously promote the overall good, and also be concerned with the quality of my own existence in a way that differs from my concern for that of others, in cases where these two principles conflict?

Sidgwick examines a number of possible answers to this question, even considering whether the prospect of divine reward and punishment can play a role in providing the egoist with reasons to promote the overall good. Ultimately, he believes he is unable to find an argument that is capable of accomplishing this task. Lazari-Radek and Singer hope to succeed where Sidgwick considers himself to have failed.

Debunking rational prudence

Interestingly, as Lazari-Radek and Singer point out, Sidgwick himself considered the merits of evolutionary debunking explanations.[14] Sidgwick noted that the mere fact that we can tell a story about where our judgements come from does not undermine them, for then that argument would undermine itself, as it would involve judgements whose origins we could explain.[15] If knowing how we came to have our judgements undermined our justification for having them, then the judgement that this state of affairs would undermine our justification would also have its justification undermined. Indeed, it seems that any of our judgements are subject to an explanation of how they came about.

It is not that knowing how our judgements came about is problematic, Sidgwick argues, but rather that it would be problematic for a particular judgement if it could be shown to have come about as a result of a process that we believe to lead to erroneous judgements.[16] Although Sidgwick admits this is a possibility, his strategy is to focus on the self-evidence of the axioms he presents, and he believes that they are not undermined by evolutionary accounts. He agrees that we can be sceptical of other intuitions, those not involved in our judgements about the axioms, but this is beside the point

---

[12]Lazari-Radek and Singer, *Point of View*, p. 150.
[13]Sidgwick, *Methods*, p. 498.
[14]Lazari-Radek and Singer, *Point of View*, p. 175.
[15]Sidgwick, *Methods*, pp. 212–13.
[16]Lazari-Radek and Singer, p. 176.

since he has already rejected the idea that we should heed these other intuitions.[17] He does not need them in order to arrive at his three principles.

Some of us might suspect that since Sidgwick's time the axioms that he claimed are not subject to evolutionary debunking arguments have since been the target of debunking. It would seem a rather fascinating characteristic of our moral reality if his axioms were not even challenged by debunking arguments while other moral judgements are. Remarkably, Lazari-Radek and Singer defend this very claim in the case of the axiom of rational benevolence.[18] They believe that rational benevolence is not subject to evolutionary debunking arguments. In contrast, with regard to the principle of rational prudence, they argue that a debunking explanation can be given, such that utilitarianism is left standing while egoism falls.[19]

After a review of contemporary perspectives on evolutionary debunking arguments Lazari-Radek and Singer develop Sidgwick-inspired criteria for the reliability of intuitions. If all three of their proposed criteria are met, the intuition in question has the highest reliability. The criteria are:

1. Careful reflection leading to a conviction of self-evidence.
2. Independent agreement of other careful thinkers.
3. The absence of a plausible explanation of the intuition as a non-truth-tracking psychological process.[20]

They believe that the intuitions supporting the principle of rational benevolence meet these criteria, while the principle of rational egoism fails at least in terms of the third criterion. Particularly important for their defence of the principle of rational benevolence seems to be their emphasis on its universal aspect. They write that 'it is … difficult to see any evolutionary forces that could have favoured universal altruism of the sort that is required by the axiom of universal benevolence. On the contrary, there are strong evolutionary forces that would tend to eliminate it.'[21] It seems that they want to require that any evolutionary debunking explanation not just explain away some inclination towards accepting a moderate altruism, but rather the kind of universal altruism of impartial benevolence reflected in the principle of rational benevolence. While it is easy to offer a story of why rational prudence, and so self-interest, would benefit us evolutionarily, how could evolutionary pressures lead us to accept a kind of impartial benevolence? I take up one possible response to this question in the following section.

Debunking rational benevolence

The most straightforward objection to Lazari-Radek and Singer's solution is to offer a debunking argument for rational benevolence. Indeed, they acknowledge that although they claim no such debunking argument has so far been offered, this position could be challenged by new results and is exposed to possible falsification in the future.[22]

---

[17]Lazari-Radek and Singer, *Point of View*, pp. 176–7.
[18]Lazari-Radek and Singer, *Point of View*, p. 185.
[19]Lazari-Radek and Singer, *Point of View*, p. 190.
[20]Lazari-Radek and Singer, *Point of View*, p. 196.
[21]Lazari-Radek and Singer, *Point of View*, p. 186.
[22]Lazari-Radek and Singer, *Point of View*, p. 194.

I believe that such a debunking argument could be made in terms of contemporary moral psychology, drawing on Jonathan Haidt's work.[23] Haidt argues that moral reasoning comes after initial affective reactions to moral situations, making such reasoning *post hoc*.[24] For Haidt, our intuitions about what is right and wrong are emotionally driven, and then, after arriving at a decision based on this affective response, we can subsequently offer reasons that would justify it. At least, this is the case for most of us, most of the time. He allows that some small number of us (moral philosophers perhaps) do employ reason at the outset when confronted with a moral situation.

If we can make Haidt's case that considerations of plausibility and self-evidence in moral judgements are quite often the result of affective responses, with reasoning only entering at a later stage in this process, then it seems a challenge has arisen that rational benevolence, and more broadly rational intuitionism as a method, must face. If we evolved a capability to deal with situations with a set of possible affective reactions, and these reactions shape our view on a host of moral issues, then this would seem to call into question our ability to consider any of our intuitions, including those of universal benevolence, insulated from evolutionary debunking.

I argue that Haidt's approach can indeed give us a plausible explanation of our caring about the general good of others such that we would endorse the principle of rational benevolence, and that it also calls into question our ability to employ the methodology of rational intuitionism. I begin with the claim that Haidt's approach could provide an account of our general concern for others in a way that debunks rational benevolence. Recall that rational benevolence is understood in terms of the idea that 'the good of any one individual is of no more importance, from the point of view (if I may say so) of the Universe, than the good of any other'.[25] Haidt offers an approach that he and his colleague Craig Joseph developed known as the Moral Foundations Theory, which proposes that there is a certain number of moral foundations from which our intuitions are said to arise.[26] It was created by 'identifying the adaptive challenges of social life that evolutionary psychologists frequently wrote about and then connecting those challenges to virtues that are found in some form in many cultures'.[27] The foundations in question are understood to be 'universal cognitive modules upon which cultures construct moral matrices'.[28]

Thus, on Haidt's view, we can think of our moral judgements as arising from one of six foundations which are cognitive modules that statistically typical humans have across cultures. These modules are triggered by certain environmental inputs, generate emotions, and can be expressed in different cultures as different understandings of a number of virtues relating to the module triggered. Haidt offers an example in terms of his Care/Harm foundation to illustrate further. He writes that if we see our own son under general anaesthesia having his appendix removed, we will not respond favourably to the scalpel piercing his skin, even though we know that this is saving

---

[23]Lazari-Radek and Singer are certainly aware of Haidt's work and cite it at several points but do not comment on its relevance to debunking rational benevolence. They discuss particular cases of moral dumbfounding, but not Haidt's general theory of morality as affective responses followed by *post hoc* judgements.

[24]Haidt, 'The Emotional Dog and its Rational Tail'. See also Haidt, *The Righteous Mind*.

[25]Sidgwick, *Methods*, p. 382.

[26]See Haidt, *The Righteous Mind*, p. 146, for a chart detailing the initial foundations proposed by Moral Foundations Theory.

[27]Haidt, *The Righteous Mind*, p. 146.

[28]Haidt, *The Righteous Mind*, p. 146.

his life.[29] Rather, we respond emotionally based on our care for our child and our desire to keep him from harm. Those foundations most relevant for explaining the kind of intuitions behind the principle of rational benevolence are probably the Care/Harm foundation and the Fairness/Cheating foundation.[30] A full explanation of how these foundations could offer an adaptive explanation for rational benevolence is beyond the scope of this article, but I attempt to offer some idea of how its initial steps might go.

If we wanted to use Haidt's theory to explain why we might think that the good of one individual is no more important than the good of any other, we might emphasize the importance of our intuitions that are linked to the emotions of compassion and sympathetic concern, as well as anger, gratitude, and guilt. Haidt thinks these two sets of emotions lead to the kinds of intuitions that we have when we face moral situations that invoke caring and kindness or fairness and justice.[31]

If we find it plausible that our own interests are not, in some impartial way, more important than those of others, then this could probably stem from our caring for others or our sense of fairness. It is arguably plausible that statistically typical humans have some amount of sympathy for the rest of humankind that could cause them to de-emphasize their own interests. Moreover, how would it be fair that my interests be considered more important than another's, simply because they are my own? In line with Sidgwick's own principle of justice, do we not tend to have emotional reactions to instances of injustice even when they happen to other people? By thinking in these kinds of ways we could adapt Haidt's ideas to explain the appeal of the intuitions that lead to the principle of rational benevolence.

No doubt some serious empirical and theoretical work would be needed to offer a complete and comprehensive explanation of rational benevolence in terms of Haidt's theory, such that it would be convincing to those not already sympathetic to his approach. But I think that the initial moves and plausibility of such an explanation are understandable here. The claim that Lazari-Radek and Singer want to make, that rational benevolence is not plausibly debunked by evolutionary factors, needs to face explanations of our moral intuitions in terms of moral psychology, and specifically an understanding of our moral intuitions as emotionally driven, the result of moral foundations common to all humans. If we evolved a capability of affective response to moral situations with later *post hoc* justification, and we understand that this capability might make us receptive to the idea that no one person's interests are, from an impartial standpoint, more important than any other person, then we might also be left to wonder about rational benevolence. Perhaps we only find rational benevolence plausible in terms of the methodology of rational intuitionism because of emotional influences inherited from our evolutionary past.

Recall that the third criterion Lazari-Radek and Singer suggest as important for the reliability of intuitions is 'the absence of a plausible explanation of the intuition as a non-truth-tracking psychological process', and that they agreed their debunking explanation is open to the possibility of empirical falsification. The principle of rational benevolence appears at least open to this charge involving the third criterion on Haidt's account, if not entirely susceptible to it. It is remarkable that Lazari-Radek and Singer would allege that the principle of rational benevolence could not be debunked

---

[29]Haidt, *The Righteous Mind*, p. 147.
[30]Haidt, *The Righteous Mind*, pp. 146, 178.
[31]Haidt, *The Righteous Mind*, p. 146.

when we can offer such an explanation in terms of Haidt's account. Perhaps Haidt's particular explanation fails, but such explanations of our moral intuitions in emotionally driven terms certainly merit a response from the likes of Lazari-Radek and Singer's rational intuitionist camp. The so-called sentimentalist approaches to our moral intuitions need to be dealt with before we can be so sure that our intuitions are capable of meeting the criteria that Lazari-Radek and Singer wish to endorse.

Perhaps Lazari-Radek and Singer could reply to theories like Haidt's. They could emphasize that endorsing universal benevolence stands apart from other kinds of moral judgements in a number of ways. They might argue that because of the kind of impartial commitment to the overall good that it seems to require, rather than merely some affective response that shows some amount of consideration for the good of others, it differs from the kind of explanations Haidt is giving. Yet I have attempted to offer here some idea of how a Haidt-style explanation not just of moderate altruism might go, but of how we could understand the appeal of Sidgwick's principles on Haidt's account. Also, it is worth emphasizing again that even if this fails in terms of Haidt's account, it is not clear why a modification of the account, or another account entirely, could not be made so as to better capture the universal aspect of rational benevolence.

Lazari-Radek and Singer might also want to say that Haidt's theory only applies to our judgements about particular cases, not to judgements about general principles, given that they cite his work on moral dumbfounding as an instance of evolutionary debunking in specific cases.[32] Yet Haidt's theory does not stop with the moral dumbfounding in specific cases that he considers. Rather, it works to build an understanding of our morality in general. His Moral Foundations Theory is an effort to explain how we come to judgements about not only particular cases, as our adaptive modules are triggered, but also how these develop into virtues in different cultures. Haidt's account of the development of virtues can be understood as providing an ultimately adaptive explanation not just of the emotional triggers produced by certain cases but also of the kind of judgements we have about virtues in general. Judgements about virtues bring us close to judgements about general principles, which are the kind of judgements in question in Sidgwick's rational intuitionism. No doubt more work needs to be done to explain how Haidt's account of virtues relates to judgements about general principles, but I think that it is sufficient here in responding to this possible objection to emphasize that his account reaches well beyond judgements about specific cases.

Lazari-Radek and Singer might instead argue that as a methodology rational intuitionism is exactly the kind of rare instance in which we do not engage in emotionally driven intuitional responses, followed by *post hoc* reasoning. It involves the kind of careful reflection that is not taking place in Haidt's general characterization of moral behaviour. If they could demonstrate this convincingly then they might be able to say that a Haidt-style challenge fails to threaten their view that we reason towards the principle of rational benevolence. I have been sketching the initial steps of an argument for specifically debunking the principle of rational benevolence in terms of our evolved moral psychology. In order to address this possible response, I now turn to consider how Haidt's account relates to the methodology of rational intuitionism more broadly.

Perhaps we can, as I have suggested, argue that the plausibility of the principle of rational benevolence is based on our evolved moral foundations that revolve around concerns of care and harm, or fairness and cheating. Perhaps also, a stronger claim could be

---

[32]Lazari-Radek and Singer, *Point of View*, p. 188. See the explanation of debunking our judgements about incest through moral dumbfounding on p. 188 n. 18, which cites Haidt's work on dumbfounding.

made with regard to rational intuitionism, that not just our intuitions about the principle of benevolence but all or nearly all of the intuitions involved in the method might be susceptible to debunking arguments in terms of our moral psychology.

Recall that the first two criteria that Lazari-Radek and Singer offer for the reliability of intuitions involve the intuition in question being felt to be self-evident, and there being 'independent agreement of other careful thinkers'.[33] It seems that in terms of self-evidence, what this criterion amounts to is that a principle is found to make sense to a particular thinker as the kind of principle that would be appropriate as a general moral axiom. The second criterion would seem to amount to something like the requirement that humans from different backgrounds encounter a principle under examination without pressure on their decision from others, and then agree that the principle under consideration makes sense.

In light of these criteria, one can pose the following challenge: what if we find certain principles plausible and agree upon them independently because the statistically typical among us are operating from something like Haidt's Moral Foundations Theory in all our moral intuitions? Perhaps if we can all find the same kinds of principles plausible and agree on them independently it is because we are generally employing the same kind of emotionally driven intuitions in order to do so, intuitions that are a result of our shared evolutionary history. To approach rational intuitionism in these terms is to offer a challenge from the standpoint of the issue raised in the third criterion, albeit for the method as a whole: perhaps not just a particular intuition but actually all our intuitions are susceptible to a Haidt-style account, and so an explanation in terms of 'a non-truth-tracking-psychological process'.[34]

The rational intuitionist might respond by emphasizing that the kinds of intuitions that are employed in the calm reflection of the rational intuitionist methodology are specifically those that are not emotionally valenced. Yet, at the very least, one could probably legitimately wonder how we could really know whether our intuitions are emotionally driven or not. If we think that one group of people, such as philosophers, tends to achieve dispassionate reactions to possible moral principles, we need to consider whether this is not merely to fetishize or privilege as dispassionate the still emotionally driven intuitions of one group of people. The intuitions of philosophers and the way philosophers come to moral judgements might be just as emotionally driven as those outside philosophy, albeit altered by the kind of thinking that philosophers tend to do. In defending the rational intuitionist method as the kind of careful philosophical thinking that those in philosophy engage in, are we privileging the emotionally driven judgements of the philosophical subculture over other subcultures in a way that could really be justified?

Fascinatingly, Haidt takes up this very question, regarding whether careful philosophical thinking is exempt from his account, in his discussion of classic moral theories. Controversially, to say the least, he offers the view that two of the great rationalist moral thinkers, Jeremy Bentham and Immanuel Kant, were arguably abnormally lacking in empathy, certainly differing from statistically typical humans in this respect. He even goes so far as to suggest that Bentham probably had Asperger's syndrome, citing evidence to that effect, and suggesting that this was also a possibility in Kant's case, though there such a diagnosis is perhaps less likely.[35]

---

[33]Lazari-Radek and Singer, *Point of View*, p. 195.
[34]Lazari-Radek and Singer, *Point of View*, p. 195.
[35]Haidt, *The Righteous Mind*, pp. 138–40.

If Haidt is onto something in his charge that those who engage in rationalist moral philosophy are statistically atypical in their lack of empathy, then we might wonder why we should think that rational intuitionism in these terms could be what we are after as the methodology for arriving at human moral truths. If rational intuitionism really is not subject to explanation in terms of Haidt's theory, then why should the statistically typical (whose moral behaviour *is* explained by Haidt's account) listen to the exhortations of a small subgroup of oddly unempathetic and specifically encultured individuals? If we are to take the second criterion seriously, such that we should achieve independent agreement about our principles, what good is it to have only a statistically atypical subset of the population be able to agree about such principles?

Of course, Haidt's theory is not the only game in town for characterizing our moral psychology. I consider how we might understand and defend the methodology of rational intuitionism against these questions by invoking the results of Joshua Greene's research programme in the following section.

Debunking non-consequentialist intuitions

A moral psychology that at first glance appears far friendlier to the kind of utilitarianism that Lazari-Radek and Singer endorse is available in Joshua Greene's work. Greene's research suggests that the kind of intuitions that deontologists tend to emphasize occur in a different part of the brain than the cost-benefit utilitarian thinking that we also engage in.[36] This would mean that the intuitions that we tend to think of as being involved in deontological judgements are affective in a similar way to Haidt's characterization of emotionally driven intuitions.

Probably agreeably to many consequentialists, Greene argues that consequentialist cost-benefit thinking focused on maximizing the overall good is 'more cognitive … and more likely to involve genuine moral reasoning'.[37] I am not able to offer all the details of Greene's empirical research here, but I will sketch a few of the major ideas. Greene uses 'cognitive' to mean those representations that are inherently neutral, that 'do not automatically trigger particular behavioral responses or dispositions'.[38] This is opposed to 'emotional' representations, those 'that do have such automatic effects, that are behaviorally valenced'.[39] Greene acknowledges that there is only a rough distinction between cognition and emotion, and that in fact all moral judgements may have some emotional aspect.[40] Yet he thinks even if this is the case, the kind of emotion involved in deontological intuitions is problematic in a way that the kind of emotion in consequentialist intuitions is not.

Greene offers neuroimaging studies of patients confronted by trolley problems as providing significant support for his theory. The classic trolley case that involves a decision between saving one person and saving five, to be decided by the pulling of a lever, tends to elicit consequentialist intuitions.[41] Yet the other famous trolley case, wherein subjects are asked to decide whether or not to push a large man in front of the trolley

---

[36]Greene, 'The Secret Joke of Kant's Soul'; Greene, *Moral Tribes*.
[37]Greene, 'The Secret Joke of Kant's Soul', p. 36.
[38]Greene, 'The Secret Joke of Kant's Soul', p. 40.
[39]Greene, 'The Secret Joke of Kant's Soul', p. 40.
[40]Greene, 'The Secret Joke of Kant's Soul', p. 41.
[41]Greene, 'The Secret Joke of Kant's Soul', p. 42.

(which would result in his death) in order to stop it and so save the five, tends to elicit deontological intuitions.[42]

After surveying a number of studies and detailing his own empirical research, Greene argues that the neuroimaging data from subjects point to different areas of the brain being activated in these different judgements. He offers the explanation that our proximity with regard to a person, whether or not we are 'up close and personal', influences whether we have an emotional response.[43] He describes what is going on in each situation as the generation in a subject of a 'prepotent, emotional response', and sometimes also a cognitive response that conflicts with it.[44]

In addition to considering trolley cases, Greene also examines neuroimaging results from patients confronted with Singer's classic story about walking by a pond and seeing a drowning child.[45] Singer thinks that even if we are wearing expensive shoes, we should save the child, and from studies it seems that most of us agree. However, when confronted with the fact that we buy expensive shoes rather than helping the global poor, few of us feel the same emotional pull as when agreeing that we ought to save the child.[46] As is the case of trolley problems, neuroimaging suggests that whether or not a victim is 'up close and personal', and so whether or not they are identifiable, plays a considerable role in whether or not we respond emotionally.[47]

Yet, Greene notes, this kind of factor does not seem reasonable. Why should whether a victim is up close and personal or not matter? Is this not our emotions simply driving us to think in unreasonable ways? It is also worth noting that the 'up close and personal' approach to understanding emotion and intuition seems to provide a readily available evolutionary explanation: throughout much of the history of our species we have only been able to help those close by and identifiable, allowing for the adaptation of features in our moral psychologies that generate emotional responses that are stronger under these conditions.

Greene's argument appears especially relevant to those deontologists who emphasize the role of intuitions about cases in their construction of general moral principles, such as, for example, Frances Kamm.[48] We might think that those non-consequentialist approaches that rely less on intuitions about cases are less susceptible to Greene's critique, such as those of John Rawls and T. M. Scanlon, which employ initial methods of rational reflection in order to develop our everyday moral intuitions into those that are consistent with certain rules about how morality should be understood.[49] Yet Greene offers a way in which his argument could be troubling more broadly.

Greene offers a dilemma for non-consequentialist theories that rely on rational reflection to generate foundational principles. He points out that for any such process of rational reflection there will be judgements that go into the reflection that are based on emotionally driven intuitions.[50] The question is whether these emotionally driven

---

[42]Greene, 'The Secret Joke of Kant's Soul', p. 42.

[43]Greene, 'The Secret Joke of Kant's Soul', p. 43.

[44]Greene, 'The Secret Joke of Kant's Soul', p. 45.

[45]Peter Singer, 'Famine, Affluence, and Morality', *Philosophy and Public Affairs* 1 (1972), pp. 229–43.

[46]Greene, 'The Secret Joke of Kant's Soul', p. 47.

[47]Greene, 'The Secret Joke of Kant's Soul', p. 47.

[48]Francis Kamm, *Intricate Ethics: Rights, Responsibilities, and Permissible Harm* (Oxford, 2007).

[49]John Rawls, *A Theory of Justice* (Cambridge, Mass., 1971); T. M. Scanlon, *What We Owe to Each Other* (Cambridge, Mass., 2000). For Greene's discussion of these theories see 'Reply to Mikhail and Timmons', *Moral Psychology*, 5 vols., ed. Walter Sinnott-Armstrong (Cambridge, MA., 2017), vol. 3, pp. 105–18.

[50]Greene, 'Reply to Mikhail and Timmons', p. 116.

intuitions also come out of the process. Greene refers to this in terms of what computer scientists call the 'garbage in, garbage out' problem, abbreviated as the 'GIGO' problem. If emotionally driven intuitions go into the process, then this would mean that the outcome of the process would be emotionally influenced, since it would be based in part on these kind of intuitions that are affective in origin. But then this seems to present a problem: if emotions go in and come out, then, as Greene puts it, 'the so-called "moral truth" now reflects arbitrary features of our evolutionary history'.[51]

If, on the other hand, the emotionally driven intuitions go in and do not come out of the process of rational reflection, then Greene raises a different issue. He asks: 'then in what sense is the moral truth deontological?'[52] That is to say, if arriving at our foundational theory of normative ethics involves a process that removes emotionally driven intuitions, and so those that are deontological, then the end result of that process is our being left with only utilitarian intuitions. This is amenable to Greene, but will probably not be for many non-consequentialists. It seems clear that most non-consequentialists worthy of the name would want to avoid their reasoning resulting in utilitarianism. The entire point of endorsing a particular process of rational reflection as a non-consequentialist would seem to be motivated by the expectation that it would produce a non-consequentialist outcome. What is the deontologist to do?

In a response to Greene, Mark Timmons suggests that the deontologist could embrace the emotional grounding of non-consequentialist intuitions, suggesting that we go in for the first horn of Greene's dilemma, an emotionally grounded deontology.[53] Timmons writes: 'although all the versions of deontology that I know of have been embedded in a rationalist metaethic, I don't see why one cannot embrace sentimentalism (or expressivism) and go on to defend a deontological moral theory'.[54]

This position does seem available, though it is probably unpalatable to many. As Greene notes, such a view would probably be fatal for a normative theory of morality because of the GIGO issue he raises. Our emotions seem unpredictable, dynamic, capricious, and individual: not the kind of things that we would want to be involved in shaping a moral theory meant to be objective. It may be that our everyday moral behaviour often does involve emotions, but when considering what we ought to do, offering an emotionally grounded answer would not appear action-guiding in any encouraging way, at least to those who do want to go in for a rationalist approach to moral theorizing. Whatever it told us to do, we might wonder whether its imperatives were based on unreliable emotions that might change based on seemingly erroneous factors.

The other options available for the deontologist appear to be either to challenge Greene's empirical claims about emotions and intuitions, or, as Greene would have it, to become a utilitarian. Perhaps a convincing argument could be made that Greene's approach towards non-consequentialist intuitions is flawed. A full investigation into this question cannot be carried out here. As such, I will not consider the truth of the empirical claims Greene makes but rather their implications if they are indeed largely correct about the status of our moral intuitions.

What I want to explore is whether Greene's view, which at first appears so favourable for consequentialists like Lazari-Radek and Singer, actually offers a way in which

---

[51] Greene, 'Reply to Mikhail and Timmons', p. 116.
[52] Greene, 'Reply to Mikhail and Timmons', p. 116.
[53] Mark Timmons, 'Toward a Sentimentalist Deontology', *Moral Psychology*, 5 vols., ed. Walter Sinnott-Armstrong (Cambridge, MA., 2017), vol. 3, pp. 93–104.
[54] Timmons, 'Toward a Sentimentalist Deontology', p. 102.

rational intuitionism as a viable methodology is threatened because of the very GIGO problem that is meant to threaten non-consequentialism. Like the processes of rational reflection endorsed by Rawls and Scanlon that Greene is targeting, Sidgwick's process of rational reflection could also be threatened. I also want to consider how the implications of the different theories that Haidt and Greene offer differ with regard to rational intuitionism and the prospect of formulating an evolutionary debunking argument of rational benevolence.

Reason, emotion, and rational intuitionism

Now that we have both Haidt's and Greene's ideas about the role of reason and emotion in our moral psychology on the table, we need to bring things together to understand how they square with Sidgwick's methodology of rational intuitionism. I have already considered how the specific principle of rational benevolence might be explained in Haidt's terms, and how Haidt's approach would see rational intuitionism as involving emotionally driven intuitions, at least for statistically typical humans. At that juncture I raised the question of whether rational intuitionism should be understood as dispassionate judgement, the kind that is meant to be pursued under ideal conditions by Sidgwick-style moral philosophers. I will now explore this issue in the light of how Greene understands consequentialist judgements to be more cognitive and deontological judgements to be more emotional. How does the kind of thinking going on in rational intuitionism relate to Greene's understanding of our moral judgements?

At first one might think that Sidgwick's rational intuitionism would have something to do with the cost-benefit intuitions that Greene believes underlie utilitarian thinking. At least, this thought might emerge if we trace Sidgwick's endorsement (by means of his rational intuitionist methodology) of the promotion of the impartial good in terms of the principle of benevolence. Sidgwick would seem to be capable of taking the horn of Greene's dilemma for reflective processes that characterizes the processes as resulting in utilitarianism after all emotionally driven intuitions have been eliminated.

Alternatively, if we can offer Haidt's theory as an explanation of how the endorsement of rational benevolence could be understood as driven by emotionally driven intuitions, this would seem to indicate that Sidgwick's intuitionism suffers from the GIGO problem. Emotions in the manner of Haidt's theory go in, and then come out as well, supporting rational benevolence. Rather than Sidgwick's rational intuitionism being a dispassionate endorsement of rational benevolence in cost-benefit terms, it would be an emotionally driven affair.

Can Greene's characterization of moral behaviour be used to come to the defence of rational intuitionism against the very GIGO issue he raises? I argue that it cannot. A defence of rational intuitionism in terms of Greene's understanding of cognitive cost-benefit thinking makes little sense: Greene is offering the dichotomy between cognitive and emotional judgements as applying to specific cases and to our level of thinking in normative ethics, not to a metaethical methodology that generates general principles.

One way of describing a Greene-informed utilitarianism is characterizing it as the view that the cost-benefit style of human moral decision-making is the one that we ought to endorse, at least when it generates better consequences overall that we do so. But in employing rational intuitionist methodology we are not yet engaging in moral thinking at the level of normative ethical theories like those of utilitarianism, deontology, or virtue ethics. We are not weighing the costs and benefits of adopting particular principles, for then we would have already endorsed utilitarianism insofar

as we are deciding between different principles based on their effects. Rather, in rational intuitionism we are engaging in reflection about what is plausible, probably also with Lazari-Radek and Singer's second and third criteria in mind: whether there could be independent agreement about our conclusions, and whether we can give debunking explanations for our plausibility considerations in evolutionary terms.

For the rational intuitionist, we are just supposed to consider whether the principle in question seems plausible, for if we asked about the consequences of different principles, we would already be consequentialists. What we are asking is a question much more abstract, such as if we ask whether the consequentialist principle of rational benevolence is plausible. We cannot seem to say that we should choose it because it will lead to more benefits and fewer costs, since that is partly what is at issue, whether cost-benefit thinking should guide us, and so rational intuitionism does not seem like a cognitive cost-benefit affair. Greene offers an interesting argument for cost-benefit thinking being less emotional, but that seems to be no proof of the principle of rational benevolence, since it seems that we cannot use cost-benefit thinking to judge this principle if we are to be rational intuitionists. The direct plausibility of the principle is what is supposed to justify the use of cost-benefit thinking.

It is also important that Sidgwick endorses the principle of rational prudence with the method of rational intuitionism, which seems directly contrary to utilitarian thinking. If we were to think of rational intuitionism as properly carried out by employing only the kind of cognitive, consequentialist, cost-benefit thinking that avoids emotionally driven intuitions, then how could we make sense of the fact that it generates the anti-consequentialist principle of rational prudence? It seems likely that Greene's dichotomy does not apply to the kind of reflection about plausibility that is going on in rational intuitionism, and so, I argue, cannot save it from the GIGO problem: evolutionarily adaptive and emotionally driven intuitions go in, and evolutionarily adaptive and emotionally driven intuitions come out.

In future, it might be valuable to conduct the kind of neuroimaging studies that Greene does while somehow surveying opinions about Sidgwick's principles. This might allow us to have a better understanding of the extent to which reason and emotion are operating when we make rational intuitionist plausibility judgements. Although the cases involved in Greene's neuroimaging studies can implicitly pit rational benevolence and rational prudence against one another in particular instances (such as in Singer's pond case), they do not seem to be ascertaining what is going on in our brains when we actually consider what kinds of principles we find plausible. They ask us to consider situations in which we are already implicitly trading off considerations of benevolence and prudence, not asking whether these should in fact be the principles that ultimately structure morality.

What Greene does offer us at this juncture is an understanding of the GIGO problem, which I see as a way of framing evolutionary debunking concerns. And the threat of GIGO in evolutionary debunking terms does seem problematic for Sidgwick's method. If we combine a Haidt-style emotional account of rational benevolence with Lazari-Radek and Singer's understanding of rational prudence as subject to evolutionary debunking as well, it seems like rational intuitionism suffers doubly from the GIGO problem. Evolutionarily influenced considerations go in, and they also come out, in terms of both rational prudence and rational benevolence. And as I noted earlier, it is important to remember that even if Haidt's particular account fails, it seems that the debunkers could regroup and mount another charge to explain rational benevolence in evolutionary terms. Yet, despite all this, as I will attempt to show, it is questionable

whether debunking arguments can actually touch either of these principles on the view of reason that Lazari-Radek and Singer endorse. If I am right this means that their own debunking argument of rational prudence is superfluous and a distraction.

Reason as a unity and evolutionary debunking

Lazari-Radek and Singer defend a view of reason as having evolved as a unity, coming along in one package, that once around delivers conclusions like those which rational intuitionism is supposed to deliver. I submit that on this understanding of reason, dealing with debunking arguments is actually a waste of time. If reason is really a unitary package as they want to say, then once it is operating in humans it delivers the kinds of judgements going on in rational intuitionism that they want, regardless of evolutionary influences. In order to make these claims, we first need an understanding of Lazari-Radek and Singer's take on reason.

Lazari-Radek and Singer formulate their view in response to Sharon Street's evolutionary debunking argument directed generally at the objectivity of ethics. Street offers a dilemma, based on the idea that 'evolutionary forces have played a tremendous role in shaping the content of human evaluative attitudes'.[55] The dilemma is as follows: either evolutionary pressures do not select for creatures that subscribe to objectively correct human evaluative attitudes, or they do in fact select for such creatures. If evolutionary pressures do not, then it is difficult to see how we can say that we have arrived at moral truth, other than by sheer coincidence. If they do, then this is problematic, Street argues, because although we would then have moral judgements that were selected for based on reproductive success, it would be difficult to explain why this would lead to our having objectively correct moral judgements.

Lazari-Radek and Singer believe that Sidgwick can offer a response to Street. They argue that given his rejection of many of our common-sense moral judgements, he could agree with Street that such judgements do not convey a correct understanding of objective morality. However, as far as his three principles are concerned, Sidgwick could offer a different story. Since Sidgwick believes that we arrive at his three principles via reason, Lazari-Radek and Singer contend that we can offer an account of the evolution of reason that explains how evolution could pressure us towards finding moral truth.

The story Lazari-Radek and Singer offer is that once our ability to reason evolved due to the advantages it conferred, it then was also capable of grasping certain kinds of truths.[56] Reason comes about because it enables us better to achieve certain ends necessary for survival and reproduction, but once it comes along we can use it to discover other things which do not necessarily confer reproductive advantages.

For example, an understanding of mathematical truths or theories of physics arguably would not seem to confer a reproductive advantage in the kind of environment in which our species first found itself with the capability of reason. Yet this does not call these truths into question, because they are by-products of our evolved faculty of reason, which did confer other reproductive benefits. The same could be true, Lazari-Radek and Singer believe, for morality. One challenge to this view is that there are different aspects of our reasoning ability, and some could have been advantageous while others were disadvantageous. The question is how we could have disadvantageous aspects of our

---

[55]Sharon Street, 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies* 127 (2006), pp. 109–66, at 109.

[56]Lazari-Radek and Singer, *Point of View*, pp. 182–3.

reasoning, given that it seems that we need to consider reason advantageous in order to explain how it evolved.

Lazari-Radek and Singer submit that one explanation is that reason came onto the scene as a package, which meant that both its advantageous and disadvantageous aspects came together. We could not have the beneficial parts without the detrimental parts. This is an empirical claim, and Lazari-Radek and Singer are fully aware of this fact. They admit that their reason as a package view could be challenged by future empirical results.

I want to draw out an implication of Lazari-Radek and Singer's emphasis on reason. They want to explain our moral judgements that lead to Sidgwick's principles as arising from reason, through a dispassionate method of reflection that does not involve emotionally driven intuitions. They want to explain reason as arising because it conferred an advantage on us, and then also, by coming as a package, offered the opportunity to arrive at certain truths. Yet, if the by-product capacities of reason that enable moral thinking are to be understood as the sole source of Sidgwick's principles, how can their evolutionary debunking solution succeed against his principle of rational prudence? And why would they need to address evolutionary factors at all?

We can pose this challenge in the form of a dilemma: either by-product capabilities of our reason alone generate Sidgwick's principles, or they do so in conjunction with some evolutionarily adaptive capabilities that we have. If the first option is embraced, then it seems unclear how the principle of rational prudence, one of Sidgwick's principles arrived at through reason, could be subject to any kind of evolutionary debunking explanation at all. If reason is solely responsible for developing such a principle, then it comes, so to speak, as a by-product of the by-product capabilities of reason. It is several steps removed from any evolutionarily adaptive features, and so not capable of being debunked as merely evolutionarily adaptive.

On the second option, we have the return of the GIGO problem for rational intuitionism. If the by-product capabilities of reason are responsible for Sidgwick's principles along with some evolutionarily adaptive capacities, then the methodology of rational intuitionism is producing principles subject to evolutionary debunking arguments. Street's dilemma once again rears its head, leaving us wondering how we could arrive at objectively correct moral judgements if evolutionary forces did not select for them, as well as wondering how, if moral judgements are selected for, these judgements could be the objectively correct ones, other than through sheer coincidence.

I submit that Lazari-Radek and Singer cannot debunk rational prudence without also allowing Haidt-style debunking arguments for rational benevolence through the door. If they want to block these, and general Street-style evolutionary debunking arguments, by proposing the idea that reason is a package, then they also cut themselves off from their own solution to the dualism of practical reason, that of debunking rational prudence. A dramatic conclusion appears to fall out from this discussion: either Sidgwick's principles and entire methodology of rational intuitionism are debunked via problematic GIGO concerns, or his dualism of practical reason remains despite Lazari-Radek and Singer's best efforts. Lazari-Radek and Singer, and others sympathetic to their approach, need to contend with a number of difficult dilemmas if they want to succeed in resolving Sidgwick's dualism of practical reason with an evolutionary debunking argument of rational prudence.

Conclusion

Our discussion of the role of moral psychology in evolutionary debunking arguments of ethical intuitions and methodologies has followed many intricate twists and turns. I

conclude by offering an outline of the approach to these issues that I have argued for in this article.

I began by presenting Sidgwick's problem of resolving the dualism of practical reason. How can we reconcile egoism and utilitarianism, the principles of rational prudence and of rational benevolence respectively, both of which are generated by rational intuitionist reflection? Lazari-Radek and Singer's answer is that we can debunk rational prudence in evolutionary terms, but not rational benevolence.

I disputed this solution, offering Haidt's understanding of the origin of our moral judgements as a candidate debunking explanation for rational benevolence, and also pointing to the possibility that many other such candidate theories could be developed to challenge the solution.

I then considered Greene's research programme as a possible defence in terms of moral psychology of the consequentialist thinking that Sidgwick, as well as Lazari-Radek and Singer, ultimately want to endorse. We might think that Greene could be an answer to Haidt-style moral psychology by providing support for consequentialism and cohering with Sidgwick's rational intuitionism.

I disputed a Greene-style solution, arguing that his work is only relevant to specific cases and judgements at the level of normative ethics, not to judgements about metaethical methodology, and that Greene in fact draws attention to the GIGO problem that is at issue in rational intuitionism.

I then turned to explain how Lazari-Radek and Singer's own understanding of moral judgements in rational intuitionism as coming about through a capacity for reason that evolved as a package presents a difficult dilemma for their position. Either all principles arrived at through rational intuitionism are susceptible to evolutionary debunking explanations, or none of them are. If they all are, then rational intuitionism is GIGO. If none of them are, then Lazari-Radek and Singer cannot offer their solution, which attempts to provide an evolutionary debunking argument for rational benevolence. Either rational intuitionism goes down as GIGO or Sidgwick's dualism survives.

Perhaps some kind of evolutionary debunking solution to Sidgwick's dualism of practical reason could yet be salvaged. Lazari-Radek and Singer's attempt is a valiant and innovative effort, but I suspect it would be easier and more profitable going were we to look elsewhere for a solution that undermines the principle of rational prudence while leaving the principle of rational benevolence unscathed.[57]