How to Debunk Moral Beliefs

Victor Kumar & Joshua May[1]

Published in *Methodology and Moral Philosophy*,
ed. by Jussi Suikkanen & Antti Kauppinen, Routledge (2019), 25-48.

Abstract: Arguments attempting to debunk moral beliefs, by showing they are unjustified, have tended to be global, targeting all moral beliefs or a large set of them. Popular debunking arguments point to various factors purportedly influencing moral beliefs, from evolutionary pressures, to automatic and emotionally-driven processes, to framing effects. We show that these sweeping arguments face a debunker's dilemma: either the relevant factor is not a main basis for belief or it does not render the relevant beliefs unjustified. Empirical debunking arguments in ethics can avoid this predicament, but only if they are refocused on highly selective classes of moral belief. Experimental data can combine with familiar consistency reasoning to reveal that like cases are not being treated alike. Selective debunking arguments are unlikely to yield sweeping sceptical conclusions, but they can lead to rational moral change.

Word count: 9,409 (main text and footnotes)

Keywords: genealogical debunking, moral scepticism, moral knowledge, consistency reasoning

A growing body of empirical research is unearthing the grounds of our moral beliefs. There are ample reasons to be concerned about what we might find. Many moral beliefs are shaped by unconscious psychological mechanisms that are part of our biological and cultural heritage. We cannot know purely through introspection all the influences on our moral beliefs, and the actual reasons for our beliefs could be poor—the reasons that we consciously endorse mere rationalizations.

Empirical debunking arguments in ethics have tended to be *global* (or *wide-ranging*), targeting all moral beliefs or a large class of them. Indeed, Guy Kahane argues that debunking arguments (evolutionary ones at least) are all-or-nothing. It 'seems utterly implausible,' he concludes, that such arguments can 'have a legitimate piecemeal use in normative debate' since 'to work at all' they are 'bound to lead to a truly radical upheaval in our evaluative beliefs' (2011, pp. 120-1; cf. also White 2010; Rini 2016). Many have recently argued in exactly this fashion that we cannot remain justified in our ordinary moral beliefs after realizing that evolutionary forces have substantially influenced them

[1] Authorship is equal. Surnames are presented in alphabetical order.

(e.g. Joyce 2006). But one needn't appeal to evolution to generate such sweeping scepticism. Walter Sinnott-Armstrong (2008), for example, argues that most ordinary moral intuitions are unreliable because they are distorted by cognitive biases, such as the order in which information is presented. Absent evidence for one's moral beliefs that is independent of such intuitions, one might conclude that moral knowledge is out of reach.

Some empirical debunking arguments target only certain classes of moral beliefs, but the classes are apparently rather large. Daniel Kelly (2011), for example, argues that disgust is an unreliable emotion in ethics, and therefore that all moral beliefs based on disgust are unjustified (see also Nussbaum 2004). Peter Singer (2005) and Joshua Greene (2014) argue that many intuitively compelling moral beliefs are based on automatic, unconscious, emotion-driven heuristics. Deontologists mistakenly offer elaborate, sophisticated justifications for these non-utilitarian beliefs, when in fact we have them for very simple, unsophisticated reasons. Singer and Greene both argue that once we purge ourselves of the irrational moral beliefs that animate deontology, only utilitarianism remains standing.

We have two main aims, one negative and one positive. First, we argue that each global debunking argument reaches far beyond what the evidence can hope to support. We agree with those who have expressed doubt that empirical facts can debunk ethics in a global fashion (cf. Lillehammer 2003; Vavova 2014), but we develop a precise diagnosis for why all global debunking arguments fail. They systematically face a *debunker's dilemma*: the arguments either (a) identify an influence on moral beliefs that we do not have reason to believe is distorting or they (b) identify a distorting influence that is relatively minor, one that fails to constitute a main basis for the targeted beliefs. In general, extant research suggests a *trade-off*: Identifying a main basis for a large class of moral beliefs tends to implicate a reputable process; but more clearly disreputable processes tend to exert an insubstantial influence on large, heterogeneous classes of moral beliefs. Our second aim is to articulate an approach that avoids the problems that beset global debunking arguments. Highly *selective* debunking arguments that appeal to consistency reasoning offer a more fine-grained and less sweeping empirical evaluation of our moral beliefs. Their upshot is not global moral scepticism but piecemeal and rational moral change.

1. Debunking in Ethics

Much naturalistic ethics is negative, aiming to debunk one or another class of moral beliefs. The word 'debunk' has special meaning here. Sometimes 'debunk' just means to prove wrong (cf. Lillehammer 2003). For example, someone debunks astrology, phrenology, or a religious doctrine by showing that its predictions are not borne out. But we'll focus on an epistemological notion of debunking: To undermine the grounds for belief, to show that our existing reasons for accepting a view are bad reasons. The view itself is not being attacked as false; rather, our belief in it is shown to be *unjustified* (see Wielenberg 2010; Kahane 2011; Nichols 2014). Rather than leading us to deny the view once believed, the conclusion of a debunking argument is that we must *withhold judgment*. (Debunking arguments thus target 'doxastic justification,' not 'propositional justification.')

We can illustrate with an example. Some of the most striking research in psychology over the last few decades targets implicit attitudes and biases. An emerging

literature suggests that implicit biases connected to race, gender, class, age, disability and sexual orientation influence us unconsciously in countless ways (see Brownstein 2015 for review). Researchers employing the implicit association task, for instance, find that participants of varying racial and ethnic backgrounds more easily associate African American faces with negatively valenced words than with positively valenced words, as compared to Caucasian faces (e.g., Nosek et al. 2007).

The associative psychological link between images of African Americans and negative valence likely drives a wide range of judgment, decision-making and behaviour. It is possible that further research will uncover precise links between racial biases and moral belief, in particular. African Americans suffer a disproportionate burden in the legal system, for many reasons of course, but we may discover empirically that their burden stems in part from implicit biases that affect moral judgments about their culpability and blameworthiness. Explicit, 'colour-blind' principles intended to justify existing legal practices threaten to be mere rationalizations of the status quo. Thus, empirical work on the psychological mechanisms underlying moral cognition might help to show that certain moral beliefs about those in the criminal justice system are unjustified and should be abandoned. This is how empirical debunking arguments in ethics can have normative significance.

Our focus will be on such debunking arguments that target ordinary moral beliefs. Thus, we'll set aside those empirically-driven arguments attempting to undermine belief in a particular metaethical theory, such as moral realism (e.g. Street 2006). The genealogical challenges are pressing regardless of whether ordinary moral beliefs presuppose some sort of objectivity. Even if the truth of any moral belief is response-dependent or relative, any plausible metaethical account will allow room for error and groundless belief.

A general worry about empirical debunking arguments in ethics is that they illicitly jump the is-ought gap, that they attempt to derive normative conclusions from purely empirical premises. To understand the structure of debunking arguments, it is important to see that they require a normative premise: Roughly, that some basis for moral belief is morally irrelevant (Kumar & Campbell 2012). For example, the debunking argument from implicit racial bias rests on the normative premise that race by itself is—of course—a morally irrelevant basis for beliefs about culpability and blameworthiness. The justification for normative premises like this is not empirical. A debunking argument is successful, in part, to the extent that its normative premise is more plausible than the moral beliefs that the argument attempts to debunk. Plainly, then, empirical debunking arguments needn't attempt to bridge the is-ought gap, much less jump it.

The best debunking arguments combine empirical claims about the sources of moral beliefs with one or more normative premises to draw a normative conclusion, which is that certain target moral beliefs are unjustified. Note that the conclusion is epistemic and second-order—yielding a verdict about the justification of one's beliefs. Nonetheless, the conclusion can have first-order moral implications. First of all, discovering that some moral belief is unjustified motivates abandoning it. Furthermore, if there is a tension among a set of beliefs and we find out that one subset is unjustified, then that lends support to the other subset. In light of further research on racial biases, we may be led to abandon beliefs about the culpability and blameworthiness of African

Americans caught up in the criminal justice system. This realization could justify embracing other beliefs (e.g. that we're morally obligated to enact racially-sensitive policies), which conflict with our previous way of thinking. Empirical research, then, does not simply tell us what our moral beliefs are; it can also offer suggestions about what they ought to be.

2. Global Debunkers

Theorists working in naturalistic ethics have recently deployed debunking arguments that are quite ambitious. Their targets tend to be global (or wide-ranging): All, or a large group of, our moral beliefs are unjustified because of their unrespectable grounds. The apparent conclusion, then, is that people should withhold judgment about not just certain moral matters but all of them or a great many. We will begin by reviewing four extant global debunking arguments that have garnered much attention. But we will not offer extended discussion of any single one of them since we are interested in bringing to light a difficulty that is common to all.

(1) A number of philosophers attempt to use evolutionary theory to debunk all moral beliefs as a mere biological adaptation (e.g. Ruse 1986; Joyce 2006; Rosenberg 2011). Some formulations of the evolutionary argument turn on *explanatory dispensability*: The best evolutionary explanation of the existence of the moral beliefs we have does not presuppose their truth (cf. Joyce 2006, p. 211). Presumably one should then give up commitment to moral facts because one lacks evidence for their existence. But this does not amount to an epistemological debunking argument of the sort at issue in this essay (cf. Wielenberg 2010, §8). Instead, this is an iteration of a longstanding debate about whether moral facts can really explain anything and, if they can't, whether they deserve a place in our ontology (see Harman 1977). Perhaps one's beliefs become unjustified upon learning that their contents are explanatorily dispensable, as Joyce and his followers allege, but that's rather controversial (Clarke-Doane 2015).

Our concern, at any rate, is with debunking arguments that locate a *defective process* of belief-formation (cf. Nichols 2014). The idea is that if morality exists merely because it enhanced survival and reproduction, then it may seem at best a mere coincidence that moral beliefs would happen upon moral truths. We shouldn't expect that our moral beliefs are true, since natural selection heavily influences their formation, favouring moral judgments that are expedient, not those that correspond to moral facts and properties. There are plausible evolutionary influences on various ordinary moral beliefs—e.g. regarding incest, reciprocity, special duties to one's kin, loyalty to one's group, retribution and so on (Sober & Wilson 1998; Street 2006; Haidt 2012). According to evolutionary debunkers, natural selection 'can't be a process that's reliable for providing us with what we consider correct moral beliefs' (Rosenberg 2011, p. 221). Even if we can have true moral beliefs, it is a matter of luck, akin to forming a belief about some historical fact by swallowing a pill (Joyce 2006, p. 179). There may well be moral truths, for all a debunking argument says, but the reasons we believe there to be any are poor. As Joyce puts it, this 'forces the recognition that we have no grounds one way or the other for maintaining these beliefs' (p. 211). One's moral beliefs may then be unjustified if one is aware of the illicit source (cf. Kahane 2011) or perhaps should be aware of it (cf. Sinnott-Armstrong 2008).

(2) Sinnott-Armstrong (2008) appeals to psychological research on moral cognition, rather than evolutionary biology, but his discussion can likewise lead to sweeping conclusions. His concern is the class of moral beliefs based on intuitions, which is arguably a rather large class, perhaps all of them. Some experimental evidence suggests that moral intuitions are subject to *framing effects*: The way that a moral problem is framed can affect intuitive beliefs about which solutions are appropriate. For example, participants' responses can be influenced by the order in which moral dilemmas are presented or by differences in words that are equivalent in meaning. Framing effects are, of course, a source of unreliability: Neither order nor wording is a morally relevant basis for belief. Now, Sinnott-Armstrong concludes that moral beliefs can be justified but only if based on more than just intuitions (yielding inferential justification). However, this non-sceptical conclusion is difficult to maintain if many or nearly all of our moral beliefs are based on intuitions. As Sinnott-Armstrong puts it: 'We could never get started on everyday moral reasoning about any moral problem without relying on moral intuitions' (p. 47). One might expect that we'll eventually get evidence that certain moral intuitions are reliable, but it's unclear how we could be confident that some evidence confirms the reliability of some moral intuitions without relying on intuitions about what counts as moral accuracy. In any event, anyone who is (or perhaps should be) aware of such framing effects arguably lacks moral knowledge, unless one can acquire the elusive, independent confirmation required for justification.

(3) Greene mounts a more focused empirical attack, targeting all 'characteristically deontological intuitions' (Greene 2014; see also Singer 2005). These are intuitions that, by Greene's own stipulation, typically go along with deontological theories, such as the intuition that lying or breaking a promise in a particular case is wrong even though it has better consequences, or that sacrificing one person as a means to saving five others is wrong. Consider, for example, two of the most famous trolley cases. In *Switch*, the protagonist can save five innocent people from death by diverting a runaway trolley onto a side track where one innocent person is stuck. Many believe it's morally acceptable to sacrifice the one to save the five. Not so, however, for *Footbridge*, in which the protagonist can save five innocents only by pushing a large man off of a bridge down onto the tracks—stopping the trolley, saving the five, but killing the man.

Greene has conducted empirical work which suggests that these intuitions are generated by a quick-and-dirty system that is heavily influenced by emotion. Key to this automatic system are brain regions like the ventromedial prefrontal cortex (with connections to other regions associated with emotion, like the amygdala). These areas apparently give rise to an array of heuristics that evolved to work in situations common long ago in the environment of evolutionary adaptedness when our ancestors developed the ability to cooperate in small groups. But in novel situations such moral heuristics are likely to lead us astray, according to Greene. Especially for the complex and controversial issues that plague our modern world, deontological heuristics are operating out of their element. Characteristically deontological intuitions are therefore unjustified because they are primarily based on an unreliable automatic system unsuited for addressing many moral problems. If a deontological system of moral beliefs is

unjustified, then we should give it up and look for a plausible alternative, which Greene thinks is utilitarianism.[2]

(4) Kelly's target is narrower than Greene's, but it appears to be almost as sweeping: All moral beliefs based on disgust (see also Nussbaum 2004). The consensus among scientists is that disgust arose as a biological adaptation for detecting and avoiding pathogens that cause disease and infection (Rozin et al. 2008; Tybur et al. 2013; see also Strohminger and Kumar, forthcoming). Only later on was disgust 'co-opted' in moral cognition. Ultimately, many people find themselves disgusted by 'impure' behaviour and the violation of taboos, by cheating and hypocrisy and by a number of other moral violations and vices (see Kumar 2017a). However, the mechanisms underlying moral disgust seem to inherit the functionality of its pathogen-oriented precursor. In general, Kelly (2011) argues, it is much better to be oversensitive to germs than undersensitive. As he puts it, disgust has been designed to follow the rule 'better safe than sorry.' That is, disgust is a mechanism that frequently generates false positives, for the sake of minimizing false negatives, and is therefore unreliable. Kelly concludes that all moral beliefs driven primarily by feelings of disgust are unjustified: 'repugnance is simply irrelevant to moral justification' (p. 148).

3. The Debunker's Dilemma

These approaches in naturalistic ethics suggest a general schema for empirical debunking arguments. All of the arguments reviewed above attack one or another class of moral beliefs as unjustified on the grounds that they are based on an epistemically defective process: Evolutionary influences are uncorrelated with moral truth; automatic affective heuristics were once adaptive but no longer reliable; disgust issues in misplaced prohibitions; and framing effects are evidence that moral judgments are susceptible to 'noise.' The process is *epistemically defective* in the sense that it is unreliable, insensitive to evidence, or otherwise yields beliefs that are unjustified or unwarranted (Nichols 2014). As many have pointed out, this is similar to arguments that debunk a belief by arguing that it is based on wishful thinking, guesswork, motivated reasoning, rationalization, or paranoia. A process is typically epistemically defective in ethics if it is a poor indicator of moral rightness or wrongness. This includes general processes like wishful thinking, but it may also include other processes specific to forming moral beliefs, e.g. egocentricity, prejudice, favouritism, jealousy, narrow-mindedness.

Thus, for a given class of moral beliefs, B, and an epistemically defective process, P, a *process debunking schema* can be constructed as follows (compare Kahane 2011; Nichols 2014):

1. B is mainly based on P. (empirical premise)
2. P is epistemically defective. (normative premise)
   So:
3. B is unjustified.

[2] Greene's latest version of his argument does explicitly appeal to something like the debunking strategy we recommend below (see Kumar & Campbell 2012; Greene 2014, p. 713). However, as he recognises, that argument alone (the 'direct route') does not debunk deontological intuitions. He also needs the argument we summarise in the text (the 'indirect route'), which is why it's the argument we attack.

Arguably, one lacks justification for the targeted beliefs only if one is aware of these premises, or perhaps if one should be aware of them (Sinnott-Armstrong 2008). On that assumption, the conclusion applies only to those people who are aware or should be aware. For simplicity's sake, we won't make this explicit in the schema.

Once we understand the structure of debunking arguments, we begin to see their shortcomings. Global debunkers confront a kind of dilemma or predicament, due to a tension between their premises. First, debunkers must identify a process of moral belief formation that is epistemically defective. If we are not confident that the process is genuinely defective, then we cannot use it to challenge moral beliefs about which we are relatively more confident (the normative flaw). The second obstacle is to identify the main basis for belief. If some genuinely defective process is a cause of belief, but it is not the main cause of belief, and the belief is also based on other processes that do not seem defective, the debunking argument is weak (the empirical flaw). It may lower the justification of the relevant moral beliefs, but only to a degree that does not render them unjustified overall. Those convinced by the argument might be pressured to re-examine their moral beliefs, but they need not abandon the beliefs altogether. Each global debunking argument struggles with one of these aspects of the *debunker's dilemma*. It will become clear that this struggle is largely due to a *trade-off*: establishing a plausible empirical premise leads to a corresponding normative premise that is implausible (and vice versa).

3.1 Framing Effects: The Empirical Flaw

Sinnott-Armstrong's (2008) argument, as we've seen, might be extended to show that moral beliefs as a class are unjustified because cognitive biases distort the intuitions on which the beliefs are ultimately based. Like other intuitions, moral intuitions are in general 'subject to framing effects.' However, this phrase is ambiguous, as it leaves the extent of the effect unspecified. It could mean that moral intuitions are only slightly affected by framing effects—e.g. a small proportion of responses change or overall confidence changes to a small degree. What a debunking argument requires, however, is that framing effects alone alter moral beliefs, such that people regularly tend to lose their belief or change its content (compare Shafer-Landau 2008).

The evidence, however, fails to establish that framing effects are a main basis for belief, for two main reasons. First, in the vast majority of studies that Sinnott-Armstrong cites, the mean response to the dilemmas does not change when order or word choice changes. Some experiments report no effect whatsoever. Others find only a slight shift on the same side of the scale of measurement, suggesting that the polarity or valence of the relevant beliefs doesn't tend to change across conditions (May 2018a). Some results do straddle the midpoint, but only barely, suggesting that on average participants were ambivalent anyway and mean responses already lack confidence. This assessment of the evidence fits with a recent meta-analysis that suggests framing effects do not generally exert a substantial influence on moral intuitions. Roughly 80% of people's moral intuitions subject to framing effects do *not* change, and that figure excludes studies that found no effect (Demaree-Cotton 2016). A second issue is that some studies are not well suited to drawing conclusions about moral judgment. For example, some involve tricky dilemmas (e.g. involving probabilities of policies) that are not representative of moral judgment generally. Moreover, researchers do not always measure moral judgment

specifically, asking only which of various policies participants 'would prefer' or which action a participant 'would perform.'

Numerous other experiments suggest that rather different, and seemingly morally relevant, factors are the central determinants of many moral beliefs. In particular, at least for many aspects of morality, judgments change dramatically with intuitively relevant changes in the intentions of the parties involved and the outcomes of their actions (for review, see Young & Tsoi 2013). For example, people overwhelmingly condemn intentional harms but not those brought about accidentally. Consider an example of condemning someone for intentionally and successfully poisoning an innocent co-worker. To our knowledge, no experimental evidence suggests that the valence of this moral judgment is sensitive to wording or order of presentation. There may be a select set of moral beliefs for which framing effects play a substantial role, inducing a reversal of valence. But moral judgment research, taken as a whole, suggests this is a restricted class: Unconfident assessments about tricky dilemmas.

Sinnott-Armstrong may welcome this result, since he admits that one can get independent confirmation that some moral intuitions are reliable. His concern is primarily to attack *moral intuitionism*, which claims that some moral beliefs are justified non-inferentially. Rightly or wrongly, he considers nearly any additional evidence to be independent confirmation (rather than simply undermining the debunking challenge). Our main concern here, however, is with whether framing effects debunk moral beliefs, and careful attention to the relevant empirical studies suggests that they do not. The experimental evidence may make us slightly less confident in some or all of our moral beliefs; we might even have less justification if justification comes in degrees. But awareness of the influence of framing effects does not render one's moral beliefs unjustified.

3.2 Disgust: The Empirical Flaw

Similar issues arise for the appeal to disgust (Nussbaum 2004; Kelly 2011). Plausibly, we're unwarranted in forming moral beliefs based on *incidental* feelings of repugnance, which do not draw one's attention to morally relevant information. Such feelings may influence moral judgment to some extent, but empirical research suggests the effect is ever so slight. Incidental disgust only *sometimes* makes moral beliefs *slightly* harsher, consistently failing to alter the valence of moral judgments, whether concerning moral violations or morally neutral scenarios (May 2014; 2018a). The studies all use fine-grained scales to measure moral judgment. For example, on a 100-point scale (1=not at all morally wrong, 100=extremely morally wrong), the average morality rating for an action by the control group may be 5 while the disgusted group's average rating is 15. The researchers sometimes—in fact, rarely—find that these differences are statistically significant, but that just means, roughly, that the difference is not likely due to chance. On the face of it, however, such small differences are not substantial for moral judgment. Particularly, for our purposes, they do not provide strong evidence that incidental disgust alone can be a main basis for moral belief.

Sometimes an experimental effect is substantial even if it is only a small, but statistically significant, shift on the given scale of measurement. But it matters greatly what is being measured and what questions researchers are attempting to address. For example, relatively small movement on a scale can be substantial if researchers are

measuring rates of infant mortality because any decrease of infant mortality would be important no matter how slight. In experimental research on moral judgment, however, very small shifts on a fine-grained scale are not clearly substantial. The effects of disgust cannot support the empirical premise of a process debunking argument, which seeks a main basis for belief.

So extant evidence suggests only that incidental disgust may sometimes make one think an action is slightly worse, but one judges the action as right or wrong regardless of incidental feelings of disgust. Indeed, a recent meta-analysis confirms this trend, as the size of the effect of disgust on moral judgment is officially 'small' (Landy & Goodwin 2015). Moreover, the authors find that 'the extent to which incidental disgust amplifies moral judgments appears to be overestimated in the published literature, and when this is corrected for, no significant effect is present' (p. 528). While feeling incidental disgust may be a common *consequence* of judging an action immoral (May 2018b), we do not have evidence that it's a main basis for a large class of moral beliefs.

Of course, non-incidental feelings of disgust may have a much stronger influence on moral belief. Recent work suggests that disgust can be flexibly attuned by learning mechanisms and therefore is not generally unreliable in the way that Kelly claims (Kumar 2017a, 2017b). Now we have a plausible empirical premise, but the normative premise consequently suffers. When disgust is non-incidental and carries with it morally-relevant information, it is not generally an epistemically defective basis for one's beliefs. The tension between the debunker's two premises rears its ugly head.

3.3 Evolution: The Normative Flaw

The previous two psychological debunking arguments can establish a plausible normative premise but at the expense of the empirical premise. Both framing effects and incidental disgust seem to be epistemically defective, but neither exert a substantial influence on moral beliefs. Evolutionary debunking arguments, by contrast, can establish a plausible empirical premise but then struggle with their normative premise: we have no good reason to accept that evolutionary forces fail to track the moral truth.

To evaluate the debunker's claim that evolutionary processes fail to track the moral truth, we need some rough, agreed upon conception of human evolution and moral truth. Of course, it is tendentious, to say the least, which is the correct moral theory. But debunkers need only appeal to an uncontroversial aspect of moral truth that we have reason to expect evolutionary processes did not track. In particular, we need a story about how natural selection (the ultimate cause) generated psychological mechanisms (proximate cause) that, along with environmental factors (e.g. cultural transmission), lead humans to form moral beliefs (compare Shafer-Landau 2012). What uncontroversial conceptions of human evolution, environment and moral truth render these in tension?[3]

Our best theories of the evolution of human psychology suggest that our deeply social living conditions required navigating cooperation with others in groups not

[3] Joyce (2006) rests his debunking argument on the claim that Darwinian forces merely explain our having moral *concepts*, rather than tendencies to form particular moral *beliefs* (contrast Street 2006). But this is a threat only if it's problematic that the explanation of our having moral concepts nowhere appeals to the existence of moral properties (or the truth of the moral beliefs in which those concepts figure). We set aside that style of debunking argument in §2.

exclusively comprised of kin. This gave rise to a host of concerns about the well-being of not only oneself but also others, especially those one considers part of one's group. Over time, it was plausibly fitness-enhancing for individuals and their groups to have various altruistic tendencies, compassion for the suffering of others, and importantly concerns about reciprocity, fairness, cheater detection, loyalty, harm or violence reduction, flourishing of one's group and the like (see e.g. Sober & Wilson 1998; Haidt 2012; Henrich 2015; Kumar and Campbell in prep).

Would such a genealogy lead us to form moral beliefs that fail to track the moral truth? One thread in common sense morality and ethical theory is something like the Golden Rule. If anything like it is a core element of moral truth, then Darwinian processes are not necessarily disconnected from the moral facts. For example, if contractualism (or contractarianism) is correct, then it will be far from a coincidence that evolutionary forces (e.g. reciprocal altruism) nudged human moral beliefs to track facts about what rules others can reasonably reject (see e.g. James 2009; Gaus 2011). The same goes for theories, like virtue ethics or an ethics of care, that focus on valuing prudence, justice, loyalty, benevolence, honesty, courage, temperance and so forth. Consider even utilitarianism: Combined with individual reasoning, experience and cultural transmission, our evolved tendencies may well provide a process by which to form justified beliefs about what will tend to promote aggregate happiness, well-being, or preference satisfaction. This is especially plausible for 'indirect' consequentialism, which holds that actions often maximise well-being if they proceed from rule-based reasoning and deliberation prevalent in ordinary moral thought (see Kumar forthcoming). Even many act consequentialists believe, as Greene does, that as 'private individuals we should nearly always respect the conventional moral rules' (2014, p. 717), since this will typically maximise overall well-being.

Thus, given plausible conceptions of our evolutionary history and more proximate causes, there is no reason to think that these forces would lead to beliefs that are sufficiently off track, distorted, or disconnected from moral truth such that our moral beliefs are unjustified (see Campbell 2014). Compare the process of punishing a child when she makes her parent angry. This process alone seems utterly blind to the moral truth. How could angering someone have anything to do with the moral facts? But this process amounts to *moral learning* for the child if combined with the further fact that her parent regularly becomes angry when people lie, cheat, steal, free ride, assault others, treat people unfairly and so on. Of course, on some ethical theories, survival of one's group and oneself can't possibly have any connection to the moral truth, even when combined with exercises of individual experience, reasoning and transmission of cultural wisdom. But then the debunking argument is persuasive only if we accept such implausible conceptions of ethics.

Of course, this might seem question-begging. In this dialectic, evolutionary debunkers often try to help themselves to the assumption that moral truths could turn out to be radically unlike how we ordinarily conceive of them. It could turn out that the morally right thing to do is to stare at the sun all day while blinking every other second. However, then, as Katia Vavova says, 'we have no idea what morality is about. So we have no idea if evolutionary forces would have pushed us toward or away from the truth. So we have no reason to think we are mistaken' (2015, p. 112). Consider the simple case of debunking belief in a god based on wishful thinking. The debunker must assume

something about what would make true the thing that's believed. How could we know whether wishful thinking fails to track the truth about the existence of a god if we don't take a stance on what it would be for that being to exist? If, for example, 'god' simply meant 'whatever makes me think there is meaning in life,' then wishful thinking isn't a defective process. Of course, this is an outlandish proposal of what 'god' means, but the point is not about the meaning of the word but simply that a process can be defective for one kind of belief but not another, and whether it's defective depends in part on some minimal characterization of what's believed. Similarly, the evolutionary debunker must either take a stance on what makes moral beliefs true or show that evolutionary pressures fail to track any account of moral truth worth its weight in salt.[4]

The main problem for evolutionary debunking arguments, then, is simply that they offer no reason to believe that a plausible account of human evolution will lead to unjustified moral beliefs. Evolutionary debunkers take on a hefty burden of spelling out exactly what the evolutionary forces are and why they can't possibly lead to beliefs that are well-grounded or track the moral truth (see Vavova 2014). One might retort that moral beliefs are debunked unless we can show that they aren't based on defective processes (cf. Rini 2016). However, this models debunking arguments on the kinds of sceptical hypothesis arguments one encounters about the external world, where the sceptic merely raises the possibility of error. Debunking arguments are instead tasked with providing evidence of actual unreliability, not merely challenging us to provide justification for our moral beliefs (May 2013). At any rate, our concern is to evaluate the premises of arguments that take on the burden of showing moral beliefs to be based on defective processes. In the case of evolutionary debunking arguments, this burden has not been met: The normative premise is unsupported.

The tension inherent in the debunker's dilemma is on full view here. Evolutionary debunkers could establish a plausible normative premise: It's surely epistemically defective to form one's moral beliefs merely on the basis of what's fitness-enhancing. But then the empirical premise is implausible: We do not form our moral beliefs merely on this basis. While fitness-enhancing considerations are part of the ultimate explanation (cf. Mogensen 2015), the complete proximate explanation appeals also to our concerns about reciprocity, harm reduction, loyalty and so forth. Now we have a plausible empirical premise. What about its corresponding normative premise? Are these considerations morally irrelevant or epistemically defective? Certainly not. Again, an evolutionary debunker might try to charge us with begging the question, for we're making substantive assumptions about moral truths. But, as we've seen, both we and the debunker must do this; and we have not helped ourselves to any contentious assumptions about moral truth. We see the best evolutionary debunking argument as taking on a plausible empirical premise about evolution and its role in moral belief formation. In that

[4] Perhaps any first-order ethical theory is problematic if married with a certain brand of moral realism. Evolution may well provide a particular challenge for anyone who thinks that moral truths are *entirely* independent of how moral agents think, feel and behave. This provides a way to argue that evolution and moral truth are disconnected without making any controversial assumptions about moral truth (compare Street 2006). Our aim, however, is to assess epistemological debunking arguments targeting ordinary moral beliefs, regardless of whether they are construed as presupposing realism of any sort, especially such an extreme form.

case, the major flaw is the argument's normative premise, but it can be improved only at the expense of the empirical premise.

3.4 Automatic Emotional Heuristics: The Normative Flaw

Greene argues that (characteristically) deontological intuitions are influenced by morally irrelevant factors. At times, he suggests that this factor in many cases is *personal force*—we intuitively treat bodily harm as worse just because 'the agent directly impacts the victim with the force of his/her muscles' (2014, p. 709).

We agree that personal force seems morally irrelevant. However, Greene recognises that this factor only substantially influences moral judgments about harm when it interacts with others, namely intention and action (as opposed to omission). For example, flipping a switch, as opposed to pushing, is not only impersonal; it's a matter of harming as a side effect via omission. When being more precise, Greene (2013) unifies these under the heading of *prototypically violent acts* (p. 245ff): We have a 'gizmo' that makes us more likely to condemn a harm if it involves personal and active harm intended as a means to an end (regardless of whether the act maximises utility). This more refined empirical premise, however, identifies a factor that is not obviously morally irrelevant. We may be warranted in distinguishing unintentionally harming by omission, say, from harming another purposefully, actively and in a personal way. Often researchers overlook how such acts relate to the legal notion of battery (or assault), which is an important element of moral and legal reasoning (Mikhail 2014).

Suppose, though, that we grant that it's morally irrelevant whether or not a harm results from a prototypically violent act. This only shows that we are not warranted in distinguishing between pairs of cases on this basis (see Kumar & Campbell 2012). We can't justifiably hold deontological judgments about one set of cases but utilitarian judgments about another similar set, since the experimental research suggests that we don't distinguish between such cases for good reasons. So far, we do not have evidence that we should treat both types of cases as a utilitarian would (or as a deontologist would, for that matter). Consistency only requires that we treat like cases alike: Either count the actions in both sets of cases as wrong (the non-utilitarian resolution) or both as morally permissible (the utilitarian resolution). Compare the (wide-scope) rational requirement to not believe incompatible propositions: Consistency alone only tells us to stop believing in one or both (more on this in Section 4).

Greene (2014) recognises this issue and attempts to tip the scales in favour of utilitarian judgments by arguing that the affective system underlying deontological intuitions is unreliable in the relevant contexts. On Greene's view the affective system is an adaptive module, rigidly suited to a world that no longer exists. It would thus be a 'cognitive miracle' if we 'had reliably good moral instincts' (p. 715) about what he terms 'unfamiliar' moral problems—that is, 'ones with which we have inadequate evolutionary, cultural, or personal experience' (p. 714). Surely we shouldn't trust intuitions about which we have inadequate experience. Now we're back to an eminently plausible normative claim, but as usual the requisite empirical claims become dubious.

The crucial question now is whether the relevant intuitions lack adequate experience with the relevant moral problems. The answer depends on the case. Some deontological intuitions may sometimes be obstacles to forming rational beliefs about how to resolve current crises. Automatic intuitions about property rights, for example,

may hinder us from addressing the yawning wealth gap. But we can accept such a limited critique without impugning other deontological intuitions about, say, autonomy and self-respect that arguably pinpoint the immorality of slavery (regardless of such an institution's effects on overall utility). There is no reason to think that *in general* automatic moral intuitions are inadequately attuned to the problems to which they are typically brought to bear.

Certainly passions sometimes lead us astray, but unconscious emotional processes are often sensitive to good reasons. As Peter Railton has recently argued, an affective system underlying moral intuition is a 'manifestation of underlying competencies and implicit knowledge that cannot readily be brought to mind or articulated' (2014, p. 816). Consider an example that Greene and others think clearly reflects the irrationality of automatic intuition: The case of Julie and Mark, adult siblings who decide to have sex with each other. Many people believe that Julie and Mark act immorally, even though they use protection and even though the one-off encounter does not affect their healthy sibling relationship (Haidt 2012). On Greene's view, this judgment stems from an automatic, emotional aversion to incest that is part of a generally unreliable guide to morality (2014, p. 712). However, as Railton argues, the affective system is plausibly tracking the potential harms that incest regularly inflicts, the threat of which is very real in Julie and Mark's case and averted only by luck. And there is some empirical evidence that most participants do not think actions like Julie and Mark's are harmless (Royzman et al. 2015). This all fits with a growing literature in 'moral learning theory' showing that affective intuitions are flexibly shaped by local material and social conditions (see Kumar 2017b; Cushman et al. 2017). In particular, a network of findings in cognitive neuroscience suggest that sophisticated learning mechanisms attune moral intuition, and development of computational models is well underway (Crockett 2013; Cushman 2013).

Thus, even in what seems the most obvious case in which our moral intuitions are inflexible and untrustworthy, matters are more complicated than would-be debunkers might have hoped. One cannot cast doubt on automatic intuitions on the grounds that they are generally unable to be attuned to today's moral problems through evolutionary, cultural, or personal experience.

3.5 What Explains the Debunker's Dilemma

Global debunkers face a dilemma. Either they do not identify a genuinely defective belief forming process (e.g. natural selection, automatic emotional heuristics). Or they do, but the defective process is not a sufficiently central factor in the genesis or maintenance of moral beliefs (e.g. framing effects, incidental disgust). Either way, these ambitious arguments fail to establish one of the premises of the debunking schema. We have shown that there is a *tension* between the two premises: When one premise is well-established, the other becomes much less plausible.

The lesson to take from all of this is not that experimental research is good for nothing in ethics. Rather, it's that experimental research is not very good at providing a simple and complete story about the sources of all of our moral beliefs or a large, heterogeneous class of them, such that the sources are defective. There is an inherent *trade-off* between targeting a large class of moral beliefs and identifying a defective process that influences them all. In general, moral beliefs are based on many factors,

some legitimate and some not. Experimental research will not reveal a single cause of our moral beliefs, so it will not reveal a single defective cause.

The problem likely arises because it's implausible that there is a single kind of process that both substantially influences a heterogeneous class of beliefs but is also defective across this diverse class. Whether a process is defective depends greatly on the content of the beliefs and how exactly this process influences them. Processes that are plausibly defective are thus especially fit to indict a specific kind of belief, not a large and diverse class. Of course, we can describe a single kind of cause of all moral beliefs or a large class of them, such as evolutionary pressures or affective heuristics. But, as we've seen, such causes will be too general to be uniformly debunking, because whether an influence is epistemically defective depends on how substantially it influences the relevant belief and what other influences are in place. Thus, given the variety and complexity of moral beliefs, it is no coincidence that we find this trade-off yielding a dilemma for global debunking arguments. While we grant the logical possibility of successfully navigating the debunker's dilemma, it is a general problem that will likely plague all wide-ranging debunking arguments.[5]

However, experimental research *is* good for something: finding differences between similar kinds of beliefs. That is, it can reveal not why we hold a large class of moral beliefs, but why our beliefs about similar issues diverge. As we'll see, these sorts of findings inform ethics in a rather different way than many authors have foreseen. Rather than debunking all of morality, or large swaths of it, the findings can offer more focused debunking arguments.

4. Debunking Consistency Arguments

We began this essay by articulating the aim and structure of debunking arguments in ethics. Empirical investigation can explain why we hold the moral beliefs we do and therefore has the potential to expose defective bases for belief. Debunking arguments tend to be broad in scope, but for this reason they face a dilemma. There is, however, a more promising form of debunking argument that is highly selective. To see this we must review a familiar type of moral reasoning and the methodology of experimental moral psychology.

Consistency reasoning is a form of moral and legal reasoning that involves, roughly, treating like cases alike (Campbell & Kumar 2012). Commonly, you try to change someone's mind about one case by arguing that it is no different from another case about which they have a different opinion. For example, Dana thinks that dog fighting is morally wrong and she may try to convince Frank that factory farming inflicts similar harms on animals destined for his dinner plate. Faced with a consistency argument, one has two options. One can change position on one of the cases, or one can attempt to identify a morally relevant difference between the cases that one's interlocutor hasn't noticed. In Frank's case, he can accept that factory farming is wrong, or deny that

[5] Katia Vavova (2014, sect. 8.2) briefly argues that evolutionary debunking arguments in particular are more problematic the more sweeping their targets are. However, our support for this similar conclusion is different. Vavova thinks global debunking is problematic because it entails less common ground between the debunkers and their targets. Our idea is that such arguments succumb to a specific trade-off, and both the empirical details and normative assumptions matter greatly.

dog fighting is wrong, or he might find an important difference between the two. If Frank acknowledges that there is no relevant difference between the two cases and yet continues to treat them differently, his position is inconsistent.

Consistency reasoning is familiar in moral philosophy and in everyday moral discourse too. More importantly, as we'll see, it fits nicely with the methodology and results from experimental psychology. Psychologists do not typically search for the categorical basis for some class of moral beliefs. It would just be too hard, since most moral beliefs are influenced by a large range of factors. What's more experimentally tractable is the search for *difference effects*. Psychologists take a concrete case, manipulate one of its features and then find out whether this affects participants' moral judgments. If it does, then although we can't conclude that the factor is *the* cause of one of the moral beliefs, we can more plausibly conclude that it explains why people respond differently to the two cases.

Experimental research that plumbs difference effects provides an input to consistency reasoning. In particular, it has the potential to cast doubt on the attempt to reconcile apparently conflicting beliefs about cases. If the research shows that what causes us to treat two cases differently is morally irrelevant, upon reflection, then we draw a distinction between them for no good reason. We thus have prima facie reason not to treat the cases differently, and to withhold from accepting principles that justify treating them differently (Kumar & Campbell 2012). The immediate conclusion of this sort of debunking argument is that we ought to treat like cases alike. However, if we are antecedently more committed to one of the beliefs, then we should revise the other. For example, suppose Frank discovered that he has different moral beliefs about dog fighting and factory farming only because of the familiarity of dogs and the unfamiliarity of factory farm animals. Since familiarity is morally irrelevant in this context, the immediate conclusion is that he should believe either that both dog fighting and factory farming are wrong or that neither is (if he believes anything). However, if Frank has more independent reason to believe in the wrongness of dog fighting than in the permissibility of factory farming, then it seems he should revise his belief about the latter.

In Section 3 we laid out a schema for global debunking arguments. We argued that it is difficult to fill out the schema so that both the empirical premise and the normative premise are true. A schema for consistency debunking is more promising. The target is a pair of moral beliefs, each of which concern a concrete case. The cases are similar, but one difference between the two explains why we form different moral beliefs about them. We can construct a schema for *Debunking Consistency Arguments* as follows:[6]

1. D is the main basis for why subject S has two different moral beliefs about similar cases. (empirical premise)
2. D is a morally irrelevant difference. (normative premise)

So,

3. S is unjustified in holding one or both moral beliefs.

Like all good debunking arguments, a normative premise is required, and the argument is successful only if the normative premise is more plausible than the moral beliefs being targeted. That is, the argument is compelling if we can make an assessment of

[6] This schema is inspired by Kumar & Campbell (2012, p. 322), but improves on it.

comparative confidence. The conclusion is that the pair of beliefs are together unjustified, and therefore at least that one of them must be abandoned. Let's illustrate with an example grounded in experimental research.

Judges, of course, are supposed to make decisions based solely on the merits of the case at hand. However, a recent study by Shai Danziger and colleagues (2011) provides evidence that judges' parole decisions are based in part on whether they have eaten recently or are hungry. After a meal, judges in their study were likely to grant parole to roughly 65% of the applicants appearing before them. However, just before a meal the number of applicants granted parole is close to 0%. One difference that explains judges' beliefs about merited parole, or anyway their decisions about parole, is whether or not they are hungry (or have the negative feelings associated with hunger). Of course, this difference is utterly irrelevant. The conclusion, then, is that the judges are not justified in believing that candidate A (before lunch) doesn't deserve parole but candidate B (after lunch) does. Popular commentary on this study often infers that the judges are too harsh when they are hungry. However, our schema helps make clear that this is an illicit leap from the empirical findings. What is warranted, strictly speaking, is only the conclusion that either the judges are too harsh when they are hungry *or* that they are too lenient when they are sated. That is, the judges should treat like cases alike, and either grant or deny parole in similar cases, not differentiate them based on how irritable they are feeling at the time of decision.

Of course, an empirical debunking argument is not strong if based only on one study; the empirical premise would not be adequately supported. Ideally, the empirical premise would be supported by numerous laboratory and field experiments, carried out by different researchers, which produce data that converge and cohere with other (perhaps non-experimental) research on the topic. If we did acquire overwhelming evidence that there is an illicit influence on parole decisions, then the empirical debunking argument of this range of verdicts gains considerable power.

Now this case targets the parole decisions of judges, but much psychological research aims to draw conclusions about people generally. Let's turn to another, more controversial application of the schema for consistency debunking (Kumar and Campbell 2012). Peter Singer (1972) famously argues that if we think it is morally obligatory to save a drowning child in a pond, then we should also think it is obligatory to save starving children in developing countries. Experimental work by Jay Musen (2011) suggests that we treat these cases differently merely due to physical distance: the drowning child is near, starving children far. If so, we think differently about the two cases for what most of us agree is no good reason, and we should withhold from treating them differently. So far, this is just a debunking consistency argument. Singer of course aims to go further. If it is more plausible that we should save the child than that we may ignore starving children, then Singer's final conclusion gains support: We should treat helping in both cases as obligatory.

Consistency debunking arguments can also have important consequences for moral principles supported by pairs of intuitive judgments. Consider, for example, the Doctrine of Double Effect, a core element of which is the claim that harming as a means is worse than harming as a side-effect. Like many ethical principles, Double Effect has been defended by appeal to intuitions about contrasting cases, such as the Switch and Footbridge cases. Proponents of Double Effect argue that the principle can capture the

contrasting intuitions about such cases, since Footbridge plausibly involves harming the man as a means to an end while Switch does not. However, the principle is unsupported if belief in it is defended ultimately by appeal to ordinary intuitions and intuitions about the cases in fact diverge for morally irrelevant reasons, as Greene (2013; 2014) argues. Belief in the principle is then unjustified, and precisely because consistency reasoning challenges holding the pair of judgments on which the principle rests—and all without taking a stance on what intuitions we should have about such pairs of cases other than that they should be the same. Now, it's controversial whether intuitions apparently supporting Double Effect are driven by irrelevant factors (Mikhail 2011; Feltz & May 2017). But the point is that debunking consistency arguments can be a powerful weapon to wield throughout moral philosophy and a potential source of rational change in moral belief.

Debunking consistency arguments are not always straightforward. The relevant difference-making factor might be misidentified if described at the wrong level of explanation or in overly simplistic terms. For example, suppose we treat as morally acceptable employing affirmative action policies for blacks but not whites. One way to describe the difference-maker in our moral judgments might be merely skin colour, which might seem morally irrelevant. But presumably our judgments differ here not just in terms of skin colour *as such*. In fact, describing the relevant factor as race might be insufficient, since the relevant difference between the judgments seems best explained by the different histories of treatment faced by each group or the challenges they are likely to confront in a given society. What might seem like a morally irrelevant difference (skin colour) might be more relevant if properly described (historical mistreatment or susceptibility to discrimination). Since we are dealing with mental phenomena, it is absolutely imperative that we properly describe the relevant factor and from the individual's perspective.[7]

When well-constructed, debunking consistency arguments can be powerful. While consistency reasoning has a long history in moral philosophy, what's novel is harnessing the power of empirical, especially experimental, research to amplify their force. Finding difference effects from the armchair is difficult. Introspection is limited in its ability to determine commonsense intuitions and the unconscious influences on them. Moreover, mountains of converging empirical evidence are difficult to ignore. Consider again the research on implicit racial and gender biases. Perhaps armchair reflection could spot such biases, but no one could have predicted how pervasive they are in the absence of scientific evidence. Opinions about police brutality, the criminal justice system and hiring decisions have been radically changed for those aware of the overwhelming evidence of implicit biases.

[7] Compare Singer's (2005, p. 348) claim that there is no morally relevant difference between killing someone in a way that was 'possible a million years ago' (e.g. personal assault) and doing so in a way that 'became possible only two hundred years ago' (e.g. an impersonal drone strike). This seems like a morally irrelevant difference in the abstract, but perhaps not if it corresponds to killing by refusing to share limited food (an omission with no assault) versus killing by advanced torture techniques (e.g. active assault). Morally relevant differences can too easily be redescribed so that they appear morally irrelevant.

5. Conclusion

In general, global debunking arguments in ethics share a common ambition, and for that reason face the debunker's dilemma. Targeting a large set of moral beliefs involves the daunting task of identifying general processes that are defective across the board and substantially influence a motley set of attitudes. The examples of evolution, emotion and framing effects are general processes to be sure, but they either hardly influence moral beliefs or aren't necessarily defective. In general, our moral beliefs and their causes are too diverse to hope for a sweeping evaluation of them.

A more promising kind of debunking argument is highly selective and relies on moral consistency reasoning along with experimental findings. Emerging empirical research is especially suited to debunking relevantly similar pairs of moral beliefs, along with any general moral distinctions based upon them. Pending independent support for the beliefs, we should treat like cases alike. This approach offers our best hope for debunking moral beliefs empirically.

Acknowledgements: For helpful feedback, we thank: Matthew Braddock, Richmond Campbell, Andrew Cullison, Matt King, Dustin Locke, Theresa Lopez, Bryan Lueck, R. Campbell Mackenzie, Colin Marshall, Kevin McCain, Andrew Moon, Shaun Nichols, Walter Sinnott-Armstrong, Robin Zheng, Aaron Zimmerman, and audiences at the Alabama Philosophical Association, the Midsouth Philosophy Conference, Oakland University, the Southern Society for Philosophy and Psychology, and the University of Birmingham. Kumar completed some of this work during a post doc funded by the John Templeton Foundation, and May developed some of the ideas during two summer workshops on moral epistemology: one at the Central European University in Budapest and one at the Prindle Institute for Ethics at DePauw University. We are grateful for their support.

References

Brownstein, M. (2015) 'Implicit Bias', in E. N. Zalta, Ed., *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), URL = http://plato.stanford.edu/archives/spr2015/entries/implicit-bias.

Campbell, R. (2014) 'Moral Epistemology', *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), E. N. Zalta (Ed.).

Campbell, R. & Kumar, V. (2012) 'Moral Reasoning on the Ground', *Ethics* 122 (2):273-312.

Clarke-Doane, J. (2015) 'Justification and Explanation in Mathematics and Morality', In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics Vol. 10*. Oxford University Press.

Crockett, M. (2013) 'Models of Morality', *Trends in Cognitive Sciences* 17(8):363-6.

Cushman, F. (2013) 'Action, Outcome, and Value: A Dual System Framework for Morality', *Personality and Social Psychology Review* 17(3):273-92.

Cushman, F, Kumar, V., & Railton, P. (2017) 'Moral Learning: Psychological and Philosophical Perspective', *Cognition* 167: 1-10.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011) 'Extraneous Factors in Judicial Decisions', *Proceedings of the National Academy of Sciences* 108(17):6889-6892.

Demaree-Cotton, J. (2016) 'Do Framing Effects make Moral Intuitions Unreliable?', *Philosophical Psychology* 29(1): 1-22.

Feltz, A. & May, J. (2017) 'The Means/Side Effect Distinction in Moral Cognition: A Meta-Analysis', *Cognition* 166: 314–327.

Gaus, G. (2011) *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World* (Cambridge University Press).
Greene, J. D. (2013) *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin Press.
Greene, J. D. (2014) 'Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics', *Ethics* 124(4): 695-726.
Haidt, J. (2012) *The Righteous Mind*. Pantheon Books.
Harman, G. (1977) *The Nature of Morality* (New York: Oxford University Press).
Henrich, J. (2015) *The Secret of Our Success*. Princeton University Press.
James, S. M. (2009) 'The Caveman's Conscience: Evolution and Moral Realism', *Australasian Journal of Philosophy* 87(2):215-233.
Joyce, R. (2006) *The Evolution of Morality*. MIT Press.
Kahane, G. (2011) 'Evolutionary Debunking Arguments', *Noûs* 45(1):103-125.
Kelly, D. (2011) *Yuck!: The Nature and Moral Significance of Disgust*. Cambridge, MA: MIT Press.
Kumar, V. (2017a) 'Foul Behavior', *Philosophers' Imprint* 17(15): 1-16.
Kumar, V. (2017b) 'Moral Vindications', *Cognition* 167:124-134.
Kumar, V. (Forthcoming) 'Empirical Vindication of Moral Luck', *Nous*.
Kumar, V. & Campbell, R. (2012) 'On the Normative Significance of Experimental Moral Psychology', *Philosophical Psychology* 25 (3):311-330.
Kumar, V. & Campbell, R. (In prep) *Why We Are Moral.*
Landy, J. F. & Goodwin, G. P. (2015) 'Does Incidental Disgust Amplify Moral Judgment? A Meta-analytic Review of Experimental Evidence', *Perspectives on Psychological Science* 10(4): 518-536.
Lillehammer, H. (2003) 'Debunking Morality', *Biology and Philosophy* 18(4):567-581.
May, J. (2013) 'Skeptical Hypotheses and Moral Skepticism', *Canadian Journal of Philosophy* 43(3): 341–359.
May, J. (2014) 'Does Disgust Influence Moral Judgment?', *Australasian Journal of Philosophy* 92 (1): 125-141.
May, J. (2018a) *Regard for Reason in the Moral Mind*. Oxford University Press.
May, J. (2018b) 'The Limits of Appealing to Disgust', *The Moral Psychology of Disgust*, Nina Strohminger & Victor Kumar (eds.), Rowman & Littlefield.
Mikhail, J. (2011) *Elements of Moral Cognition*. Cambridge University Press.
Mikhail, J. (2014) 'Any Animal Whatever? Harmful Battery and its Elements as Building Blocks of Moral Cognition', *Ethics* 124(4): 750-786.
Mogensen, A. L. (2015) Evolutionary debunking arguments and the proximate/ultimate distinction. *Analysis* 75(2): 196–203.
Musen, J. (2010) 'The Moral Psychology of Obligations to Help Those in Need', Honours thesis, Harvard.
Nichols, S. (2014) 'Process Debunking and Ethics', *Ethics*, 124: 727-49.
Nosek, B., Greenwald, A., & Banaji, M. (2007) 'The Implicit Association Test at Age 7: A Methodological and Conceptual Review', in J. A. Bargh, Ed., *Automatic Processes in Social Thinking and Behavior* (Philadelphia: Psychology Press).
Nussbaum, M. C. (2004) *Hiding from Humanity: Disgust, Shame, and the Law*. Princeton, NJ: Princeton University Press.
Railton, P. (2014) 'The Affective Dog and its Rational Tale: Intuition and Attunement', *Ethics*, 124: 813-59.
Rini, R. A. (2016) 'Debunking debunking: a regress challenge for psychological threats to moral judgment', *Philosophical Studies* 173(3): 675–697.
Rosenberg, A. (2011) *The Atheist's Guide to Reality: Enjoying Life without Illusions* (Norton).

Royzman, E. B., Kim, K., & Leeman, R. F. (2015) The curious tale of Julie and Mark: Unraveling the moral dumbfounding effect. *Judgment and Decision Making*, 10(4), 296–313.

Rozin, P., Haidt, J., & McCauley, C. (2008) 'Disgust', in M. Lewis, J. Haviland-Jones, * L. Barrett, Eds., *Handbook of Emotions* (Guilford Press).

Ruse, M. (1986) *Taking Darwin Seriously*. New York: Blackwell.

Shafer-Landau, R. (2008) 'Defending Ethical Intuitionism', W. Sinnott-Armstrong (ed), *Moral Psychology, Vol. 2*, MIT Press.

Shafer-Landau, R. (2012) 'Evolutionary Debunking, Moral Realism and Moral Knowledge', *Journal of Ethics & Social Philosophy* 7(1):1-37.

Singer, P. (1972) 'Famine, Affluence, and Morality', *Philosophy and Public Affairs* 1 (3):229-243.

Singer, P. (2005) 'Ethics and Intuitions', *Journal of Ethics* 9 (3-4):331-352.

Sinnott-Armstrong, W. (2008) 'Framing Moral Intuitions', W. Sinnott-Armstrong (ed), *Moral Psychology, Vol. 2*, MIT Press.

Sober, E. & Wilson, D. S. (1998) *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Street, S. (2006) 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies* 127(1):109-166.

Strohminger, N. & Kumar, V., Eds. (Forthcoming) *The Moral Psychology of Disgust* (Rowman & Littlefield).

Tybur, J., Lieberman, D., Kurzban, R., & DeScioli, P. (2013) 'Disgust: Evolved Function and Structure', *Psychological Review*, 120 (1): 65-84.

Vavova, K. (2014) 'Debunking Evolutionary Debunking', *Oxford Studies in Metaethics* Vol. 9, OUP, pp. 76-101.

Vavova, K. (2015) 'Evolutionary Debunking of Moral Realism', *Philosophy Compass* 10(2):104-116.

White, R. (2010) 'You Just Believe that Because….', *Philosophical Perspectives* 24(1):573-615.

Wielenberg, E. J. (2010) 'On the Evolutionary Debunking of Morality', *Ethics* 120(3):441-464.