ARTICLE

Can Rational Reflection Save Moral Knowledge from Debunking?

Noah McKay

University of Edinburgh, Edinburgh, Scotland
Email: n.d.mckay99@gmail.com

Abstract

In this essay, I reply to an influential objection to evolutionary debunking arguments against moral realism. According to this objection, our capacity for autonomous rational reflection allows us to grasp moral truths independently of distorting evolutionary influences, so those influences do not prevent us from having moral knowledge. I argue that rational moral reflection is not, in fact, autonomous from evolutionary influences, since it depends on our evolved, pre-reflective grasp of moral properties. I then consider and reject the suggestion that realists can supply an autonomous foundation for rational moral reflection or do without any such foundation. Next, I address the allegation that my arguments have skeptical implications for rational reflection in non-moral domains. Finally, I conclude with a gesture toward a more promising route for realists who oppose debunking arguments.

1. Evolutionary debunking arguments

Many philosophers contend that evolutionary theory casts doubt on our ability to know objective moral truths. They reason, in rough outline, like this: our moral-belief-forming capacities are, like our other cognitive capacities, the products of evolution by natural selection. And natural selection favors traits – including predispositions to form certain beliefs – that maximize differential reproductive success. But whether moral beliefs are *true* seems irrelevant to whether they maximize differential reproductive success. Consider, for instance, the belief that caring for one's young is morally good. Natural selection favors this belief because animals who hold it leave more healthy offspring than those who do not, but that has nothing to do with whether caring for one's young *actually is* morally good. This belief is adaptive because it motivates animals to look after their genes, not because it accurately represents moral facts (even if it happens to be true). So, natural selection is blind to moral truth – it does not discriminate between true moral beliefs and false ones. But that means it would be an incredible coincidence if our moral-belief-forming capacities evolved in a way that

enabled us to reliably discern moral truths. Since such a coincidence probably has not occurred, we lack justification for all of our moral beliefs (Cofnas 2022; Joyce 2006, 2016; Kahane 2011; McKay 2023; Morton 2016; Ruse 1986, 2006; Street 2006, 2008).

In the pages that follow, I will defend this kind of argument – known as an "evolutionary debunking argument," or an "EDA" – against a formidable objection. But before I do so, several caveats are necessary. First, EDAs are considered by most to pose a unique challenge to moral realism. Realists hold (again, roughly) that moral judgments express beliefs, that these beliefs are often true, and that their truth is an objective matter, in the sense that it is (in relevant respects) independent of the psychological attitudes of actual or hypothetical agents (Cuneo 2007: 20–51). EDAs are supposed to show that, if this is the right account of moral judgments, we cannot have moral knowledge. In this paper, I will tailor my arguments to moral realism, and I will only consider objections that I think a moral realist would raise.[1]

Second, the genealogical story I outlined in the opening paragraph is obviously oversimplified. Even according to debunkers, natural selection does not favor some beliefs over others *directly* – beliefs themselves are not units of selection. Rather, natural selection favors predispositions to form certain moral judgments, or, to put it another way, biases toward certain moral judgments (Dale 2022; Street 2006: 118–20). These predispositions usually do not issue in moral beliefs independently of socialization. Furthermore, some of our moral predispositions may be almost entirely the products of cultural influences, rather than innate biology (Arvan 2020, 2021; Braddock 2021; Kitcher 2011; Levy and Levy 2020). However, though biological evolution does not *determine* the contents of our moral beliefs, it constrains them heavily, e.g. by constraining our innate biases, our conceptual schemes, and our forms of cultural expression. After all, human thought and culture are ultimately manifestations of human biology, albeit indirect and flexible ones.[2] And if the contents of our moral beliefs have been heavily constrained by factors that are blind to moral truth, that is plenty of cause for worry. Moreover, there is reason to doubt that other, non-Darwinian forces at work in the evolution of moral thinking, such as exaptation and cultural evolution, are likely to be truth-tracking. So, EDAs do not depend for their cogency on a strong link between evolved biology and moral beliefs. (I will address this in more detail in §3.)

Third, EDAs differ in the sense in which they allege that evolution is blind to moral truth. Some versions say that our moral beliefs are insensitive in the modal-epistemic sense: in other words, because of how evolution works, we would still hold the same moral beliefs we actually hold (e.g. "Caring for one's young is morally good") even if they were all false (e.g. even if caring for one's young were morally evil or indifferent).[3] Others say that our moral beliefs are unsafe in the modal-epistemic sense – in other

---

[1]For reasons to think that some versions of anti-realism are in similar trouble, see Tropman (2014).

[2]I am not suggesting that thought and culture are *reducible* to biology. My point is that the way we are cognitively and culturally is, to a great extent, due to the way we are biologically.

[3]It is not perfectly clear that this version of the argument has yet been espoused by any particular debunker, though Street (2006: 132) gestures toward it, as do others occasionally. This is likely because moral naturalists, like Sturgeon (1985), have made it famously difficult to sustain the insensitivity charge when it comes to moral beliefs. In part of my (2023), I appeal to a kind of conceptual insensitivity rather than metaphysical insensitivity. More precisely, I argue that our moral intuitions are invariant across all conceptually possible worlds which are identical to the actual world with respect to natural properties but different from it with respect to moral properties. I then argue that this shows that moral truth was causally irrelevant to the development of our moral intuitions, since counterfactual reasoning about conceptual possibilities is integral to establishing causal dependencies, *even when those conceptual possibilities*

words, because of the way evolution works, we easily could have been wrong about the moral truths. If we had evolved to be more like bees, for example, we would all have been collectivists who felt duty-bound to kill unproductive males (Ruse 1986; cf. Braddock 2021). Yet others say that moral facts played no causal role in the development of our moral beliefs, or, perhaps more modestly, that we can explain all of our moral beliefs in terms of evolutionary forces, without appealing to any moral facts at all (Joyce 2006, 2016; Street 2006). Despite their differences, these formulations are all intended to show that evolution is, in some knowledge-undermining way, disconnected from moral facts.[4] The objection to EDAs I will consider below, and my rejoinders to it, are general enough to apply to any of these versions, as far as I can tell.

Fourth, unlike many familiar skeptical arguments, EDAs do not impose stringent requirements for epistemic justification. In fact, EDAs are consistent with extremely permissive epistemic standards – debunkers may grant that moral beliefs are *prima facie* justified merely because they seem, on reflection, to be true. The real problem, according to EDAs, is not that realists are unable to offer justification for their moral beliefs, but that they have *undercutting defeaters* for those beliefs. An undercutting defeater is a belief that undermines one's (otherwise adequate) justification for another belief: for example, my belief that I am red-green color blind is an undercutting defeater for my belief that the wall before me is green, since it undermines my (otherwise adequate) justification for that belief, namely that the wall looks green to me. Debunkers claim that realists' beliefs about evolution undercut their *prima facie* justification for their moral beliefs: though it is entirely reasonable, under normal circumstances, for realists to be confident in their moral judgments, maintaining such confidence is *un*reasonable once they realize that evolution is blind to moral truth. Notice that realists cannot defeat this defeater merely by appealing to their moral intuitions, any more than a color-blind person could reinstate her color beliefs by appealing to her color vision (see Vavova 2015).

Fifth, realists have raised a staggering number of objections to EDAs, most of which are beyond the scope of this essay (see Wielenberg 2016 for an excellent overview). My present (modest) aim is to address just one of these objections, in my view the most important one. According to this objection, our capacity for autonomous rational reflection allows us to grasp moral truths independently of distorting evolutionary influences, so those influences do not prevent us from having moral knowledge (Fitzpatrick 2014, 2015, 2016, 2017). This objection is important for two reasons. First, it is highly plausible at first glance, given the role that reasoning often plays in moral evaluation, and given that reason *does* seem to allow us to achieve autonomy from our evolutionary predispositions in a range of cases. Second, it is broadly appealing, since it is independent of many of the methodological[5] disputes that dominate the relevant literature and applies with equal force to most (or all) extant EDAs.

---

are not metaphysical possibilities. So my preferred version of the argument is a kind of amalgam of the first and third formulations.

[4]See Shafer-Landau (2012) for a more detailed attempt to tease out the various versions.

[5]I call these disputes methodological because they are mostly about whether debunkers are within their epistemic rights to make certain dialectical moves, rather than about whether the substantive premises of EDAs are true or false. Mostly, these disputes are about whether debunkers or their opponents are begging the question, or about whether debunkers are allowed to cite natural selection as an explanation of moral beliefs, or about whether it is reasonable for debunkers to expect realists to bracket their moral assumptions while considering EDAs.

Nevertheless, I will try to show that the objection fails. In §2, I will show that rational moral reflection is not, in fact, autonomous from evolutionary influences, since it depends on our evolved, pre-reflective grasp of moral properties. In §3–4, I will consider – and reject – the suggestion that realists can supply an autonomous foundation for rational moral reflection or do without any such foundation. In §5, I will turn to the allegation that my arguments have skeptical implications for rational reflection in non-moral domains. Finally, in §6, I will conclude with a gesture toward a more promising route for realists who oppose EDAs.

2. Autonomous rational reflection

Though it may not be obvious, EDAs depend on a rather strong claim about the explanation of our moral beliefs, namely that evolution has universally shaped the contents of those beliefs in some deep and ineliminable way. It is not enough that evolution merely played *some* role in fixing these contents. For, as Guy Kahane (2011) points out, practically *all* justified beliefs are at least partly explained by factors that are blind to their truth values. For example, my belief that I am sitting at a red table is partly explained by the fact that the lights are on, and although this fact is blind to the color of the table (the lights would be on *whatever* color the table was), my belief is still justified. This is because the content of my belief is shaped by other processes that are sensitive to its truth value. Thus, in order to defeat realists' justification for their moral beliefs, debunkers must show that evolution has influenced or constrained the contents of those beliefs in such a way as to prevent other, truth-sensitive factors from guiding their moral judgments toward the truth.

William Fitzpatrick argues that realists can and should reject this "Extreme Explanatory Claim" (2014: 247) and instead affirm that our moral beliefs are principally the upshots of autonomous rational reflection – that is, "thinking that transcends the micromanaging influences of natural selection in the distant past, proceeding independently of such evolutionary shaping of the content of our thinking, following standards internal to developed methods of inquiry" (2014: 243). According to Fitzpatrick, we can train ourselves, through careful moral reflection and instruction, and using the raw cognitive materials provided by evolution, to grasp moral truths by grasping the *reasons why* certain actions are morally right or wrong – that is, by grasping their right-making or wrong-making properties. It is our grasp of independent moral facts as such that explains the contents of our moral judgments:

> If asked why we believe that [the Taliban's] practices are morally wrong we will cite *reasons* that we take to support the *truth* of the belief, not merely psychological or sociological causes for it that operate independently of such reasons. Such practices, we'll point out, are unjust, cruel, demeaning, and sexist, violating human rights and dignity by depriving these girls of central human capabilities and goods, based on arbitrary considerations. We cite these reasons as *wrong-making features* of these practices, against our background view of the *standards of moral excellence* for human beings and action. And in cases like this what is plausibly happening is that we are *correctly* grasping *the wrong-makingness* of these factors, and this is precisely what leads to our moral judgment: we believe the Talibanic practices to be wrong *because* they *are* wrong and, being morally competent, we've *recognized* this evaluative fact by grasping the *reasons* why they're wrong, as such. And that is to say that the moral properties and facts come straightforwardly into the explanation of such moral beliefs. (Fitzpatrick 2015: 896)

Thus, even if evolution partly explains the contents of our moral beliefs, it does not do so exhaustively: a complete explanation of our moral beliefs will cite our reasons for holding those beliefs. And even if evolutionary pressures have had some distorting influence on our moral thinking, rational reflection enables us to achieve *autonomy* from those influences, in the same way that it (hopefully) enables us to transcend our natural cognitive limitations when it comes to mathematics, logic, and philosophy.

Fitzpatrick is certainly right to point out that our moral beliefs are not exhaustively explained by evolutionary processes. Having a moral belief is not like being two-legged or warm-blooded. But no debunker, to my knowledge, has claimed otherwise. Debunkers recognize that moral beliefs are significantly shaped by rational reflection (along with a host of other developmental and cultural factors), but they contend that such reflection depends ultimately on *pre-reflective* evaluative tendencies shaped by evolution. More precisely, moral reflection depends on our pre-reflective dispositions to form moral concepts – like <good>, <bad>, <worthy>, <guilty>, etc. (Joyce 2006, 2016; Morton 2016) – and to apply those concepts to certain actions and agents (Dale 2022; Street 2006). Together, these dispositions constitute what I will call our "pre-reflective grasp of moral properties," or, more concisely (if slightly less appropriately[6]), our "moral intuitions." The challenge EDAs pose to our pre-reflective grasp of moral properties is twofold: first, given their evolutionary origins, our moral concepts are unlikely to pick out any mind-independent moral properties; and, second, given their evolutionary origins, our dispositions to apply those concepts are unlikely to correspond to the actual distribution of mind-independent moral properties (if there are any such properties).

This presents a difficulty for Fitzpatrick's view: in order to engage in rational moral reflection, we must, on pain of regress, rely ultimately on our pre-reflective grasp of moral properties. But this means that rational moral reflection is *not* in fact autonomous from evolutionary influences if, as debunkers allege, that pre-reflective grasp has been shaped by evolution.

In fewer words, the problem is this: We cannot begin to reflect about whether an action is right or wrong unless we *already have* some notion of what makes actions right or wrong. And where would this notion come from, if not from evolution?

It is not clear how Fitzpatrick would answer this question. In the passage quoted above, he seems to suggest that we identify properties as right-making or wrong-making by setting them again a full-blown set of background *beliefs* about standards of moral excellence. But how do these background beliefs originate? Presumably, Fitzpatrick would want to say that they are the upshots of autonomous rational reflection. This is, after all, his reason for claiming that beliefs like these are safe from debunking in the first place. But this is a non-starter, since it would make the beliefs in question items of reflective moral knowledge, rather than deliverances of moral intuition. Moreover, this answer betrays a vicious kind of explanatory circularity, since it appeals to rational moral reflection in order to explain how rational moral reflection is possible.

---

[6]This term is most often used to denote pretheoretical moral *beliefs*, or, if not full-blown beliefs, intellectual seemings of some sort (see Huemer 2008: 370–1). I am using it in a broader way to denote, not just those propositional attitudes, but dispositions to have them and the conceptual materials necessary to form them. Though I will continue to say that rational moral reflection depends on moral intuition, my arguments do not require that each of us has a stock of pre-given beliefs about moral norms on which we draw when forming reflective moral judgments. Though I think moral reflection sometimes works that way, it needn't always.

Elsewhere, Fitzpatrick suggests that we encounter moral properties through "ongoing evaluative experience," and it is this experience that provides fodder for moral reflection (2014: 245). How does evaluative experience put us in touch with moral properties? Fitzpatrick's answer seems to be that it is emotionally laden (2015: 902). But surely our *emotional* dispositions are a part of our evolutionary heritage if anything is, so this answer fails to preserve the autonomy of rational reflection from evolution. If Fitzpatrick were to reply that we can achieve emotional autonomy from evolutionary influences through "the right kind of emotional development and training" (2015: 902), we could justly ask: How do we know which kinds of emotional development and training are the *right* ones? Where could this knowledge come from, if not from moral intuition? If we have no hold on the moral truth *prior to* our pre-reflective grasp of moral properties, how could we know whether our grasp has improved?

Perhaps Fitzpatrick would reply that we can correct our moral intuitions *post-hoc* once we have arrived at a mature moral understanding, just as we often correct our folk-scientific and folk-philosophical intuitions as our theoretical understanding matures. Considered on its own, this suggestion is plausible – at the very least, those of us who believe that moral progress is possible had better hope it is true! Notice, however, that this response does nothing to establish the *autonomy* of rational moral reflection from evolution. If such *post-hoc* revision of our moral intuitions is possible, it depends on higher-level moral beliefs *which in turn* depend on the very intuitions they are invoked to revise. Though we could, perhaps, put some distance between ourselves and our evolved moral intuitions this way, we could not achieve independence from them. And if our evolved moral intuitions are radically misleading, mere distance will not be enough.

At this point, one might ask why the *autonomy* of rational reflection is necessary for the success of Fitzpatrick's account. Why not hold that rational reflection based on the moral intuitions with which evolution has actually furnished us can lead us to the truth? The answer is that if our moral intuitions are misleading – that is, if what we pre-reflectively *take to be* good moral reasons are not, in fact, good moral reasons – then trying to evaluate the reasons for our moral judgments is futile. We will just end up producing *more bad reasons* for those judgments. To put the point a slightly different way, if the premises of our moral reasoning are suspect, then rational moral reflection is simply the business of deriving dubious moral conclusions from dubious moral premises. We have what computer scientists call a "garbage-in-garbage-out" (GIGO) problem (cf. Street 2006: 123–4; Schafer 2010: 474–5).

Perhaps an analogy will clarify this worry. Suppose a cadre of incompetent lab assistants presents a scientist with an inaccurate batch of experimental data. Suppose, further, that when the scientist discovers he is working with a bad batch, one of his assistants reassures him, "Not to worry – true, the data is quite inaccurate. But I have just read a spectacular essay by a philosopher who thinks that rational reflection allows us to transcend distortions in our most basic empirical beliefs and concepts. So if you reflect carefully on the data, I have no doubt you will come to the right conclusions." The obvious problem with the lab assistant's proposal is that no amount of careful reflection will turn bad data into good data. In the same way, no amount of careful reflection will turn a bad batch of moral intuitions into a good batch of moral beliefs. True, perhaps our imagined scientist could get *something* out of his bad data-set by relying heavily on empirical knowledge he had acquired elsewhere. But if rational moral reflection is not autonomous, there is no "elsewhere" from which to draw such corrective knowledge in the moral case: if, as I have argued, all of our moral thinking

is dependent on moral intuition, then if evolution has set our moral intuitions on the wrong track, it has set *all* our moral thinking on the wrong track.

Fitzpatrick alleges that the GIGO objection is question-begging against his view, since it assumes precisely what is in dispute, namely that our moral beliefs are mostly "garbage" to begin with (2015: 900). However, this is simply mistaken: GIGO does not depend on the assumption that our moral beliefs are mostly misleading. Rather, it depends on the much weaker claim that the *inputs* to rational moral reflection – namely, our *pre-reflective* moral notions – are likely misleading. This begs no questions against Fitzpatrick's view, since his view concerns the *outputs* of rational moral reflection – namely, our *reflective* moral beliefs. GIGO is intended to *show* that our reflective moral beliefs are likely misleading, but it does not assume this. Of course, realists can resist GIGO by denying that rational reflection must take our evolved moral intuitions as inputs. But that shows that the success of Fitzpatrick's account does, after all, hinge on whether rational moral reflection is autonomous from those intuitions. And, as I argued above, this is doubtful.

Let us take stock: the problem I have outlined thus far is that evolution has heavily shaped our pre-reflective grasp of moral properties, according to EDAs. Since rational moral reflection must, on pain of regress, depend on this pre-reflective grasp somewhere down the line, it is not autonomous from evolutionary influence, *contra* Fitzpatrick. Thus, the contents of our reflective moral beliefs have likewise been shaped by evolution in a deep and ineliminable (even if indirect) way. So, there is little hope so far that an appeal to rational moral reflection will free realists from debunking worries.

All hope is not lost, however. There are two ways in which realists may yet establish that rational moral reflection is autonomous. First, they can try to supply a foundation for moral reflection that is itself autonomous from evolution by identifying some subset of our moral intuitions that has been insulated from evolutionary distortion. Second, they can try to show, *contra* my arguments above, that it is possible to engage in rational moral reflection without reliance on any intuitive foundation – that is, without recourse to any pre-reflective grasp of moral properties. I will address these two alternatives below.

3. Establishing autonomous foundations

Let us begin with the first possibility. Russ Shafer-Landau (2017) proposes a general strategy for distinguishing moral beliefs that are the products of evolution from those that are not. If, as debunkers assume, we can identify which moral beliefs have been shaped by selective pressures by identifying which ones are adaptive, then we must be able to identify which ones have *not* been so shaped by identifying which ones are *not* adaptive (2017: 178–80). For example, since the belief that all people have equal moral value confers no clear selective advantage on those who hold it, this belief is likely undistorted by evolution. And if we can identify enough of these undistorted beliefs, they can serve as a new foundation for rational moral reflection (cf. Huemer 2008; Berker 2014: 246–8).[7]

---

[7]Shafer-Landau appears to think that EDAs proceed by debunking *particular* moral beliefs by showing that they had adaptive value for our evolutionary ancestors, and that opponents of EDAs need only employ this strategy in reverse. Though some EDAs take this form (see Kelly 2014), I think the reasoning behind most EDAs is more general than this: the claim is that moral cognition *in general* is an adaptation, and that means it is tailored to bolster differential reproduction, not to discern moral truth.

The most obvious difficulty with this proposal is that it is unclear what *is* supposed to explain this special class of moral beliefs. They are, *ex hypothesi*, not rooted in any cognitive adaptation, and to be useful to the realist they must not be the deliverances of rational reflection. (Recall that they are supposed to be our new reflective starting-points.) But other candidate explanations, such as genetic drift or exaptation, threaten to make these beliefs random, and that is even worse (Kahane 2011: 111–12).

However, there is at least one explanatory alternative to biological evolution that is decidedly non-random: realists may hold that the origins of our moral intuitions are (primarily) *cultural*, rather than biological, and that they are passed from generation to generation via socialization, rather than genetic inheritance. Marcus Arvan (2020, 2021) has defended a detailed account along these lines. Drawing on a wide range of recent neuroscientific evidence, Arvan argues our capacity for thinking in moral terms is undergirded by a battery of cognitive capacities – for example, the ability to deliberate about the future and take account of others' perspectives – that evolved at different times in our species' history and that were originally selected for their roles in simple means-ends reasoning. Moral cognition is thus a case of evolutionary co-option: it emerged from the joint operation of capacities that originally evolved to solve problems unrelated to morality. Moral norms themselves, on Arvan's view, have a different etiology that is even more remote from Darwinian forces – they are the products of our ancestors' conscious rational deliberations aimed at prudence (i.e. maximal lifetime utility), and we internalize them through socialization. Moral intuition, therefore, is *learned*, not innate, and even our ability to learn it is an exaptation, not an adaptation.[8]

Whether or not Arvan's story is correct in its details, many philosophers and scientists concur that moral thinking is largely the upshot of cultural development, and there is significant empirical evidence for this (Braddock 2021; Levy and Levy 2020; but see Cline 2015 and Dale 2022 for dissent). In light of this, it seems probable that at least *some* of our pre-reflective moral notions arose independently of natural selection. Perhaps realists favorable to Shafer-Landau's strategy can draw on theories like Arvan's to identify moral intuitions that are free from Darwinian influence, and perhaps these will be enough to construct an autonomous foundation for rational moral reflection.

Alas, I don't find this alternative promising, for two reasons. First, as Nathan Cofnas (2022: 6–8), Andreas Mogensen (2016: 1808), and Matthew Braddock (2021) point out, the influence of selection pressures is not limited to genetically heritable traits. Selection is sensitive to any traits that are reliably transmitted from one generation to the next, and these include traits that are passed down through socialization. Just as natural selection favors traits that lead to greater differential genetic propagation, cultural selection favors ideas that lead to greater differential doxastic propagation. As Braddock (somewhat simplistically) puts it, if natural selection favors organisms who make more babies, cultural selection favors beliefs that gain more adherents (2021: 182). And there is little reason to think that cultural selection pressures are any more sensitive to mind-independent moral facts than biological ones.

On the contrary, Braddock (2021) argues convincingly that cultural evolution is highly contingent, in the sense that it easily could have given rise to moral norms incompatible with the ones we now hold, given the vast diversity and complexity of

---

[8]I cannot do justice to all the nuances of Arvan's account here (see Arvan 2020, especially chapters 1–3). Fortunately, however, the details should not make an important difference to my arguments below.

the selection pressures at work in the transmission of ideas. The best evidence for this is that many of the moral judgments realists in western liberal democracies take to be obvious – e.g., that women are entitled to education, opportunity, and sexual freedom – don't catch on easily in many parts of the world, even among the morally reflective. Furthermore, there is good reason to believe that the differential propagation of beliefs is heavily influenced by factors that realists consider morally indifferent or morally backwards (see Henrich 2017: 167–9). For example, some moral beliefs, like the belief that using contraception is immoral or that a woman's dignity is contingent on her fertility, lead to greater differential reproduction and consequently propagate more quickly across generations. Other beliefs, like the belief that taking vengeance on enemy tribes is morally good, or that a group has a moral responsibility to impose its way of life on other groups, motivate those who hold them to destroy or assimilate those who do not. Even beliefs that most realists take to be obviously true, like the belief that people should be evaluated on the basis of virtue and competence rather than on the basis of race or class, or the belief that all people should have equal access to education and property, plausibly survive and spread because they come with tremendous economic advantages. And economic, reproductive, and colonial viability are not determined by moral facts. Thus, whether the evolution of moral cognition was driven primarily by genetic forces or by cultural ones is beside the point: either way, our moral belief-forming tendencies are the upshots of selection pressures that are blind to moral truth.

Now, Fitzpatrick and others sympathetic to his view will insist that their moral beliefs are not attributable to factors like these. They hold their moral beliefs for good reasons, not because those beliefs are psychologically contagious or conducive to economic or military success. This may very well be true, but note that it does nothing to resolve the problem at hand. In the present section, we are considering ways in which some moral intuitions might have arisen independently of Darwinian forces, in hopes of furnishing an autonomous foundation for rational moral reflection. The point of the preceding paragraphs is that we cannot reasonably expect cultural evolution to have supplied us with a reliable *pre-reflective* grasp of moral properties.

Now for the second problem with the appeal to cultural transmission: even if the moral norms we have internalized were not deeply shaped by cultural selection pressures, if we are to hold that these norms give us insight into moral truth, we must presuppose that our ancestors *got the moral facts right* when they decided which norms to hand down. But they evolved, just like we did, so if we cannot grasp moral truths without the help of enculturation and socialization, neither could they. The norms they chose were likely mere reflections of their desires and practical interests, and we have already seen why we cannot trust those (Cofnas 2022: 13–15). Note that it will not help in this context to insist, as Arvan does, that our ancestors were guided by careful rational deliberation, since the question at issue is whether rational deliberation is a reliable way to get at the moral truth in the first place.[9]

To clarify: though I have argued that realists lack viable non-selectionist, non-random explanations for their moral intuitions, and though I think this is problematic,

---

[9]I should emphasize that I have no reason to think that Arvan himself would endorse the strategy for responding to EDAs discussed here. In fact, Arvan is probably not a realist in the sense at issue in this paper, since he holds that morality derives from prudence (Arvan 2020) and thus presumably takes moral facts to depend on (contingent) facts about our practical interests. That puts his view closer to Street's (2006) constructivism than to Cuneo's (2007) realism.

it is not part of my argument that, *in general*, we must explain the reliability of our cognitive faculties in order to have justified beliefs. As I noted in §1, EDAs do not impose any such requirements for epistemic justification; rather, they advance undercutting defeaters for realists' moral beliefs. It is in order to avoid such defeaters that realists must supply an explanatory alternative to natural or cultural selection. Otherwise, they are not justified in rejecting the selectionist explanation, and the defeater for their moral beliefs stands.[10]

Before leaving this section, consider the following, more general worry for the strategy that Shafer-Landau endorses. According to our best theories, new pre-reflective, truth-tracking cognitive capacities arise in something like the following way: random variation in a population of animals eventually gives rise to novel cognitive traits, and some of these happen to enable the animals who bear them to grasp independent truths. Usually, this helps the animals in question survive and reproduce – after all, having true beliefs about the world is generally better, practically speaking, than having false ones. Eventually, through natural selection, the entire animal population comes to have the novel cognitive trait. After a long while, some animal populations (humans, say) accumulate enough of these novel traits to grasp all sorts of truths, even quite exotic ones. The most important element in this story, for our purposes, is the role of natural selection in picking out the cognitive capacities that are truth-tracking from those that are not. It is ultimately because of natural selection – or, perhaps in the case of more sophisticated kinds of knowledge, cultural selection – that truth-tracking capacities dominate in our species, as opposed to non-truth-tracking ones.[11]

But if that is right, then, contra Shafer-Landau, the fact that a pre-reflective cognitive process was not shaped by selection pressures is a strike *against* it. So the moral realist is faced with the following dilemma: (1) if the contents of our moral intuitions have been heavily shaped by natural or cultural selection, then they likely do not track the truth. And (2) if they have *not* been so shaped, then they likely do not track the truth. So, either way, (3) our moral intuitions likely do not track the truth. This dilemma shows that the realist must do more, in order to provide an autonomous foundation for moral reflection, than simply identify some particular non-adaptive moral beliefs. She must refute either (1) or (2). The former strategy, if successful, would undermine EDAs from square one and thus vitiate the need for an appeal to autonomous rational reflection.[12] The latter strategy would require identifying some process other than

---

[10]Unless, of course, the Darwinian explanation is fatally flawed for independent reasons. In fact, I think Darwinism is vastly overrated. But that is another topic entirely.

[11]An anonymous reviewer objects to this claim on grounds that many of our cognitive capacities, like our capacities for reading, writing, and doing science, are truth-tracking despite not having been directly selected. I have three responses to this objection: First, reading, writing, and doing science are reflective activities (at least initially), not pre-reflective ones, and my arguments in this section apply only to pre-reflective cognition. Second, it is not clear to me that these cognitive capacities are distinct from others that were plausibly subject to direct selection: reading and writing is a refined exercise of our capacity for language; doing science is a refined exercise of our capacities for observation, induction, and mathematics (the latter of which is, in turn, a refined exercise of our capacities for simple arithmetic and deduction); and so on. Third, even if our capacities for reading, writing, and doing science are distinct, pre-reflective cognitive capacities, their reliability depends on the reliability of the more rudimentary capacities aforementioned, so it is still ultimately dependent on the influence of selection pressures. See §5 for related arguments.

[12]See Copp (2008), Enoch (2011), Skarsaune (2011), Brosnan (2011), Bloomfield (2018), and Wielenberg (2014) for examples of this approach. See Street (2008), Fitzpatrick (2014), Shafer-Landau (2017), and Vavova (2015) for, in my view, decisive criticisms.

evolution by natural or cultural selection that could cause truth-tracking, pre-reflective doxastic dispositions to arise and predominate in animal populations, and that is bound to be difficult, if not impossible.[13]

4. Reflection with no foundation

Until now, I have been arguing that moral reflection must, on pain of regress, depend ultimately on our pre-reflective grasp of moral properties. But this claim is disputable: some realists think that certain moral truths are rationally self-evident in a way that does not depend on substantive moral assumptions. Terrence Cuneo and Russ Shafer-Landau (2014), for example, have argued that our systems of moral belief are bounded by "fixed points," and that these fixed points are conceptual truths. More precisely, the fixed points are true in virtue of the relations between the essences of the concepts which constitute them. For example, the proposition "cruelty is wrong" is true in virtue of the fact that it is of the essence of the concept <wrong> that it necessarily applies to whatever falls under the concept <cruelty>. Cuneo and Shafer-Landau argue that the moral fixed points exhibit several characteristic features of conceptual truths and can be known *a priori* by those who understand their constituent concepts and their interrelations.

Realists who buy this story might object that, since the moral fixed points are conceptual truths, they can be known through rational reflection independently of more basic moral judgments. Thus, they will argue, the problems I raised for Fitzpatrick's account are no problems at all. Notice that this is not the kind of circular appeal to moral intuition that I warned against in §1; this is an appeal to conceptual mastery and logical intuition, two cognitive abilities that EDAs leave unchallenged.[14]

I doubt this move will work, however, for at least two reasons. First, it requires that our moral concepts – <cruelty> and <wrong>, for instance – apply to things in the world. But, as I noted above, some EDAs call this assumption into question. Even if it is a conceptual truth that cruelty is wrong, if cruel acts essentially have certain normative properties, it is still an open question whether anything really satisfies the concept <cruelty>. Maybe what seems to us like cruelty is really pseudo-cruelty, something empirically indistinguishable from cruelty but lacking the normative properties necessary to count as genuine cruelty. If so, then the conceptual truth that cruelty is wrong is no more theoretically interesting than the conceptual truth that pegasuses have wings. This is a special case of a more general limitation on the usefulness of conceptual knowledge: conceptual mastery typically suffices only for knowledge of conditionals of a certain kind.[15] For example, mastery of the concept <mermaid> affords knowledge of the proposition, "For any x, if x is a mermaid, x has a tail," but it does not afford knowledge that there are mermaids; it is still an open question whether anything satisfies the concept <mermaid>. In the same way, although mastery of the concepts <cruelty> and <wrong> may afford knowledge of the proposition, "For any x, if x is

---

[13]There is an important exception here for theists like me, who have the option of appealing to divine creative activity, or to divine coordination of evolutionary forces, or some such process. In fact, I think this is the right way to handle the problem, but I cannot argue for that here (see Baggett and Walls 2016; Braddock 2021; Crummett and Swenson 2020; Linville 2009; McKay 2023).

[14]Some EDAs challenge the *veracity* of our moral concepts (Joyce 2006; Morton 2016), but they do not challenge our ability to form, understand, or manipulate moral concepts.

[15]The most plausible exception to this rule is mathematics. I think there are stark disanalogies between mathematical knowledge and conceptual moral knowledge, but I cannot discuss these in detail here.

an act of cruelty, then x is wrong," it does not afford knowledge that there really are cruel acts.[16] Insofar as <cruelty> is a morally laden concept, we may doubt that any action satisfies it, and EDAs arguably give us reason to doubt this.

Perhaps Cuneo and Shafer-Landau would deny that the concept <cruelty> is itself morally laden and hold instead that it is a purely empirical concept. In other words, it has no normative properties built in, and yet it somehow entails normative properties. This implies that normative concepts are *analytically* reducible to empirical ones, and I find it very hard to believe that this is what Cuneo and Shafer-Landau have in mind, especially since they make no attempt to tease out the conceptual links between empirical and normative properties. I grant that, if such a reduction is possible, then EDAs are toothless, along with virtually all other arguments for moral skepticism. But I am entirely unconvinced, for the standard set of reasons,[17] that any such reduction will be found, and until one is, EDAs remain in force.

The second reason this approach will not work is that it is doubtful on empirical grounds that moral knowledge is conceptual. The most immediate reason why is that error theorists[18] disbelieve the moral fixed points despite having a subtle grasp of all the relevant concepts, so conceptual mastery apparently does not suffice for knowledge of these propositions. Cuneo and Shafer-Landau reply that conceptual truths need not be obvious: for example, it is not evident to everyone that God's possible existence entails his actual existence (see Plantinga 1974: 197–217). So it is possible that error theorists simply fail to see that the moral fixed points are conceptual truths, even though they are.

I am skeptical of this reply. Granted, some conceptual truths are not obvious to laypeople, but advanced theoreticians have no problem grasping them. That is why no modal logicians doubt that, given Plantinga's assumptions about God and the property of maximal greatness, God's possible existence entails his actual existence (in S5, at least). And error theorists are a sharp lot, even if they are wrong about metaethics, so it is extremely unlikely that they are wrong about the moral fixed points in the way that an undergraduate logic student might be wrong about the validity of Plantinga's ontological argument (see Ingram 2015).

Moreover, the moral fixed points *are* obvious: practically everyone believes them, and no formal proofs or technical philosophical arguments are necessary to show they are true, unlike the obscure logical and metaphysical beliefs to which Cuneo and Shafer-Landau analogize them. So it seems highly implausible that the moral fixed points are obscure conceptual truths – obscure enough, that is, for error theorists

---

[16]I don't mean to suggest that knowledge of conditionals like the ones aforementioned is *useless*. On the contrary, as Michael Huemer (2008) points out, conceptual moral knowledge of this sort allows us to identify formal constraints on moral reasoning and theorizing. For example, it plausibly follows from the essence of the concept <good> that if x is better than y and y is better than z, then x is better than z. It plausibly also follows from the essence of the concept <wrong> that, if it is wrong to do y given x, then it is wrong to do both x and y. Note, however, that none of this is sufficient to establish any substantive moral claims independently of moral intuition.

[17]The best reason for denying that moral concepts are analytically reducible to natural concepts is still, in my view, Moore's Open Question Argument (2012: 19–27). Note that conceptual irreducibility is compatible in principle with ontological reducibility. Even if moral concepts cannot be analyzed in naturalistic terms, moral properties may still turn out to be natural properties of some special sort.

[18]That is, philosophers who think that moral judgements express beliefs but that these beliefs are always false, because there are no moral properties.

to be so confidently wrong about them. Unfortunately, it appears we will have to lean on our pre-reflective grasp of moral properties after all. And if my arguments in the last two sections are on target, that spells trouble for moral realism.

5. Do these arguments overgeneralize?

At this point, the realist might complain that my arguments prove too much. After all, if rational reflection does not enable us to achieve autonomy from our evolved intuitions when it comes to the moral domain, why think it enables us to do so when it comes to the scientific, mathematical, and philosophical domains (Clarke-Doan 2012; Fitzpatrick 2015)? And if, as seems plausible, evolution is blind to truths about quantum physics, modal logic, and differential calculus, why not conclude that our beliefs about these topics are undermined?

I think this question is rather easily answered. Recall that moral beliefs are susceptible to debunking because their truth values do not determine their adaptive utilities: whether moral beliefs accurately represent the distribution of mind-independent moral properties seems irrelevant to whether they motivate behavior that is advantageous from an evolutionary point of view. In contrast, the truth values of rudimentary non-normative beliefs typically *do* determine their adaptive utilities. Primates who are mathematically inept, prone to illogic, or unable to understand causal regularities in their environments will die more quickly and leave fewer progeny on average *because* they are unable to accurately model the world. Primates with a great preponderance of true non-normative beliefs will live longer and enjoy greater reproductive success on average *because* their world-models are accurate. Thus, natural selection heavily favors predispositions to form true pre-reflective, non-normative beliefs over predispositions to form false ones. Unlike in the moral domain, therefore, there is no reason to think that our reflective starting-points in non-normative domains are suspect.

The realist might object that this reasoning does not extend to highly abstract theoretical knowledge, since none of our distant evolutionary ancestors knew quantum physics or modal logic. These branches of knowledge were developed through rational reflection on a set of rudimentary empirical, logical, and mathematical judgments, in a way similar to our moral theories. So, why think that rational reflection is unreliable in the moral case, but not in all the other cases? This objection rests on a misunderstanding: nothing I have said implies that rational reflection is unreliable per se (that would be an exceedingly odd thing for a philosopher to say). The alleged problem is not with rational reflection itself, but rather with the raw materials on which it operates in the moral domain. My claim is that the *starting points* of rational reflection *about morality* – namely, our moral intuitions – are suspect because of their evolutionary origins. And there is little reason to think that the starting points of rational reflection about other things, including mathematics, science, and logic, are similarly suspect. That is good news, of course, for the mathematicians, scientists, and logicians, but bad news for the realists seeking companions in guilt.

6. Conclusion

To sum up: I have argued, first, that appeals to rational moral reflection cannot defuse EDAs directly, since it is doubtful that such reflection enables us to achieve autonomy from our evolved moral intuitions. I have argued, second, that any attempt to furnish autonomous foundations for rational moral reflection by identifying a special class of

uncorrupted moral beliefs is bound to fail, since these beliefs would have to be explained by some mechanism other than evolution by natural or cultural selection, and no such mechanism is likely to be truth-tracking. I have tried, third, to refute the view that our most basic moral beliefs are items of conceptual knowledge by showing that mastery of moral concepts is not sufficient for holding these beliefs. And I have replied, fourth, to the worry that my arguments over-generalize by pointing out that selection pressures are sensitive to the truth values of rudimentary non-normative beliefs, so we do not have parallel reasons to think that our reflective starting points in non-normative domains are suspect.

The moral of this essay is that appeals to rational reflection are unlikely to neutralize EDAs. A related point is that simply identifying moral beliefs that are not the products of natural selection is not enough to vindicate moral knowledge. Instead (or in addition), realists should focus their efforts on one of the two horns of the dilemma I outlined in §3. I doubt this approach will succeed in the long run,[19] but it seems to me the most promising way forward.[20]

References

Arvan M. (2020). *Neurofunctional Prudence and Morality: A Philosophical Theory*. New York: Routledge.

Arvan M. (2021). 'Morality as an Evolutionary Exaptation.' In J. De Smedt and H. De Cruz (eds), *Empirically Engaged Evolutionary Ethics*, pp. 89–109. Synthese Library. Cham: Springer International Publishing.

Baggett D. and Walls J.L. (2016). *God and Cosmos: Moral Truth and Human Meaning*. Oxford: Oxford University Press.

Berker S. (2014). 'Does Evolutionary Psychology Show That Normativity Is Mind-Dependent?' In J. D'Arms and D. Jacobson (eds), *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics*, pp. 215–52. Oxford: Oxford University Press.

Bloomfield P. (2018). 'Tracking Eudaimonia.' *Philosophy, Theory, and Practice in Biology* **10**(2).

Braddock M. (2021). 'The Contingency of the Cultural Evolution of Morality, Debunking, and Theism Vs. Naturalism.' In J. De Smedt and H. De Cruz (eds), *Empirically Engaged Evolutionary Ethics*, pp. 179–201. Cham, Switzerland: Springer.

Brosnan K. (2011). 'Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?' *Biology and Philosophy* **26**(1), 51–64.

Clarke-Doane J. (2012). 'Morality and Mathematics: The Evolutionary Challenge.' *Ethics* **122**(2), 313–40.

Cline B. (2015). 'Nativism and the Evolutionary Debunking of Morality.' *Review of Philosophy and Psychology* **6**(2), 231–53.

Cofnas N. (2022). 'How Gene-Culture Coevolution Can–But Probably Did Not–Track Mind-Independent Moral Truth.' *Philosophical Quarterly* **73**(2), 14–434.

Copp D. (2008). 'Darwinian Skepticism about Moral Realism.' *Philosophical Issues* **18**, 186–206.

Crummett D. and Swenson P. (2020). 'God and Moral Knowledge.' In K. Vallier and J. Rasmussen (eds), *A New Theist Response to the New Atheists*, pp. 33–46. New York, USA: Routledge.

Cuneo T. (2007). *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.

Cuneo T. and Shafer-Landau R. (2014). 'The Moral Fixed Points: New Directions for Moral Nonnaturalism.' *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* **171**(3), 399–443.

Dale M.T. (2022). 'The Evolution of Moral Belief: Support for the Debunker's Causal Premise.' *History and Philosophy of the Life Sciences* **44**(2), 1–18.

Enoch D. (2011). *Taking Morality Seriously*. Oxford: Oxford University Press.

---

[19]Again, without an appeal to theism (see fn. 13).

[20]I am indebted to Michael Cholbi, John Wingard, William Fitzpatrick, Daniel Kelly, and an anonymous reviewer for their input.

FitzPatrick W. (2014). 'Why There Is No Darwinian Dilemma for Ethical Realism.' In M. Bergmann and P. Kain (eds), *Challenges to Moral and Religious Belief: Disagreement and Evolution*, pp. 237–55. Oxford University Press.

FitzPatrick W. (2015). 'Debunking Evolutionary Debunking of Ethical Realism.' *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* **172**(4), 883–904.

FitzPatrick W. (2016). 'Scientific Naturalism and the Explanation of Moral Beliefs: Challenging Evolutionary Debunking.' In K.J. Clark (ed.), *The Blackwell Companion to Naturalism*, pp. 386–400. Hoboken, NJ: John Wiley & Sons, Incorporated.

FitzPatrick W. (2017). 'Why Darwinism Does Not Debunk Objective Morality.' In M. Ruse and R.J. Richards (eds), *The Cambridge Handbook of Evolutionary Ethics*, pp. 188–201. Cambridge Handbooks in Philosophy. Cambridge: Cambridge University Press.

Henrich J. (2017). *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making us Smarter*. Princeton, NJ: Princeton University Press.

Huemer M. (2008). 'Revisionary Intuitionism.' *Social Philosophy and Policy* **25**(1), 368–92.

Ingram S. (2015). 'The Moral Fixed Points: Reply to Cuneo and Shafer-Landau.' *Journal of Ethics and Social Philosophy* **9**(1), 1–6.

Joyce R. (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.

Joyce R. (2016). 'Reply: Confessions of a Modest Debunker.' In U.D. Leibowitz and N. Sinclair (eds), *Explanation in Ethics and Mathematics: Debunking and Dispensability*, pp. 124–5. Oxford: Oxford University Press.

Kahane G. (2011). 'Evolutionary Debunking Arguments.' *Nous* **45**(1), 103–25.

Kelly D. (2014). 'Selective Debunking Arguments, Folk Psychology, and Empirical Moral Psychology.' In H. Sarkissian and J.C. Wright (eds), *Advances in Experimental Moral Psychology*, pp. 130–47. London: Bloomsbury Academic.

Kitcher P. (2011). *The Ethical Project*. Cambridge, MA: Harvard University Press.

Levy A. and Levy Y. (2020). 'Evolutionary Debunking Arguments Meet Evolutionary Science.' *Philosophy and Phenomenological Research* **100**(3), 491–509.

Linville M.D. (2009). 'The Moral Argument.' In J.P. Moreland and W.L. Craig (eds), *The Blackwell Companion to Natural Theology*, pp. 391–448. John Wiley & Sons, Ltd.

McKay. (2023). 'Moral Knowledge and the Existence of God.' *International Journal for the Philosophy of Religion*. https://doi.org/10.1007/s11153-023-09868-z.

Mogensen A.L. (2016). 'Do Evolutionary Debunking Arguments Rest on a Mistake about Evolutionary Explanations?' *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* **173**(7), 1799–817.

Moore G.E. (2012). *Principia Ethica*. Stilwell: Neeland Media.

Morton J. (2016). 'A New Evolutionary Debunking Argument Against Moral Realism.' *Journal of the American Philosophical Association* **2**(2), 233–53.

Plantinga A. (1974). *The Nature of Necessity*. Oxford: Oxford University Press.

Ruse M. (1986). *Taking Darwin Seriously: A Naturalistic Approach to Philosophy*. Oxford: Blackwell.

Ruse M. (2006). 'Is Darwinian Metaethics Possible (And If It Is, Is It Well Taken)?' In G. De Anna and G. Boniolo (eds), *Evolutionary Ethics and Contemporary Biology*, pp. 13–26. Cambridge Studies in Philosophy and Biology. Cambridge: Cambridge University Press.

Schafer K. (2010). 'Evolution and Normative Skepticism.' *Australasian Journal of Philosophy* 88(3), 471–88.

Shafer-Landau R. (2012). 'Evolutionary Debunking, Moral Realism and Moral Knowledge.' *Journal of Ethics and Social Philosophy* **7**(1), 1–38.

Shafer-Landau R. (2017). 'Moral Realism and Evolutionary Debunking Arguments.' In M. Ruse and R.J. Richards (eds), *The Cambridge Handbook of Evolutionary Ethics*, pp. 175–87. Cambridge Handbooks in Philosophy. Cambridge: Cambridge University Press.

Skarsaune K.O. (2011). 'Darwin and Moral Realism: Survival of the Iffiest.' *Philosophical Studies* **152**(2), 229–43.

Street S. (2006). 'A Darwinian Dilemma for Realist Theories of Value.' *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* **127**(1), 109–66.

Street S. (2008). 'Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About.' *Philosophical Issues* **18**, 207–28.

Sturgeon N. (1985). 'Moral Explanations.' In J. Rachels (ed.), *Ethical Theory 1: The Question of Objectivity*, pp. 49–78. Oxford: Oxford University Press.

Tropman E. (2014). 'Evolutionary Debunking Arguments: Moral Realism, Constructivism, and Explaining Moral Knowledge.' *Philosophical Explorations* **17**(2), 126–40.

Vavova K. (2015). 'Evolutionary Debunking of Moral Realism.' *Philosophy Compass* **10**(2), 104–16.

Wielenberg E.J. (2014). *Robust Ethics: The Metaphysics and Epistemology of Godless Normative Realism.* Oxford: University Press.

Wielenberg E.J. (2016). 'Ethics and Evolutionary Theory.' *Analysis* **76**(4), 502–15.

Noah McKay is a graduate student at the university of Edinburgh. His published work touches primarily on epistemology, metaethics, and philosophy of religion. Email: n.d.mckay99@gmail.com