A New Evolutionary Debunking Argument Against Moral Realism

Justin Morton
The University of Texas at Austin
mortonjj@utexas.edu

Abstract: Evolutionary debunking arguments (EDAs) claim that evolution has influenced our moral faculties in such a way that, if moral realism is true, then we have no positive moral knowledge. I present several popular objections to the standard version of this argument, then give a new EDA that has clear advantages in responding to these objections. Whereas the Standard EDA argues that evolution has selected for many moral beliefs with certain contents, this New EDA claims that evolution has selected for *one* belief: belief in the claim that categorical reasons exist. If moral realism is true, then this claim is entailed by all positive moral claims, and belief in it is defeated due to evolutionary influence. This entails that if realism is true, then we have no positive moral knowledge. While there may be objections against this New EDA, it is much stronger than the Standard EDA, and one realists ought to worry about.

Introduction

Recently, philosophers have been especially concerned with the role that evolution has

played in shaping our moral faculties. None deny that evolution has had some influence. But some

argue that, if moral realism were true, such evolutionary influence would undermine our moral

knowledge. I, and others, find such "evolutionary debunking arguments" (EDAs) to be deeply

flawed, as they are normally formulated. In this paper, I formulate a new EDA, which targets moral

knowledge indirectly, by providing a defeater for belief in categorical reasons. But if realism is true,

then all positive moral beliefs entail the existence of at least one categorical reason. I argue from this

that realism entails moral skepticism. One major virtue of this "New EDA" is that it does a much better job at avoiding the deepest problems with standard EDAs.

The structure of this paper is as follows. In §1, I formulate what I call the Standard EDA. This argument is meant to put in its most plausible form what is common to all (or at least most) of the EDAs on the market today. In §2, I summarize four popular objections to the Standard EDA. In §3, I defend the New EDA. Finally, in §4, I revisit the objections to the Standard EDA, showing how two of them don't threaten the New EDA, while with regard to the other two, the New EDA has distinct advantages over the Standard EDA. I conclude that there is a strong case to be made that, if moral realism is true, then we rationally ought to be moral skeptics. This is, at the very least, a high price for realists to pay.

1. The Standard EDA

Moral realism, as I define it, is the thesis that (i) sincere moral judgments express beliefs, (ii) some of those beliefs are true, and (iii) the truth of moral beliefs does not constitutively depend on the attitude of any actual or hypothetical agent (Shafer-Landau 2012: 1). There have been many objections to moral realism on evolutionary grounds (Street 2006; Joyce 2007: ch. 6; Horn forthcoming; Greene 2008: 35-80; Kitcher 2007; Ruse and Wilson 1986: 173-192).[1] Here I will focus on what I consider the most popular type of evolutionary objection—what I'll call the Standard EDA:

The Standard EDA
1. **Epistemological Premise:** If (a) moral realism is true, (b) evolution has strongly influenced our moral faculties in ways that are doxastically discriminating, and (c) there is no independent confirmation of the reliability of those faculties, then we have no positive moral knowledge.
2. **Empirical Premise:** Evolution has strongly influenced our moral faculties in ways that are doxastically discriminating.
3. **Autonomy:** There is no independent confirmation of the reliability of our moral faculties.

---

[1] I make no claims to perfectly summarize any author's views; I attempt here to construct the best possible objection that makes sense of what such authors say.

4. Therefore, if moral realism is true, then we have no positive moral knowledge.

I'll explain (2) and (3) before explaining (1)—but first, a terminological point. "Positive moral knowledge" is knowledge of a positive moral claim. A positive moral claim is one that attributes a moral predicate to something. Examples include "Stealing is wrong" and "It would be generous to buy this man a meal." Likewise, when I later talk of "positive moral beliefs," I mean only "belief in a positive moral claim."

Regarding (2), a faculty is influenced by X in a "doxastically discriminating" way iff the faculty is, in virtue of X, disposed to produce beliefs with certain propositional contents rather than others. Consider the belief that incest is wrong. A human who believes that incest is wrong is more likely to have offspring who can pass on their genetic material than a human who doesn't have this belief. This is because incest increases the chances of sterility or deformation in one's offspring, and believing that incest is wrong will make one less likely to engage in it. So, evolution selects for the belief that incest is wrong (or something like it).[2]

Premise (3) is an autonomy thesis about the moral: we cannot confirm the reliability of our moral faculties except by showing that they have generated (mostly) true moral beliefs. But this cannot be done without assuming either the reliability of our moral faculties or the truth of our moral beliefs. For example, there is no moral almanac by which we can check the moral facts.

Premise (1) is the most obscure, partly because, as Shafer-Landau (2012) points out, there are so many ways of understanding why the consequent might follow from (a-c). Here is one way it has been defended (Joyce 2007: ch. 6; Bedke 2009; Clarke-Doane 2012):

Insensitivity
5. If (a-c) hold, then we would have the positive moral beliefs we do regardless of whether they are true.
6. If we would believe that P regardless of whether P, then we do not know that P.

---

[2] Some might wonder at the idea that evolution could influence our mental faculties at all. But evolutionary psychologists propose that evolution could explain both physiological *and* psychological phenomena. For a fuller account of this proposal, see James (2011): 18-19.

7. Therefore, if (a-c) hold, then we have no positive moral knowledge.

The plausibility of Insensitivity is not very important here: I give it just as an example of how (1) is defended. But briefly, (6) is supposed to be an intuitively plausible epistemological principle. (5) is claimed to hold because evolution selects for the content of our beliefs, not their truth. It is adaptive for you to believe that you ought to take care of your children regardless of whether it is true.

However, let me emphasize that there are many ways of arguing for (1), and no part of my argument here depends on any one in particular. Furthermore, there may be reasons to doubt (5)—for example, Fitzpatrick (2014) argues that (5) presupposes that evolution is the *only* influence on our moral judgments. But I offer Insensitivity simply to aid in understanding some of the typical rationales that underlie the Standard EDA.

2. Objections to the Standard EDA

The Standard EDA, however appealing, has its share of problems. In this section I will summarize four popular, quite strong objections to it. I do not claim that proponents of the Standard EDA have no reply to any of these objections. However, in §4, I'll argue that the New EDA fares better with regard to all of them.

2.1   *The Limited Explanation Objection*

The first objection, one that has received wide support, I will call the limited explanation objection (Shafer Landau 2012: 5-8; Fitzpatrick 2014: 241-246; Parfit 2011: 534-538; Huemer 2008b: 368-392; James 2011: 79-81; Copp 2008: 194; Street 2006: 155). It essentially consists of a denial of the empirical premise, (2). Proponents of the objection grant that there are *some* moral beliefs that have clear and plausible evolutionary explanations (like the belief that incest is wrong). But there are other moral beliefs that do not have such clear evolutionary explanations (e.g., the belief that all human persons have equal and inalienable rights).

With such undebunked beliefs in hand, the realist can take the limited explanation objection in two different directions. She can argue—as Michael Huemer (2008b), Joshua Greene (2008), and Peter Singer (2005) each do—that we should abandon our debunked beliefs while retaining our undebunked beliefs. Though (on a realistic picture) we might have to abandon many of our moral beliefs, we still have moral knowledge, and so the Standard EDA fails. We may even end up with a more coherent set of moral beliefs.

On the other hand, it may be possible to gain back knowledge of our previously debunked beliefs from the undebunked ones. For example, there is a good evolutionary explanation of why we believe that we ought to take care of our children. But there isn't as clear of an explanation of the belief that we ought to take care of any helpless person whose existence resulted partially from our voluntary action. And, in combination with some uncontroversial empirical premises—that my child is helpless, a person, and resulted partially from my voluntary action—this broader principle entails that I ought to take care of my children. So, I can know the latter, even though belief in it has an evolutionary explanation, because I can infer it from undebunked beliefs. (Maybe there are problems with the example; it is meant only as an illustration.) Whichever of these two strategies the realist takes, she can retain a substantial amount of positive moral knowledge.

Now, the standard debunker does have some available replies. She might argue, e.g., that while the supposedly undebunked beliefs don't admit of *direct* evolutionary explanations, they are *indirectly* explained by evolution (see James 2011: 2.4). But as mentioned earlier, I will not flesh out the dialectic any further. Hopefully, it is clear that the standard debunker has her work cut out for her, and it is unclear whether she will be able to answer the objection.

2.2   *The Independent Confirmation Objection*

Russ Shafer-Landau (2012: 33-35) targets premise (3) of the Standard EDA, claiming that there may be a source of independent confirmation of the reliability of our moral faculties. He

claims that we can show that a doxastic faculty is reliable by showing that it is either identical to or a species of a doxastic faculty that we have independent warrant for believing to be reliable. In the moral case, this would be *independent* confirmation in that it does not assume the reliability of our moral faculties, or the truth of any given moral belief. Furthermore, this confirmation is compatible with the autonomy of the moral domain—i.e., that no moral beliefs can be derived from non-moral ones.

Shafer-Landau suggests that whatever faculties generate our non-moral synthetic *a priori* knowledge might also generate some set of our moral beliefs. And we have good reason to trust these faculties, which generate judgments such as that justified true belief is insufficient for knowledge, and that nothing can be red all over and green all over at the same time. But plausibly, those very same faculties also generate our *a priori* moral beliefs, such as that it is *pro tanto* wrong to cause an innocent person to suffer. So we have good reason to trust (at least) the faculties that generate our *a priori* moral beliefs. And it stands to reason that, in the way mentioned in 2.1, from this base we could derive knowledge of many *a posteriori* moral claims. Again, I do not claim here that the standard debunker has no possible reply to this objection—he may, e.g., argue that the faculty that generates our *a priori* moral beliefs is not identical to/a species of any faculty that reliably generates non-moral synthetic *a priori* beliefs. But he has his work cut out for him here.

2.3   The Overgeneralization Objection

Some have argued that the Standard EDA—particularly the epistemological premise—is overbroad (Huemer 2008a: 218-219; Enoch 2011: 175-176; Shafer-Landau 2012: 22; Vavova 2014: 82-83; Bedke 2009; Clarke-Doane 2012; see also Plantinga 1993: ch. 12). It seems to entail that if realism about a domain D is true, then strong evolutionary influence on D-faculties entails D-skepticism. But our perceptual faculties have been strongly influenced by evolution, as have our mathematical faculties. So, if realism in these domains is true, then we have no perceptual or

mathematical knowledge. But if so, the objectors claim, there is certainly something wrong with the Standard EDA: certainly perceptual (if not mathematical) realism is true. Since we have perceptual (and mathematical) knowledge, we can justifiably dismiss the Standard EDA, since it implies that we do not. Once again, the standard debunker may have a plausible response. The standard debunker could, e.g., deny mathematical realism and argue that her reasoning does not extend to the perceptual domain since evolution selects for *true* perceptual beliefs (but for a problem with the latter, see Street (forthcoming b: 25-28). However—the old refrain—my point is simply that the path to a satisfactory response is not clear, and may be impassible.

2.4   *Third Factor Responses*

A fourth type of objection to the Standard EDA is what has been called a *third factor response*. Such responses target the epistemological premise, claiming that even if evolution has influenced our moral faculties and there is no independent confirmation of them, we could still be justified in our moral beliefs. Third factor responses generally start by assuming the truth of a substantive moral principle (though for a notable exception, see Behrends 2013). They then show how, on this assumption, evolution predictably brings us to have (at least some) true moral beliefs, even though it doesn't select for the *truth* of those beliefs. Many third factor responses have been proposed (Behrends 2013; Brosnan 2011; Enoch 2010 and 2011: ch. 7; Skarsaune 2011; Wielenberg 2010 and 2014: ch. 4). It will suffice for my purposes to briefly outline two prominent ones—David Enoch's and Erik Wielenberg's.

Enoch asks us to assume (plausibly) that our survival—or whatever evolution "aims" at—is at least somewhat good. This makes it plausible that anything that promotes our survival (or whatever) is good. But now consider how evolution influences us to act in ways that promote our survival: often when Φing promotes survival, it is adaptive for us to believe that Φing is good—this makes us more likely to Φ. But now it's much less of a surprise that many of our moral beliefs are

true: when Φing promotes survival, then (a) it is good (by our assumption) and (b) we are likely to believe that it is good. That our beliefs about goodness coincide with the facts about goodness is no longer a mystery. (As Enoch acknowledges, there may still be some explaining to do. This is unimportant here.)

Wielenberg attempts to defend (some of) our knowledge of rights. Assume, he says, the substantive moral principle that one has rights whenever one has certain mental capacities (call them "C"). The precise nature of C is unimportant, except for the following sufficient condition: whenever one has the capacity to form beliefs, one has C. Now imagine that I believe that I have rights. Because I have a belief, I have C. And because I have C, I have rights. So, on the assumption of one moral principle, we get the following result: whenever I believe I have rights, I actually do. Again, on one assumption, it's no longer surprising that a large class of our moral beliefs are true.

That's a very abbreviated introduction to the third factor response to the Standard EDA. There is much more to be said about each of the two exemplar views: for example, in each case, our knowledge of a limited class of moral claims might be expanded (via, e.g., coherence relations) to other moral claims. Furthermore, there may be ways that the standard debunker can reply to the likes of Enoch and Wielenberg (I discuss one such way in §4.3). But once again, it won't be easy.

This concludes my consideration of common objections to the Standard EDA. I do not claim that there are no other strong objections. These four, however, seem to me to be four of the strongest. And, as I'll show in §4, the New EDA fares better in replying to each of them.

3. The New EDA

I think that there is a distinct EDA—the New EDA—that is plausible in its own right, in addition to its advantages vis-à-vis the above objections. In this section, I'll outline the New EDA and defend its premises. Aside from the incorporation of a more complex epistemology in the New EDA, there is one basic difference between that argument and the Standard EDA. While the

Standard EDA debunks moral beliefs on the basis of their *contents*, the New EDA debunks moral beliefs on the basis of a claim that they all *entail*. The Standard EDA, for example, claims that evolution selects for the disposition to judge that it's wrong not to take care of one's children, and it's on this basis that we have no knowledge of that claim. The New EDA, on the other hand, will claim that the judgment that it's wrong not to take care of one's children entails a claim belief in which is defeated. In fact, this claim, which turns out to be that categorical reasons exist, is entailed by all positive moral claims. From this, given some plausible epistemological principles, it follows that if moral realism is true, then we have no positive moral knowledge.

Before I turn to the New EDA, let me introduce and define an integral term: "categorical reason." Defining this term is somewhat complicated by the fact that it is commonly defined in two different ways. According to the weaker definition:

CR$_1$: A has a categorical reason to Φ in circumstances C iff A has a reason to Φ and any agent in C has a reason to Φ.

And according to the stronger definition:

CR$_2$: A has a categorical reason to Φ in circumstances C iff A has a reason to Φ and that reason obtains regardless of what desires (broadly construed) A has or what judgments she makes.

On this second reading, categorical reasons are just what many call "external reasons" (as in Williams 1981). I believe that the New EDA can be soundly formulated on either reading, though I have CR$_2$ in mind in what follows.

Here is the New EDA:

The New EDA
8. Assume that moral realism is true.
9. **Empirical Premise:** Evolution has strongly influenced our belief in categorical reasons.
10. **Autonomy:** There is no independent justification for belief in categorical reasons.
11. If (9) and (10), then our belief in categorical reasons is defeated.
12. So, our belief in categorical reasons is defeated. (9,10,11)
13. All positive moral claims entail that at least one categorical reason exists, and we are justified in believing that this entailment holds.

14. If P entails Q, we are justified in believing that P entails Q, and our belief that Q is defeated, then we do not know that P.
15. So, we do not have any positive moral knowledge. (12,13,14)
16. Therefore, if moral realism is true, then we do not have any positive moral knowledge. (8,15)

In the rest of this section, I will defend (9), (10), (11), (13), and (14).

3.1   Defense of (9): The Empirical Business

(9) is an empirical premise. Since I am just a humble philosopher, I cannot mount a full defense of this premise. What I can do, however, is show why there is at least enough support for (9) for us to worry about what would follow from it. This is, after all, all the standard debunker has (Street 2006: §3).

Let me start with a clarification. To say that evolution has strongly influenced our belief in categorical reasons does not mean that evolution has caused us to believe that categorical reasons exist. "Categorical reason," after all, is a philosophical term of art. However, suppose that we had reason to think that evolution had favored humans who have a sense that there are some actions (or desires, etc.) that are favored *no matter what*. This would count as evidence of a strong evolutionary influence on our belief in categorical reasons, and thus would support (9). Henceforth, let "our belief in categorical reasons" be understood in this pre-theoretical way.

There is some reason to think such influence has occurred. The basic idea behind (9) is that those humans are more adaptive who believe that they have a reason to (e.g.) take care of their children *no matter what they desire*. Why is this? Because they will be more likely to take care of their children than if they merely desired to do so. As Richard Joyce points out, a mere *desire* to do something can easily be overridden by stronger desires, and long-term desires can be hastily re-evaluated in light of short-term desires. Desires are, in Joyce's (2001: 136-137) terms, "unreliable things" (see also Olson 2011). We will be far more adaptive if we believe that some things are favored/required *no matter what we desire*.

Further support for (9) is based on thinking about cooperative situations that can be modelled on the Prisoner's Dilemma (James 2011: §2.6). These are cases in which multiple organisms would be better off (reproductively speaking) helping each other, but in which, if they deliberate from the standpoint of their self-interest, they rationally ought to play the free-rider. As James (2011: 59) notes, the problem in need of a solution, in such cases, is to "design individuals to establish and preserve cooperative alliances *despite* the temptation not to cooperate." A community of individuals who have some mechanism that *trumps* their desires in favor of cooperative action will do better, reproductively speaking, than a community of individuals that does not. Such a mechanism is provided by a sense of having *categorical reason* to follow through on promises, to help those who help you, etc. With such a sense, even where you believe that it would be in your best interest to defect, you won't. In sum, we seem to have enough reason to believe (9) that we should worry about what follows from it.

3.2 Defense of (10): No Independent Justification

So let's assume that evolution has strongly influenced our belief in categorical reasons. Now I want to show that there is no independent justification for belief in categorical reasons. First off, what does it mean for a justification to be independent? A justification for belief in categorical reasons is independent iff it doesn't rely on either (a) our intuitions about what categorical reasons we have, or on (b) our belief or intuition that P, where P would entail the existence of a categorical reason.

And, as it turns out, there is no such justification. For brevity's sake, I'll consider here just one well-known argument for categorical reasons. I footnote a second, for variety's sake, at the end of the section. Consider David Enoch's (2011: 261-262) "Sufficiently Bad Bad-Guy" argument against existence-internalism—the thesis that one has a reason to Φ only if there is a sound deliberative route from one's existing motivations to one's Φing. Imagine that Sufficiently Bad Bad-

Guy has no desire that would be promoted, even under ideal deliberative conditions, by refraining

from harming Victim. Enoch argues:

17. If existence-internalism is true, then Sufficiently Bad Bad-Guy has no reason not to hurt
    Victim.
18. But clearly Sufficiently Bad Bad-Guy *does* have a reason not to hurt Victim.
19. Therefore, existence-internalism is false.

(17) follows from our supposition that Sufficiently Bad Bad-Guy has no desire which would be

served by refraining from hurting Victim, even under ideal deliberative conditions. But, so (18)

claims, he *does* have a reason not to perform said heinous act. So, he must have a reason that obtains

regardless of his desires, and that entails the existence of at least one categorical reason.

The problem, however, is apparent. The plausibility of (18) depends on our intuition that

Sufficiently Bad Bad-Guy has a categorical reason not to hurt Victim. The intuition that supports

(18) is the same intuition that I would defend by saying "No, Sufficiently Bad Bad-Guy has a reason

not to hurt Victim, *no matter what*!" In fact, if I understand who Sufficiently Bad Bad-Guy is, and I

understand what a categorical reason is, then I can't believe (18) without believing that his reason

not to hurt Victim is a categorical one. So this argument for categorical reasons doesn't give us

independent justification for belief in them, whatever its other merits are as an argument.[3]

Perhaps there are some arguments for the existence of categorical reasons that are entirely

distinct from both that given in the text and in footnote 3, but I am unaware of them. Such

arguments, I suspect, would not constitute independent justification of belief in categorical reasons;

but then again, I would have to see the arguments first.

---

[3] Shafer-Landau (2009) gives an independent argument for categorical reasons: the "fanatic" has no commitment
promoted by not perpetrating an act of terror, and so has no non-categorical reason not to do so. However, since he is
intuitively blameworthy, he must have a categorical reason to refrain. Yet, whereas Enoch's argument violates (a) in the
definition of independence, Shafer-Landau's relies on an intuition that something obtains (the fanatic's blameworthiness)
that would entail the existence of a categorical reason, violating (b) in the definition of independence.

3.3   Defense of (11): Why Evolutionary Pressures Defeat Belief

So now let's assume that (9) and (10) are plausible. What follows from their conjunction? In (11) we get an answer analogous to (1), the epistemological premise of the Standard EDA: if (9) and (10)—and assuming moral realism—then our belief in categorical reasons is defeated. All I mean by such "defeat" is the classic notion from epistemology. A defeater D for A's belief that P is a true proposition such that if A justifiably believed that D, then A would be unjustified in believing that P. When D obtains, we can say that A's belief that P is defeated. One standard example is a Gettier case: you have a true, justified belief that it is 10:39, based on your view of a clock that reads "10:39." But your belief may even now be defeated, since the clock you're looking at is broken and you just happened to look at it when it read correctly. The proposition "The clock is broken" is a defeater for your belief that it is 10:39 because, if you justifiably believed that the clock was broken, you would not be justified in the belief that it is 10:39.

I said above that there are many ways of arguing for the epistemological premise of the Standard EDA, giving *Insensitivity* as a notable example. The same is true regarding (11), I suspect. (Since the New EDA is, well, new, I cannot appeal to notable defenses of that premise.) In this section I will offer just one defense of (11), based on an argument from insensitivity. Though I do in fact think that this argument is sound, there may well be other appealing defenses of the premise.

Here is the argument, which runs along the same lines as *Insensitivity*:

New Insensitivity
20. If (9) evolution has strongly influenced our belief in categorical reasons and (10) there is no independent justification for belief in categorical reasons, then for any agent A, A would believe in categorical reasons regardless of whether they exist or not.
21. If A would believe that P regardless of whether P, then A's belief that P is defeated.
22. Therefore, if (9) and (10), then for any agent A, then A's belief in categorical reasons is defeated.

The justification for (20) runs roughly as follows. Evolution selects for adaptive traits, and your belief that you have a categorical reason to (say) take care of your children would be adaptive

regardless of whether you *actually do* have such a reason. The belief that there are any categorical

reasons at all is similarly adaptive (and thus selected for) because without it, you could not believe

that you have any *particular* categorical reason. Now, a belief that P can be strongly influenced by

such a truth-insensitive cause and yet still be sensitive to the truth—if the agent in question has

some independent justification for the belief that P (e.g., God has told him that P), and his belief is

also strongly influenced by this consideration. However, when there is a truth-insensitive cause of

our belief that P—as evolution is, in this case—and there is no independent justification for

believing that P, we would believe it regardless of whether P.[4]

(21) seems to me an eminently plausible epistemological principle. When I would believe

that P no matter whether P, this counts as a defeater for my belief that P. That is, if I would believe

that P regardless of its truth, and if on top of that I justifiably believed this about myself, then I

would not be justified in my belief that P. Suppose I believe that I'm a great basketball player. Then

I find out (with justification) that, due to my immense self-confidence, I would believe this

regardless of whether it was true or false. I would no longer be justified in my belief that I am a great

basketball player. (Notice that I am not committed to the claim that, when A's belief that P is

insensitive, it is *unjustified*.)

3.4   *Defense of (13): Moral Claims Entail Categorical Reasons*

Call the thesis that all positive moral claims entail that *some particular* categorical reason exists

"the entailment claim." The entailment claim entails (13), since (i) if all positive moral claims entail

the existence of some particular categorical reason, then they entail that at least one categorical

reason exists, and (ii) an argument that all positive moral claims entail the existence of at least one

categorical reason is also an argument that we are justified in believing this to be true. So I need to

---

[4] To avoid problems resulting from the metaphysical necessity of some moral claims, we should understand the
counterfactuals here to quantify over all *conceptually* possible worlds, as in Clarke-Doane (2012): 320-321.

show here that, if moral realism is true, then the entailment claim is true. I'll give two arguments to this effect. But first let me clarify the entailment claim. Consider the following four attenuations of it:

- The entailment claim does not mean that each positive moral claim entails a categorical reason to *do* anything. I might have a categorical reason to *feel* a certain way, or to *desire* something in particular. E.g., that some act is generous may entail that I have a categorical reason to like when people perform that act.

- The entailment claim does not mean that each attribution of a given moral property M entails a categorical reason to do (or feel, desire, etc.) the same thing in every situation. That some act is M may give me a categorical reason to Φ in one situation and Ψ in another. E.g., that some act is generous may in one instance give me categorical reason to perform it; in another circumstance it may give me categorical reason to help someone perform it.

- The entailment claim does not mean that each positive moral claim entails an *all-things-considered* categorical reason. Positive moral claims often entail only *pro tanto* reasons. E.g., that some act is generous may give me *pro tanto* categorical reason to perform it, although I have *trumping* reasons to refrain.

- The entailment claim does not mean that any agent with a positive moral belief also believes that some categorical reason exists. For all it says, even agents who are fully competent with moral concepts may be unaware of the entailment claim.

So in sum, all positive moral claims entail a *pro tanto* categorical reason to Φ, where Φ is not necessarily an action and can vary across situations for a given moral claim. Finally, even where a positive moral claim is true, it is possible that not all agents who are competent with moral concepts will be aware of this entailment. Since each positive moral claim entails that *some particular* categorical

reason obtains, each also entails that at least one categorical reason obtains. So, if the entailment claim is true, then (13) is true.

There has been some dispute over the entailment claim—mostly within the debate over moral error theory. There are many who think that the entailment claim is true—that, in fact, it is a conceptual truth (Olson 2011; Joyce 2001: 175-177).[5] Yet there are also some who deny it (Foot 1972; Finlay 2008). Such denial is not worrisome for my argument here, because the arguments against the entailment claim rely on premises inimical to moral realism (as in Finlay 2008). My claim is that *if moral realism is true*, then the entailment claim is true (although not necessarily conceptually true).

I'll give two arguments for the entailment claim. The first is the weakest, and it comes cheaply. Consider: what is it that makes moral claims *moral*, and not some other type of claim? It can't be their normativity: prudential claims are normative, as are some epistemic claims. Nor can it be that they are other-directed: many claims about etiquette are other-directed. The entailment claim provides an answer: what makes moral claims moral is that they entail categorical reasons. I will leave this as a speculation (although see Joyce 2011, who agrees with me on this point)—whether the distinction between moral and non-moral claims is important, and whether there are other plausible ways to draw the distinction, are beyond my ken.

What I consider a second, stronger argument for the entailment claim relies directly on the commitments of moral realism. Without the entailment claim, we cannot explain the type of authority that moral claims have, *according to realists*. This is clearest in the types of cases that Enoch is considering. Sufficiently Bad Bad-Guy has reason not to hurt Victim. But according to the moral realist, his reason not to hurt Victim is not conditional on his having an interest in Victim's

---

[5] Outside of debates about error theory, I think Parfit (2011: 283-288) is committed to the entailment claim, and Enoch (2011: 94) expresses sympathy for it.

continued well-being, or on his desire not to see anyone suffer, or on any judgment that he makes. He shouldn't do it *no matter what*! Even claims about what would be virtuous, or what states of affairs are good, have distinctive authority over us. Whatever the generosity of an action gives you reason to do/feel/think/etc., this reason is not contingent on your having any specific desires or judgments. This authority that moral claims have over us, according to realists, is explained only by their entailing categorical reasons (for similar points, see Parfit 2011: 283-288 and Joyce 2001: ch. 2).

Think of the matter a different way, via example. At the very *least*, if moral realism is true, then when some act is generous one has categorical reason to perform it in certain counterfactual situations. Surely if some act is generous, then *ceteris paribus*—i.e., where no energy is lost on your part by performing it that wouldn't otherwise be lost, where you sacrifice nothing that you want by performing it, etc.—you have a categorical reason to perform it. In the specified situation, anyone, *no matter what they desire or judge*, has a reason to act generously. Similarly, assume that the absence of physical conflict is good. If we consider a situation where, by the press of a button, you can make it the case that physical conflict is avoided, and where it otherwise wouldn't be, and where furthermore there are no competing goods that are sacrificed by one's pushing the button—well, don't we have reason, no matter what we desire, to push the button? If so, then if moral realism is true, even such positive moral claims as these entail the existence of a categorical reason.

Anti-realists are not similarly committed to the entailment claim—at least not all of them. Consider Streetian constructivism: the fact that A has a reason to Φ is constituted by the fact that the proposition "A has a reason to Φ" is entailed by A's other judgments about reasons, the non-normative facts, and the constitutive standards for judgments about reasons (Street 2009; Street 2010; Street 2012; Street forthcoming). Street-type constructivism actually entails the falsity of the entailment claim, when that claim is read according to CR_2. According to the Street-type constructivist, if moral claims entail reasons at all, they can't be categorical reasons—all of A's

reasons are *constituted* by facts about what normative judgments she makes, and so clearly do not obtain regardless of what judgments she makes. The upshot of all this is that moral realists are committed to the entailment claim, and thus (13), while at least some anti-realist views—I suspect most—are not.

But perhaps, after all that I've said, there are still those who think that they can be moral realists without committing to the entailment claim. Yet even if so, the New EDA will simply apply to the type of realism committed to the entailment claim. And any type of realism that is not so committed will have a new problem: that of either explaining the authority of moral claims, or explaining away the appearance of such.

3.5   *Defense of (14): The Epistemological Business*

We can't get the conclusion of the New EDA without some principle linking the defeat of belief in an entailed claim with lack of knowledge of the claim doing the entailing. In other words, we need (14): If P entails Q, we are justified in believing that P entails Q, and our belief that Q is defeated, then we do not know that P. However, that claim is not only plausible on its face, but it follows from two other epistemological premises, each even more plausible than (14).

Here's the argument:

23. If my belief that P is defeated, then I do not know that P.
24. If P entails Q, we are justified in believing that P entails Q, and our belief that Q is defeated, then our belief that P is defeated.
25. Therefore, if P entails Q, and our belief that Q is defeated, then we do not know that P.

There are controversial claims that rely on (23)—e.g., attempts to provide necessary and sufficient conditions for knowledge. But (23) merely asserts a necessary condition on knowledge, and that is not controversial at all (Klein 1971; Lehrer and Paxson 1969). It serves as a good explanation of why, in at least some Gettier cases, the agent does not have knowledge: there's some true proposition that, if the agent justifiably believed it, would entail that her belief is unjustified.

(24) is simply a closure principle for defeat, and it seems as plausible as any non-trivial epistemological principle. Here's an example that may help to illustrate that plausibility. Suppose I look at the clock and, seeing that it reads "2:14," conclude that it is afternoon. But the clock is broken, which defeats my belief that it is afternoon. I also believe that it is 2:14 p.m., which entails that it is afternoon. (24) says that, in this situation, my belief that it is 2:14 p.m. is also defeated.

I see no need to defend (24) further, since it is so plausible. But it just so happens that there is a valid argument for it, which only depends on a closure principle for justification:

Closure$_J$:   If P entails Q, and I am justified in believing that P entails Q, then if my belief that Q is unjustified, then my belief that P is unjustified.

I leave the defense of Closure$_J$ to others (Hawthorne 2004: ch. 1). Suffice it to say that it is as plausible a (non-trivial) principle as they come in epistemology.

The argument for (24) is fairly complex. In what follows, recall that when A's belief that P is defeated, there is some true proposition D such that, if A justifiably believed that D, then A would be unjustified in believing that P. The following argument actually shows something more specific than (24), but which entails (24)—that the same defeater that defeats belief that Q also defeats belief that P.

26. P entails Q, I am justified in believing this, and my belief that Q is defeated.
    (Assumption)
27. P entails Q.
    (&Elim 26)
28. There is some true proposition D and if I justifiably believed that D, then I would not be justified in believing that Q.
    (from 26)
29. There is some true proposition D.
    (&Elim 28)
30. If I justifiably believed that D, then I would not be justified in believing that Q.
    (&Elim 28)
31. I am justified in believing that D.
    (Assumption)
32. I am not justified in believing that Q.
    (→Elim 30, 31)

33. If P entails Q, I am justified I believing this, and I am not justified in believing that Q, then I am not justified in believing that P.
      (Closure_J)
34. I am justified in believing that P entails Q.
      (&Elim 26)
35. P entails Q, I am justified in believing this, and I am not justified in believing that Q.
      (&Intro 27, 32, 34)
36. I am not justified in believing that P.
      (→Elim 33, 35)
37. If I am justified in believing that D, then I am not justified in believing that P.
      (→Intro 31, 36)
38. There is some true proposition D and if I am justified in believing that D, then I am not justified in believing that P (i.e., my belief that P is defeated).
      (&Intro 29, 37)
39. Therefore, if P entails Q, I am justified in believing this, and my belief that Q is defeated, then my belief that P is defeated.
      (→Intro 26, 38)

The argument is sound so long as Closure_J is true—as it almost surely is.

In this section, I've defended the premises of the New EDA. I believe that defense has been successful, though merely suggestive at points (e.g., the empirical business). So at this point, it seems that realists ought to worry about whether the empirical premise is true.

4. Revisiting the Objections to the Standard EDA

In this section, I argue that on top of the New EDA's independent plausibility, it has far better replies than the Standard EDA to the four objections above. I'll argue that the New EDA has decisive replies to the first two objections, while it has distinct advantages over the Standard EDA with regard to the latter two.

4.1   The New EDA's Replies to the First Two Objections

Recall the limited explanation objection: it hinged on the claim that there are some moral beliefs that do not easily admit of an evolutionary explanation. However, the New EDA can grant this. It's empirical premise claims only the belief that categorical reasons exist has been strongly influenced by evolution. The Standard EDA's empirical premise, on the other hand, claims that

many different (moral) beliefs have been so influenced. If evolution has strongly influenced our moral faculties *in doxastically discriminating ways*, then we should expect (at the very least) most of our moral beliefs to be adaptive. This makes it easy for the realist to find moral beliefs that are not likely to have emerged as a result of adaptive pressures, thus causing the standard debunker worry in proportion to the number of non-adaptive beliefs on display. The empirical premise of the New EDA is not open to such criticism. It relies only on the claim that *one* belief is adaptive, and it turns out to be clear how that belief would be adaptive. Let the contents of positive moral beliefs be what they may—e.g., what exactly we ought to do, or what acts are generous—the New EDA still entails that there is no moral knowledge.

The independent confirmation objection, recall, was that we can independently confirm a doxastic faculty if we can show that it is identical to or a species of a type of faculty we know to be reliable, regardless of whether our moral faculties are reliable. It is clear why this objection does not apply to the New EDA: I am not targeting our moral faculties at all! Rather, I claim that a *particular belief* is defeated, since it was formed in a way that is not sensitive to the truth, and I derive the lack of positive moral knowledge from the defeat of this one belief. So Shafer-Landau's objection just does not apply to the New EDA.

One might think that my reply here is cheap, and that regardless of whether the objection applies to the New EDA, Shafer-Landau can show that our moral faculties are reliable. Doesn't the reliability of our moral faculties outweigh (in some sense) the defeat of something entailed by our positive moral beliefs? If so, we could have positive moral knowledge even where our positive moral beliefs entailed a defeated belief. But this is no good. Even if we grant that our moral faculties are a species of some generally reliable faculty, this gives us merely *pro tanto* reason to trust them. If our moral beliefs are defeated—as I have argued they are—then we no longer have reason to trust them. As the plane crash survivor trudges through the desert, he may have *pro tanto* reason to trust his

beliefs about the convenience store he seems to see up ahead, since he knows that his faculty of vision is generally reliable. But that reason can be undermined if he learns that this particular belief is defeated—here the defeater is the claim that in circumstances such as those our survivor is in, we often seem to see things that aren't there. In the same way, even if Shafer-Landau is right, his objection does nothing to threaten the New EDA. We could start with *pro tanto* reason to trust our moral faculties, but if I'm right, we still have no positive moral knowledge.

4.2   The New EDA's Reply to the Overgeneralization Objection

The New EDA has an important advantage over the Standard EDA in replying to the overgeneralization objection. Recall the objection: if evolutionary pressures undermine moral knowledge on the assumption of moral realism, then they also undermine perceptual (or mathematical) knowledge on the assumption of perceptual (or mathematical) realism. It might seem that both EDAs are equally susceptible to the objection: after all, both claim that evolutionary influence of some kind undermines knowledge. And neither EDA has some special capacity to deny evolutionary influence on our perceptual or mathematical faculties or beliefs. So, perhaps both EDAs overgeneralize, if either does.

However, I think that the New EDA has at least the following advantage in replying to the objection: whereas the Standard EDA targets our moral faculties in general, the New EDA targets a particular belief. But it seems that, if there has been evolutionary influence in the perceptual or mathematical realm, it has largely been influence on the respective *faculties*. No particular perceptual belief seems to have been selected for (e.g., that there is a table in front of me): such beliefs seem far too fine-grained to be the object of direct evolutionary influence. Similarly for mathematical beliefs: e.g., the belief that $679 - 456 = 223$ is not directly selected for. And this means that the Standard EDA overgeneralizes to these realms, since it claims that evolutionary influence on a *faculty* means that the beliefs produced by that faculty don't count as knowledge. The New EDA, on the other

hand, claims only that *direct* evolutionary influence on a belief means that that very belief does not count as knowledge. It can allow that *indirect* evolutionary influence on a belief (via influence on the faculty that produced that belief) doesn't undermine knowledge.

One might worry that the New EDA will still overgeneralize if perceptual or mathematical claims entailed the existence of a categorical reason. But such an entailment does not obtain. That 2 + 2 = 4 does not entail the existence of a categorical reason. Neither does the claim that there is a table in front of me. These are just not the *kind* of facts that entail any kind of reasons: whereas claims about reasons are normative, these are purely non-normative claims. And of course, it's a commonplace in ethics (and normative theory in general) that a normative claim does not follow from a purely non-normative claim.

For these reasons, I think that the New EDA has a better reply to the overgeneralization objection than does the Standard EDA. But unlike the New EDA's responses to the first two objections, the response here does not seem utterly conclusive. So I claim only that the New EDA's reply is superior to the Standard EDA's, not that it is ultimately successful.

4.3   The New EDA's Reply to Third Factor Views

Recall the general form of a third factor response: each generally starts by assuming a substantive moral principle. It then shows how evolution predictably brings us to have (at least some) true moral beliefs, even though it doesn't select for the *truth* of those beliefs. I am not convinced that either EDA has a knock-down reply to this type of objection. However, it seems to me that the most promising reply available to either EDA works much better for the New EDA than for the Standard EDA, and in this respect the former is better off than the latter.

If any EDA is to stand against third factor responses, it must be because those responses *beg the question* against the relevant EDA. Several philosophers have argued that this is indeed the case, with regard to the Standard EDA (Shafer-Landau 2012: 33-34; Behrends 2013: 7-8; Horn

forthcoming; Vavova 2014: 81; Vavova 2015: 111-112). It's easiest to see why when we look at

Enoch's view, which assumes that survival is good. Belief in this claim has a clear evolutionary

explanation (this is, in fact, integral to the success of Enoch's reply). This belief is thus part of the

target of the Standard EDA: it is such claims that the argument attempts to show that we can't

know, if realism is true. But surely we shouldn't crucially rely on our belief that P, when replying to

an argument that purports to show that we don't know that P! So, Enoch shouldn't rely on the claim

that survival is good, since the Standard EDA attempts to show that he doesn't know this.

Consider the following analogy: you know that taking a certain pill will cause you to believe

that Napoleon won the Battle of Waterloo (Joyce 2007: 179-180). You know that you have taken

such a pill. Are you rationally permitted to rely on your belief that Napoleon won Waterloo in

assessing whether you know that Napoleon won Waterloo? Of course not—to do so would assume

the very knowledge in question. For the same reason, we cannot assume the truth of any positive

moral claim when assessing whether we have any positive moral knowledge.

My goal here is not to prove that this strategy for replying to third factor views is ultimately

viable. That would require fuller argument. But I do hope it's clear that this is the best option for the

standard debunker to take in replying to third factor views. The jig is up, so to speak, if the standard

debunker allows her critic to appeal to some positive moral claims in his response (although for an

independent line of objection, see Joyce 2016).

However, when applying this argumentative strategy, the standard debunker will run into a

problem that the new debunker won't. To see why, let's look to Wielenberg's third factor response.

His substantive moral principle is that any person with C has rights. But notice: this is precisely the

sort of claim that seems incapable of direct evolutionary explanation, since it's unclear why it would

be adaptive to believe it. In fact, it might be downright *dis*advantageous: if I have the belief in

question, in many circumstances I will put my kin on equal standing with those who are completely

unrelated to me, lowering the chances of passing on my genes. In this way, the standard debunker's appeal to the question-begging strategy depends crucially on her ability to reply to the limited explanation objection.

To state the point modestly: it is unclear whether evolution has had a strong influence on our belief that all beings with C have rights. And it seems to me that, in such a situation, it is permissible to rely on this belief in replying to the Standard EDA. Return to our analogy: Wielenberg's case is like not having good reason to believe that you have taken the Napoleon Pill. In such a situation, it is plausible that you *can* rationally rely on your belief that Napoleon won Waterloo. So, the Standard EDA is susceptible to at least one third factor response *even if we grant that others (like Enoch's) are question-begging.*

However, we have seen that the New EDA targets all moral beliefs and is therefore immune to the limited explanation objection. For this reason, both third factor responses under consideration will beg the question against it, if either does. So will most third factor views, since most assume a positive moral claim. But the New EDA has it that all moral claims entail a categorical reason, which (long story short) means that we can't have knowledge of any positive moral claim. So, all third factor views that assume a substantive normative claim—including both Wielenberg's and Enoch's—will beg the question against the New EDA by assuming what it purports to disprove: that we have any positive moral knowledge. So in sum, the Standard EDA has a vulnerability to third factor views that the New EDA does not. It is a derivative vulnerability— deriving from the Standard EDA's vulnerability to the limited explanation objection—but a vulnerability nonetheless.

The New EDA's reply here is not decisive. It depends on whether the question-begging reply works in general; it is also vulnerable to any third factor response which can manage not to assume a positive moral claim, such as Behrends (2013). But the New EDA has a very important

advantage over the Standard EDA in replying to third factor views, just as it does in replying to all of the objections mentioned here.

Conclusion

The fact that evolution has strongly influenced our moral beliefs seems worrying, at first glance, if the moral facts don't depend constitutively on our attitudes. However, there are some deep problems that face the standard formulation of the argument. The New EDA, which I have presented here, gives an independently plausible line of argument that avoids many of the problems with the Standard EDA. If that argument works, then moral realism entails moral skepticism—an intolerably high price for realists to pay. I cannot say here conclusively whether the argument succeeds—the state of empirical work is incomplete, and there are further questions to be answered regarding the overgeneralization and third factor objections. What I can say, however, is that this is a seriously worrisome argument for realists—far more worrisome than the Standard EDA.[6]

References

Bedke, Matthew. (2009) 'Intuitive Non-Naturalism Meets Cosmic Coincidence'. Pacific Philosophical Quarterly, 90, 188-209.

Behrends, Jeff. (2013) 'Meta-normative Realism, Evolution, and Our Reasons to Survive'. Pacific Philosophical Quarterly, 94, 486-502.

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

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

Copp, David. (2008) 'Darwinian Skepticism About Moral Realism'. Philosophical Issues, 18, 186-206.

Enoch, David. (2010) 'The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It'. Philosophical Studies, 148, 413-38.

Enoch, David. (2011) Taking Morality Seriously. Oxford: Oxford University Press.

Finlay, Stephen. (2008) 'The Error in the Error Theory.' The Australasian Journal of Philosophy, 86, 347-69.

Fitzpatrick, William. (2014) 'Why There is No Darwinian Dilemma for Ethical Realism'. In Michael Bergmann and Patrick Kain (eds.), Challenges to Moral and Religious Belief (Oxford: Oxford University Press), 237-55.

Foot, Phillippa. (1972) 'Morality as a System of Hypothetical Imperatives'. Philosophical Review, 81, 305-16.

Greene, Joshua. (2008) 'The Secret Joke of Kant's Soul'. In Walter Sinnott-Armstrong (ed.), Moral Psychology, Vol. 3: The Neuroscience of Morality, Emotions, Brain Disorders and Development (Cambridge, MA: MIT Press), 35-80.

Hawthorne, John. (2004) Knowledge and Lotteries. Oxford: Oxford University Press.

Horn, Justin. (Forthcoming) 'Evolution and the Epistemological Challenge to Moral Realism'. In Michael Ruse and Robert Richard (eds.), The Cambridge Companion to Evolutionary Ethics (Cambridge: Cambridge University Press).

Huemer, Michael. (2008a) Ethical Intuitionism. New York: Palgrave McMillan.

Huemer, Michael. (2008b) 'Revisionary Intuitionism'. Social Philosophy and Policy, 25, 368-92.

James, Scott. (2011) An Introduction to Evolutionary Ethics. Chichester: Wiley-Blackwell.

Joyce, Richard. (2001) The Myth of Morality. Cambridge: Cambridge University Press.

Joyce, Richard. (2007) The Evolution of Morality. Cambridge, MA: MIT Press.

Joyce, Richard. (2013) 'Ethics and Evolution'. In Hugh Lafollette and Ingmar Persson (eds.) The Blackwell Guide to Ethical Theory, 2nd edition (Malden: Blackwell), 123-47.

Joyce, Richard. (2016) 'Evolution, Truth-Tracking, and Moral Skepticism'. In his Essays in Moral Skepticism (Oxford: Oxford University Press), 142-58.

Kitcher, Philip. (2007) 'Biology and Ethics'. In David Copp (ed.), The Oxford Handbook of Ethical Theory (Oxford: Oxford University Press), 163-85.

Klein, Peter. (1971) 'A Proposed Definition of Propositional Knowledge'. The Journal of Philosophy, 68, 471-82.

Lehrer, Keith and Thomas Paxson, Jr. (1969) 'Knowledge: Undefeated Justified True Belief'. The Journal of Philosophy, 66, 225-37.

Parfit, Derek. (2011) On What Matters. Oxford: Oxford University Press.

Plantinga, Alvin. (1993) Warrant and Proper Function. New York: Oxford University Press.

Olson, Jonas. (2011) 'In Defense of Moral Error Theory'. In Michael Brady (ed.), New Waves in Metaethics (Aldershot: Palgrave McMillan), 62-84.

Ruse, Michael and E.O. Wilson. (1986) 'Moral Philosophy as Applied Science'. Philosophy, 61, 173-92.

Shafer-Landau, Russ. (2009) 'A Defense of Categorical Reasons'. Proceedings of the Aristotelian Society, 109, 189-206.

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

Singer, Peter. (2005) 'Ethics and Intuitions'. The Journal of Ethics, 9, 331-52.

Skarsaune, Knut. (2011) 'Darwin and Moral Realism: Survival of the Iffiest'. Philosophical Studies, 152, 229-43.

Street, Sharon. (2006) 'A Darwinian Dilemma for Realist Theories of Value'. Philosophical Studies, 127, 109-66.

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

Street, Sharon. (2010) 'What is Constructivism in Ethics and Metaethics?'. Philosophy Compass, 5, 363-84.

Street, Sharon. (Forthcoming a) 'How to Be a Relativist About Normativity'. Journal of the American Philosophical Association.

Street, Sharon. (Forthcoming b) 'Objectivity and Truth: You'd Better Rethink It'. In Russ Shafer Landau (ed.), Oxford Studies in Metaethics, vol. 11 (Oxford: Oxford University Press).

Vavova, Katia. (2014) 'Debunking Evolutionary Debunking'. In Russ Shafer Landau (ed.) Oxford Studies in Metaethics, vol. 9 (Oxford: Oxford University Press), 76-101.

Vavova, Katia. (2015) 'Evolutionary Debunking of Moral Realism'. Philosophy Compass, 5, 104-16.

Wielenberg, Erik. (2010) 'On the Evolutionary Debunking of Morality'. Ethics, 120, 441-64.

Wielenberg, Erik. (2014) Robust Ethics. Oxford: Oxford University Press.

Williams, Bernard. (1981) 'Internal and External Reasons'. In his *Moral Luck* (Cambridge: Cambridge University Press), 101-14.