Are Ambitious Evolutionary Debunking Arguments Self-Refuting?

Caner Turan
American University of Beirut

Evolutionary debunking arguments (henceforth, EDAs) present an epistemic challenge for evaluative/moral realism. The idea is that, since (i) evolutionary forces shape extensively the content of our evaluative/moral beliefs, and (ii) evolutionary mechanisms aim at reproductive success rather than attitude-independent evaluative/moral truths, evaluative/moral realism is an epistemically untenable metanormative (or metaethical) view. A possible response to the challenge is to assert that ambitious EDAs are self-refuting: since they take *all* of our evaluative/moral beliefs to be epistemically suspect, they cannot provide independent reason to believe that our evaluative/moral beliefs are debunked by the argument. My aim in this paper is to show how to plausibly respond to such an objection. To do that, I first lay out the epistemic challenge EDAs pose for evaluative/moral realism and briefly explain Sharon Street's and Richard Joyce's arguments (Section 1). I then discuss Katia Vavova's objection that ambitious EDAs are self-refuting (Section 2). I argue, contra Vavova, that the level of ambition of an EDA does *not* affect the strength of its *epistemic* premise because EDAs are essentially *inductive* arguments. Rather, the level of ambition of an EDA affects the strength of its *empirical* premise, which is the Achilles heel of any ambitious EDA.

1. Evolutionary Challenge for Realism

1.1. The Structure of an Evolutionary Debunking Argument

Evolutionary debunking arguments claim to undermine the justification of our evaluative beliefs by placing a special focus on the evolutionary origins of them.

Some of such arguments are more ambitious than the others as they try to undermine the justification of *all* evaluative beliefs,[1] while some of them are targeted at moral beliefs only,[2] and some at a certain subset of moral beliefs.[3] All EDAs, however, take a common form. They all claim that knowledge of a certain subset of beliefs is improbable, since (i) such beliefs are shaped exclusively by the mechanisms of natural selection and (ii) evolutionary processes aim at reproductive success and thus are insensitive to attitude-independent evaluative truths, if there are any. The former is the *empirical* premise, and the latter is the *insensitivity* premise. EDAs also have an *epistemic* premise, namely that if non-naturalist evaluative (or moral) realism,[4] the empirical premise, and the insensitivity premise are true, then we cannot justify the beliefs in question. These three premises constitute the blueprint of any EDA:

1. *Empirical premise.* Evolutionary mechanisms have a pervasive influence on the content of our evaluative/moral beliefs.
2. *Insensitivity premise.* Evolutionary mechanisms aim at reproductive success and not attitude-independent evaluative/moral truths.
3. *Epistemic premise.* If there are attitude-independent evaluative/moral truths, evolutionary mechanisms have a pervasive influence on the content of our evaluative/moral beliefs, and evolutionary mechanisms aim at reproductive success and not attitude-independent evaluative/moral truths, then we lack an independent reason to think that our evaluative/moral beliefs track the truth, i.e., we lack justification for our evaluative/moral beliefs.
4. *Skeptical conclusion.* We lack knowledge of attitude-independent evaluative/moral truths, if they exist at all.

The epistemic premise is the core of any debunking argument. There are many ways of forming beliefs. Think, for example, of people who rest their beliefs about an outcome of a football game, or about whether it is going to rain the following day on the behavior of animal oracles. Although their beliefs might turn out to be true, they are only incidentally true since animal behavior has nothing to do with the states of affairs in a football game or with the state of the atmosphere. Hence, we have a good reason to suppose that people who form their beliefs through a process that is not good at tracking the truth – as in the case of animal oracles – are not justified in their beliefs. Such processes are called "off-track." (Kahane 106) Hearing, on the other hand, is most of the time an epistemically reliable process, which means that it is good at tracking the truth. People whose beliefs are informed by their hearing mechanism are probably correct in their beliefs about what they hear, provided that they don't have an impaired hearing mechanism and no environmental factor is distorting their beliefs. For example, the fact that I hear Beethoven's Eighth Symphony in the radio at the moment is a good justification

for my belief that Beethoven's Eighth Symphony is now playing in the radio, provided that there is nothing unusual about my hearing mechanism and nothing in my surroundings affects its proper function. EDAs, therefore, are based on the crucial distinction between processes that track the truth and off-track processes.

1.2. Street and Joyce

According to Sharon Street's empirical premise, evolutionary processes have an enormous influence on the content of our evaluative beliefs. Although evolutionary forces do not *directly* determine the content of our evaluative beliefs, they select for "basic behavioral and motivational tendencies" (Street 113), which in turn play an important role in shaping the content of these beliefs. Street then presents her Darwinian Dilemma: either there is a relation between evolutionary influences on our evaluative beliefs and independent moral truths or there is not. If there is no causal connection between evolutionary mechanisms and moral truths, then evolutionary forces probably have a distorting influence on our evaluative beliefs, which are off-track. It would be an enormous and inexplicable coincidence to claim that our evolutionarily shaped (thus distorted) evaluative beliefs are true. Assuming a relation between selective pressures and evaluative truths is not a good strategy for the realist either, since the realist must appeal to a *tracking account* to explain the nature of that relation. Tracking account loses the scientific battle against what Street calls the *adaptive link account* because of its metaphysical commitments and its failure to explain why true evaluative beliefs promote survival (Street 126–30).

Joyce's EDA aims exclusively at our moral beliefs. His empirical premise is that we have a complete and "confirmed non-moral genealogy" (Joyce 190). Joyce's insensitivity premise claims that the best explanation of our basic moral judgments is that "they are expressions of underlying 'design features' of human psychology" (140). Just as we cannot justify our beliefs about the battle of Waterloo if they are caused by an imaginary belief pill that is insensitive to the facts about Waterloo, so we cannot justify our moral beliefs if they are generated by biological processes that are insensitive to proposed moral facts or truths. And Joyce's conclusion is that "we have no reason to believe in moral facts" (210).

The important difference between Joyce's and Street's respective EDAs is that while Joyce thinks his EDA supports moral skepticism, Street thinks her EDA ultimately supports the truth of anti-realism. Joyce adopts a realist semantics *only* about moral propositions – that is, only moral propositions refer to categorical reasons. Street, on the other hand, believes that *all* evaluative discourse should be viewed in realist terms. Since realism itself is an evaluative claim, we "must be to adjust our metaethical view so as to become antirealists" (Street 141). Thus, Street's EDA implies not only moral skepticism but also complete evaluative skepticism.

2. Ambition and Strength

It is important to note that if the debunking argument places the burden of proof on the realist, the argument then collapses into a pervasive skepticism about *all* of our beliefs.[5] The debunker's aim is to undermine a limited set of beliefs using scientific evidence. But if the debunker's argument asks the realist to provide an independent reason to think that their beliefs are not mistaken, then the empirical premise becomes superfluous and the argument's target extends to our entire body of belief.[6] We may never provide good (independent) reasons for the truth of *any* of our beliefs due to *some* possible distorting factor, but this general epistemic worry has never been the debunker's concern. Rather, the debunker is concerned with the rationality of our evaluative or moral beliefs. To this end, she gives scientific evidence that gives us reason to think that our evaluative or moral beliefs are mistaken due to the distorting influence of evolutionary forces on such beliefs. Thus, the burden of proof must be on the debunker: she needs to give us evidence of error (a good reason that we are probably mistaken in our evaluative or moral beliefs) that follows from an empirical claim (evolutionary evidence), rather than asking the realist to give independent reason that our beliefs are not mistaken. Only then she can selectively claim that we cannot rationally maintain our evaluative or moral beliefs.

When formulating the empirical premise, the debunker determines the *target* of her argument: empirical evidence could show that *all* of our *evaluative* beliefs have been shaped by evolutionary processes, or it could show that *all* of our *moral* beliefs are determined by such processes. Street's ambitious EDA chooses the former target, while Joyce's less ambitious EDA chooses the latter. On Vavova's account, Street's argument is less likely to succeed than Joyce's argument due to the epistemic principle that she calls "the Inverse Rule of Debunking." However, she thinks that both accounts are ambitious enough to fail.

2.1. The Inverse Rule of Debunking

Vavova's claim is that the more ambitious a debunking argument becomes the less prospect of success it has. This is because the debunker tries to give us good reasons to think that we are mistaken about a certain body of beliefs, and what makes a reason good is its independence from the set of beliefs that are called into doubt. For example, if the aim of your argument is to undermine all perceptual beliefs, it would become illegitimate to base any of your premises on the truth of your beliefs that are formed through your senses. Street's EDA calls all of our evaluative beliefs into doubt; thus, the independent ground that reveals our mistake in our evaluative beliefs cannot involve any evaluative claim. Joyce's EDA can employ our nonmoral evaluative beliefs as an independent ground for the evidence of error, since its target encompasses all of our moral beliefs. Vavova calls this epistemic principle "The Inverse Rule of Debunking," according to which "[t]he

potential strength of a debunking argument is inversely proportional to its ambition" (Vavova 98).

Although I don't see any compelling reason to reject the principle,[7] I argue that ambition of an EDA does not really affect the strength of its *epistemic* premise. Rather, the relation between ambition and strength of an EDA becomes relevant only with respect to the scope of evolutionary influence on the content of our beliefs, since EDAs are essentially inductive arguments. The right strategy to debunk the debunker is thus to attack the *empirical* premise of her argument, which determines inductive strength of any EDA.

2.2. Self-Refutation Argument

Since the debunker's aim is not to deductively prove that our evaluative or moral beliefs are *necessarily* wrong but merely to show that such beliefs are *probably* wrong given our evolutionary and epistemic conditions, EDAs are *inductive* arguments. An inductive argument infers from a limited number of observations to a general, probabilistic conclusion. For instance, when we reach the conclusion that the sun will probably rise tomorrow from the set of observations that sun has risen regularly so far, we give an inductive argument. Similarly, an EDA reaches the conclusion that our evaluative or moral beliefs are *probably* wrong from an evolutionary explanation of such beliefs. When we are given an EDA, we realize that there is a discrepancy between what we (or realists, to be more specific) take evaluative/moral judgments to be and how evolutionary psychology describes them.[8] We also make a distinction between objective and subjective judgments.[9] From these observations we infer that our evaluative/moral beliefs are probably mistaken, just as we infer that the sun will probably rise tomorrow. Making such inferences is just one of the things our minds naturally do.

When we infer the probabilistic conclusion of an EDA from the scientific evidence it provides, we make an assumption about epistemic reasons, namely that scientific evidence has the power to undermine our intuitions. But what makes us believe that scientific evidence is epistemically more reliable than our intuitions? It is perfectly possible that having this evaluative claim is also an adaptation. Wouldn't this then threaten the kind of EDA that calls *all* of our evaluative beliefs into doubt? This is Vavova's objection to Street's EDA (87–9). Vavova argues that Street's EDA targets both practical and epistemic reasons, both of which have been shaped by natural selection. The idea is that if we cannot trust any of our evaluative beliefs, then we cannot trust our beliefs about whether our evaluative beliefs are debunked by the argument. "[T]o evaluate we must rely on the evaluative" (Vavova 89); however, if the argument aims to undermine *all* evaluative judgments, then we lack the resources to determine whether the targeted beliefs are debunked. Hence, Vavova concludes, Street's EDA is self-refuting.

I don't think Vavova's strategy delivers a knockout blow to Street's argument. Recall that an EDA is an inductive argument and that inductive arguments allow

for their conclusion to be false due to their probabilistic nature. Street does not claim that our evaluative beliefs are *necessarily* wrong but that they are *probably* wrong given our evolutionary and epistemic conditions. The conclusion of Street's EDA allows the possibility that some of our evaluative beliefs turn out to be true, and it follows that these true evaluative beliefs could include some of our beliefs about epistemic reasons, science, mathematics, and so on.

As a matter of fact, our beliefs about epistemic reasons are more likely to be true compared to our beliefs about concrete evaluative matters, even though the former are also beliefs. Our beliefs about epistemic reasons are *beliefs about beliefs* because they are about whether our beliefs are epistemically trustworthy. But *beliefs about beliefs* are categorically different from *beliefs about specific cases*. Notice the difference between two kinds of statements: (1) "Our evaluative beliefs are probably false unless supported by empirical evidence," and "Our scientific beliefs are probably false unless supported by empirical evidence;" (2) "My wife is beautiful," and "The earth is flat." The former kind of statements is more likely to be a product of *reasoning* rather than *biological conditioning*, and thus it is more likely to be true. This is because the reasoning that is involved in the former kind of statements forces one to distance themselves from their (possibly distorted) beliefs about specific cases and make them realize that such beliefs are epistemically vulnerable. And this process allows room for belief revision.

Moreover, there is a gap between the evolutionary origins of something and its truth: the fact that our genes partly determine our acceptance of evolutionary theory does not make evolutionary theory wrong, or the fact that our genes caused us to engage in mathematics does not render mathematical propositions wrong. Similarly, conceding that our beliefs about epistemic reasons have been shaped by selective pressures does not entail that those beliefs and the probabilistic conclusion of Street's EDA are wrong.

It is difficult to declare an EDA to be self-refuting unless its epistemic premise renders our beliefs about epistemic reasons wrong. I am pointing to the difference between the statements "All *X*s are wrong" and "All *X*s are *probably* wrong," when both statements are themselves instances of *X*. It is more difficult to call the latter self-refuting because it leaves open the possibility of itself (or any other instance of *X*) being true. Also, even self-refuting statements or arguments can give us important clues about the issue being discussed. Take, for example, Mark Twain's famous statement "All generalizations are false, including this one." It is one way to approach this generalization about generalizations (or belief about beliefs) by claiming that it is self-refuting and thus it does not establish *anything*. It is also possible to take it to show ultimately that *some* generalizations are true. If the statement is true, then all generalizations but the statement itself are false. If the statement is false, then some generalizations are true. In both cases, some generalizations are true. Likewise, it is possible to take the statement "All of our

evaluative beliefs are *probably* false, including this one" to show ultimately that *some* of our evaluative beliefs are true.

What about Joyce's EDA? According to Vavova, EDAs that target moral beliefs, such as Joyce's EDA, still target too much (90–3). Such EDAs claim that we are probably mistaken about morality because our moral beliefs have been shaped by natural selection, which is an off-track process. However, Vavova argues, to show that there is no relation between adaptive moral beliefs and true moral beliefs, we first need to know something about the contents of true moral beliefs and adaptive moral beliefs. And this requires us to make assumptions about what morality is like. Otherwise, morality could be about anything and accordingly we would have no reason to think that adaptive moral beliefs and true moral beliefs do not coincide. EDAs like Joyce's make such assumptions (e.g., moral judgments are rationally authoritative) but they at the same time call our entire body of moral beliefs into doubt. They thereby render their own moral assumption illegitimate. Thus, EDAs that declare *all* of our moral beliefs to be epistemically suspect cannot give us independent reason to think that we are mistaken about morality.

I don't think such an objection refutes Joyce's EDA. Since the epistemic premise of an EDA takes the form of a *reductio*, it is essential for any EDA to assume something about morality or about normative domain in general. Otherwise, it would not go through. If your aim is to debunk moral realism, you should first assume that moral realism is true. If your aim is to debunk our moral beliefs altogether, you should first make an assumption about basic commitments and presuppositions to morality. The epistemic premise is a conditional: *If* our moral beliefs have such-and-such features, then they are *probably* wrong, considering the extensive influence of evolutionary forces on the content of those beliefs and the insensitivity of evolutionary processes to the truth of them.[10]
Moral beliefs could, of course, have different features than it is assumed by the argument. It is possible that the correct account of morality is an *anti-realist* one. In that case, only the assumed conception of morality, that is moral realism, could be debunked. This is where Joyce seems to go wrong. He takes the conclusion of his argument to have debunked morality in general. However, he thereby dismisses anti-realist conceptions of morality that could escape the evolutionary challenge.[11] Although there seems to be nothing wrong with making a metaethical *assumption* to get the argument going, dismissing alternative conceptions of morality could possibly create a problem for the debunker, if their intention is to debunk morality as a whole.

2.3. Ambition and Inductive Strength

I have argued that the level of ambition of an EDA does not affect the strength of its *epistemic* premise because EDAs are essentially inductive arguments. Focusing on the epistemic premise and declaring more ambitious EDAs to be self-defeating do not remove the skeptical worry that we might be mistaken in our

evaluative/moral beliefs. As long as one admits that our beliefs are heavily shaped by natural selection, it is natural and plausible to think that objective morality could simply be an illusion. And this worry remains even if we think the argument is self-defeating. Does this mean there is no relationship whatsoever between an EDA's ambition and strength?

There *is* such a relationship, but the level of ambition only affects the strength of the *empirical* premise. The only way to ease the skeptical worry seems to be to show that there is *no extensive* evolutionary influence on our beliefs. Many philosophers and evolutionary biologists agree that certain capacities and tendencies relevant to evaluative thought and behavior, and *some* of the content of our evaluative beliefs *can* be explained by evolution. However, there is much less agreement among them on whether evolutionary forces have a *pervasive* influence on the content of our evaluative beliefs. The idea is that the effects of *human culture* and *moral reasoning* on the contents of our evaluative beliefs can be thought of independently from the effects of biological evolution on such beliefs.[12] I cannot go into detail and review the literature due to space limitation; however, for our purposes, it will suffice to say that the more ambitious an EDA is the more difficult it is to provide a complete evolutionary explanation of the beliefs in question.

If the debunker can show conclusively that her empirical premise is true, her argument will get very strong. However, the more sets of beliefs she claims to have determined by our biological nature, the *more difficult* it gets to provide a complete evolutionary origins story. For instance, it would be more difficult for a debunker who tries to debunk all of our evaluative beliefs to prove her empirical premise than a debunker who aims only at our moral beliefs. But provided that both debunkers succeed in their respective tasks, the former debunker's argument would get inductively stronger than the latter debunker's argument, for her empirical premise would encompass a greater number of beliefs. Thus, the level of ambition of an EDA has an effect on the strength of its *empirical* premise, and accordingly on the inductive strength of the argument.

3. Concluding Remarks

I have rejected Vavova's claim that Street's and Joyce's EDAs target too much and become self-refuting. This is because EDAs are essentially *inductive* arguments with a *probabilistic* conclusion, which allows *some* of our beliefs to be true. Furthermore, even a self-refuting debunking argument could indicate the truth of a subset of what is being debunked. The "Inverse Rule of Debunking" is true, but not for the reasons Vavova provides. In other words, the level of an EDA's ambition affects the argument's strength but not because more ambition causes an internal contradiction. Rather, the more set of beliefs an EDA calls into question the harder it becomes to provide a complete evolutionary origins story. The

empirical premise is the chink in an EDA's armor.

NOTES

1. Sharon Street's EDA in her 2006 paper is generally considered as an example of such an ambitious argument (cf. Kahane 114–7; Das 430–2; Vavova 80–1).
2. See Richard Joyce.
3. See Joshua Greene.
4. Typically, non-naturalist evaluative or moral realism is committed to three claims: (1) a claim about moral judgments being truth-apt (cf. Sayre-McCord 5), (2) a claim about moral principles being attitude-independent (cf. Shafer-Landau 15), and (3) a claim about the moral facts obtaining (ibid.). The first claim distinguishes realism from non-cognitivist accounts, according to which moral judgments do not try to describe some feature of the world but they are expressions of our emotional or other non-propositional attitudes. The second claim distinguishes realism from relativist accounts, which assert that moral truth is constructed by individual preferences or social conventions. And the third claim distinguishes realism from nihilistic accounts that reject moral truth and value altogether. When I use the term 'realism' throughout the paper I refer to *non-naturalist* realism, unless stated otherwise.
5. I agree with Vavova (82–4) on this point.
6. No empirical premise is needed for the following argument to go through: (P_1) There are infinitely many possible states of belief; (P_2) One of those states of belief is correct; (P_3) It is unreasonable to adopt a belief if we don't have good reason for our belief being correct; (P_4) We have no good reason to think that we are in the correct state. (C) It is unreasonable to believe that we are in the correct state. An EDA that places the burden of proof on the realist takes this specific form and ultimately asks the realist to justify *all* of their beliefs, not just their evaluative or moral beliefs (cf. Vavova 83–4).
7. In fact, the principle reflects the nature of moral and political disagreements very well. Think of two people (person *A* and person *B*) you disagree with on a moral matter *M*. Imagine that you agree with the person *A* on many other moral matters, whereas you disagree with the person *B* on *all* moral matters. Your disagreement with *A* has a significant debunking potential since you have a common independent ground from which you can provide evidence of error. Your disagreement with *B* has no such debunking potential. It is not even possible to evaluate your disagreement with *B* as you lack a common independent ground to move forward (cf. Vavova 99).
8. According to the realist, evaluative/moral claims refer to *categorical* (attitude-independent) reasons. However, evolutionary psychology tells us that such claims are based on evolved psychological dispositions that favor adaptive behaviors. That is, we form our evaluative/moral beliefs not as a response to attitude-independent evaluative/moral truths but simply because they promote reproductive success. This conflicts with the presupposed categorical nature of morality.
9. Any argument that moves from scientific evidence to evidence of error in our beliefs indicates the distinction between objective and subjective judgments or standpoints. For example, there is an important difference between the statements "It is cold out here"

and "It is 20 degrees Fahrenheit," even though both refer to the same state of affairs. While the former reflects a *subjective* outcome of an interaction between one's body and a state of affairs (some species could feel warm in those conditions), the latter reflects an *objective* fact. Facts about temperature and the fact that evolutionary forces shape psychological dispositions (both in humans and other primates) that favor adaptive behaviors are thus *objective*, whereas beliefs such as "It is cold out here" or "It is wrong not to help distant strangers" are *subjective* in this context.

10. To reiterate, the epistemic premise and the conclusion it leads to allow the possibility of some of our moral beliefs' being true. And these true moral beliefs could include beliefs about the nature of moral beliefs. Nothing in the argument precludes this possibility.

11. Street does not fall into this trap and acknowledges that anti-realist conceptions of morality are safe against her EDA (152–4).

12. Cf. Nagel 142; Sober 93–113; Ayala 258–60; FitzPatrick; Prinz.

WORKS CITED

Ayala, Francisco J. *Evolution, Explanation, Ethics and Aesthetics: Towards a Philosophy of Biology*. Academic Press, 2016.

Das, Ramon. "Evolutionary Debunking of Morality: Epistemological or Metaphysical?" *Philosophical Studies*, vol. 173, no. 2, 2016, pp. 417–35. *SpringerLink,* doi: 10.1007/s11098-015-0499-9.

FitzPatrick, William J. "Why There Is No Darwinian Dilemma for Ethical Realism." *Challenges to Moral and Religious Belief*, edited by Michael Bergmann and Patrick Kain, Oxford University Press, 2014, pp. 237–55.

Greene, Joshua. "The Secret Joke of Kant's Soul." *Moral Psychology, vol. 3*, edited by W. Sinnott-Armstrong, MIT Press, 2007, pp. 35–79.

Joyce, Richard. *The Evolution of Morality*. MIT Press, 2006.

Kahane, Guy. "Evolutionary Debunking Arguments." *Noûs*, vol. 45, no. 1, 2011, pp. 103–25. *Wiley Online Library*, doi: 10.1111/j.1468-0068.2010.00770.x.

Nagel, Thomas. "Ethics without Biology." *Mortal Questions*, Cambridge University Press, 1979, pp. 142–46. *Cambridge Core*, doi: 10.1017/CBO9781107341050.012.

Prinz, Jesse J. "Is Morality Innate?" *Moral Psychology, The Evolution of Morality: Adaptations and Innateness*, vol. 1, edited by Walter Sinnott-Armstrong, MIT Press, 2008, pp. 367–406.

Sayre-McCord, Geoffrey. "The Many Moral Realisms." *Southern Journal of Philosophy*, vol. 24, Supplement, 1986, pp. 1–22. *Wiley Online Library*, doi: 10.1111/j.2041-6962.1986.tb01593.x.

Shafer-Landau, Russ. *Moral Realism: A Defence*. Oxford University Press, 2003.

Sober, Elliott. *From a Biological Point of View: Essays in Evolutionary Philosophy*. Cambridge University Press, 1994.

Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies*, vol. 127, no. 1, 2006, pp. 109–66. *SpringerLink*, doi: 10.1007/s11098-005-1726-6.

Vavova, Katia. "Debunking Evolutionary Debunking." *Oxford Studies in Metaethics*, vol. 9, edited by Russ Shafer-Landau, Oxford University Press, 2014, pp. 76–101.