Philosophia
https://doi.org/10.1007/s11406-024-00725-5

Evolutionary Debunking and the Folk/Theoretical Distinction

M. Scarfone[1]

Received: 8 March 2023 / Revised: 12 February 2024 / Accepted: 20 February 2024

Abstract
In metaethics, evolutionary debunking arguments combine empirical and epistemological premises to purportedly show that our moral judgments are unjustified. One objection to these arguments has been to distinguish between those judgments that evolutionary influence might undermine versus those that it does not. This response is powerful but not well understood. In this paper I flesh out the response by drawing upon a familiar distinction in the natural sciences, where it is common to distinguish folk judgments from theoretical judgments. I argue that this in turn illuminates the proper scope of the evolutionary debunking argument, but not in an obvious way: it is a very specific type of undermining argument that targets those theories where theoretical judgments are inferred merely from folk judgments. One upshot of this conclusion is that it reveals a verboten methodology in metaethics. The evolutionary debunking argument is therefore much less powerful than its proponents have supposed, but it nevertheless rules out what is perhaps a common way of attempting to justify moral judgments.

Keywords Evolutionary debunking arguments · Moral epistemology · Folk morality

1 Introduction

Evolutionary debunking arguments (EDAs) are sceptical challenges that have captured considerable attention.[1] In metaethics, one aim of EDAs is to show that, owing to some to-be-specified evolutionary influence, our moral judgments are unjustified. Characterizing EDAs is a matter of considerable debate, but the formulation I will

[1] For an overview of this literature as it relates to morality, see FitzPatrick (2016, especially Sect. 4) and Vavova (2015).

✉ M. Scarfone
mpscarf1@gmail.com

[1] Department of Philosophy, University of Toronto, Toronto, ON, Canada

Published online: 14 March 2024

focus on goes something like this: (1) if evolution has influenced our moral faculties in a particular way, then we should not think that our moral judgments are justified; (2) evolution *has* so influenced our moral faculties; (3) so, we should not think that our moral judgments are justified. Because (1) is a claim about when justification is defeated it is an *epistemological* premise (e.g., Hanson, 2017; Isserow, 2019; Lutz, 2018; Moon, 2017; Schechter, 2018; Sterelny & Fraser, 2016; Vavova, 2015, 2018). And because (2) is a claim about our evolutionary history it is an *empirical* premise (e.g., Bloom, 2013; Churchland, 2011; de Waal, 1996, 2006; Gazzaniga, 2005; Greene, 2013; Hauser, 2006; Joyce, 2006; Kitcher, 2011; Tomasello, 2016). It is the combination of epistemological and empirical premises that has made EDAs intriguing to a wide range of academics, from ethicists to epistemologists, and from psychologists to primatologists. So-called 'debunkers' think that the evolutionary etiology of our moral judgments produces an undermining effect, while 'anti-debunkers' do not.

There is a common anti-debunker response that says though evolutionary influences would render some of our moral judgments unjustified, we can nevertheless make other moral judgments that are justified (Brosnan, 2011; Copp, 2008; FitzPatrick, 2015; Parfit, 2011, 2017; Toner, 2011). Call this the *Not All Judgments* response. In §2 I explain that while this seems to be a powerful response it is one whose structure has not been adequately fleshed out.[2] To remedy this, in §3 I suggest a plausible way of characterizing the difference between these types of judgements. On the one hand, we have *folk moral judgments*, which are the sorts of judgments that evolutionary pressures plausibly would have exerted significant distorting influence upon. And on the other hand, we have *theoretical moral judgments*, which are not. This folk/theoretical distinction may be familiar from the natural sciences, and I argue that it can help us flesh out the Not All Judgments response.

However, the analysis I offer here is not one where folk judgments are unjustified and theoretical judgments are justified. Rather, introducing the folk/theoretical distinction helps show a better way to understand EDAs. They do not actually undermine justification for our moral judgments according to all metaethical theories, but rather undermine a very specific way of explaining such justification. In §4 I show that what I call Inference Debunking constitutes a potent defeater, albeit one with a very specific scope – it undermines the inference *from* our folk judgments *to* our theoretical judgments. To fill out the picture, I show some recent metaethical theories that use this verboten inference. An important upshot of my characterization of EDAs is that it opens up future work to be done showing exactly which metaethical theories violate Inference Debunking and which do not. So in metaethics, EDAs are much less powerful than many debunkers and even anti-debunkers have supposed, but a richer understanding of their structure rules out what are perhaps common ways of attempting to justify moral judgments.

[2] The emphasis on structure is important here. As I explain below, there are a few extant accounts that make use of a Not All Judgments response, but they tend to be first-order ethical positions and not methodological accounts examining and explaining the difference between types of judgments.

2 What kind of argument is the EDA?

One way to home in on the target of EDAs is by asking *which* moral judgments in particular are suspect. It makes a difference whether EDAs call into question our ability to make any justified first-order moral judgments at all, or whether some specific subset of our moral judgments is impugned. This is because if EDAs are understood as *justification* debunking arguments, then no matter how moral judgments are construed by any metaethical theory they will be unjustified.[3] But if EDAs are understood as *theory* debunking arguments, then whether a moral judgment is impugned will depend upon how a particular metaethical theory characterizes these judgments.[4]

In addition to the distinction between theory debunking and justification debunking, work on EDAs has distinguished between the *contents* of our moral judgments and the *capacities* or abilities for making those judgments.[5] Suppose that natural selection favored those ancestors who made some specific moral judgment j. If one group of early hominids that made only moral judgment j outproduced another group of early hominids who made only moral judgment j* then natural selection might have favored the former hominids *because of* the specific content of their moral judgment. And if we are still disposed to make moral judgment j, that judgment may be suspect because it had been 'merely' advantageous and not necessarily because it tracks moral truth. There would then be an undercutting defeater for thinking that j is true.[6] When glossed in this way, the EDA is a *content etiology argument* which attempts to impugn some (but not all) of our moral judgments.

On the other hand, a *capacity etiology argument* seeds doubt concerning our ability to make any justified moral judgments. Again, if one group of early hominids that was disposed to make any moral judgments outproduced another group of early

[3] In lieu of the justification vs. theory debunking distinction, a reviewer asks instead whether something like Joyce's 'modest' debunking would be more appropriate here. As I understand it, a modest debunking argument "will allow the possibility that justification may be (re)instated" once removed, while a much stronger debunking argument would show that the "removal of justification would be permanent; nothing could reinstate it" (2016: 125). I opt for the justification vs theory debunking distinction because my Inference Debunking reading of the debunking project views theories themselves as suspect. I am drawing attention to particular metaethical theories (e.g., the moral fixed points view, and Metaethical Mooreanism), and the way they incorporate folk judgments, rather than looking at whether justification simpliciter can be restored after it has been undermined. For more on justification debunking, and how it differs from theory debunking, see Joyce (2014) and Sinclair (2018).

[4] For a paradigmatic theory debunking argument, see Street (2006). In brief, Street argues that moral realists in particular are unable to square our evolutionary history with the purported connection between our moral judgments and the moral facts. This leaves open the possibility that other metaethical views can win the day. See also Bedke (2009), whose construal of the EDA targets only nonnaturalist versions of moral realism. For a paradigmatic justification debunking argument, see Joyce (2001, 2006). Unlike Street and Bedke, Joyce's EDA targets all attempts at justifying our moral judgments, realist or antirealist, naturalist or nonnaturalist, and so on.

[5] The content/capacity distinction can then correspond to particular characterizations of an EDA, which will make salient those respective features. See for example FitzPatrick (2015).

[6] An *undercutting* defeater, rather than a *rebutting* defeater. EDAs purportedly serve to undermine thinking 'j is true', but it does not directly support thinking that 'j is true' is false. See Pollock (1986).

hominids *without* that disposition, then natural selection might have favored the ability in the former hominids for making any moral judgments in the first place. So explained, EDAs seem to impugn all of our moral judgments.[7]

Combined with the distinction between theory debunking and justification debunking, casting the purported undermining effect of EDAs in terms of content or else capacity helps to narrow our focus.[8] But while important, these distinctions alone still cannot settle the question about which first-order moral judgments are problematic. And there is a common anti-debunking strategy that exploits this lacuna. What we can call the *Not All Judgments* (NAJ) response begins by granting to the debunker that some to-be-specified evolutionary influence has probably played a distorting role on our moral judgments, in virtue of their content or else our capacity to make them at all. But, the NAJ response continues, even accounting for such distortion, we might nevertheless possess the ability to make justified moral judgments.

I think we can identify three key claims of the NAJ response. The first claim is that there are different first-order moral judgments that we can make, some of which might be unjustified but others which are justified.[9] The second claim is that we should focus on the basic mental capacities for moral judgments, as opposed to the contents of those judgments.[10] And the third claim is that philosophical training (or something like it) can allow us to make justified first-order moral judgments using our evolved capacities.[11]

[7] Moreover, when EDAs are glossed in terms of an epistemological premise, an empirical premise, and a conclusion, it is more plausible to read the empirical premise as about *capacity* rather than *content*. See Kahane (2011) and Sinclair (2018) for discussions of EDAs structured in precisely this way.

[8] Even though the EDA is a capacity etiology argument, it can nevertheless implicate contents in a derivative way (and the reverse isn't true). One way would be if an impugned capacity always issued contents. I thank a reviewer for drawing this to my attention. Here two things are important: first, the evolutionary influence works on the capacity, not the contents; and second, this derivative effect will only spell trouble for a particular way of interpreting the EDA (i.e., what in §4 I call Folk and Theoretical Judgment Debunking) which I don't think we have reason to accept.

[9] For example, Derek Parfit writes that while our capacities for first-order moral judgment "were partly produced by evolutionary forces […] these abilities later ceased to be governed by these forces, and had their own effects" (Parfit 2011: 520). For Parfit, some of our judgments are made owing to pushes from our past, while other judgments are made because our abilities for such judgments have been sufficiently modified. Similarly, William FitzPatrick says that "it's enough if natural selection has given us general cognitive capacities that we can now develop and deploy in rich cultural contexts, with training in relevant methodologies, so as to arrive at justified and accurate beliefs in that domain" (FitzPatrick 2015: 5–6).

[10] In brief, think here of approaches from Singer (2005), de Lazari-Radek & Singer (2012), and Greene (2008): they argue that deontological judgments are produced by an off-track process. This makes clear that they are offering *capacity*-debunking accounts: in their view, the capacity that produces deontological judgments is not a capacity that tracks moral truth. This view is then accompanied by a positive proposal that is similar to a NAJ, but one that is instead positioned squarely within first-order ethics—namely, the capacity that produces utilitarian judgments is not similarly off-track. In essence, then, if our moral judgments are produced by 'capacity x' then they are unjustified, but if they are instead produced by 'capacity y' then they can be justified. See footnote 12 for a problem with these attempts.

[11] These latter two features are brought out best by FitzPatrick. He writes: "the basic mental capacities that enable us to sit around worrying about things like metaphysical modality are part of our evolutionary heritage: they didn't appear by chance and they weren't designed by God; they evolved through natural selection. But natural selection did not design our cognitive capacities to track truths about metaphysical

The NAJ response allows anti-debunkers to grant the empirical premise, namely that evolution has influenced our capacities for making moral judgments. It also allows them to accept that there is some truth to the epistemological premise, namely that some such influence can undermine justification. Nevertheless, despite those concessions, anti-debunkers argue that we are able to make justified moral judgments. But anti-debunkers who rely on the NAJ response owe us a richer account of how this will work.[12] Aside from suggestions about the possibility of philosophical training allowing us to have justified moral judgments, the details here are thin. The above characterizations do not offer a principled way of determining which judgments are justified and which are not. I will attempt to remedy this in the next section.

The case I will make is as follows. By availing ourselves of the distinction between folk and theoretical judgments, we can see the proper scope of EDAs: they are not *justification* debunking arguments, but rather are *theory* debunking arguments. In particular, they undermine those theories where theoretical judgments are inferred merely from folk judgments. For that reason, I characterize EDAs as an example of Inference Debunking. This allows us to pick out the unjustified moral judgments more precisely – namely, they are those judgments that we have in virtue of the verboten inference identified. This analysis brings two important upshots: first, EDAs are shown to be much less powerful than some debunkers have supposed (e.g., those with aims of justification debunking); and second, EDAs are shown to be much more powerful than some anti-debunkers have supposed, because it rules out what might be a common way of attempting to justify moral judgments (by inferring them merely from our folk judgments).

Footnote 11 (continued)

necessity … [But] we're able to deploy those capacities, in the cultural context of philosophical training, to think intelligently and often accurately about things like metaphysical necessity or countless other arcane topics such as differential geometry and relativistic quantum theory, the facts of which are equally irrelevant to the etiology of the capacities we use in thinking about them" (FitzPatrick 2015: 887). See also Toner (2011), who suggests that our capacity for moral judgments has been "co-opted… for purposes other than natural selection" (Toner 2011: 529). Cf. Copp (2008) and Brosnan (2011). Vlerick and Broadbent (2015) suggest something similar regarding evolutionary influence on epistemic reliability and justification, which is more general than justification regarding moral judgments.

[12] To be sure, there are some extant proposals that make use of an NAJ response regarding first-order ethics. In footnote 10, I referenced arguments that say if our moral judgments are produced by 'capacity x' then they are unjustified, but if they are instead produced by 'capacity y' then they can be justified, e.g. Singer (2005), de Lazari-Radek & Singer (2012), and Greene (2008). The claim here is that deontological judgments are off-track because they respond to morally irrelevant features (like emotions, or proximity to victims, etc.), but utilitarian judgments can be on-track because they do not. But as Berker (2008) convincingly shows, such 'argument[s] from morally irrelevant features' beg the question at a crucial juncture, because "what's doing all the work in the argument from morally irrelevant factors is… [an] invocation, from the armchair, of a substantive intuition about what sorts of factors out there in the world are and are not morally relevant" (2008: 326). That is, it is the invocation that some factors are 'morally irrelevant' that allows one to conclude that some capacity for moral judgment is off-track. Berker's counterargument is that there is no non-question-beginning way to show this. Thus, relying on an 'argument from morally irrelevant features' cannot be of help here.

3 The Folk/Theoretical Distinction

The NAJ response says that, though our capacities for moral judgment evolved, and seem to be subject to being undermined by EDAs, when combined with philosophical training those capacities can be used to arrive at justified moral beliefs. To spell out this response, I suggest we begin by drawing a distinction between our *folk moral judgments* and our *theoretical moral judgments*. The folk/theoretical distinction here mirrors the familiar one from the natural sciences, where it is common to distinguish between *pre-theoretical* or *folk judgments* and *theoretical* or *formal judgments*. For example, we now readily differentiate folk judgments about the movement of heavenly bodies and theoretical astrophysical judgments, between folk judgments about animal behaviour and theoretical judgments about biology, between folk science and science proper, and so on. I argue that we should also avail ourselves of this common distinction to characterize different types of moral judgments, and in doing so I think we in turn help explain the NAJ response.

In the natural sciences, folk judgments are typically characterized as "aris[ing] more informally and not as direct reflections of formal instruction in scientific principles" (Keil, 2010: 826; cf. Carey, 1988). Similarly, folk sciences are said to be those that "without explicit instruction in such areas, [lead] people… to develop domain-specific ways of thinking about relatively bounded sets of phenomena" (Keil, 2010: 826). We can capture these 'ways of thinking' about certain phenomena by referring to the folk theories or folk sciences from which our folk judgments issue. As Peter Keil notes, we have domain-specific ways of thinking about "the behavior of solid objects, living kinds, and the minds of others", or what we would commonly refer to as folk physics, folk biology, and folk psychology, respectively (Keil, 2010: 826; see also Medin & Atran, 2004: 926). Work has also been done to uncover judgments within folk chemistry (Au, 1994), folk cosmology (Siegal et al., 2004), folk economics (Lakshminaryanan et al., 2008), and so on. In brief, these folk sciences are those ways of thinking about physics, biology, psychology, and so on, without explicit instruction in those areas, not as direct reflections of formal instruction, while our folk judgments are just specific instances of thinking about those bounded sets of phenomena.[13]

What folk science and theoretical science (or science proper) both share is "a common goal of explaining real-world phenomena and of making predictions" (Keil, 2010: 834). But where folk sciences and folk judgments struggle is when it comes to articulation. Keil notes that people "are frequently unable to come up

[13] The existence of folk sciences and their accompanying folk judgments is well-established. There is an "emerging consensus about the existence of many folk sciences across all cultures that lead both to real successes at understanding the world and to misconceptions" (Keil, 2010: 828). Since by and large people are not trained physicists, biologists, or psychologists (let alone polymaths trained in all three disciplines), we do not often rely on our own robust theoretical knowledge of physical objects, plants and animals, and so on when navigating our environments. Instead, we rely on folk sciences in order to make judgments about these phenomena. While there is no exact label for this type of thinking, I use the phrase 'folk science'. Others refer to 'intuitive theories' or 'naive theories' (see Carey, 1985; Carey & Spelke, 1996; Slaughter & Gopnik, 1996; etc.), or "framework theories" (see Band et al., 2007).

with complete explanations of mechanisms, even for surprisingly simple systems" (Keil, 2010: 829). Regarding folk mechanics, when pressed we are often unable to explain the mechanical workings of bicycles, toilets, and other basic systems of which we may believe ourselves to have an intuitive and adequate grasp (Rozenblit & Keil, 2002). So the judgments within folk sciences are often "plagued with the problems of gaps, inconsistencies, and contradictions" (Keil, 2010: 830). Identifying the borders between folk sciences and formal or theoretical sciences proves to be a difficult task. In part owing to the likelihood of folk sciences being 'gappy' and relying on large hunches, there is motivation to move on and articulate a formal theory for certain domains — hence, formal or theoretical sciences. So we may distinguish folk sciences from theoretical sciences by noting that the latter are concerted efforts at reducing the gaps, taming the guesses, and minimizing deference within the former (cf. Keil, 2010: 834).

But in what sense does this count as 'moving on' from folk science? David Braddon-Mitchell suggests that folk theories are not theories that agents can necessarily write down. Rather, we have "evidence for the [folk or] tacit theory from the practice of agents in making their judgments and navigating around the world" (Braddon-Mitchell, 2004: 278). So *implicitness* can be taken as typical of a folk science, and making commitments *explicit* would then represent a move towards a formal science. This represents progress because when a theory is made explicit the judgments and commitments of that theory can be scrutinized. For example, we can examine whether our folk judgment about bird flight conflicts with our folk judgment about which animals (e.g. penguins) are birds. Subjecting claims to explicit scrutiny assists in our updating and correcting for the common errors of our folk judgments and theories. The more explicit the commitments, the more readily such scrutiny can take place.

Drawing our attention to this commonly used distinction in the natural sciences can be of help in analysing the NAJ response to EDAs. With this aim in mind, if there is an emerging consensus about the existence of folk biology, folk chemistry, and so on, might there also be *folk morality*? And if so, what would be the difference between folk morality and its accompanying folk moral judgments and a more *theoretical morality* and its accompanying theoretical moral judgments?

There has been some limited attention given to certain 'folk notions' being serviceable to philosophy in general, and to moral philosophy in particular. Frank Jackson, for one, offers a distinction between *folk morality* and what he calls *mature folk morality* (Jackson, 1998). For Jackson, 'folk morality' marks something like the characterization of folk science above: it regards those domain-specific judgments that we implicitly make absent explicit instruction. An example of a folk moral judgment might be of the form 'it would be right to φ in circumstance c'. And this folk judgment can have a particular conceptual profile, perhaps something like 'the rightness of φ-ing can be true or false'. So we might make a folk moral judgment and also implicitly think that it is truth-apt – just like we might make the folk biological judgment that 'this is a white oak tree' and the accompanying folk biological taxonomic judgment that white oak trees are a type of oak tree, which are a type of tree. In the moral case, both the folk judgment

itself and the conceptual profile for that judgment would be 'ways of thinking' about right and wrong action without training in moral philosophy.

On the other hand, 'mature folk morality' would represent an advance on those folk moral judgments because it springs up after reflection and negotiation regarding our folk judgments. In particular, mature folk morality would be a departure from folk morality insofar as it works towards settling the facts (in Jackson's view) or at least making explicit the commitments (in my view) regarding folk moral judgments. In our toy case, mature folk morality is a maturation of folk morality insofar as it can articulate and explain *whether* moral judgments are indeed truth-apt. This would of course require deliberate reflection about our moral discourse and practice, in much the same way that a formal or theoretical science is a deliberate examination of the relevant bounded set of phenomena.

It is important to recognize that it would be a serious mistake to assume that making explicit the commitments of folk morality will entail the vindication of those commitments. If we have folk moral judgments of the form that 'the rightness of φ-ing can be true or false', we should not expect that that our theoretical moral judgments will necessarily incorporate the claim that moral judgments *are* truth-apt. Doing so would radically underestimate the way in which theorizing can allow us to depart from folk morality. Recall that the move from folk science to formal science involves a concerted effort to eliminate gaps in understanding, reduce mere hunches, and so on. If that is right, then we should expect some daylight between folk science and formal science. Expecting formal biology to confirm those folk judgments underestimates the power of the theoretical sciences. Similarly, in the moral case, we should expect that some aspects of folk morality might not be preserved in our theoretical judgments — and it may be that the supposed truth-aptness of our moral judgments is one such judgment that gets left behind.[14]

When it comes to folk morality, suppose that we believe that 'hurting people merely for fun is morally wrong'. This would be a folk moral judgment if we did not have any specific moral theory in mind when holding this judgment (e.g. we did not have it because we thought hurting people reduces overall utility, or violates the inherent dignity of persons, etc.). It would be a folk moral judgment if we did not actually know that there are different moral theories that seek to explain why this judgment is justified (e.g. utilitarianism, deontology, etc.). We would instead have this judgment independent of explicit instruction in moral theorizing. Such folk beliefs about right and wrong "make up the core we need to share in order to count as speaking a common moral language" (Jackson, 2000: 132).

[14] For his part, Jackson does not appear to be drawing upon the common folk/theoretical distinction within the natural sciences. And while he doesn't spell out how we should understand folk morality in particular (in terms of, say, the necessity of folk judgments being implicit and absent explicit characterization for the relevant domain), he is nevertheless committed to the existence of folk moral judgments. Indeed, for Jackson such judgments are necessary in order to engage in moral philosophy at all. He suggests that our folk moral judgments are those pre-philosophical judgments that are "part and parcel of having a sense of what is right and wrong, and of being able to engage in meaningful debate about what ought to be done" (Jackson, 2000: 130). Note how this echoes the earlier characterization of folk sciences, which are said to arise informally and yet are heavily relied on for navigating our environments.

So what then is the role of moral philosophy? I think that moral philosophy is in large part an attempt to bring clarity and rigor to folk moral judgments, in much the same way as the sciences bring clarity and rigor to folk judgments. Because they constitute the shared basis for moral discussion, our folk moral judgments are what we often appeal to in debating moral matters. We ask and answer questions about them: are any of our folk moral judgments in tension? which judgments should we refrain from making? what follows from our moral judgments? and so on. Notice again the similarity between this reflection and negotiation and that of formal or theoretical biology, for example: we start with the bounded set of phenomena (say, plants and animals) and attempt to bring rigor and clarity to our thoughts about that domain by applying the scientific method. And yet, even if we perhaps by necessity appeal to our folk moral judgments when negotiating our lives with others, that does not render our folk judgments incapable of being revised — indeed, of even radical revision. This is because, due to our engaging in moral philosophy, "folk morality is currently under negotiation: its basic principles, and even many of its derived ones, are a matter of debate and are evolving as we argue about what to do" (Jackson, 2000: 132). That is, moral philosophy can offer a corrective or update on our folk moral judgments, just as easily as it might justify or support those judgments.[15]

Thus, the account for folk moral judgments looks similar to the account of folk judgments more generally. However, because Jackson does not mark the distinction between folk and theoretical judgments, some work needs to be done to extend the folk/theoretical distinction from the natural sciences to moral philosophy. Here we are helped by the initial characterization I offered above. While our folk moral judgments are domain-specific judgments absent explicit instruction, engaging in moral philosophy is itself a way of coming to make theoretical moral judgments — and that can only take place after explicit instruction. This also neatly echoes what both Parfit and FitzPatrick suggest is a way of correcting for evolutionary distortions on our moral judgments, i.e., the third critical feature of the NAJ response. While our folk moral judgments are implicit, our theoretical moral judgments are necessarily explicit. Because of this explicitness, the precepts of morality are readily available to reflection, analysis, and updating. In this way, folk moral theories and theoretical moral theories share a commonality with folk theories and formal theories more generally. As a result, there is no reason to suspect that all aspects of a folk theory will be preserved by the more theoretical one.

If the folk/theoretical distinction can indeed be extended to moral philosophy, then I think it can be used to help explain the NAJ response. The next section explores this suggestion.

[15] I offer some suggestions of what this corrective might look like at the end of §4.

4 Diagnosing the EDA

Debunkers think EDAs threaten our moral judgments. The challenge is then to explain why we should think that any of our moral judgments are true if those beliefs are likely to have been evolutionarily advantageous. If such advantage tracks evolutionary fitness and not necessarily truth, then we have an undercutting defeater for thinking that our moral judgments are in fact true. In response, anti-debunkers have urged that while *some* of our moral judgments seem susceptible to this undermining argument, *other* moral judgments are not. This NAJ response is a helpful start in answering the challenge posed by the EDA, but more needs to be said about the structure so we can explain which judgments are debunked and which are not.

With the folk/theoretical distinction in hand, there seem to be various ways to disambiguate our debunked/undebunked moral judgments. I will canvass three ways of drawing conclusions about the undermining effect of EDAs. That will allow us to see the strength of the NAJ response. I will present two flawed characterizations before settling on a plausible third.

First, some anti-debunkers may be interpreting EDAs as targeting only our folk moral judgments, and as a result they respond by saying that though our folk moral judgments might be unjustified our theoretical moral judgments are (or can be) justified. Call this interpretation:

> **Folk Judgment Debunking**: Our folk moral judgments (but not our theoretical moral judgments) are unjustified because of evolutionary considerations.

The NAJ response can initially seem effective here. These anti-debunkers might suggest that our folk moral judgments are not of central importance for warding off the undermining effect of EDAs: though our folk moral judgments might be unjustified, we may have other theoretical judgments that are not similarly impugned. Since there is some daylight between our folk judgments and our theoretical ones (as we saw in §3), there is a gap between impugning our folk moral judgments and undermining all moral judgments. If our theoretical moral judgments are what are *really* important, then characterizing EDAs as Folk Judgment Debunking poses little threat to finding some ultimate justification for our moral judgments. Or so an anti-debunker might argue.

But I think there is room for the debunker to respond to Folk Judgment Debunking. While one can accept that EDAs are *supposed* to target our folk moral judgments, and that it would in principle leave untouched our theoretical moral judgments, this ends up creating the following problem: even if the undermining effect is restricted to our folk moral judgments this would still be a very significant loss. Any list of our justified moral judgments would seem to be emaciated if it cannot include *any* of our folk moral judgments. Recall Parfit's examples: while evolutionary considerations might undermine some folk judgments, like 'that we have reason to care more about good or bad experiences when these experiences are in the future rather than the past', those same considerations do

not undermine other judgments like 'that everyone's well-being matters equally'. So characterizing EDAs as leading to Folk Judgment Debunking, results in the following problems: anti-debunkers would not be able to say, for example, that we truly have reason to care more about future experiences than past ones, nor would anti-debunkers be able to take for granted *any* of our folk moral judgments. In giving up our folk moral judgments anti-debunkers would be giving up too much. It would lead them to have an impoverished view of morality and of our moral judgments. They could only justifiably accept some claims, such as 'everyone's well-being matters equally', but not seemingly basic claims, such as 'we have more reason to care about future experiences than past ones'.

In light of this worry, the anti-debunker might respond in the following way: the folk/theoretical distinction does not necessarily rule out the *content* of our folk moral judgments, but rather the way in which one *arrives* at those judgments. Per Folk Judgment Debunking, though EDAs might undermine our folk moral judgments, precisely because they are the sort of judgments that evolutionary influences plausibly have exerted a distorting influence upon, our theoretical moral judgments might be similar or identical to those judgments, but which we arrive at in a justified way. For example, if we come to learn that a belief is the result of an unreliable formation process, that gives us a reason to be sceptical of the truth of that belief. But it of course doesn't show the falsity of that belief. Nor does it preclude our using a more reliable belief formation process to arrive at an in content similar yet justified belief.[16]

Another option might be that it is not the folk moral judgments that are unjustified, but rather both the folk *and* theoretical ones because the evolutionary distortion of our folk judgments necessarily carries through to our theoretical judgments. Contra NAJ anti-debunkers, there would be no way to correct for evolution's distorting influence. Characterizing the EDA in this way would lead to the following view:

> **Folk and Theoretical Judgment Debunking:** Both our folk moral judgments and our theoretical moral judgments are unjustified because of evolutionary considerations.

In particular, the idea here would be that *because* our folk judgments are debunked our theoretical judgments will be too. This would be a 'garbage-in, garbage out' objection, because our theoretical moral judgments would just be gussied up versions of debunked folk moral judgments. According to the debunker, evolutionary considerations reveal that our folk moral judgments are produced by an off-track process, and despite what NAJ proponents suggest there is no way that philosophical theorizing could correct for that.

[16] For example, if you learn that that the only reason you believe that Mercury is the closest planet to the Sun is because you were hypnotized to think so, you should not think that your belief is true. You should instead think that your belief is unjustified. But if you later learn, via a more truth-sensitive belief formation process, that Mercury *is* the closest planet to Sun, then you should think that your belief is true. Your initial unreliability would not necessarily preclude your later reliability.

But while characterizing the EDA as Folk and Theoretical Judgment Debunking avoids the limitation of Folk Judgment Debunking, this comes at the expense of scope. The problem for Folk and Theoretical Judgment Debunking is that either the undermining effect applies across the board of our cognitive capacities, or else it is restricted perhaps just to the capacities relevant to moral judgments. There are legitimate questions one could ask about whether our capacities or abilities for first-order moral judgments could truly 'cease to be governed' (as Parfit says) by evolutionary forces in the relevant way. Here one might think that if our relevant capacities are subject to EDAs, then they are now and forever rendered suspect (e.g., Greene, 2008; Kahane, 2011; Street, 2006). But 'garbage in, garbage out' objections quickly generalize, capturing not only the faculties that allow us to make moral judgments, but also those that allow us to think accurately about issues like metaphysical modality, differential geometry, relativistic quantum theory, or any complex issue (cf. FitzPatrick, 2015: 887).

Perhaps evolution does indeed have such a powerful and pervasive undermining effect on all our basic cognitive capacities. But I do not explore that much larger, more complicated thesis here. I restrict my analysis to our moral capacities alone, because of the following dilemma: either a wide range of judgments are impugned, or else moral judgments are arbitrarily impugned. If we embrace the first horn, we have to use a similarly impugned capacity in order to make that very argument; if we embrace the second horn, we are (currently) lacking an empirically informed account of how this works.

Regarding the first horn, let us suppose that the EDA undermines both our moral capacities and our capacities for higher-order mathematics. In that case the anti-debunker could say something like "Maybe this challenge applies to all of our capacities, I don't know. I'm just concerned with how it applies to our specifically moral ones." In this case, the EDA is simply a special case of a more general phenomenon, and restricting analysis to it alone would be fine. But consider a different higher-level capacity, namely our ability to figure out which arguments are valid and sound. Is this capacity subject to the EDA as well? That is, does the garbage-in, garbage-out objection apply broadly enough to capture this capacity for assessing arguments? The problem with thinking that it *does* apply to this capacity is this: since this is the very capacity one is relying on in making their argument, it seems to me that one cannot coherently say "Maybe the EDA applies to the very same capacities I'm using to determine whether the debunking argument is valid and sound. I don't know. I'm just concerned with whether it applies to our specifically moral capacities." So the first horn is problematic, because if the EDA applies too widely then it affects the very capacity used to make that argument. In this case, it's not that the EDA is a special case of a more general phenomenon but rather that the success of the EDA is being assessed by a faculty that has *ex hypothesi* been impugned.

Now, regarding the second horn, the question is whether we can say that our moral capacities alone have been impugned. To make the case for there being a non-arbitrary answer here, I think we need an empirically informed account of the evolutionary influence, one that spells out the specific details of how our capacity for

moral judgment alone has been undermined. This account would have to be careful not to be too broad as to be captured by the first horn of the argument. In short, we will need to avoid giving a 'just-so story' about a capacity that might be thought in play.[17] This is why the second horn is about arbitrariness. I think that we can stipulate or argue from the armchair that our moral judgments are thus-and so, but we cannot similarly stipulate or argue from the armchair about the specific details of evolutionary influence on our capacity for moral judgment.

All of our cognitive capacities have an evolutionary history, and with them the potential for distortions. Again, take our capacities for thinking about which arguments are valid and sound: it seems that we should accept that whatever capacities allow us to engage in that sort of thinking were also produced by evolution. But (plausibly) we can use our philosophical training to form justified judgments about which arguments are both valid and have true premises. If we cannot, then we are hopelessly out to lunch, having undermined our abilities for assessing the preceding arguments about which moral judgments are debunked/undebunked. This is why the NAJ response distinguishing between folk and theoretical judgments serves us well. That distinction allows us to acknowledge that some our folk judgments might be unjustified, and it also stresses that we can use our philosophical training to pick up the slack. Since this training is presumably what both debunkers and anti-debunkers are relying on, both should want it to escape the distorting influence of evolutionary pressures.

Of course, this does not yet show that philosophy *can* indeed pick up that slack. But that is not my aim here. My aim is to understand the structure of the NAJ response and what that reveals about the target of the EDA. And the two preceding options seem to show that either EDAs are bad arguments or else their strength lies elsewhere. Anti-debunkers suspect the former. But in the remainder of the paper I explore the latter.

While the folk/theoretical distinction is a helpful entryway for understanding the NAJ response, I think it also has the upshot of revealing a better way to characterize the target of EDAs. Rather than targeting our folk judgments (i.e., Folk Judgment Debunking), or both our folk and theoretical judgments (i.e., Folk and Theoretical Judgment Debunking), I suggest that EDAs are better understood as targeting the *inference from* our folk moral judgments *to* our theoretical moral judgments. That is, the undermining effect is not on a particular *set of judgments*, nor a *particular capacity* for making moral judgments, but instead on a particular *type of inference*. Consider the following gloss:

> **Inference Debunking:** If our theoretical moral judgments are inferred *merely from* our folk moral judgments then those theoretical moral judgments are unjustified.

Characterizing EDAs in terms of Inference Debunking has the following advantage: it makes salient that there should be a measure of independence between our

[17] This is often recognized as a verboten move not only in the empirical sciences (Hubalek 2021; Smith 2016) but also in the EDA literature (Joyce 2006; Kahane 2014; Shafer-Landau 2012).

folk moral judgments and our theoretical moral judgments. And this, I suggest, is exactly as things should be. Incorporating the folk/theoretical distinction from the natural sciences showed us that thinking otherwise would underestimate that difference between our folk theories and our more formal or theoretical sciences. Just as it would limit the epistemic power of our scientific theories if they were expected to merely preserve and confirm all and only our folk judgments about biology, chemistry, etc., we should also think that it limits our theoretical moral theories if they merely preserve our folk moral judgments. As we saw above, folk theories are 'gappy' and often get things wrong, which is part of the reason for developing more formal theories in the first place. So if we incorporate the folk/theoretical distinction into moral philosophy then we should allow for the possibility that some of our folk judgments are mistaken. But it would be inappropriate to conclude that philosophical training cannot escape these distortions. Again, since both debunkers and anti-debunkers are relying on their capacities for philosophical theorizing, both should want these capacities to escape the distorting influence of evolutionary pressures. Instead, we should view the undermining effect of evolutionary considerations as providing a block on the easy inference from folk judgments to theoretical ones.

When EDAs are understood in terms of Inference Debunking they are illuminating in the following way. Recall that there are at least two ways of understanding EDAs: as *theory* debunking arguments or as a *justification* debunking arguments. According to the account I have developed here, EDAs in metaethics are a specific type of *theory* debunking argument. This is because Inference Debunking does not rule out all attempts at explaining the justification of our moral judgments. Instead, EDAs as Inference Debunking only undermine a specific way of justifying moral judgments: it rules out those metaethical theories that say judgments are justified merely by inference from our folk moral judgments. This does not imply that there are no justified moral judgments, nor adequate metaethical theories. It only blocks one way of getting your metaethical theory: namely, by straightforwardly inferring from our folk moral judgments.[18]

The preceding thus identifies a particular type of verboten inference. While I think the question of which metaethical theories are targeted by Inference Debunking is open, here is an example of the kind of inference that I think is ruled out. Suppose that one has the folk moral judgment that 'harming people merely for fun is wrong', and that accompanying that judgment is the belief that even if no one believed it to be so, harming people merely for fun would *still* be wrong.[19] A

[18] There are other responses to EDAs that might look like Inference Debunking. For example, Bogardus (2016) argues that EDAs show how only moral beliefs based on some representational intermediary are unjustified, but since one need not go in for representationalism the scope of debunking is far narrower than debunkers suppose. According to Egeland, (2022), EDAs have an undermining effect when one's motivating reason for a particular moral belief is undercut by a normative reason which that person also genuinely possesses. The novelty of Inference Debunking, however, is the way in which it cuts across the pre-philosophical to post-philosophical divide, and by doing so it makes salient that there should be a measure of independence between our folk moral judgments and our theoretical moral judgments.

[19] There is some evidence from the social sciences that suggests non-philosophers understand moral judgments in this way. See Goodwin & Darley (2008).

metaethicist might point to these judgments and conclude that, therefore, moral facts are mind- and attitude-independent. But I suggest it is *this* sort of inference that is targeted by Inference Debunking: the inference *from* the folk moral judgment *to* the theoretical moral judgment.

Some metaethical accounts use this sort of verboten inference. One example is the 'moral fixed points' view developed by Cuneo and Shafer-Landau (2014). The view is a species of nonnaturalist moral realism, and their main thesis is that there are substantive moral claims which are also conceptual truths (Ibid., 400). As they explain, "these truths not only constitute any reasonably comprehensive moral system for beings such as us in a world such as ours, but also fix the boundaries of moral thought: one could not engage in competent moral thinking while rejecting them" (Ibid., 401). They give the following as some examples of such truths: it is wrong to break a promise simply for convenience's sake; the interests of others are sometimes morally weightier than our own; it is wrong to satisfy a mild desire if doing so requires killing many innocent people; and so on. The moral fixed points view goes on to say that metaethical theories must accommodate such truths within their moral systems. Because Cuneo and Shafer-Landau think that such truths fix the boundaries of morality, any system (for creatures like us, in a world like ours) must incorporate them to qualify as a moral system.

With that brief summary in mind, we can see that the moral fixed points view suggests that metaethicists must accommodate at least some folk moral judgments within their theoretical moral judgments. In particular, Cuneo and Shafer-Landau think metaethicists have to develop accounts that accommodate folk moral judgments, like the fact that 'it is wrong to break a promise simply for convenience's sake' because failing to do so entails that the account does not capture anything distinctively moral. But if this is the methodology suggested by the moral fixed points view then it seems subject to the EDA glossed as Inference Debunking. I've argued that there should instead be an important measure of independence between folk judgments and theoretical judgments, just as there is in the natural sciences, because doing so respects both the point and power of the theoretical sciences. A proposal to develop metaethical theories that simply accommodate folk judgments ignores the benefits gained by making our commitments explicit and thus open to negotiation. Again, moral philosophy can offer a corrective or update on our folk moral judgments just as readily as it might attempt to justify and support those judgments. But ignoring the 'corrective' or 'updating' part of moral philosophy and focusing only on the 'justifying' and 'supporting' downplays the power of philosophical theorizing. It risks turning metaethics into a matter of special pleading, an enterprise engaged in simply to confirm that which we already believe. This is not to suggest that a metaethical theory *cannot* accommodate any folk judgments (that would be the EDA construed as Folk Judgment Debunking). Rather, I am suggesting that a metaethical theory must do more than simply declare that certain judgments must be preserved just in virtue of them being folk judgments.

This verboten inference also shows up in Moorean metaethical theories. By 'Moorean' I mean those metaethical theories that argue that particular commonsensical moral claims are immune from sceptical challenges. While common across philosophy, there has been increased attention given to these sorts of arguments

from moral philosophers (e.g., Fuqua, 2021, Sampson forthcoming). In brief, if a sceptical challenge from moral nihilists concludes that there are no moral truths, a Moorean response would be to insist that there is at least one moral truth, such as 'recreational genocide is morally wrong' or 'it is good to save a friend from drowning'. The key Moorean move here is to argue that these commonsensical moral claims are far more plausible than any of the claims within the sceptical arguments. And since an argument is only as good as its weakest claim, Mooreans conclude that we always have less reason to believe the conclusions of sceptical arguments like moral nihilism as compared to particular commonsense moral claims.

However, notice that these 'commonsensical' moral claims will often be examples of folk moral judgments. Or, at least, a moral claim will likely *appear* more commonsensical if it is a folk moral judgment, i.e., the sort of judgment one has absent explicit instruction in moral philosophy. The further along the spectrum we go from *folk* to *theoretical*, the less commonsensical a claim will seem. If Mooreans say that a claim like 'it is good to save a friend from drowning' will inevitably be epistemically superior to any sceptical challenge, insofar as we will always have more reason to believe the former, then it is straightforwardly incorporating a folk judgment into a more theoretical account. Like Cuneo and Shafer-Landau's moral fixed-point view, Mooreans seem to think particular moral claims will necessarily be preserved in one's moral theory. Metaethical Mooreanism thus seems to erase any measure of independence between folk judgments and theoretical judgments. In doing so it overlooks the benefits to be gained by opening up our beliefs to scrutiny and negotiation. As before, this sort of approach to theorizing looks like special pleading, which to be sure is one important way of doing moral philosophy. But just as important are those modes of theorizing that attempt to offer an update or corrective for our folk beliefs. Moral philosophers should not rule out this latter approach *ex ante*.

This opens up an avenue for some interesting future work. Towards that end, it might be helpful to see what it could look like when theorizing could have a corrective effect on our folk judgments. While I am not committing to this having already happened in the moral domain, here's a non-moral example of what I have in mind. Let's suppose that your car's speedometer does not work properly, and trying to gauge your speed merely from the speedometer will not give you an accurate result. However, there is a way to correct for this problem if you know two other things: the motor's current RPM, and the size of the tires on the car.[20] In this case, I think our judgment about the car's speed has the right independence because the wonky speedometer does not directly figure into our judgment about the speed. Once we know the car's speed in that circuitous way, we might then be able to see whether the speedometer is useful in a secondary way.

But what does this look like in a moral case? Suppose that we have a commonsense judgment that 'it is good to save a friend from drowning', and a metaethical view that says we always have more reason to believe a commonsense claim than

[20] A vehicle's speed in MPH is equal to RPM x tire diameter x $\pi \times 60$ (i.e., minutes in an hour) / 63,360 (i.e. inches in a mile).

any argument seeking to undermine that claim. If the Inference Debunking construal of the EDA is correct, then we cannot make this easy inference. But this doesn't rule out any other way of getting that same result. Perhaps there is some work around for this claim, similar to the conversion from RPM and tire size to speed. What we're looking for is some other faculty, a non-moral one, to play a role in helping us figure out whether our moral judgments are on-track.

So it will be productive to examine exactly which theories violate Inference Debunking and which do not. For now, the above offers a way of filling out one common anti-debunker response to the EDA. The NAJ response suggests that, while some of our moral judgments are subject to the EDA and its undermining defeater, we might have other moral judgments that are not. I have argued that a plausible way of fleshing out this response is by distinguishing between our folk moral judgments and our theoretical moral judgments. Relying on this folk/theoretical distinction also has the upshot of showing that EDAs are a very specific type of theory debunking argument. They prohibit just those metaethical theories that rely on theoretical moral judgments inferred merely from our folk moral judgments.

Declarations

Conflict of interest The author did not receive support from any organization for the submitted work. The author declares that they have no conflict of interest.

References

Au, T. K. (1994). Developing an intuitive understanding of substance kinds. *Cognitive Psychology, 27*, 71–111.

Band, M., Medin, D. L., & Atran, S. (2007). Cultural mosaics and mental models of nature. *PNAS, 104*(35), 13868–13874.

Bedke, M. S. (2009). Intuitive non-naturalism meets cosmic coincidence. *Pacific Philosophical Quarterly, 90*(2), 188–209.

Bedke, M. (2014). No coincidence? *Oxford Studies in Metaethics, 9*, 102–125.

Berker, S. (2008). The normative insignificance of neuroscience. *Philosophy and Public Affairs, 37*, 293–329.

Berlin, B., Breedlove, D. E., & Raven, P. H. (1973). General principles of classification and nomenclature in folk biology. *American Anthropologist, 75*, 214–242.

Bertamini, M., Spooner, A., & Hecht, H. (2004). *The representation of naive knowledge about physics. Multidisciplinary approaches to visual representations and interpretations.* In G. Malcolm (Ed.). Elsevier.

Bloom, P. (2013). *Just babies: The origins of good and evil*. Crown Publishers.

Bogardus, T. (2016). Only all naturalists should worry about only one evolutionary debunking argument. *Ethics, 126*(3), 636–661.

Boulter, S. J. (2007). The 'evolutionary argument' and the Metaphilosophy of commonsense. *Biology and Philosophy, 22*(3), 369–382.

Braddon-Mitchell, D. (2004). Folk theories of the third kind. *Ratio, 17*(3), 277–293.

Brosnan, K. (2011). Do the evolutionary origins of our moral beliefs undermine moral knowledge? *Biology and Philosophy, 26*(1), 51–64.

Carey, S. (1985). *Conceptual change in childhood*. MIT Press.

Carey, S. (1988). Conceptual differences between children and adults. *Mind & Language*, *3*(3), 167–181.

Carey, S., & Spelke, M. (1996). Science and core knowledge. *Philosophy of Science, 63*(4), 515–533.

Churchland, P. (2011). *Braintrust: What neuroscience tells us about morality*. Princeton University Press.
Copp, D. (2008). Darwinian skepticism about moral realism. *Philosophical Issues, 18*, 184–204.
Cuneo, T., & Shafer-Landau, R. (2014). The moral fixed points: New directions for moral nonnaturalism. *Philosophical Studies, 171*(3), 399–443.
de Waal, F. (1996). *Good natured: The origins of right and wrong in humans and other animals*. Harvard University Press.
de Waal, F. (2006). *Primates and philosophers*. Princeton University Press.
de Lazari-Radek, K., & Singer, P. (2012). The objectivity of ethics and the unity of practical reason. *Ethics, 123*(1), 9–31.
Egeland, J. (2022). The epistemology of debunking argumentation. *Philosophical Quarterly, 72*(4), 837–852.
Fitz Patrick, W. (2015). Debunking evolutionary debunking of ethical realism. *Philosophical Studies., 172*(4), 883–904.
Fitz Patrick, W. (2021). Morality and evolutionary biology. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. Springer.
Fuqua, J. (2021). Ethical mooreanism. *Synthese, 199*(3–4), 6943–6965.
Gazzaniga, M. S. (2005). *The ethical brain*. Dana Press.
Greene, J. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. Penguin.
Green, J. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology volume 2: The cognitive science of morality: Intuition and diversity* (pp. 35–79). MIT Press.
Hanson, L. (2017). The real problem with evolutionary debunking arguments. *Philosophical Quarterly, 67*(268), 508–533.
Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. Harper Collins.
Hubalek, M. (2021). A brief (Hi)Story of just-so stories in evolutionary science. *Philosophy of the Social Sciences, 51*(5), 447–468.
Isserow, J. (2019). Evolutionary hypotheses and moral skepticism. *Erkenntnis, 84*(5), 1025–1045.
Jackson, F. (1998). *From metaphysics to ethics: A defence of conceptual analysis*. Oxford University Press.
Jackson, F. (2000). Psychological explanation and implicit theory. *Philosophical Explorations, 3*(1), 83–95.
Joyce, R. (2001). *The myth of morality*. Cambridge University Press.
Joyce, R. (2006). *The evolution of morality*. MIT Press.
Joyce, R. (2014). The evolutionary debunking of morality. In J. Feinberg & R. Shafer-Landau (Eds.), *Reason and responsibility: Readings in some basic problems of philosophy* (15th ed., pp. 589–597). Cengage Learning.
Joyce, R. (2016). *Essays in moral skepticism*. Oxford University Press.
Kahane, G. (2011). Evolutionary debunking arguments. *Nous., 45*(1), 103–125.
Kahane, G. (2014). Evolution and impartiality. *Ethics, 124*(2), 327–341.
Keil, P. C. (2010). The feasibility of folk science. *Cognitive Science, 34*(5), 826–862.
Kitcher, P. (2011). *The ethical project*. Harvard University Press.
Lakshminaryanan, V., Chen, M. K., & Santos, L. R. (2008). Endowment effect in capuchin monkeys. *Philosphical Transactions of the Royal Society B, 363*, 3837–3844.
Lutz, M. (2018). What makes evolution a defeater? *Erkenntnis, 83*(6), 1105-1126.
Malle, Knobe, & Nelson. (2007). Actor–observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology, 93*(4), 491–514.
Medin, D. L., & Atran, S. (2004). The native mind: biological categorization and reasoning in development and across cultures. *Psychological Review, 111*(4), 960–983.
Moon, A. (2017). Debunking morality: Lessons from the EAAN literature. *Pacific Philosophical Quarterly, 98*(S1), 208–226.
Nagel, T. (2012). *Mind and cosmos*. Oxford University Press.
Parfit, D. (2011). *On what matters: Volumes I & II*. Oxford University Press.
Parfit, D. (2017). *On what matters: (Vol. III)*. Oxford University Press.
Pollock, J. (1986). *Contemporary theories of knowledge*. Rowman and Littlefield Publishers.
Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26*(5), 521–562.
Sampson, E. (2023). Moorean arguments against the error theory: A defense. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 18, pp. 191–217). Oxford University Press.

Schechter, J. (2018). *Explanatory challenges in metaethics. Routledge handbook of metaethics*. In T. McPherson, & D. Plunkett (Eds.), (pp. 443–459). Routledge.
Shafer-Landau, R. (2012). Evolutionary debunking, moral realism and moral knowledge. *Journal of Ethics and Social Philosophy, 7*(1), 1–38.
Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science, 32*(6), 1049–1062.
Siegal, M., Butterworth, G., & Newcombe, P. A. (2004). Culture and children's cosmology. *Developmental Science, 7*(3), 308–324.
Sinclair, Neil. (2018). *Belief pills and the possibility of moral epistemology Oxford Studies in Metaethics 14*. Oxford University Press.
Singer, P. (2005). Ethics and intuitions. *The Journal of Ethics, 3–4*, 331–352.
Slaughter, V., & Gopnik, A. (1996). Conceptual coherence in the child's theory of mind: Training children to understand belief. *Child Development, 67*(6), 2967–2988.
Smith, R. J. (2016). Explanations for adaptations, just-so stories, and limitations on evidence in evolutionary biology. *Evolutionary Anthropology, 25*(6), 276–287.
Sterelny, K., & Fraser, B. (2016). Evolution and moral realism. *British Journal for the Philosophy of Science, 68*(4), 981–1006.
Street, S. (2006). A darwinian dilemma for realist theories of value. *Philosophical Studies, 127*(1), 109–166.
Street, S. (2008). Reply to Copp: Naturalism, normativity, and the varieties of realism worth worrying about. *Philosophical Issues, 18*(1), 207–228.
Tomasello, M. (2016). *A natural history of human morality*. Harvard University Press.
Toner, C. (2011). Evolution, naturalism, and the worthwhile: A critique of Richard Joyce's evolutionary debunking of morality. *Metaphilosophy, 42*(4), 520–546.
Vavova, K. (2014). Debunking evolutionary debunking. *Oxford Studies of Metaethics, 9*, 76–101.
Vavova, K. (2015). Evolutionary debunking of moral realism. *Philosophy Compass., 10*(2), 104–116.
Vavova, K. (2018). Irrelevant influences. *Philosophy and Phenomenological Research,* 134–152. https://doi.org/10.1111/phpr.12297
Vlerick, M., & Broadbent, A. (2015). Evolution and epistemic justification. *Dialectica, 69*(2), 185–203.