

Evolution and the Missing Link (in Debunking Arguments)

Uri D. Leibowitz and Neil Sinclair

Penultimate version of paper appearing in Michael Ruse & Robert J. Richards (eds.), *Cambridge Handbook to Evolutionary Ethics*. Cambridge University Press (forthcoming 2017)

Introduction

What are the consequences, for human moral practice, of an evolutionary understanding of that practice? By “moral practice” we mean the way in which human beings think, talk, and debate in moral terms.

The first type of answer holds that moral practice – or some feature of it – is vindicated by an evolutionary account of its origins. One might argue, for example, that evolutionary considerations support substantive moral claims. Richards, for instance, holds that an evolutionary account of moral practice justifies the claim that each of us “ought to act for the community good” (1986, 289; compare Casebeer 2003). Alternatively, evolutionary theorizing might vindicate moral practice insofar as it makes true or justifies a particular view of the nature of that practice. Harms (2000), for instance, argues that the fact that moral practice is an adaptation supports a version of moral realism – the view that moral judgments are maps of a moral reality, made true independently of human attitude. Finally, evolutionary theorizing might vindicate moral practice insofar as it helps

demonstrate how actual human beings have sufficient reason to go on engaging in that practice (Campbell 1996).

The second type of answer to the initial question is less optimistic. It takes moral practice – or some of its features – to be debunked by evolutionary understanding. There are several different types of debunking, and several possible targets. First, evolutionary considerations might show particular moral theories to be false or unjustified. Here “moral theories” cover both substantive first-order moral theories (such as utilitarianism) and second-order metaethical theories of the nature of moral practice (such as realism). Greene (2008) and Singer (2005), for example, take evolutionary accounts of the origins of moral intuitions to place pressure on first-order deontological theories. Street, by contrast, argues that “realist theories of value prove unable to accommodate the fact that Darwinian forces have deeply influenced the content of human values” (2006, 109). According to both, evolutionary considerations are a threat to the validity of moral theories. A different type of debunking has a more down-to-earth target: moral judgments of the ordinary folk, who may embrace no clear theory. According to the first instance of this type, evolutionary considerations show certain folk moral judgments to be false (or incoherent). This is truth-debunking (Joyce 2013). Ruse (1986, 254), for example, argues for the falsity of moral judgments based on the explanatory redundancy of the supposed moral facts they concern. According to the second instance of this type, evolutionary considerations show folk moral judgments to be unjustified, not knowledge, or otherwise lacking in epistemic status. This is epistemic debunking. Joyce, for example, argues that “we have an empirically confirmed theory about where our moral judgments come from ... which forces the recognition that we have no grounds one way or another for maintaining those [judgments]” (2006a: 211).

The most interesting types of truth- and epistemic-debunking have global targets: they urge that all moral judgments are false or unjustified. It was the possibility of such global debunking that so alarmed many of Darwin's contemporaries. In what follows, we focus on global epistemic debunking arguments.

Evolutionary Explanations

Evolutionary debunking arguments (EDAs) begin with the claim that moral practice can be given an evolutionary explanation. An evolutionary explanation here is (roughly) an account of how the possession of a trait, or another trait that is in some way connected to it, increased the relative evolutionary fitness of some ancestral population. The "relative evolutionary fitness" is a direct measure of the ability of individuals of a population to pass their genes onto the next generation and is always relative to the competition and to a given environment.

When it comes to evolutionary explanations of moral practice, there are several distinct phenomena one might take evolution to explain. First is the capacity to think, talk, and debate in moral terms. Call this the "capacity hypothesis." Second is a population-wide tendency to make particular types of moral judgment, distinguished by their contents. Call this the "tendency hypothesis." Whereas the capacity hypothesis explains our possession of moral concepts, the tendency hypothesis explains common patterns in the way these concepts are deployed. Neither hypothesis entails that the relevant trait will develop independently of appropriate environmental conditions. Joyce (2006a, 108–142) defends the capacity hypothesis. According to his account, the capacity to make moral judgments

can aid the reproductive fitness of an individual by providing a motivational bulwark against weakness of will. By categorizing certain paths of action under moral labels (e.g., as right or required), the individual is more inclined to pursue those paths and less inclined to eschew them for short-term gain. Where the paths of action thus preferred are prosocial behaviors (e.g., keeping a promise), pursuing them leaves the individual better able to reap the reproductive benefits of cooperation. Further, the capacity to make moral judgments can aid the evolutionary fitness of a population by enabling members of that population to negotiate, test, refine, and sustain cooperative social structures. By contrast, Street adopts the tendency hypothesis. She notes that “one enormous factor in shaping the content of human values has been the forces of natural selection” (2006, 114, our emphasis). For Street, the explananda of evolutionary explanations are the “deep and striking patterns, across both time and cultures, in many of the most basic evaluative judgements that human beings tend to make.” Street goes on to list particular moral judgments that seem to admit this kind of explanation. For example: “it is clear how beneficial (in terms of reproductive success) it would be to judge that the fact that something would promote one’s survival is a reason to do it” (2006: 115).

One feature of these evolutionary hypotheses is worth emphasizing. Both can be viewed as offering a type of functional explanation – that is, they specify a selected-for effect of a given trait (Wright 1976). According to Joyce’s hypothesis, the capacity to make moral judgments was selected for insofar as past instances of that type had the effect of stabilizing mutually beneficial cooperative structures. According to Street’s hypothesis the tendency to make particular moral judgments was selected for insofar as past instances of that type had the effect of producing (or making more likely) survival-promoting

behaviors. The important feature of both hypotheses is that the selected-for effect concerns the actions and behavioral tendencies of agents who possess the trait (and perhaps a derivative effect on the actions and tendencies of other agents). Call these “downstream motivational effects.” Just as importantly, it seems that according to these hypotheses the effect for which moral traits were selected was not the effect of successfully tracking explicitly moral facts. These moral traits were selected for not because they somehow enabled agents to “hook up” with a realm of moral facts, but simply because they prompted fitness-enhancing behaviors. More controversially, it is often held that the effect for which moral traits were selected does not involve those traits tracking facts of any kind – moral or otherwise.¹ In both ways, the evolutionary explanation of moral traits seems importantly different to the evolutionary explanation of our perceptual capacities.² Human beings have the capacity to perceptually detect the layout of objects in their immediate environment, in part because previous instances of that capacity successfully hooked up with the layout of objects in the ancestral environment. It is, of course, only because such capacities went on to prompt fitness-enhancing behaviors that they were selected for, but, importantly in this case, the fact that the capacities tracked some independent truths is part of the explanation of why the behaviors they prompted were fitness-enhancing. Nothing comparable seems to apply in the moral case.³

¹ For doubts, see Sinclair 2012 and White 2010, 588.

² See Joyce 2006a, 182–184; Fraser 2014, 463; and Kitcher 2005, 176.

³ Another comparison is with the capacity to make mathematical judgments. Although it seems a stretch to say that this capacity was selected for so as to track mathematical truth, it does seem that any evolutionary explanation of this capacity must at least assume that mathematical truths obtain – again nothing similar seems to apply in the moral case (Joyce 2006a, 182).

It is noteworthy that neither the capacity hypothesis nor the tendency hypothesis (directly) explains individual moral judgments, what might be called token judgments (i.e., instances of a general type). It is one thing to explain why humans have the capacity to make moral judgments, or why we find “widespread human acceptance” of certain moral judgments, but quite another to explain a token moral judgement. One feature that is often overlooked in the evolutionary debunking literature is that if the explananda are token judgments, then it is far from obvious that evolutionary accounts can perform the explanatory work required. Sober (1984), for example, argues that evolution is suited to explain why a population consists of one set of individuals with certain traits rather than another set with different traits, but not to explain why a particular individual has the traits it does, rather than other traits.⁴ If Sober is right, evolution is likewise not suited to explain why a particular individual makes a certain moral judgment rather than another. It might be tempting to insist that an evolutionary account will be part of the explanation of token moral judgments. Whether this temptation is to be resisted depends on one’s views on what complete explanation consists in.⁵ For present purposes we needn’t resolve this complex issue. Whether or not an evolutionary account will partake in the explanation of a token moral judgement, it is clear that other ingredients will be required.⁶ Indeed, the striking diversity in actual moral judgments seems to count against the view that all such judgments can be fully explained by evolution (Kahane 2011, 118; Fraser 2014, 469). So

⁴ For criticisms, see Neander 1988 and Nanay 2005.

⁵ See Leibowitz 2011.

⁶ For debunkers, these additional elements will likely be causal-psychological-sociological factors. The question then arises: what, if anything, does evolution add to the mix? We partially address this question in what follows.

even if debunkers show that the evolutionary component of the explanation of token judgments is not truth-tracking in the relevant sense, there is still work to be done to establish that the best explanation of token judgments need not invoke the truth of the judgments in question.⁷

In the remainder of this chapter we focus on how the absence of an explicitly moral tracking element in the evolutionary explanation of morality is supposed to undermine the justification for moral judgments.

A Generic Debunking Argument

How might the foregoing considerations feature in a debunking argument? Consider the following:

- (1) The best explanation of moral practice is an evolutionary explanation.
- (2) An evolutionary explanation of moral practice does not rely on moral practice tracking moral truths.

Hence,

- (C) all token moral judgments are unjustified.

This argument is not formally valid: there is a logical gap between the premises and conclusion – this is the “missing link” of our title. We think it is elusive.

Consider (2). What does “does not rely on moral practice tracking moral truths” mean? Synonyms include: “an off-track process” (Kahane 2011), “tracking failure”

⁷ Perhaps “third-factor” explanations can be understood as exploiting this possibility, although they are not typically presented in this way. See Enoch 2010, Wielenberg 2010, and Brosnan 2011.

(Lillehammer 2010), and “a truth-mooting genealogy” (Mason 2010). Precisely what this tracking failure amounts to is a key question, and different answers generate different debunking arguments. But there seems to be considerable agreement on the following, minimal account: evolutionary explanations do not invoke or assume the truth of moral judgments.⁸ As Ruse (1986, 254) notes: “You would believe what you do about right or wrong, irrespective of whether a ‘true’ right or wrong existed.” This claim is supported by the fact that evolutionary explanations of moral traits are consistent with supposing that all moral judgments are false – the so called-error theory of morality (Joyce 2013).

Consider next (1). As previously noted, neither the capacity nor tendency hypotheses (completely) explain token moral judgments. Hence, if we are careful about what the “explanation of moral practice” in (1) means, we can see that (C) does not follow from (1) and (2). Toner, for example, raises the possibility that while the moral capacity might have originally been selected “for the broad purpose of increasing genetic fitness by enabling reciprocity,” it later came to be “pressed into the service of other purposes ... which do require sensitivity to evidence and truth” (2010, 533). On this view, the evolutionary explanation of the capacity to make moral judgments is not truth tracking, but the manifestation of this capacity is sensitive to moral truths. We do not have the space to delve into the details of this view here. Instead we note that debunkers usually overcome this hurdle by making the following assumption: that whatever non-evolutionary factors are involved in the explanation of token moral judgments, these are also truth-mooting with respect to moral truth (Joyce 2006a, 181; Mason 2010, 774). To incorporate this assumption, we can define an “evolutionary+” explanation of a moral judgement as one

⁸ See Mason 2010, 773; Fraser 2014, 459; Joyce 2006a, 183–184, 211; and Street 2006, 155.

that involves the capacity or tendency hypothesis together with these additional morally truth-mooting factors. This generates the following argument:

(1*) The best explanation of each token moral judgment is an evolutionary+ explanation.

(2*) An evolutionary+ explanation of a token moral judgment does not rely on this judgment tracking moral truths.

(3) If the best explanation of a token judgment does not rely on this judgment tracking the truth of its content, then that judgment is unjustified.

Hence,

(C) all token moral judgments are unjustified.

This is a valid argument. And granting both (1*) and (2*), the question is whether (3) is true or well supported.

Specific Debunking Arguments

One surprising feature of EDAs of this kind is the minimal role played by evolution.⁹ That is, it may be possible to explain token moral judgments in morally truth-mooting yet non-evolutionary ways. Thus one might explain token judgments psychologically in terms of the “moral sensibility” of the judger (Harman 1977, 7), psychoanalytically (e.g., Freud), sociologically (e.g., Durkheim, Nietzsche), historically (e.g., Marx), or microphysically (when science is sufficiently advanced). If such explanations do, in fact, explain token moral judgments without relying on these judgment tracking moral truths, then each can be used

⁹ See Lillehammer 2003, 570; Mason 2010, 770–771; and Brosnan 2011, 53–54.

to generate a debunking argument. If “the best explanation” in (1*) is a uniquely referring expression, then only one of these arguments will be sound. But note that it is far from obvious that an evolutionary explanation is necessarily superior to, say, a microphysical explanation. And most physicalists will probably think that in principle there should be a microphysical explanation of each token moral judgment. Whether the availability of an explanation of one kind excludes the viability of other kinds of explanation of the same phenomenon is a question we must put aside here. Nevertheless, it is worth noting that on this construal of EDAs, these arguments don’t seem to advance the dialectic over and above, say, non-evolutionary arguments such Harman’s famous challenge, which urges moral skepticism on the basis that “we can have evidence for a hypothesis of a certain sort only if such hypotheses sometimes help explain why we observe what we observe” (1977, 13).

Despite its plausibility, the previous argument faces significant challenges. The notion of an explanation of a judgment that does not rely on it tracking relevant truth is central. One way to understand this notion is as the claim that a judgment can be explained without invoking its truth (this is the minimal sense of “tracking failure” noted earlier). On this understanding (3) claims the following:

(3*) If the best explanation of a token judgment does not invoke the truth of its content, then that judgment is unjustified.

But (3*) is controversial. First, there is the possibility of reduction. Suppose that property A is reducible to property B. Suppose further that a particular judgment concerning property A can be explained by citing property B. Then it seems that the judgment can be explained without invoking its truth, but it is far from obvious that the judgment is, therefore,

unjustified. In particular, it may well be justified because the judgment is explained by invoking the property that is the reductive basis of the property that features in the judgment. This point is quite general. For example, physicists sometimes suggest that all other properties can be reduced to the properties quantified over by fundamental physics. If so, it seems that every judgment that we make (that isn't about physics) can be explained without explicitly invoking its truth. But, of course, this doesn't threaten the justification of all such judgments. So long as Bob's judgment about the location of his cat, say, is responsive to, and explained by, the reductive basis of catness, it will be justified (Quinn 1986, 537–549).

There are at least two possible replies to the reduction problem. The first is to deny that a reduction is available. This is the option preferred by Joyce in the moral case (2006a, 190–209). But once we realize that reduction threatens (3*), we can see that weaker metaphysical relations may do so as well. For example, suppose the property of being black locally supervenes on (without being reducible to) a particular surface spectral reflectance property "SSR." On one common understanding of the supervenience relation, this is to say that the blackness of an object cannot change without its SSR-ness changing. Now suppose that Bob's judgment that his cat is black is explained by his cat's possession of SSR. This seems grounds for thinking that Bob's judgment is justified. Again the point is general: it seems that a judgment can be justified even if it can be explained without invoking or assuming its truth, in particular when it can be explained by the relevant supervenience base.¹⁰

¹⁰ Joyce seems inclined to reject supervenience claims along with reductionism, at least for moral properties (2006a: 209). Yet the basis for this view is an application of Ockham's

The second possible reply to the problem of reduction accepts that a reduction (or supervenience claim) may be available, but insists that, as a matter of fact, the relevant judgments are explained neither by the properties referred to be the judgment itself, nor by any plausible reductive or supervenience basis for those properties. But though this response delivers a more plausible version of (3*), it does so only by rendering the similarly modified versions of (1*) and (2*) less plausible. As mentioned previously, neither the capacity nor the tendency hypotheses show that the best explanation of each token moral judgment needs not invoke the truth of this judgment or the truth of any plausible relevant reductive or supervenient base. Indeed, it is consistent with both hypotheses that some moral judgments are explained in part by invoking a plausible reductive or supervenient base for the properties to which they refer (an example follows).

The second general problem with (3*) is that there are plausible counterexamples. Bob judges that the sun will rise tomorrow and that all men are mortal, both on the basis of previously observed instances, and both judgments are justified. But in neither case does the relevant truth seem to play a role in explaining Bob's judgment (see White 2010, 583; and Dancy 1985, 34).

Razor. In that case the epistemic debunking argument is not independent of a theory debunking argument. Further, one might argue that there is, after all, something that moral properties are required to explain – namely the intelligibility of our moral practices. See McDowell 1985; Shafer-Landau 2003, 13–52; and Enoch 2011.

Further Arguments

What other ways might there be to capture the notion of an explanation of a judgment that does not rely on this judgment tracking moral truths? This section considers several recently discussed options.

(Lack of) Sensitivity

The first option for understanding the notion of an explanation of a judgment that does not rely on this judgement tracking moral truths is that the best explanation of token moral judgments reveals that these judgments are insensitive. A moral judgement is insensitive just in case we would make the same judgment that we make now even if the presumptive moral truths were actually false. More formally, the following arguments are generated:

- (1*) The best explanation of each token moral judgment is an evolutionary+ explanation.
- (2**). An evolutionary+ explanation of a token moral judgment reveals that this judgment is insensitive.
- (3**). If the best explanation of a token judgment reveals that this judgment is insensitive, then this judgment is unjustified.

Hence,

- (C) all token moral judgments are unjustified.¹¹

¹¹ For discussion, see Joyce 2016; Clarke-Doane 2012; Wielenberg 2010; Lillehammer 2010, 365; and White 2010, 581.

Sensitivity is sometimes represented formally as “ $\neg p \square \rightarrow \neg Bap$,” where “p” is the content and “Bap” is a’s judgment (or belief) that p and $\square \rightarrow$ stands for the subjunctive conditional.¹² Premise (3**) claims that insensitivity undermines justification.

Sensitivity is not all it’s cracked up to be. The second premise is questionable. Consider Bob’s judgement that Jones’s torturing of Smith is wrong, which Bob possesses partly because Jones’s torturing causes unwanted agony. Grant that we can explain this judgment without invoking or assuming its truth – Bob makes this judgment partly because evolution has endowed him with a moral capacity and partly because this capacity is sparked into life by the more proximal fact that Jones’s torturing causes unwanted agony. Still, Bob’s judgment can be sensitive to moral truth so long as the wrongness of Jones’s act supervenes on (or is reducible to) its causing unwanted agony. Were Jones’s act not wrong, it wouldn’t be an act of causing unwanted agony, and hence Bob would not have judged it wrong.¹³ In this case, Bob’s judgment is sensitive to moral truth even though it can be explained in evolutionary(+) terms.¹⁴

In reply, the debunker might say: “You’ve only managed to show that some moral judgments are sensitive by assuming a substantive moral truth – namely that the wrongness of Jones’s action supervenes on it causing unwanted agony; but appeal to such claims is illegitimate, since it is precisely their justification that is at issue.” To which the

¹² Strictly speaking, sensitivity needs to be relativized to a judgment as formed by a particular method. See Nozick 1981, 179.

¹³ See Sturgeon 1985.

¹⁴ Another point made by Sturgeon is that, if moral truths are necessary then if one’s moral judgment that p is true, it is trivially sensitive because there are no worlds in which the antecedent of the counterfactual ($\neg p \square \rightarrow \neg Bap$) is true.

reply runs: “I may appeal to substantive moral claims to make a case for thinking that some moral judgments are sensitive, but by the same token you have to appeal to (or assume) substantive moral claims to make the case that moral judgments are insensitive to moral truth. To consider what would happen were a given moral ‘truth’ to turn out false, one needs to make assumptions about what that truth supervenes on.” This point highlights a more general worry with many debunking arguments: there is clearly something problematic in trying to debunk all moral judgments on the basis of an argument that makes moral assumptions (Lillehammer 2010, 374–375).

A second set of problems for this argument is that sensitivity does not seem necessary for justification. If, as many believe, false judgments can be justified, and since false judgments are necessarily insensitive, we have many counterexamples to (3**).¹⁵ Further, insofar as it is expressed in terms of counterfactual conditionals, sensitivity inherits well-known problems about the assessment of such claims. According to one popular view, for instance, the counterfactual, “if p were true, then q would be true,” is true in the actual world just in case in all close possible worlds where p is true, then so is q. However, explicating a notion of “closeness” here can prove tricky (Lewis 1973).

(Lack of) Safety

A second option for understanding the notion of an explanation of a judgment that does not rely on this judgment tracking moral truths is that the best explanation of token moral judgments reveals that these judgments are unsafe. Here, to say that a judgment that p (as

¹⁵ See White 2010, 580; and Dancy 1985, 39. It may be more plausible to suppose that sensitivity is necessary for knowledge; see Nozick 1981.

formed by method M) is safe is to say that in all nearby possible worlds where a subject (utilizing M) comes to judge that p, then p (i.e., $Bap \Box \rightarrow p$). The thought is that justification requires that where the exercise of our cognitive capacities generates a particular judgment, then, even granting that this judgment is true, it better not be true by mere luck – that is, it better not be that the very same judgement, in a nearby scenario, would have been false. More formally, the arguments are as follows:

(1*) The best explanation of each token moral judgment is an evolutionary+ explanation.

(2***) An evolutionary+ explanation of a token moral judgment reveals that this judgment is unsafe.

(3***) If the best explanation of a token judgment reveals that this judgment is unsafe, then this judgment is unjustified.

Hence

(C) all token moral judgments are unjustified.

Here we can be fairly quick as the difficulties mirror those of the preceding argument. The same example (Bob's belief about Jones) shows that (2***) is questionable. Further, (3***) seems false because safety is too strong as a necessary condition for justification: all false judgments are unsafe, but some false judgments are justified.¹⁶ And again, insofar as safety involves appeal to a vague notion of “nearby” worlds, claims of safety remain difficult to evaluate (Williamson 2009).

¹⁶ Again, safety may be more plausible as a necessary condition for knowledge, see Sosa 1999.

Luck

One way of understanding the sensitivity argument is relying on the thought that evolutionary explanations demonstrate the inflexibility of our moral judgments. In particular, that argument suggests that moral judgments are inflexible in the face of changes in moral truth. But there is a distinct sense in which moral judgments seem too flexible – namely, they seem to be beholden to morally irrelevant and highly contingent evolutionary factors (White 2010, 586–587). This suggests another way to understand the notion of an explanation of a judgment that does not rely on this judgment tracking moral truths: that the best explanation of token moral judgments reveals that these judgments are lucky. To say that a's judgment that p (as formed by method M) is lucky is to say that it is not the case that: were p to be true, then a would (employing M) judge that p. Again employing the possible worlds analysis, this is to say that: it is not the case in all nearby worlds were p, a (employing M) judges that p. Formally, $\neg(p \square \rightarrow Bap)$. So we get the following arguments:

(1*) The best explanation of each token moral judgment is an evolutionary+ explanation.

(2****) An evolutionary+ explanation of a token moral judgment reveals that this judgment is lucky.

(3****) If the best explanation of a token judgment reveals that this judgment is lucky, then this judgment is unjustified.

Hence

(C) all token moral judgments are unjustified.

Premise (3****) claims that luck undermines justification. Suppose – the argument can be put – that some moral claim, p, is true. Suppose also that evolution(+) has endowed us with the judgement that p. If it is the case that the same processes could have easily have endowed us with the judgment that not-p (or no judgment concerning p), then our initial judgment cannot be justified. This is to say, even granting the initial judgment is true, it better not be that it is held by luck – that is, it better not be that the same moral truth, in a nearby scenario, would not have been believed.

Unfortunately for the debunker, this argument faces familiar problems. Where a moral judgment is partly explained by a supervenient basis for the relevant moral property, that judgment will not be lucky. Here it is the case that in all nearby worlds where the moral truth obtains, the judgment occurs.¹⁷ Furthermore it is doubtful that luck undermines justification. Suppose, as a schoolboy, Jones was given a copy of *Principia Ethica* as a prize. Subsequently, he came to form many true beliefs about what Moore says in that book. But in some nearby worlds the prize was *Principia Mathematica* and in these worlds no judgments about Moore resulted. Jones's judgments about the contents of *Principia Ethica* are lucky but still (it seems) justified.¹⁸ Finally, insofar as luck is defined in the same counterfactual-conditional terms as sensitivity and safety, this argument faces the same problems of evaluating claims about distance in possible-world-space.

¹⁷ Clarke-Doane (2016) launches a bolder defence of lack of luck. See also Wielenberg (2016) and Joyce (2016) for doubts.

¹⁸ See Ichikawa and Steup 2014, sec. 8.

Diagnosis

Once is happenstance; twice is coincidence; a third time is somewhat dull. These last three arguments fail for similar reasons. Each claim that moral judgments exhibit a certain type of “tracking failure” (lack of sensitivity, lack of safety, luck) and that this type of tracking is necessary for justification. But it is not clear that these types of tracking failure are established by evolutionary explanations of morality, and there seem to be plausible examples of moral judgments that might track moral truths in the relevant ways that are consistent with evolutionary accounts. In any case, the proposed necessary conditions on justification are questionable. Perhaps other debunking arguments fail for similar reasons, but rehearsing these moves now seems redundant.¹⁹ Might there be a different approach that will bring a great debunking dividend?

(Lack of) Method

One common response to the previous dialectic is as follows: Sure, it might be that some moral judgments are sensitive, safe, or what-have-you, and it might be that one can provide particular examples of moral judgments that satisfy these conditions. But this is not to provide any general account of the mechanisms by which humans form moral judgments such that those judgments might end up being sensitive, safe, or otherwise justified. In particular, it is not to provide any general account of how the mechanisms for forming moral judgments that we actually employ (or some suitably refined version of them) might

¹⁹ See, for example, the (lack of) reliability argument discussed in Wielenberg 2010. See White 2010, 584–585, for a diagnosis of why such arguments might seem attractive.

constitute ways of getting into contact with moral truth. According to this line of thought, evolutionary explanations of our moral judgments debunk their justification insofar as they emphasize the absence of a positive epistemological story for those judgments (Fraser 2014, 470).

This line of thought has been expressed in several ways. Blackburn puts it in terms of whether we are able to explain why the “correspondence conditionals” given by sensitivity and safety obtain (1993, 161, 167). Joyce, citing Harman, puts the point thus: “we do *not* have a believable account of how moral facts could help explain the mechanisms and forces that give rise to moral judgements” (2016, 135). More formally, the argument might be as follows:

(4) We do not have a plausible general account of how moral truths could help explain token moral judgments.

(5) If we do not have a plausible general account of how certain truths in a domain could help explain token judgments in this domain, then those judgments are unjustified.

Hence,

(C) all token moral judgments are unjustified.

But this argument too is questionable. Premise (5), as stated, might lead to skepticism outside of morality in a way unpalatable to debunkers. For example, it is doubtful that we have a plausible general account of how truths about other minds could help explain token judgments about other minds, but few participants in this debate would like to endorse

skepticism about other minds.²⁰ Likewise, it is doubtful that, say, Aquinas and his contemporaries had a plausible general account of how truths about perceptual objects could help explain token perceptual judgments, but few would be comfortable endorsing the view that all of Aquinas' perceptual judgments were unjustified.

Moreover, premise (4) is highly contested. Many philosophers of a realist persuasion, for instance, have put forward "moral epistemologies," that is, accounts of how our moral judgments hook up with moral truth. According to the moral coherentist, for example, moral judgments are justified insofar as they are part of a coherent system of beliefs, including particular and general moral beliefs as well as second-order beliefs about the nature of moral practice (Brink 1989, 100–143). According to foundationalists, on the other hand, certain moral judgments are justified non-inferentially insofar as they are self-evident (Audi 1993) or intuited (Huemer 2005). Such epistemologies are not obviously refuted by evolutionary considerations. Hence Joyce notes: "No one... thinks that genealogical empirical data alone can secure a sceptical victory; at most it battles alongside sceptical arguments of an a priori metaethical nature" (2016, 136).

But for our purposes the most salient feature of this argument is that it does not mention evolution at all – evolutionary considerations play no role in supporting its conclusion. So what, if anything, do evolutionary considerations bring to the table?

Our proposal is that the mistake in this argument, as in the preceding arguments, is a misidentification of the appropriate target of EDAs. All of the arguments considered here, like much of those discussed in the literature, aim to establish a first-order epistemological result: that all token moral judgments are unjustified. For the reasons previously canvassed

²⁰ See Leibowitz 2014.

it seems doubtful that evolutionary considerations can partake in successful arguments for this conclusion. Instead, the true force of evolutionary considerations for our understanding of moral practice may be located elsewhere. In particular, such considerations highlight the potential success of a naturalist, anti-realist, skeptical research program. This is the program that takes the central task in understanding moral practice to be locating that practice within a wider scientific view of human beings, which rejects the postulation of any judgement-independent realm of moral facts and which considers all moral judgments to be unjustified. Instead of trying to locate evolutionary explanations in a premise for an argument for this latter claim, we propose to view evolutionary explanations as aiming to show that if one adopts a naturalist, anti-realist approach, one has a better overall account of morality.

On this view, evolutionary considerations are held to be consistent with the existence of objective moral truths and justified moral judgments. But, the argument runs, a realist, non-naturalist, optimist research program is not as successful as the competing naturalist, anti-realist, skeptical one. The latter offers a clean, coherent, and satisfactory explanation of phenomena that the former struggles to address. On this view the debunker need not claim that all moral epistemologies are hopeless or that evolutionary considerations can establish that moral knowledge is impossible. Instead her goal is to show that the research program she favors is progressive, whereas the alternative is degenerative (for this terminology, see Lakatos 1970). The force of EDAs, we propose, is at the meta-level and not in first-order epistemology.

To see this in more detail, consider two competing theories of the nature of moral practice, initially held before any evolutionary account of that practice is known or

believed. According to the former – a caricature of realism – human moral judgments are more-or-less reliable cognitions of an independently existing moral reality. According to the latter – a caricature of anti-realism – moral judgments answer to no independent reality but have a primary role in guiding action and behavior (error theory is one version of this anti-realism, but there may be others: see Blackburn 1993). Now suppose we add to this situation an evolutionary account of moral practice roughly along the lines Joyce suggests: where this practice involves traits selected for their downstream motivational effects. Which theory does this evolutionary appendix help the most? Arguably, it leads to an increased degree of coherence for the anti-realist theory – for it provides important detail as to the purpose of moral practice, if it is not to track an independently existing moral reality. Conversely, while the evolutionary story is consistent with the realist theory, it fails to add to its coherence. One may, thus, reasonably view developments in evolutionary psychology as an indication that the anti-realist program is progressive and, consequently, that one has good (pro tanto) reason to favor the anti-realist program to its nonprogressive realist competitor.

How damaging is an evolutionary account of morality to those pursuing the realist research program? The answer is not obvious. First, even granting that evolutionary considerations demonstrate that the anti-realist program is progressive whereas its realist competitor is degenerative, it does not follow that it is irrational to pursue the realist program. There are ample examples from the history of science in which major breakthroughs were made by scientists pursuing degenerative programs (see Matheson and Dallmann 2015). Moreover, the extent to which evolutionary considerations reveal that the anti-realist program is progressive whereas the realist one is degenerative is not

obvious. Proponents of EDAs will claim that evolutionary considerations increase the coherence of their overall account of morality while leaving their realist opponents with unanswered questions about the mechanism for tracking moral truth. But a committed moral intuitionist, for example, who believes that intuitionism provides a satisfactory answer to the mechanism question, might not feel the bite of this argument. At best she will grant that recent evolutionary findings do more explanatory work for anti-realists than they do for her, but she will resist the claim that such considerations show that there are legitimate questions that her own research program leaves unanswered. Finally, the move to research programs suggests that we need to consider the successes and failures of such programs in all domains. For example, if, as some realists claim, there are successful indispensability arguments for realism – that is, arguments that purport to identify phenomena for which the best explanation must appeal to moral facts – then EDAs can hardly be claimed to have established that the anti-realist program is progressive. Expanding the scope in this way may also require us to consider the successes and failures of the competing research programs outside of morality, for example, in mathematics and religion.²¹ If so, there is still much work to do before proponents of EDAs can celebrate the victory of the anti-realist agenda.

²¹ For discussions of indispensability arguments, their relation to EDAs, and their application to mathematics and other domains, see Leibowitz and Sinclair 2016.

Conclusion

If the preceding reflections are correct, then evolutionary explanations of moral practice do add something distinctive to debates about the nature of morality (over and above other morally truth-mooting explanations). But it is not the knock-down argument for skepticism that some debunkers seem to think. The availability of evolutionary accounts of moral practice that nowhere invoke moral truth seems to increase the coherence of anti-realist positions in metaethics, whereas the coherence of realist positions at best remains unchanged and possibly decreases. But even on the most favorable interpretation EDAs can hardly be claimed to have established that the realist program should be abandoned.

References

Audi, R. 1993. 'Ethical Reflectionism' *The Monist* 76: 295-315.

Blackburn, S. 1993. *Essays in Quasi-Realism*. Oxford: Oxford University Press.

Brink, D. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Brosnan, K. 2011. 'Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge' *Biology and Philosophy* 26: 51-64.

Campbell, R. 1996. 'Can Biology Make Ethics Objective?' *Biology and Philosophy* 11: 21-31.

Casebeer, W.D. 2003. *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition*. Cambridge, Ma.: MIT Press.

Clarke-Doane, J. 2012. 'Morality and Mathematics: The Evolutionary Challenge' *Ethics* 122: 313-340.

Clarke-Doane, J. 2016. 'Genealogy and Reliability' in U. Leibowitz and N. Sinclair, eds., *Explanation in Ethics and Mathematics: Debunking and Dispensibility*. Oxford: Oxford University Press.

Dancy, J. 1985. *An Introduction to Contemporary Epistemology*. Oxford: Blackwell.

Enoch, D. 2010. 'The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It' *Philosophical Studies* 148: 413-438.

Enoch, D. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

Fraser, B. 2014. 'Evolutionary Debunking Arguments and the Reliability of Moral Cognition' *Philosophical Studies* 168: 457-473.

Greene, J. 2008. 'The Secret Joke of Kant's Soul' in W. Sinnott-Armstrong, ed., *Moral Psychology: The Neuroscience of Morality*, vol. 3: 35-79.

Harman, G. 1977. *The Nature of Morality*. Oxford: Oxford University Press.

Harms, W.F. 2000. 'Adaption and Moral Realism' *Biology and Philosophy* 15: 699-712.

Huemer, M. 2005. *Ethical Intuitionism*. Palgrave Macmillan.

Ichikawa, J. and Steup M. 2014. 'The Analysis of Knowledge' *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2014/entries/knowledge-analysis/>.

Joyce, R. 2006. *The Evolution of Morality*. Cambridge, Ma.: MIT Press.

Joyce, R. 2014. 'The Evolutionary Debunking of Morality' in J. Feinberg and R Shafer-Landau, eds., *Reason and Responsibility: Readings in Some Basic Problems of Philosophy*, 15th edition.

Joyce, R. 2016. 'Reply: Confessions of a Modest Debunker' in U. Leibowitz and N. Sinclair, eds., *Explanation in Ethics and Mathematics: Debunking and Dispensibility*. Oxford: Oxford University Press.

Kahane, G. 2011. 'Evolutionary Debunking Arguments' *Noûs* 45:103-125.

Kitcher, P. 2005. 'Biology and Ethics' in D. Copp, ed., *The Oxford Handbook of Ethics*. Oxford: Oxford University Press.

Lakatos, I. 1970. 'Falsification and the Methodology of Scientific Research Programmes' in I. Lakatos & A. Musgrave, eds., *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press. 91-195.

Leibowitz, U. D. 2011. 'Scientific Explanation and Moral Explanation' *Noûs* 45: 472-503.

Leibowitz, U. D. 2014. 'Explaining Moral Knowledge' *Journal of Moral Philosophy* 11: 35-56.

Leibowitz, U. D. & Sinclair, N. eds. 2016. *Explanation in Ethics and Mathematics: Debunking and Dispensability*. Oxford: Oxford University Press.

Lewis, D. 1973. *Counterfactuals*. Cambridge, Ma.: Harvard University Press.

Lillehammer, H. 2003. 'Debunking Morality: Evolutionary Naturalism and Moral Error Theory' *Biology and Philosophy* 18: 567-581.

Lillehammer, H. 2010. 'Methods of Ethics and the Descent of Man: Darwin and Sidgwick on Ethics and Evolution' *Biology and Philosophy* 25: 361-378.

Mason, K. 2010. 'Debunking Arguments and the Genealogy of Religion and Morality' *Philosophy Compass* 5: 770-778.

Matheson, C. and Dallmann, J. 2015. 'Historicist Theories of Scientific Rationality', *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2015/entries/rationality-historicist/>>.

McDowell, J. 1985. 'Values and Secondary Properties' in T. Honderich, ed., *Morality and Objectivity*. London: Routledge and Kegan Paul: 110-129.

Nanay, B. 2005. 'Can Cumulative Selection Explain Adaptation?' *Philosophy of Science* 72: 1099-1112.

Neander, K. 1988. 'What Does Natural Selection Explain? Correction to Sober' *Philosophy of Science* 55: 422-426.

Nozick, R. 1981. *Philosophical Explanations*. Cambridge Ma.: Harvard University Press.

Quinn, W. 1986. 'Truth and Explanation in Ethics' *Ethics* 96: 524-544.

Richards, R. 1986. 'A Defense of Evolutionary Ethics' *Biology and Philosophy* 1: 265-293.

Ruse, M. 1986. *Taking Darwin Seriously*. Oxford: Basil Blackwell.

Shafer-Landau, R. 2003. *Moral Realism: A Defense*. Oxford: Oxford University Press.

Sober, E. 1984. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. University of Chicago Press.

Sosa, E. 1999. 'How to Defeat Opposition to Moore' *Noûs* (Supplement: *Philosophical Perspectives* 13, *Epistemology*) 33: 141-153.

Sinclair, N. 2012. 'Metaethics, Teleosemantics and the Function of Moral Judgements' *Biology and Philosophy* 27: 639-662.

Singer, P. 2005. 'Ethics and Intuitions' *The Journal of Ethics* 9: 331-352.

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

Sturgeon, N. 1985. 'Moral Explanations' in D. Copp and D. Zimmerman, eds., *Morality, Reason and Truth*. Totowa, N.J.: Rowman and Allanheld. 49-78.

Toner, C. 2010. 'Evolution, Naturalism, and the Worthwhile: A Critique of Richard Joyce's Evolutionary Debunking of Morality' *Metaphilosophy* 42: 520-546.

White, R. 2010. 'You Just Believe That Because....' *Philosophical Perspectives* 24: 573-615.

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

Wielenberg, E. 2016. 'Evolutionary Debunking Arguments in Religion and Morality' in U. Leibowitz and N. Sinclair, eds., *Explanation in Ethics and Mathematics: Debunking and Dispensibility*. Oxford: Oxford University Press.

Williamson, T. 2009. 'Reply to Alvin Goldman' in P. Greenough and D. Pritchard, eds., *Williamson on Knowledge*. Oxford: Oxford University Press.

Wright, L. 1976. *Teleological Explanations*. Berkeley: University of California Press.