Koon, J. (2021). “The Epistemology of Evolutionary Debunking,” *Synthese* 199(5-6): 12155-12176. <https://doi.org/10.1007/s11229-021-03327-w>.

The Epistemology of Evolutionary Debunking

Justis Koon

1. Introduction

Moral realism, as I will understand it here, is the conjunction of four theses: that moral language should be interpreted literally, that moral claims express beliefs, that there are at least some moral truths, and that these truths are mind- and language-independent.[1] Sharon Street (2006; 2015) and Richard Joyce (2006; 2013; 2016) have both advanced evolutionary debunking arguments which purport to show that, if moral realism is true, our moral beliefs are systematically unjustified.[2] These arguments are motivated by recent empirical work on the evolution of morality, work which suggests that the human moral sense was selected chiefly to promote cooperation among small tribes of hunter-gatherers in our distant evolutionary past.[3] If, however, our moral sense evolved due to the positive contribution that cooperation made to our ancestors' reproductive fitness, it becomes something of a mystery how it could also succeed in tapping into a well of mind-independent moral truths. It seems

[1] For comparison, Ayer-style emotivists will reject all four theses, error theorists will reject the third, and moral relativists and constructivists will reject the fourth. There are a few meta-ethical views – I am thinking especially of thin forms of reductive naturalism, like those defended by Copp (2008) and Sterelny and Fraser (2017) – where it is unclear whether we should categorize them as realist or anti-realist. I will not be addressing these sorts of views here.

[2] See Horn (2017) and Lutz (2018) for more recent presentations of evolutionary debunking arguments, and Korman (2019) for a review of the evolutionary debunking literature.

[3] See Section 3.5 and the references therein.

like it would be an extraordinary coincidence – in Street's words, nothing short of a miracle – if evolutionary forces indifferent to the moral truth somehow shaped our faculties to be appropriately sensitive to it.

Contrast the situation for vision. Any plausible account of the evolution of our visual faculties will make it clear that they were selected to capture information about the color, shape, texture, brightness, and relative distance of objects in our visual field, and to produce beliefs that accurately reflect these features of our surroundings.[4] We should not expect natural selection to have made our vision perfectly reliable, of course, both because selection is not all-powerful and must work within existing physical and biological constraints, and because visual illusions may, in rare circumstances, be adaptive. But, generally speaking, it will be an enormous boon to an organism's fitness for it to have an accurate picture of its environment, rather than being left in the dark about what goes on around it, blind not only to the presence of food, water, and potential mates, but also to the threats posed by predators and other hazards. So evolutionary theory gives us every reason to think that our visual faculties were selected primarily to produce true beliefs about the sources of the light waves impinging on our retinas.

But scientists tell a completely different sort of story when it comes to the evolution of our moral sense, one on which it was selected not to produce true moral beliefs, but to enable us to reap the benefits of cooperating with other members of our species.[5] Joyce and Street have seized on this peculiar feature of the evolutionary history of our moral sense and developed it into an argument against moral realism. For reasons others have cataloged, however, both Joyce and Street's accounts of evolutionary debunking contain serious defects, especially in the epistemic principles they rely on to

[4] Yong (2016) gives a nice popular overview of the evolution of vision.

[5] Some philosophers writing on evolutionary debunking arguments, beginning with Street, contrast the thesis that our moral sense was selected to acquire true beliefs with the thesis that it was selected for survival and reproduction, or the thesis that it was selected to promote fitness. This is a confusion; trivially, all selection favors organisms who are fitter or more successful at surviving and reproducing than their conspecifics, so it makes no sense to say that a trait was selected for survival and reproduction or for fitness. What we are interested in when we inquire what a trait was selected for is which (if any) of its effects boosted our ancestors' reproductive fitness, and thereby caused the genes associated with that trait to proliferate throughout our species.

generate their conclusions.[6] This paper is an attempt to address some of the problems that have been identified for Joyce and Street's accounts, and, by so doing, move the evolutionary debunking project they initiated onto surer footing.

Here is how my approach differs from theirs: first, to set the stage, I introduce a thought experiment that presents a striking analogy for the evolution of morality, and which demonstrates how compelling evolutionary-debunking-style reasoning can be when applied to a case where we are not antecedently invested in the outcome (Section 2). Second, I carefully formalize a new evolutionary debunking argument (Section 3.1) and show how the success of the evolutionary debunking project depends, to a large extent, on the truth of calibrationist views of higher-order evidence (Section 3.2). Third, I offer a new rationale for why learning that our moral sense was selected to facilitate cooperation should undermine the justification for our moral beliefs (Section 3.3). Finally, I respond to three of the most compelling objections which have been raised against Street and Joyce's debunking arguments (Section 4); employing the resources of calibrationism, and referring back to the thought experiment from Section 2, will help to illuminate why these objections are unsuccessful.[7]

One caveat before we proceed: for ease of exposition, I will be assuming throughout most of this paper that human beings have a moral sense or faculty whose function is innately specified, comparable, in this respect, to our faculty for processing natural language.[8] A number of philosophers

[6] See White (2010), Shafer-Landau (2012), Vavova (2014; forthcoming), Bogardus (2016), Clarke-Doane (2016), Sinclair (2018), and Clarke-Doane and Baras (2021). The most important points of criticism, to my mind, are that Joyce's version of the argument depends on a causal epistemic principle that is widely believed to be false (although see Korman and Locke [2020] for a defense), while the epistemic principle Street invokes is not clearly spelled out but implausible on most interpretations. The argument I develop in this paper replaces these principles with a version of the calibrationist view of higher-order evidence, which enjoys substantial (although by no means universal) support in the literature.

[7] I should note that it seems likely to me that any evolutionary debunking argument is liable to struggle with Moorean responses (Section 4.1) and third-factor explanations (Section 4.2) unless it avails itself of calibrationism and its associated independence requirement.

[8] I understand the moral sense to be the faculty that generates our gut reactions or intuitions in ethics, both about particular cases (real or hypothetical) and about general principles. For instance, it is the moral sense which intimates to us that torturing children is wrong, that generosity is a virtue, and which makes “maximize the amount of well-being in the world” – but not “maximize the amount of injustice in the world” – seem like a plausible moral principle. I suggest, moreover, that the moral sense plays an indispensable epistemic role in justifying our moral beliefs (inasmuch as they are justified at all); none of our moral beliefs could lay claim to any positive epistemic status if not for the base-level infusion of evidence supplied by the moral sense. Note, though, that nothing in this paper hangs on how I am conceiving of the moral sense. In the end, I will argue that evolutionary debunking arguments succeed even if our brains turn out not to contain any kind of specialized faculty for moral cognition.

and cognitive scientists have disputed this assumption on empirical grounds;[9] towards the end (Section 4.3) we will see whether relaxing it damages the debunking argument's prospects for success.

2. Evolutionary Debunking Not Ruled Out *A Priori*

Joyce and Street's debunking arguments have faced a large number of objections. A few authors have even suggested that facts about our evolutionary history can never – not even in principle – affect the epistemic status of our beliefs.[10] Should we follow these philosophers in thinking that debunking arguments are totally misconceived, that events from our distant evolutionary past just have no bearing on whether our beliefs today are justified? I do not think so. To the contrary, I believe we can be confident that some account of evolutionary debunking or other must succeed. The proof of this is that we are able to construct cases where evolutionary-debunking-style reasoning seems quite compelling. Here is one:

Fools Rush In

In the year 2988, humanity makes first contact with an alien race, inhabitants of a lush and hospitable planet orbiting a dim star in the Arcturus Stream. In an unlikely case of convergent evolution, these Transarcturians, as we come to call them, are nearly identical to humans of the early 21st century, with our computers and automobiles, our nation-states and mixed economies, our universities and hospitals, and our science, philosophy, and ethics. Scans of the Transarcturian planet reveal only one anomaly: huge regions of the globe have been left to the virgin forest, completely devoid of cities, roads, and all other signs of civilization. The explorers inquire with the Transarcturian ambassador after this curiosity, who, with some bewilderment, informs them that these are the planet's forbidden zones, where no Transarcturian may tread. When pressed for an explanation why, the ambassador responds, "Isn't it obvious? Because they are forbidden."

The explorers drop the matter to avoid a diplomatic incident, but on the next expedition, a

[9] See Machery and Mallon (2010), FitzPatrick (2015), and Levy and Levy (2020).
[10] White (2010), for one, defends this view.

science team is sent to one of the forbidden zones to investigate, mindful of the perils that may await them there. They find bizarre and variegated vegetation, unlike anything else on the planet, and vast underground deposits of lead, but nothing to account for the fear and reverence the forbidden zone inspires among their hosts. Their questions unanswered, they return to discuss their findings with a physicist at one of the leading Transarcturian universities. He shows little surprise or interest, and, when asked to clarify what feature of the zones places them off-limits, replies, "The fact that they are forbidden, of course! Can't you tell? Perhaps you had better direct your inquiries to a philosopher instead, or a priest. This is not really my area of expertise."

Stunned by this dismissal, the Earth scientists retreat to their spacecraft to try and puzzle out what they have observed. After much discussion, they hit on the hypothesis that, in the distant evolutionary past of the Transarcturians, the regions of the planet now known as the forbidden zones played host to vast deposits of radioactive thorium, since decayed into lead. Although the Transarcturians' physiology would have been hardy enough to protect them from any ill health effects of the radiation, their gametes could not have resisted its mutagenic properties, and over hundreds of thousands of years (the Earth scientists conjecture), the Transarcturians were selected to carry an innate psychological predisposition to avoid the forbidden zones, to reduce the risk of causing mutations to their germline DNA.[11] Strangely, the Transarcturians appear to experience this aversion in normative terms – they don't just feel an urge to stay out of the forbidden zones, they perceive themselves as having an obligation to do so. Although wary of interfering in the development of a less technologically advanced civilization, the Earth scientists ultimately resolve to share their conclusions with the Transarcturians, wishing to liberate the race from its ancient and congenital superstition.

To flesh out the story further, we can assume that whenever a Transarcturian approaches a

[11] Just so the thought experiment is not confounded by ethical concerns, I ask the reader to assume, somewhat implausibly, that the radiation would not have caused harmful congenital disorders in the Transarcturians' offspring, and that their disposition to avoid trespassing in the forbidden zones, if the Earth scientists' hypothesis is correct, is the result of selection operating directly on the Transarcturians' genes. The idea is that any genes which predisposed the Transarcturians to steer clear of the forbidden zones were favored by selection just because those genes were less likely to be altered by the mutagenic effects of the radiation.

forbidden zone, she has an experience with a distinct phenomenal character – something like an intuition – with the content that she ought not proceed any further.[12] Let's suppose, moreover, that the Transarcturians have also devised complex systems of norms surrounding forbiddenness, and their philosophers are continually disputing whether these norms should be understood in consequentialist terms, as a matter of minimizing the aggregate amount of time any Transarcturian spends in a forbidden zone, or deontologically, as imposing a *pro tanto* duty not to encroach on the forbidden zones which must be balanced against competing obligations.

It seems clear to me that as soon as the Transarcturians are informed of the Earth scientists' hypothesis and its supporting evidence, they are no longer justified in retaining their normative beliefs about the forbidden zones, not, at least, if those beliefs continue to be construed realistically. Intuitively speaking, when the Transarcturians learn that their forbiddenness faculty may have been selected to carry out a function unrelated to producing true beliefs, they can no longer trust that its outputs are accurately representing features of the world around them. It just does not seem plausible to suggest that the Transarcturians should feel free to shrug off this revelation about their evolutionary history, and carry on with their lives exactly as before.[13] If they are rational, they must take seriously the possibility that all of their attitudes towards the forbidden zones, along with the whole edifice of norms they've constructed around them, are an elaborate sham or illusion foisted on them by their genes. In other words, when the Transarcturians learn that their forbiddenness beliefs may have a deviant causal history, one not properly aimed at truth, this appears to defeat or undermine the justification for those beliefs.[14]

[12] Note that the Transarcturians do not believe that the forbidden zones are, in general, dangerous; they see “forbiddenness” as an intrinsic normative feature of certain areas of their world, just as we see goodness as an intrinsic normative feature of certain states of affairs.

[13] Of course, if the Transarcturians could offer an alternative account of the etiology of their forbiddenness faculty suggesting that it was selected to produce true beliefs, and this account was clearly better-supported by the available evidence than the Earth scientists' conjecture, that would change their epistemic situation substantially. But we are assuming they have no such account to offer.

[14] Some externalists about justification may insist that the Transarcturians' forbiddenness beliefs, because they were unreliably formed, were never even *prima facie* justified, and hence cannot be undermined or defeated. Proponents of radically aprioristic moral epistemologies which hold that false moral beliefs can never be justified may be inclined to say the same thing. There are two questions to separate here: the first is whether it makes sense to describe beliefs that are not *prima facie* justified as being undermined or defeated, while the second concerns how the Earth scientists' discovery should affect the epistemic status of the

The key point is that it is exceedingly difficult to see what chain of reasoning could lead us to this conclusion, if not for an evolutionary debunking argument roughly along the lines of those advanced by Joyce and Street. Hence, if we hope to capture our intuitive verdict about the Transarcturians, we must accept that information about our distant evolutionary past does have the potential to affect the epistemic status of our present-day beliefs after all. We will return to this thought experiment later on, in Sections 4 and 5, but for now, let's get the debunking argument itself on the table.

3. Evolutionary Debunking Redux

3.1 Overview of the Argument

Neither Joyce nor Street, in their original presentations of their debunking arguments, made any attempt to formalize them. Unfortunately, this has generated a large amount of confusion about how their arguments are supposed to be structured. Let me begin, then, by laying out what I take to be the best way of formalizing the core evolutionary debunking argument against moral realism:

Empirical Premise: We have good reason to think that our moral sense was selected principally for functions other than producing true moral beliefs.

Etiological Principle: If moral realism is true, and if we have good reason to think that our moral sense

Transarcturians' beliefs, assuming they were never justified to begin with. The first of these questions strikes me as a semantic matter of limited significance, so I will focus on the latter. Even if we say the Transarcturians' forbiddenness beliefs were unjustified to begin with, there is still a clear sense, I think, in which they held those beliefs rationally, or reasonably, or blamelessly, prior to the revelation about their evolutionary history. It's hard to fault them for trusting intuitions that are built into their minds from birth, and widely shared throughout their species. My contention is that, after hearing about the Earth scientists' discovery, intuitively, their forbiddenness beliefs cease being rational, or reasonable, or blameless in this sense, and that this is the sort of change to a belief's epistemic status that would normally defeat its justification, had it been justified in the first place. And, if this subjunctive or counterfactual claim about the Transarcturians' beliefs is true, that's all that will be needed to underwrite the intended analogy between the forbidden zones and morality. Of course, realists have the option of biting the bullet and insisting that hearing about the Earth scientists' discovery should have no effect whatsoever on the epistemic status of the Transarcturians' beliefs. But it seems to me that this still comes at the cost of saying something counter-intuitive about the case.

was selected principally for functions other than producing true moral beliefs, then, unless we are able to corroborate that our moral sense is reliable through the use of some other belief-forming mechanism whose etiology is not subject to similar doubts, the balance of independent evidence suggests that our moral sense is unreliable.

Autonomy Clause: We are unable to corroborate that our moral sense is reliable through the use of some other belief-forming mechanism whose etiology is not subject to similar doubts.

Epistemic Principle: If the balance of independent evidence suggests that our moral sense is unreliable, the justification for our moral beliefs is defeated.

Conclusion: If moral realism is true, the justification for our moral beliefs is defeated.[15]

In brief: the empirical premise observes that our moral sense has a suspicious evolutionary history, having been selected for purposes other than producing true moral beliefs; the etiological principle posits that this sort of evolutionary history, under the assumption that moral realism is true, gives us reason to suspect that our moral sense is unreliable; the autonomy clause establishes that our moral beliefs cannot be rehabilitated by some other faculty whose epistemic credentials are above reproach; and the epistemic principle asserts that having reason to think that our moral sense is unreliable strips our moral beliefs of their justification. By successive applications of *modus ponens*, this leaves us with the conclusion that, if moral realism is true, none of our moral beliefs are justified.

Two notes are in order. First, nothing hangs on the choice of justification as the epistemic status targeted by the argument, and a similar evolutionary debunking argument could just as easily place warrant or knowledge on the chopping block instead. Second, the argument is intended as a *reductio* – as Shafer-Landau (2012: 1) puts it, the combination of moral realism and moral skepticism

[15] Compare the formalizations in Shafer-Landau (2012) and Morton (2016).

implied by the conclusion is a “logically coherent position that contains about zero appeal.” The hope is that those persuaded by the argument will be more inclined to jettison one of the realist's package of metaphysical and linguistic theses, and adopt an anti-realist view instead, than to stick to their guns about realism but abandon all claims to moral knowledge.

The remainder of this paper will be devoted to motivating the premises of this argument and defending it from objections. Let's begin with the epistemic principle.

3.2 The Epistemic Principle

Contemporary theories of justification almost universally incorporate a theory of defeat, an account of how a belief that is *prima facie* justified – by virtue of being supported by the evidence, for instance – can come to lose its justification. A schematic version of the debunking argument's epistemic principle spells out a sufficient condition for defeat:

Epistemic Principle (Schema): If the balance of independent evidence suggests that the belief-forming mechanism which generates our beliefs in domain *D* is unreliable, the justification for our *D*-related beliefs is defeated.[16]

This principle is not intended to cover all instances of defeat; it deals exclusively with cases where we come across evidence calling our own reliability into question. Evidence that one of our own belief-forming mechanisms is unreliable is a type of higher-order evidence – unlike first-order evidence, it does not bear directly, as it were, on the truth of our beliefs, but tells us that our capacity to judge the evidence available to us has been compromised.[17]

The epistemic principle is a consequence of standard formulations of calibrationism, the most prominent theory of the epistemic significance of higher-order evidence.[18] Calibrationists hold that

[16] Note that – as their names indicate – the schematic versions of both the epistemic principle and the etiological principle (Section 3.3) are supposed to be true for all substitution instances of their variables. This is not the case for the schematic versions of the autonomy clause (Section 3.4) or the empirical premise.

[17] For helpful discussion of higher-order evidence, see Christensen (2010) and DiPaolo (2018).

[18] Sliwa and Horowitz (2015) present a clear and accessible defense of the view. See also White (2009),

higher-order evidence is relevant to the epistemic status of our ordinary, first-order beliefs, that information about how reliable we are can affect whether our first-order beliefs are justified. Suppose, for instance, that you believe that *p* is true, but you also have good reason to suspect that your judgments on the subject are unreliable. According to calibrationism, it is irrational to maintain your belief under these circumstances. The right thing to do, says the calibrationist, is to heed your higher-order evidence of unreliability, and abandon your first-order belief that *p*. Thus, if you have higher-order evidence that one of your belief-forming mechanisms is systematically unreliable, as the debunker claims is true of our moral sense, calibrationism implies that you are no longer justified in retaining any of the beliefs that it produces.[19]

Of special note is the independence requirement built into the epistemic principle, one of calibrationism's distinctive features. When higher-order evidence calls the reliability of one of our belief-forming mechanisms into question, the independence requirement tells us we must bracket off the mechanism's outputs when figuring out how to respond, and make that determination on the basis of the independent evidence alone.[20] The idea is that the higher-order evidence indicts not only the belief-forming mechanism itself, but its outputs as well, leaving them unsuitable to be cited as evidence or reasons for belief. Here is a case which will help to illustrate the need for this requirement:

Delusions of Gander

Christensen (2016), Schoenfield (2018), Vavova (2018), and Kappel (2019), along with Schoenfield (2015) and Isaacs (2021) for criticism. Calibrationism is usually formulated in terms of credences, but for the sake of simplicity, I will stick to all-or-nothing beliefs here.

[19] For ease of exposition, I will be a bit careless in what follows about the distinction between (on the one hand) psychological mechanisms that produce beliefs as outputs and (on the other) psychological mechanisms that produce other cognitive states as outputs, such as perceptions or intuitions. I will also occasionally run together the distinction between belief-forming mechanisms that are unreliable *tout court* and belief-forming mechanisms that are unreliable when used within some circumscribed domain. Nothing of substance hangs on these distinctions, at least so far as this paper is concerned.

[20] The independence requirement first emerged from the literature on peer disagreement, and most discussion of it has been restricted to that context. Elga (2007) and Christensen (2007; 2009; 2011; 2018; 2019) defend the requirement, while Arsenault and Irving (2012), Kelly (2013), and Lord (2014) number among its detractors. Vavova (2014) discusses the independence requirement in connection with evolutionary debunking.

A researcher working at DARPA informs her coworker Alex that, as a prank, she spiked his morning coffee with a psychotropic medication. In clinical trials, she tells him, the drug had no effect on half of subjects, while the other half experienced vivid hallucinations of geese for a day or two. On Alex's way home from work, he spies three geese frolicking in a field along the side of the road. Disregarding his colleague's warning, he comes to believe that there are three geese in the field.

Intuitively, I take it, Alex's belief is unjustified, whether or not the drug has actually affected him – knowing that there's a 50% chance that he's hallucinating means that he can no longer trust his visual experiences of geese. Suppose, however, that Alex decides to stick to his guns, and offers the following explanation for why: "My coworker warned me that I might suffer from hallucinations of geese, but I know that I'm not hallucinating right now. Evidence: I see three geese in the field. If I were hallucinating, I wouldn't be seeing three geese in the field, I'd only be imagining that I see three geese. But I'm actually seeing three geese. So I must not be hallucinating." On its face, this response seems unacceptably question-begging. Even if Alex is one of the lucky subjects who is immune to the effects of the drug, and his perceptions are veridical, there still seems to be something wrong with Alex citing his visual experiences as a reason to think that he's not hallucinating. The purpose of the independence requirement is to rule out this sort of response: it forbids Alex to appeal to his visual experiences as evidence, because they are not appropriately independent of the belief-forming mechanism whose reliability has been called into question. To avoid violating the independence requirement, Alex must bracket off or prescind away from his visual experiences when determining what he ought to believe. And, if he does so, he will surely recognize that he is not justified in retaining his belief that there are three geese in the field, in light of the risk that his perceptions have been altered by the drug.

Let me emphasize that there is no danger that this epistemic principle will lead to radical skepticism. To see why, it is helpful to distinguish it from a superficially similar-looking principle in the vicinity:[21]

Strong Internalist Condition on Justification (SICJ): If we lack independent evidence that the belief-forming mechanism which generates our beliefs in domain *D* is reliable, the justification for our *D*-related beliefs is defeated.

Effectively, (SICJ) requires that we have independent confirmation of a faculty's reliability before it can be used to form justified beliefs. But (SICJ) is too strong, and does have skeptical consequences, because we have no way of acquiring evidence which will underwrite the reliability of the five senses, taken as a suite, whose warrant does not ultimately trace back to the senses. The debunking argument's epistemic principle, in contrast, places a much weaker condition on justification: it requires only that we lack independent evidence, on balance, for thinking that the mechanism generating our beliefs is unreliable. And, while it may be impossible to independently verify that our senses are reliable, we certainly have no independent reason for thinking they are not. Moreover, as we saw in the introduction, there is no way to mount a parallel debunking argument against vision or the other senses, which were, according to the best available accounts of how these faculties evolved, selected to accurately represent the world around us.

This provides a basic overview of the calibrationist account of higher-order evidence, the source of the debunking argument's epistemic principle. Let me flag, however, that of all of the argument's premises, it is the epistemic principle – more specifically, the independence requirement embedded in the principle – which seems least secure to me. There are a great many cases, like *Delusions of Gander*, where the independence requirement appears indispensable for reaching the correct verdict. Nevertheless, other plausible and well-motivated perspectives on higher-order evidence, like Thomas Kelly's (2010) total-evidence view, do away with the requirement, and I cannot offer a decisive argument against Kelly's view here. To a large extent, then, the success or failure of the evolutionary

[21] Vavova (2018) draws this distinction clearly.

debunking project will depend on the outcome of these outside debates about how we should respond to higher-order evidence.[22]

3.3 The Etiological Principle

The etiological principle asserts that one type of higher-order evidence which can call the reliability of a belief-forming mechanism into question is evidence concerning that faculty's evolutionary history. Schematically:

Etiological Principle (Schema): If *D*-realism is true, and if we have good reason to think that the belief-forming mechanism, *M*, which generates our beliefs in domain *D* was selected principally for functions other than producing true *D*-related beliefs, then, unless we are able to corroborate that *M* is reliable through the use of some other belief-forming mechanism whose etiology is not subject to similar doubts, the balance of independent evidence suggests that *M* is unreliable.

In other words, if the best scientific explanation of how a belief-forming mechanism evolved suggests that it was selected primarily for functions other than acquiring true beliefs in its target domain, this gives us *prima facie* reason to think that that mechanism is unreliable, at least if we choose to remain realists about the domain.[23] *Prima facie* reason, rather than decisive reason, because a belief-forming mechanism with a dubious etiology might still be rehabilitated if its outputs can be corroborated using paradigmatically reliable faculties like vision or memory. More on this in the next section; for now, let's focus on the claim that learning that a belief-forming mechanism was selected for a function other than producing true beliefs gives us some reason to mistrust it.

[22] Note, though, that even if a view like Kelly's turns out to be correct, this is not necessarily a fatal blow for the evolutionary debunking project. Calibrationism undoubtedly makes things much easier for the debunker, but the total-evidence view still implies that higher-order evidence of unreliability will often (although not always) serve to defeat the justification for our first-order beliefs.

[23] Although the principle, as presented, concerns natural selection and belief-forming mechanisms, due to the well-known conceptual parallels between functions conferred by selection and functions conferred by design, it can readily be extended to apply to scientific instruments and other information-gathering artifacts as well. I will exploit these parallels in one example later in this section.

Here's a case which will help to illustrate the intuitive plausibility of the principle:

Bump in the Night

Jane is a precocious eight-year-old who is deeply afraid of the dark. Whenever her old house creaks and groans during the night, she comes to believe that there are insidious creatures lurking in its shadows. To ease Jane's fears, one day, her mother tells her that her beliefs have a perfectly reasonable explanation: far back in our evolutionary past, intrepid children who ventured out to investigate strange noises at night often fell prey to leopards and other predators, while timider children who remained safely tucked into bed survived and went on to have children of their own. As a result, natural selection favored children whose nighttime-monster-related beliefs caused them to remain in bed after dark, regardless of the truth of those beliefs, until the genes for nocturnal childhood fearfulness swept throughout the population.

Intuitively, once Jane is taught that the belief-forming mechanism generating her fears was selected to keep her in bed at night no matter what, not to produce true beliefs, this gives her good reason to suspect that it's unreliable, defeating the justification for her nighttime-monster-related beliefs. Notice, though, that the mother does not comment directly on the reliability of Jane's belief-forming mechanism, only on its evolutionary history. Hence, if the mother's story puts any pressure on Jane to reduce her confidence in her beliefs, this must be because evidence about the evolutionary history of our belief-forming mechanisms can affect the epistemic status of the beliefs they produce.

Bump in the Night is an imperfect analogy for morality, however, because Jane has independent background knowledge of what it would mean for a house to be populated by dangerous predators after dark. We do not have this sort of independent background knowledge when it comes to morality: the only epistemic access we have to the moral truths (if such there be) is by way of the faculty targeted by the debunking argument. At the same time, calibrationism's independence requirement prohibits us from appealing to the beliefs and intuitions generated by our moral sense once its reliability has been called into question by higher-order evidence. But if the moral sense's outputs are off-limits, and if we have no other method for getting at the moral truths, this means the debunker is claiming that we can determine that our moral judgments are unreliable from a position of complete ignorance about morality. Katia Vavova (2014: 92) argues that this is impossible. She writes:

> [W]e cannot determine if we are likely to be mistaken about morality if we can make no assumptions at all about what morality is like… [T]he debunker's challenge threatens anyone who holds that the attitude-independent moral truths do not, in any helpful way, coincide with the evolutionarily advantageous beliefs… But even to make this crucial judgment, that these two sets do not have the same contents, we need to know something about the contents of those sets—what they are or what they are like… If we can make no moral assumptions, then we cannot [establish] that the true evaluative beliefs and the adaptive evaluative beliefs come apart.

Vavova thinks the debunker will need to make at least some minimal assumptions about the nature of morality in order to establish that our moral sense is unreliable – she does not see how it could be possible to prove that a belief-forming mechanism consistently gets things wrong if we know nothing about the domain it's operating in. I believe, however, that Vavova is incorrect on this score: information about a belief-forming mechanism's etiology can establish that it's likely to be unreliable in a given domain, even if we have no background knowledge of the domain in question, and make no assumptions about what it's like. To see why, consider the following case:

Metronome

An experimental physicist presents Pete with a non-descript black box with an attached pair of headphones. "This is a cutting-edge morphic resonance field detector," the physicist tells him, "Works just like a Geiger counter. Try it out!" Pete, who has never heard of morphic resonance fields before, takes the device for a whirl, and is pleased when it gives off a satisfying clicking noise at intermittent intervals. Later, Pete notices a serial number (M24601-D) and a phone number stamped on the underside of the device. Pete calls the number, and inquires after the device's provenance. The clerk on the other end of the phone line, after looking up the serial number, informs Pete that it was designed to be a metronome, but was subsequently marked "D" for defective and discarded because it failed to keep the time.

Once Pete discovers that the device was designed as a metronome, I take it, he should abandon his belief that it reliably detects morphic resonance fields, and retreat to a position of agnosticism on the subject instead. Pete, in other words, ought to be skeptical whether the device really works as a morphic resonance detector, while at the same time suspending judgment about what morphic resonance fields are, what they are like, and whether they even exist in the first place. This suggests that Vavova's objection is mistaken – learning that a belief-forming mechanism has a deviant etiology can give us reason to think that it's unreliable at getting at the *D*-related truths, even if we make no assumptions about the nature of *D*, and approach *D* from a position of total ignorance. If we apply this lesson from *Metronome* to the moral case, it follows that information about our evolutionary history does have the power to establish that our moral sense is likely to be unreliable, even if we bracket off all of our pre-existing beliefs about the nature of morality, and adopt an attitude of agnosticism towards the subject instead.

In both *Bump in the Night* and *Metronome*, an agent acquiring information about the causal history of her beliefs suggests that the mechanism producing those beliefs is unreliable. But we are still in need of an explanation for why this is so, for why a connection like this should hold between a belief-forming mechanism's etiology and its reliability in the present day. Here is what I think is going on in these cases: take a trait or artifact which was selected or designed for some function, φ-ing, and pick some other use it might be put to, ψ-ing. In the abstract, what are the chances that it will be successful at ψ-ing? They are slim. We know, from our familiarity with the natural world and with human inventions, that traits or artifacts selected or designed for one function will generally be incapable of carrying out most other tasks you can name. To be sure, there are many cases in biology of adaptations being successfully repurposed, just as there are many examples of humans putting artifacts to creative, off-label uses. But, for an adaptation or artifact operating outside of its area of functional specialization, the realm of tasks that it's unable to perform will inevitably dwarf the realm of tasks that it's able to perform successfully. This means that it's highly unlikely that an adaptation will succeed at performing some arbitrarily-chosen task for which it was not selected.

To illustrate, take the heart, which was selected to circulate the blood. It also has a small number of other effects, for instance, it generates a nice percussive rhythm, and is a rich source of vitamins if eaten. But it is hopeless at composing sonatas, at shielding us from the rain, at performing arithmetic, at grasping and manipulating objects, at filtering toxins from the body, and so on, for virtually any other use we might wish to put it to. The same is true for all adaptations we are acquainted with: their usefulness seldom extends far beyond the function or set of functions for which they were selected. This, I think, is the intuition at the heart of evolutionary debunking arguments: evidence that the function of a belief-forming mechanism or artifact is to φ is equally evidence that it will not serve to ψ.[24] Thus, evidence that a device was designed as a metronome is evidence that it does not detect morphic resonance fields, and evidence that our moral sense was selected to promote cooperation is evidence that it does not reliably produce true moral beliefs.[25]

[24] Note that this rule is only intended as a statistical generalization, which means that evidence that a belief-forming mechanism was selected to φ has the potential to be screened off – rendered probabilistically irrelevant – by more specific information about the nature of φ-ing and ψ-ing. The rule is important here because of the context created by the independence requirement, where we are bracketing off the outputs of the belief-forming mechanism in question, and evaluating its reliability from a position of relative ignorance. This will often leave us without much admissible information about the nature of ψ-ing, allowing facts about the belief-forming mechanism's causal history to take on a larger evidential role. A slightly different type of example will help to illustrate: knowing nothing about pharmacognosy, you should think it unlikely that a piece of tree bark would be effective at curing a headache, because bark is selected to protect the tree's inner layers from the elements, not to produce medically valuable effects in humans. But the evidential significance of this fact about the evolved function of tree bark would be screened off by knowledge that the bark in question comes from the willow tree, and so contains the compound salicin, from which aspirin was derived.

[25] Several authors, including Berker (2014), have wondered whether the success of evolutionary debunking arguments really does depend on the finer details of human evolution, or whether the debunking argument's force instead comes from the fact that our moral beliefs have any causal history at all. The rationale for the etiological principle I have presented here suggests that the details of how we evolved do matter. For the

The realist has an obvious rejoinder here: perhaps a trait selected for one function is likely to be useless at some wholly distinct task. But what if the two tasks are not distinct, or if we are unsure whether they are distinct? Morality, after all, presents itself as being centrally concerned with cooperation, so if we trust what our moral intuitions tell us about the nature of morality, we should expect there to be substantial overlap between the true moral beliefs and those beliefs which make us into better cooperators. And, if morality is systematically linked to prosocial behavior in this way, a faculty selected to facilitate cooperation might well succeed, as a side effect, at reliably generating true moral beliefs.

This response is seductive, but it begs the question. So far as I can tell, we believe that morality is connected to cooperation only because our moral sense intimates to us that this is so, because cooperative actions seem intuitively right to us, while antisocial actions seem intuitively wrong. It's just not clear what genuinely independent epistemic path there could be to reaching this conclusion. But this means the realist is attempting to use the outputs of a faculty called into question by higher-order evidence in order to vindicate the epistemic credentials of that very same faculty. In other words, she is relying on our moral intuitions in order to prove that our moral intuitions are reliable. This is a paradigmatic violation of the independence requirement, no different in principle than Alex citing his visual experiences of geese as proof that he is not suffering from goose-related hallucinations. By contrast, if we bracket off our moral intuitions and attend to the independent evidence alone, as calibrationism requires, we will then be left with no reason to think there is any particular connection between morality and cooperation. We cannot, of course, rule out the possibility that they're somehow related – but anyone who invests more than a little credence in this possibility is almost certainly allowing her judgment to be illicitly swayed by her moral intuitions. Accordingly, since we have no independent grounds for thinking that beliefs which promote cooperation will also

debunking argument to work, it is essential that (to the best of our knowledge) our moral sense was selected for a function other than generating true moral beliefs. Other possible etiologies will not be as congenial to the debunker. For instance, were we to discover that our moral beliefs have never been shaped by selection at all, and are instead the product of a completely domain-neutral information-processing mechanism in the brain, it is not clear that an evolutionary debunking argument targeting moral realism would be viable.

tend to coincide with the moral truths, the realist's objection is unsuccessful.[26]

To sum up: I have used *Bump in the Night* and *Metronome* to motivate the schematic version of the debunking argument's etiological principle, and offered a rationale for why we should think the principle holds, a rationale that draws on inductive reasoning from our past experience with artifacts and adaptations. But, even supposing I were mistaken about this rationale, *Bump in the Night* and *Metronome* would still present a compelling case for the etiological principle on their own. When Jane and Pete learn that their beliefs have a deviant etiology, this certainly seems to suggest that the mechanism generating their beliefs is unreliable, and it is difficult to see why the same reasoning should not apply to our moral sense as well. At minimum, the realist owes us an explanation for why we should reject the etiological principle when it comes to the evolution of morality, when it seems to reach the right verdict across a range of similar cases.

3.4 The Autonomy Clause

The etiological principle leaves open a way for a belief-forming mechanism with a dubious etiology to be rehabilitated, if its outputs can be corroborated through the use of some other faculty whose epistemic credentials are above reproach. The purpose of the autonomy clause is to close off this escape hatch. Here is a schematic version of the autonomy clause:

Autonomy Clause (Schema): We are unable to corroborate that our *D*-related beliefs are reliable through the use of some other belief-forming mechanism whose etiology is not subject to similar doubts.

For a case where the autonomy clause comes out false, take literacy. We have good reason to think that the cognitive mechanism we employ in reading or writing was selected principally for functions other than producing true beliefs about the written word, namely, that literacy emerged too

[26] I discuss a related objection, involving so-called third-factor explanations, in Section 4.2.

recently in our evolutionary history to have engaged natural selection to any significant degree. The best explanation of our literary competence is that it is what Stephen J. Gould and Richard Lewontin (1979) call a spandrel, a byproduct of our faculties for processing the spoken word, rather than an adaptation. Happily, though, literacy's reliability can readily be corroborated through the use of the senses, spoken language, and memory. Each time we inscribe a message on paper, ask a friend or colleague to read it, and observe that their interpretation of its contents, repeated back to us, jibes with our own, we independently confirm that our faculties for understanding the written world are reliable. Consequently, the autonomy clause is false when it comes to literacy, which means that an evolutionary debunking argument targeting our ability to read would fail at this stage.

I do not believe it will be possible to rehabilitate our moral sense by a similar procedure, however, because morality presents itself as being independent from other domains of inquiry.[27] Moral truths, I take it, cannot be verified empirically, and play no role in our best scientific theories. This is as it should be. If empirical evidence could be brought to bear on foundational moral claims, it would be possible for some future scientific discovery to overturn our most deeply-held moral convictions, just as past scientific discoveries forced us to accept that the apparent motion of the heavens is an illusion produced by the diurnal rotation of the Earth, and that the apparent acceleration of objects in free fall is an illusion created by the curvature of spacetime around the Earth's mass. This would mean that empirical evidence could one day show, for instance, that we in fact have a *pro tanto* duty to torture children, or that comforting the afflicted is a great moral evil. Surely, though, these claims are false, and no scientific discovery could or should ever convince us otherwise. But the price of insulating our moral beliefs from scientific refutation is that they cannot lay claim to empirical confirmation. Hence, if we cannot trust the deliverances of our moral sense, none of our moral beliefs will be justified, because there is nowhere else we can turn to for aid.

[27] There is a large volume of literature on this topic (see, for instance, Larmore [2008] and McPherson [2008]) that I cannot hope to adequately summarize here. In what follows I will focus on presenting what I take to be the most compelling reason for thinking that the autonomy clause is true of morality.

3.5 The Empirical Premise

I have little to say here about the empirical premise. At this point, a large number of scientific and philosophical treatises have been written on the evolutionary history of morality in humans.[28] Most assign selection for prosocial attitudes and cooperation a pivotal role in this history, while none, to my knowledge, suggest that our moral sense was selected to produce true beliefs about a mind-independent domain of morals. Readers who still wish to reserve judgment on the matter, however, are free to set aside the empirical premise and read the debunking argument as (doubly) conditional instead: if both the empirical premise and moral realism are true, then the justification for our moral beliefs is defeated.

3.6 Putting It All Together

Let's recap. Research on the evolutionary history of the human moral sense suggests that it was selected not to produce true moral beliefs, but to allow us to reap the benefits of cooperating with other members of our species. Because it's unlikely that a faculty selected to promote cooperation will also succeed at reliably producing true moral beliefs, this gives us higher-order evidence that our moral sense is unreliable. Our moral beliefs might still be rehabilitated if it were possible to corroborate them through the use of some other faculty with impeccable epistemic credentials, but, due to morality's autonomy from other domains of inquiry, no outside help is forthcoming. The only basis we have for thinking our moral sense is reliable is that it is self-certifying, that our moral beliefs present themselves as being true to us, but the independence requirement forbids cognitive faculties from vindicating themselves in the face of higher-order evidence of unreliability. Hence, learning that the human moral sense was selected in our ancestors for purposes other than acquiring true moral beliefs systematically defeats the justification for our moral beliefs, and forces us to embrace moral skepticism – so long, that is, as we insist on remaining moral realists.

[28] For a variety of perspectives, see, in addition to Joyce (2006), Alexander (1987), Richerson and Boyd (2005), Hauser (2006), Bowles and Gintis (2011), Kitcher (2011), Baumard et al. (2012), Boehm (2012), Tomasello (2016), and Sterelny (2021).

4. Objections and Replies

4.1 Moorean Responses

A natural reply to evolutionary debunking arguments, developed by Jonathan Fuqua (forthcoming), takes its inspiration from G.E. Moore. Many of the realist's putative moral truths, like that it is wrong to torture small children regardless of whether or not anyone believes that it is, seem self-evident, indeed, practically indisputable. As Fuqua (forthcoming: 274) puts it, this is a claim that "nearly every sane person would believe were it brought before the mind's eye." The most that can be said for the premises of an evolutionary debunking argument, in contrast, is that they enjoy modest plausibility upon reflection. Consequently, since we are vastly more confident that our moral beliefs are true than we are in any of the evolutionary debunking argument's premises, a realist might insist that all the debunking argument can really succeed at proving is that one of its premises is false for some subtle reason which eludes us at present.

Unfortunately, the connections between evolutionary debunking and other cases involving higher-order evidence and defeat make it clear that this response won't do. Alex, from *Delusions of Gander*, might find his visual experiences of geese utterly compelling, but he still is not justified in trusting those experiences once he finds out there's a good chance that he's hallucinating. Similarly, Jane, the child protagonist from *Bump in the Night*, is no longer justified in believing that sinister creatures are lurking in her house's shadows at night once she hears about the evolutionary origins of her beliefs, no matter how vivid her fears may seem to her. We would not accept a Moorean argument from Alex or Jane as an appropriate reason for them to ignore the higher-order evidence they've been presented with, and consistency demands we say the same thing about evolutionary debunking and moral realism as well. Higher-order evidence has the power to defeat even our most deeply-held beliefs, and a subject who sticks to her guns in the face of higher-order evidence that the faculty producing her beliefs is unreliable is simply being irrational.[29] Hence, because evolutionary debunking

[29] According to calibrationism, that is. Philosophers who reject calibrationism, or who wish to restrict its scope, may insist that there is a privileged set of beliefs, perhaps including some moral beliefs, which are so obvious

arguments purport to offer higher-order evidence that our moral sense is unreliable, it is no help just to point to the manifest wrongness of torturing children. Moral realists who wish to resist the debunking argument's conclusion must come up with a compelling reason to reject one of its premises.

4.2 Third-Factor Explanations

One popular response to evolutionary debunking arguments is to invoke so-called third-factor explanations, which deny the argument's etiological principle, and insist that our moral sense could turn out to be reliable even if it was selected for purposes other than producing true beliefs.[30] Although the details of these accounts differ, I will follow Selim Berker (2014) in interpreting them as making claims about the grounding of moral facts. The idea is this: while our moral sense may not have been selected to acquire true moral beliefs *per se*, it was selected to track some set of natural facts – perhaps natural facts connected to survival (Enoch 2010), consciousness (Wielenberg 2010), pain (Skarsaune 2011), or well-being (Brosnan 2011) – which (at least partially) ground the moral facts. If some such grounding relationship holds, there is little mystery in how our moral sense could turn out to be reliable. Selection predisposes us to believe that survival and well-being are good, and the survival and well-being facts ground the facts about moral goodness, so our moral beliefs will, by and large, tend to come out true.

Calibrationists will reject third-factor explanations as violations of the independence requirement introduced in Section 3.2. The reason why the realist's claims about grounding seem plausible to us is because they agree with our moral intuitions, which tell us that survival is good, that pain is bad, and so on, but the independence requirement obliges us to bracket off our moral intuitions when evaluating higher-order evidence that our moral sense is unreliable. And it is hard to see how we could have any insight into what grounds the moral facts that is genuinely independent of our first-order moral beliefs and intuitions. Once we prescind away from the outputs of our moral sense, we

or self-evident that they can never be defeated by higher-order evidence.

[30] For compelling criticism of third-factor explanations along different lines than those pursued here, see Lutz (2018). Tersman (2017) and Klenk (2020) discuss third-factor explanations in the context of moral disagreement.

have no more reason to think that facts about moral goodness are grounded in facts about survival and well-being than to think they are grounded in facts about death and misery. Indeed, without our moral intuitions to rely on as evidence, it's not clear we have any basis for thinking that there ever were any moral facts in the first place. As a result, because third-factor explanations covertly depend on the outputs of our moral sense for their justification, these explanations run afoul of the epistemic principle's independence requirement, and so beg the question in favor of moral realism.

4.3 Debunking Without A Moral Sense

Edouard Machery and Ron Mallon (2010) challenge the assumption, built into the debunking argument's empirical premise, that we, as human beings, have a cognitive faculty which can reasonably be described as a moral sense. They begin by distinguishing three different theses about the evolution of morality, and then evaluate the empirical evidence for each. A first thesis, which Machery and Mallon (2010: 5) take to be relatively uncontroversial, is that some components of moral cognition – they suggest the grab bag of "emotions, dispositions, rule-based reasoning systems, or concepts" – have been shaped by natural selection. The second thesis they consider is that normative cognition, a general faculty for reasoning about normative concepts like obligation and permission, is an adaptation. This claim, they argue, has a fair amount of empirical evidence in its favor. Machery and Mallon are more skeptical of the third and strongest thesis, that our capacity for specifically moral cognition is an adaptation. And it is this third thesis, they believe, that is needed to underwrite evolutionary debunking arguments against moral realism.

There is plenty of room to find fault with Machery and Mallon's interpretation of the science. In particular, it is not clear to me how much difference there really is between their third thesis and the conjunction of the first two. But I will not pursue this line of argument here. Instead, I wish to push back against their contention that only the third thesis is strong enough to serve as the basis for evolutionary debunking arguments.

Let us suppose that only the weakest and least controversial of Machery and Mallon's theses, that certain components of moral cognition have been shaped by natural selection, is true. A lot hangs on which components, exactly, fit the bill. Consider the following two claims:

The Lives of Others: The lives, interests, and well-being of others have value, and should be given some independent weight in our deliberations.

Bidding from the Outside: Sometimes we ought to do things we do not want to do and that will not advance our interests.

I take The Lives of Others and Bidding from the Outside to be fundamental presuppositions of all ethical theories, all ethical theories, that is, aside from some vulgar forms of egoism with negligible appeal. If natural selection did not predispose us to accept these claims, or, at least, to reason and deliberate as though we do, it seems fair to say that morality is in no real sense an adaptation. If this is the case, the philosophers and scientists cited in Section 3.5 are badly mistaken about how morality evolved, and the debunking argument's empirical premise is false. If, on the other hand, natural selection did predispose us to think that the lives of other human beings have value, and that we should sometimes act against our own interests – not because these beliefs are true, but because they promote cooperation, and cooperation improved our ancestors' chances at survival and reproduction – this will be sufficient to get an evolutionary debunking argument off the ground. If we have good reason to suspect that we accept the core presuppositions of ethical thought only because they made our ancestors into better cooperators, that will undermine the justification for our beliefs in those presuppositions, and no project in ethics can succeed if beliefs as foundational as The Lives of Others and Bidding from the Outside turn out to be unjustified.[31]

Consider, also, that worries analogous to Machery and Mallon's do not strike us as the slightest bit compelling when it comes to the Transarcturians from *Fools Rush In*. Our verdict about their case,

[31] Compare Morton (2016) on this point.

that learning of the Earth scientists' discovery undermines the justification for their forbiddenness beliefs, does not seem to depend on which of Machery and Mallon's three theses is true of them. Perhaps selection endowed the Transarcturians with a domain-specific forbiddenness faculty, or perhaps only with a general faculty for normative cognition together with some basic forbiddenness components. It is difficult to see how this makes any difference. Once the Transarcturians find out about the possible evolutionary origins of their forbiddenness beliefs, their underlying conviction that there are any forbiddenness facts at all loses its justification and must be discarded, which casts everything else they believe about the forbidden zones into doubt as well.

By parity of reasoning, the success of evolutionary debunking arguments does not depend on how much of our moral cognition is an adaptation; it is enough that selection, in order to facilitate cooperation in our ancestors, shaped our minds to be predisposed to accept The Lives of Others and Bidding from the Outside without regard for whether they are true. And, if our confidence in these theses turn out to be unjustified, the rest of our moral beliefs will be subject to defeat as well, as the fruit of the poisonous tree.

5. Conclusion

Let's suppose that the debunking argument developed in this paper ultimately proves to be flawed. Does that mean the realist is out of the woods? Not quite – the realist still needs to contend with the bare analogy presented by *Fools Rush In*. After all, it seems clear that the justification for the Transarcturians' forbiddenness beliefs is defeated once they hear the Earth scientists' revelation about their evolutionary history, and it is difficult to see why, other than pure chauvinism, we should think our moral beliefs are any better off epistemically. Hence, no response to the evolutionary debunking argument developed in this paper can be considered complete unless it also explains why the apparent analogy between morality and the forbidden zones does not, in fact, hold.

This concludes my case for the thesis that evolutionary debunking arguments, properly formulated, present a powerful challenge to moral realism. I have said little, however, about how I think we should conceive of morality, if it is not to be construed realistically. Although I do not have space to discuss my own views at any length here, I suggest we should take seriously the proposal that morality is an adaptive illusion, one built into our minds by natural selection in order to facilitate cooperation among our hunter-gatherer ancestors.[32] Thus, the metaphysics of morality is the metaphysics of illusions, the epistemology of morality is the epistemology of illusions, and the semantics of morality is the semantics of illusions. I do not believe morality is unique in this respect – I follow Daniel Dennett (1991; 2013; 2016; 2017) in thinking that much of our conscious interface with the world, what he calls the manifest image, is an illusion created by selection to aid us in navigating our physical and social environment. It takes only a little reflection on the aim and workings of natural selection to convince yourself that this might be so. Selection's focus on survival and reproduction is single-minded and absolute; it has no special love for truth, and it will eagerly pack our minds with illusions, other evolutionary constraints permitting, whenever doing so contributes to our reproductive fitness. It should come as no surprise, then, if the moral sense, which presents itself as a window onto a mind-independent domain of morals, instead turns out to be a sham mirror pointed squarely back into our evolutionary past.

[32] To the best of my knowledge, this view originates with Ruse (1986).

Acknowledgments

I would like to thank Dallas Amico, Louise Antony, Philip Bricker, David Christensen, Patrick Grafton-Cardwell, Sophie Horowitz, Tim Juvshik, Hilary Kornblith, Christopher Meacham, Keehyuk Nahm, Ryan Olsen, and Alejandro Perez Carballo, along with several anonymous referees, for their helpful comments on previous drafts of this paper. I have also benefited from discussions with Patrick Forber and Said Saillant Montisano.

References

Alexander, R. (1987). *The Evolution o f Moral Systems*. New Brunswick: AldineTransaction.

Arsenault, M. and Irving, Z. (2012). "Aha! Trick Questions, Independence, and the Epistemology of Disagreement," *Thought: A Journal of Philosophy* 1(3): 185-194.

Baumard, N., Andre, J., and Sperber, D. (2012). "A Mutualistic Approach to Morality: The Evolution of Morality By Partner Choice," *Behavioral and Brain Sciences* 36(1): 59-78.

Berker, S. (2014). "Does Evolutionary Psychology Show That Normativity is Mind-Dependent?" in J. D'Arms and D. Jacobson (eds.) *Moral Psychology and Human Agency: Essays on the New Science of Ethics*. Oxford: Oxford University Press.

Boehm, C. (2012). *Moral Origins: The Evolution of Virtue, Altruism, and Shame*. New York: Basic Books.

Bogardus, T. (2016). "Only All Naturalists Should Worry About Only One Evolutionary Debunking Argument," *Ethics* 126(3): 636-661.

Bowles, S. and Gintis, H. (2011). *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton: Princeton University Press.

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

Christensen, D. (2007). "Epistemology of Disagreement: The Good News," *Philosophical Review* 116(2): 187-217.

Christensen, D. (2009). "Disagreement as Evidence: The Epistemology of Controversy," *Philosophy Compass* 4(5): 756-767,

Christensen, D. (2010). "Higher-Order Evidence," *Philosophy and Phenomenological Research* 81(1): 185-215.

Christensen, D. (2011). "Disagreement, Question-Begging, and Epistemic Self-Criticism," *Philosophers' Imprint* 11(6): 1-22.

Christensen, D. (2016). "Disagreement, Drugs, Etc.: From Accuracy to Akrasia," *Episteme* 13(4): 397-422.

Christensen, D. (2018). "On Acting as Judge in One's Own Epistemic Case," *Proceedings and Addresses of the American Philosophical Association* 93(1): 207-235.

Christensen, D. (2019). "Formulating Independence," in M. Rasmussen and A. Steglich-Petersen (eds.) *Higher-Order Evidence: New Essays*. Oxford: Oxford University Press.

Clarke-Doane, J. (2016). "Debunking and Dispensability," in U. Leibowitz and N. Sinclair (eds.) *Explanation in Ethics and Mathematics: Debunking and Dispensability*. Oxford: Oxford University Press.

Clarke-Doane, J. and Baras, D. (2021). "Modal Security," *Philosophy and Phenomenological Research* 102(1): 162-183.

Copp, D. (2008). "Darwinian Skepticism About Moral Realism," *Philosophical Perspectives* 18(1): 186-206.

Dennett, D. (1991). *Consciousness Explained*. New York: Little, Brown and Company.

Dennett, D. (2013). "Bestiary of the Manifest Image," in D. Ross, J. Ladyman, and H. Kincaid (eds.) *Scientific Metaphysics*. Oxford: Oxford University Press.

Dennett, D. (2016). "Illusionism As the Obvious Default Theory of Consciousness," *Journal of Consciousness Studies* 23(11-12): 65-72.

Dennett, D. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. New York: W. W. Norton.

DiPaolo, J. (2018). "Higher-Order Defeat is Object Independent," *Pacific Philosophical Quarterly* 99(2): 248-269.

Elga, A. (2007). "Reflection and Disagreement," *Nous* 41(3): 478-502.

Enoch, D. (2010). "The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope With It," *Philosophical Studies* 148(3): 413-438.

FitzPatrick, W. (2015). "Debunking Evolutionary Debunking of Moral Realism," *Philosophical Studies*

172(4): 883-904.

Fuqua, J. (forthcoming). "Metaethical Mooreanism and Evolutionary Debunking," *Proceedings of the American Catholic Philosophical Association*. <https://doi.org/10.5840/acpaproc2020917110>.

Gould, S. and Lewontin, R. (1979). "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme," *Proceedings of the Royal Society of London, Series B: Biological Sciences* 205(1161): 581-598.

Hauser, M. (2006). *Moral Minds: The Nature of Right and Wrong*. New York: HarperCollins.

Horn, J. (2017). "Evolution and the Epistemological Challenge to Moral Realism," in M. Ruse and R. Richards (eds.) *The Cambridge Handbook of Evolutionary Ethics*. Cambridge: Cambridge University Press.

Isaacs, Y. (2021). "The Fallacy of Calibrationism," *Philosophy and Phenomenological Research* 102(2): 247-260.

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

Joyce, R. (2013). "The Evolutionary Debunking of Morality," in J. Feinberg and R. Shafer-Landau (eds.) *Reason and Responsibility*. Boston: Cengage.

Joyce, R. (2016). "Evolution, Truth-Tracking and Moral Skepticism," in *Essays in Moral Skepticism*. Oxford: Oxford University Press.

Kappel, K. (2019). "Escaping the Akratic Trilemma," in M. Skipper and A. Steglich-Petersen (eds.) *Higher-Order Evidence: New Essays*. Oxford: Oxford University Press.

Kelly, T. (2010). "Peer Disagreement and Higher-Order Evidence", in R. Feldman and T. Warfield (eds.) *Disagreement*. Oxford: Oxford University Press.

Kelly, T. (2013). "Disagreement and the Burdens of Judgment," in D. Christensen and J. Lackey (eds.) *The Epistemology of Disagreement: New Essays*. Oxford: Oxford University Press.

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

Klenk, M. (2020). "Third Factor Explanations and Disagreement in Metaethics," *Synthese* 197(1):

427-446.

Korman, D. (2019). "Debunking Arguments," *Philosophy Compass* 14(12): 1-17.

Korman, D. and Locke, D. (2020). "Against Minimalist Responses to Moral Debunking Arguments," in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics Volume 15*. Oxford: Oxford University Press.

Larmore, C. (2008). *The Autonomy of Morality*. Cambridge: Cambridge University Press.

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

Lord, E. (2014). "From Independence to Conciliationism: An Obituary," *Australasian Journal of Philosophy* 92(2): 365-377.

Lutz, M. (2018). "What Makes Evolution a Defeater?" *Erkenntnis* 83(6): 1105-1126.

Machery, E. and Mallon, R. (2010). "Evolution of Morality," in J. Doris (ed.) *The Moral Psychology Handbook*. Oxford: Oxford University Press.

McPherson, T. (2008). "Metaethics and the Autonomy of Morality," *Philosophers' Imprint* 8(6): 1-16.

Morton, J. (2016). "A New Evolutionary Debunking Argument Against Moral Realism," *Journal of the American Philosophical Association* 2(2): 233-253.

Richerson, P. and Boyd, R. (2005). *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: The University of Chicago Press.

Ruse. M. (1986). "Evolutionary Ethics: A Phoenix Arisen," *Zygon* 21(1): 95-112.

Schoenfield, M. (2015). "A Dilemma for Calibrationism," *Philosophy and Phenomenological Research* 91(2): 425-455.

Schoenfield, M. (2018). "An Accuracy-Based Approach to Higher-Order Evidence," *Philosophy and Phenomenological Research* 96(3): 690-715.

Shafer-Landau, R. (2012). "Evolutionary Debunking, Moral Realism, and Moral Knowledge," *Journal of Ethics and Social Philosophy* 7(1): 1-37.

Sinclair, N. (2018). "Belief Pills and the Possibility of Moral Epistemology," in R. Shafer-Landau (ed.)

Oxford Studies in Metaethics Volume 13. Oxford: Oxford University Press.

Skarsaune, K. (2011). "Darwin and Moral Realism: Survival of the Iffiest," *Philosophical Studies* 152(2):229-243.

Sliwa, P. and Horowitz, S. (2015). "Respecting All the Evidence," *Philosophical Studies* 172(11): 2835-2858.

Sterelny, K. (2021). *The Pleistocene Social Contract: Culture and Cooperation in Human Evolution*. Oxford: Oxford University Press.

Sterelny, K. and Fraser, B. (2017). "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. (2015). "Does Anything Really Matter or Did We Just Evolve to Think So?" in A. Byrne, J. Cohen, G. Rosen, and S. Shiffrin (eds.) *The Norton Introduction to Philosophy*. New York: Norton.

Tersman, F. (2017). "Debunking and Disagreement," *Nous* 51(4): 754-774.

Tomasello, M. (2016). *A Natural History of Human Morality*. Cambridge: Harvard University Press.

Vavova, K. (2014). "Debunking Evolutionary Debunking," in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics Volume 9*. Oxford: Oxford University Press.

Vavova. K. (2018). "Irrelevant Influences," *Philosophy and Phenomenological Research* 96(1): 134-152.

Vavova, K. (forthcoming). "The Limits of Rational Belief Revision: A Dilemma for the Darwinian Debunker," *Nous*. <https://doi.org/10.1111/nous.12327>.

Wielenberg, E. (2010). "On the Evolutionary Debunking of Morality," *Ethics* 120(3): 441-464.

White, R. (2009). "On Treating Oneself and Others As Thermometers," *Episteme* 6(3): 233-250.

White, R. (2010). "You Just Believe that Because…" *Philosophical Perspectives* 24(1): 573-615.

Yong, E. (2016). "Inside the Eye: Nature's Most Exquisite Creation," *National Geographic* February

2016 Edition.