Evolution and moral naturalism

Richard Joyce

Penultimate draft of paper appearing in K. J. Clark (ed.),
The Blackwell Companion to Naturalism (2016) 369-385.

ABSTRACT:
Moral naturalism is the view that moral properties exist in a manner that fits with our scientific worldview. Might empirical discoveries about the genealogy of moral judgments (that, for example, they issue from an evolved psychological faculty) serve to undermine moral naturalism? One way of undermining moral naturalism is to show that moral properties do not exist at all. The possibility of genealogical considerations supporting this conclusion are examined and found to be weak. Alternatively, might empirical discoveries about the genealogy of moral judgments serve to *vindicate* some form of moral naturalism? This possibility is also explored and found to be unconvincing.

1. Introduction

Methodological naturalism is the principle that requires of anything whose existence we acknowledge that it fit (in some manner to be specified) with our naturalistic scientific worldview. Cats and couches pass this test; ghosts and gods do not. What of morality? There are two ways one might seek a "naturalization" of morality. First, one might investigate how moral judgments, moral practices, and moral institutions fit into the scientific worldview. One might call this the "psychology and anthropology of morality." Second, one might investigate how moral properties—like goodness, evil, praiseworthiness, and so forth—fit into the scientific worldview. That these are very different types of inquiry is best brought out by analogy: it is one thing to try to understand, in scientifically respectable terms, why people believe in ghosts; it is quite another thing to try to understand *ghosts* in scientifically respectable terms.

Methodological naturalists are likely to think that we are well down the road to providing the former kind of naturalization of morality, though many controversies and puzzles remain. But methodological naturalists are considerably less likely to think that the second kind of naturalization of morality is forthcoming. Many of them think that moral properties really are analogous to ghosts: entities that by their very nature do not fit into the scientific worldview. A methodological naturalist, encountering the idea of entities that by nature are non-natural, has straightforward advice: "Don't believe in them." This denial of belief can take different forms. The *error theorist* thinks that the discourse employing these concepts embroils us in falsehood, and that we should henceforth disbelieve and cease to assert propositions that imply that the entities in question exist. (The atheist is a familiar kind of error theorist about theistic discourse.) By comparison, the *noncognitivist* takes a step back and denies that the discourse under scrutiny was ever really committed to the existence of *entities* at all. While the sentence "Stealing is wrong" appears to ascribe the property of wrongness to stealing, noncognitivists hold that appearances are deceptive; they might hold that this judgment amounts to nothing more than "Boo to stealing!"—in which case, asking whether the property of wrongness fits into the scientific worldview betrays conceptual confusion. Both the error theorist and the noncognitivist may be methodological naturalists, and both are likely to think that the former kind of moral naturalization is desirable and forthcoming, but neither of them thinks that a naturalization of moral *properties* (the second kind of naturalization) is possible.

Not all methodological naturalists, however, take this attitude toward morality; many think that the second kind of naturalization of morality is worth pursuing and is forthcoming. The *moral naturalist* holds that moral properties—like goodness, evil, praiseworthiness, and so forth—do fit within the scientific worldview. In order to have a simple version of moral naturalism to hand, consider utilitarianism. A certain kind of utilitarian holds that the only thing of intrinsic moral value is *happiness*, and maintains that we are thus morally obligated to produce the maximal amount of happiness. According to this view, moral properties are identical to or supervene upon the property *being productive of happiness*, which is a causal/psychological quality that, it seems reasonable to assume, fits quite smoothly into the worldview provided by science.[1]

Metaethicists have long argued over these matters. Do moral judgments ascribe properties (as the noncognitivist denies)? If so, do these properties exist (as the error theorist denies)? If so, are these properties of a naturalistic order (as the moral naturalist affirms)? Arguments for and against these metaethical positions are numerous, and the progress remarkable only in its unimpressiveness. By contrast, there are grounds for optimism regarding the first kind of moral naturalization: the psychology and anthropology of morality. After all, questions about the nature of moral judgments, practices, and institutions seem tractable through familiar empirical means, enlisting the resources of such disciplines as psychology, sociology, anthropology, neuroscience, primatology, and economics. Given the progress one might reasonably hope for regarding the naturalization of moral judgments, an intriguing question arises: Could this progress contribute to the other kind of moral naturalization—naturalizing moral properties? Could understanding what moral judgments *are* (why we make them, where they come from, etc.) help establish whether the things that the judgments are *about*—moral goodness, evil, praiseworthiness, and so forth—actually exist? In particular, could understanding the origin of human moral judgment have metaethical implications? Even more particularly, could evidence that human morality is the product of Darwinian selection provide a premise in an argument establishing (or refuting) moral naturalism? It is the goal of the rest of this chapter to explore these questions.

2. The evolution of morality

Humans are moral creatures. In this context, this doesn't mean that humans are morally praiseworthy or admirable (or, for that matter, blameworthy or iniquitous); nor does it mean that humans are proper subjects of moral concern. Rather, it means that humans make moral judgments: We classify our world in terms of moral values, our actions in terms of moral rules,

[1] The moral naturalist need not be a methodological naturalist. She may be willing to countenance the existence of non-natural entities in her ontology; she just happens to think that moral properties are of a naturalistic kind. A utilitarian theist, perhaps?

our character traits in terms of moral virtues and vices, and so forth. Where does this way of thinking come from? One may pose this question synchronically, and it is the job of moral psychology to provide an answer —to reveal what faculties are involved in moral judgment. One may also ask the question diachronically, prompting an investigation of the processes by which humans came to make moral judgment in the first place. Recent years have seen a burgeoning of discussion about the evolutionary origins of the human moral faculty.[2] One possibility is that moral judgment is a relatively recent cultural invention, exploiting various psychological faculties that evolved for other purposes. Another possibility is that there exist in the human mind mechanisms that evolved specifically to make moral judgment possible. On the latter hypothesis—which can be called "moral nativism"—a faculty for making moral judgments is a biological adaptation that emerged because this way of thinking provided our ancestors with some sort of reproductive advantage over their competition. What sort of advantage? On this point, hypotheses diverge, but what is striking about them is that they seem entirely compatible with the error-theoretic stance; they do not appear to imply or presuppose that any of our ancestors' moral judgments were *true*. Let me explain.

Most biological traits have nothing to do with truth. It makes no sense to say that one's gall bladder, or any of its activities, is true or false. (Of course, *that one has* a gall bladder is true or false, but that's a different matter entirely.) But *judgments* can be true or false, and thus any evolved psychological faculty designed to produce a kind of judgment does have something to do with truth. It might be argued that the moral faculty (if there is such a thing) governs only feelings and emotions, and that the "judgments" it produces are not really the right kind of item to be assessed as true or false. But this extreme view is fairly implausible upon examination. It can be granted that the moral faculty has a great deal to do with emotions (like anger and guilt), but that is no reason to leap to the conclusion that truth-evaluable judgments have no place in its operations. It is very hard to see, for example, how a mere *feeling*, absent any truth-evaluable judgments, could count literally as the emotion of guilt, for guilt necessarily involves thoughts along the lines of "I have transgressed." Let us assume, then, that the moral faculty produces truth-evaluable judgments (let us, in other words, put noncognitivism to one side). However, this is not to assume that the evolutionary function of the moral faculty is to *track the truth*. The evolutionary function of a trait is the reason that it was selected for; it reflects why it was reproductively useful to our ancestors. In many cases, of course, truth is useful. Our evolved perceptual faculties, for example, are designed to give us an approximately true representation of where objects are in relation to us. There would have been no adaptive advantage for organisms to have a perceptual faculty producing false beliefs about, say, the location of food or the distance of predators. Similarly, if humans have an evolved faculty for simple arithmetic (see Dahaene 1997; Butterworth 1999), the explanation of why such an ability was useful to our ancestors presupposes the truth of the arithmetical beliefs. Again: having false beliefs about how many of your children are accounted for around the campfire, or how many lions are still chasing you after a couple have quit the chase, and so on, is likely to have been disastrous.

[2] See Krebs (2005), Joyce (2006), Machery and Mallon (2010), Mikhail (2011), and Kitcher (2011).

But it doesn't always work out this way; sometimes falsehood is useful. For example, people robustly judge themselves better than average in all sorts of ways, including supposing themselves to have an above-average ability to resist the temptation to make unrealistic positive self-evaluations (Friedrich 1996; Pronin et al. 2002). Such everyday delusions might enhance physical health or motivate confident participation in social activities, and thus it is not implausible to suppose that humans have been hard-wired by natural selection to systematically make such unrealistic self-evaluations (see Taylor and Brown 1988; McKay and Dennett 2009). But the beliefs don't need to be true in order to accomplish such adaptive ends; indeed, a great many of them must be false, since not everyone can be better than average. We must conclude that if there is an innate mechanism producing these kinds of beliefs, it does not have the function of tracking the truth: it exists not because it produces accurate self-appraisals, but rather in virtue of producing self-appraisals that benefit the agent's physical and/or psychological well-being.

Most nativist hypotheses suggest that moral thinking evolved because it played a vital role in enhancing social cohesion. An individual who judges cheating her comrades to be morally repugnant is less likely to do so—perhaps even less likely than someone who sees that cheating will harm her own long-term interests—and thus, on the assumption that cheating is frequently maladaptive, the moral judgment may be selected for. The plausibility of this hypothesis seems independent of whether cheating one's comrades (or anything else) actually *is* morally repugnant. Other moral nativists emphasize the role that moral judgments can play in signaling one's commitment to social projects (Miller 2007; Nesse 2007). Abiding by moral norms frequently involves foregoing immediate profit, meaning that morality can function as a *costly* signaling device. Costly signals correlate with *honest* signals, since the profits that can be gained by giving a dishonest signal will cease to provide a net gain if the signal is sufficiently expensive to produce (Noë 2001). Thus, if one's reproductive capacities depend on being chosen as a partner in various cooperative ventures (hunting, raising a family, etc.), and those doing the choosing will prefer those who are strongly committed to such ventures, then it may be adaptive to advertise one's prosocial allegiance in a costly fashion. Hence, making moral judgments in a sincere manner may be adaptive as a signaling device. Yet, again, there is no pressure to assume that the moral judgments need be *true* in order to play this adaptive role; one might endorse this hypothesis while maintaining an error-theoretic metaethical stance.

Here is not the place to assess the evidence regarding whether moral nativism is true or probable. The jury will be out on this matter for some time yet, and at present no one should be pressing claims either in favor of or against moral nativism with any great confidence. The main point I've stressed is that on all of the live versions of the moral nativist hypothesis, mention of *truth-tracking* is noticeably absent. Let us, then, assume for the sake of argument that one such hypothesis is true; will this have implications for moral naturalism or for metaethics more generally?

3. The case for moral nativism undermining moral naturalism

Recent years have seen a flurry of discussion surrounding what are called "evolutionary debunking arguments."[3] The basic idea is that discoveries about the evolutionary origins of some familiar mode of thinking might in some manner undermine that thinking. Nobody thinks this applies generally to any evolved psychological faculty; we've already seen that some evolutionary accounts will be *vindicating* (i.e., the cases of perceptual beliefs and arithmetic beliefs). The modes of thinking that seem particularly vulnerable to these arguments are those regarding which an evolutionary genealogy reveals the mechanisms in question not to be truth-tracking. Of course, the argument is nothing so simple (and fallacious) as saying that the faculty in question is not truth-tracking, therefore the judgments it produces are not true. The notion of "truth-tracking" in play here is one that pertains only to the evolutionary function of the faculty, and observing that a psychological faculty does not have the evolutionary function of producing truths does not imply that it *fails* to produce truths. (Human bipedalism does not have the evolutionary function of allowing us to ride bicycles, but it nevertheless allows us to do so.)

Evolutionary debunking arguments vary in the strength of their conclusions, for when we say that such arguments aim to undermine morality, the term "undermine" is intentionally indeterminate. Sharon Street (2006), for example, argues that moral nativism reveals moral realism to be probably false. Moral realism is the metaethical view that moral properties exist as objective features of the world.[4] Street argues that the moral realist, confronted with evidence confirming moral nativism, faces a dilemma concerning the relation between our moral judgments (products of the contingencies of our evolutionary ancestry) and the supposed realm of objective moral facts. On the one hand, if there is no relation, then it would be an astonishing coincidence if many of our moral judgments were even approximately true—a conclusion supposedly disagreeable to the realist. The problem with the other horn of the dilemma, according to Street, is that it is empirically dubious. I have already noted that the usual nativist hypotheses see the ancestral adaptive payoff of having a moral faculty in terms of enhancing certain cooperative tendencies, not in terms of tracking moral truths. Street thinks this "adaptive link hypothesis" is superior to any truth-tracking hypothesis for three reasons: it is more parsimonious, clearer, and more illuminating of the phenomenon it seeks to explain (2006: 129).

Street's antirealist conclusion might be put as follows: "There are no objective moral facts." Yet she doesn't deny the possibility of moral facts *per se*—they will simply be of a constructivist (i.e., non-objective) nature. For a simple example of what such non-objective normative facts might be like, think of facts about the value of money. That a given disk of metal is worth 10 cents may be a fact, but it is a fact constituted by human activity (approximately, by our collective willingness to treat it as worth 10 cents). By contrast, that a given disk of metal is

[3] See Enoch (2010), White (2010), Wielenberg (2010), Brosnan (2011), Kahane (2011), Joyce (2013a; 2016), Fraser (2014), and Kelly (2014).

[4] Moral realism cuts across moral naturalism, inasmuch as there are non-naturalist versions of realism (e.g., G.E. Moore's intuitionism) and nonrealist versions of naturalism (e.g., Ronald Milo's constructivism).

made up mostly of copper is in no sense constituted by human activity; it is an objective matter. The moral constructivist holds that moral values are, very roughly, like monetary values. But just as there's nothing weird or non-naturalistic about monetary values, so there needn't be about constructed moral values. Thus, Street's arguments do not threaten moral naturalism; she allows that moral properties may well exist and fit smoothly into the scientific world order, so long as we construe them as human-constructed features of our world (Street 2012).

A much more radical kind of undermining—and one that is inimical to moral naturalism—would be the error-theoretic conclusion that there are no moral facts at all. What are the prospects of moral nativism providing support for such a view?

Michael Ruse (1986; 2006; 2009) argues that part of what made moral thinking adaptive for our ancestors was its being imbued with a kind of practical objectivity. Judging norms that enjoin cooperation to be human constructs (albeit highly useful ones) renders them susceptible to practical sabotage: such norms seem "escapable," as if there is no good reason for following them if one can get away with secret transgressions. By contrast, judging norms that enjoin cooperation to be somehow *there*, in the nature of things—as if the world comes with obligations written into it, as if people come with rights inbuilt, as if some actions are inherently wrong and some right—can strengthen one's resolve to comply. "It is precisely because we think that morality is more than mere subjective desires," writes Ruse, "that we are led to obey it" (1986: 103).

Thinking along these lines can cast doubt on the viability of the constructivism favored by Street (and others). In order for morality to serve its main purposes, the thinking goes, it must be treated as inescapable, as authoritative, as objective. Indeed, one might go so far as to say that if some kind of objectivity is necessary for the uses to which we put moral norms, then that kind of objectivity is an essential part of morality, conceptually speaking. If a system of human-constructed norms couldn't play many of the practical roles that morality plays, then there are grounds for claiming that it could not count, literally, as a *moral* system at all. This seems to be what Ruse intends when he writes, "Ethics is subjective, but its meaning is objective" (2006: 22) and "[W]hat I want to suggest is that…the *meaning* of morality is that it is objective" (2009: 507).

Suppose, then, that Street's evolutionary argument against moral realism were convincing but that, for reasons like those just sketched, her embrace of constructivism were not. In other words, we would have a good argument to the conclusion "There are no objective moral facts," but we would also have grounds for maintaining "Moral facts are essentially objective." From these premises, the only conclusion to draw would be the error-theoretic one: "There are no moral facts at all." We therefore have the structure of an evolutionary argument against moral naturalism before us; now let us critically assess it.

An important component of Street's argument to debunk moral realism is an appeal to parsimony. Regarding explanations of why moral thinking was beneficial to our ancestors, she argues that the adaptive-link hypothesis is superior to the realist's tracking hypothesis (according to which our ancestors profited from developing accurate representations of a realm of objective moral facts) in part because the former is more parsimonious:

> The tracking account obviously posits something extra that the adaptive link account does not, namely independent evaluative truths (since it is precisely these truths that the tracking account invokes to explain why making certain evaluative judgements rather than others conferred advantages in the struggle to survive and reproduce). (Street 2006: 129)

Ruse's argument also appeals to parsimony. He often uses an analogy to make his point, referring to the spike of interest in séances in Europe in the aftermath of the First World War (Ruse 1986: 256-257; 2006: 22-23; 2009: 504-505). Imagine a grief-stricken mother attending such a séance, during which time she comes to believe that her dead son has spoken to her from beyond the grave. We can explain everything that needs explaining about this belief by reference to psychological and sociological factors; there is no need to suppose that the belief might be *true*. Similarly (Ruse thinks), moral nativism explains everything that needs explaining about why humans judge certain actions to have objective moral status; there is no need to suppose that these judgments might be *true*. To do so would be ontologically profligate.

Street's and Ruse's parsimony arguments question whether moral truths are needed to explain our moral judgments. These can be seen as local instances of a broader type of argument against moral realism, questioning whether moral truths are needed to explain *anything at all*. Is there any phenomenon encountered in the world that remains mysterious and inexplicable unless we introduce moral facts into the picture? In fact, however, it seems plausible to claim that the broader argument collapses into the local form of argument (focused on the explanation of moral judgments), for we cannot acknowledge an instance of a moral fact being necessary to explain some phenomenon without ourselves making a moral judgment. Consider, for example, the existence of Polish concentration camps in 1944. Seeking an explanation for their existence will lead to reference to (inter alia) decisions made by Hitler. Seeking to push the explanation back further—asking why Hitler made these decisions—one might make reference to Hitler's *depravity*. (This example is from Sturgeon 1985.) Here, then, we seem to have a concrete phenomenon (literally concrete: barracks and buildings), the explanation of whose existence makes reference to (inter alia) a moral property. Yet recognizing Hitler's depravity requires one to make a moral judgment—and then, of that moral judgment, the question can always be pressed: Must moral facts be invoked in order to explain it?

This form of argument against moral facts has been articulated by Gilbert Harman (1977; 1986). Harman's version of the argument doesn't focus on the evolutionary explanation for moral judgment, like Street's or Ruse's, but rather on the process of socialization. For the purposes of a debunking argument, though, this is a distraction; the crucial feature is that moral judgments are taken to issue from a process that appears not to be truth-tracking. Harman, however, doesn't push the argument through to an error-theoretic conclusion. His argument contains a crucial qualification: "In the absence of some sort of naturalistic reduction of moral claims, this kind of explanation [i.e., e.g., one that makes reference to Hitler's depravity] does not make it possible to test moral claims empirically in all the ways in which scientific claims can be empirically tested" (1986: 62). It is the conditional nature of this claim that is important to notice. If "some sort of naturalistic reduction of moral claims" is available, then this threat to the existence of moral facts (or the problem of testing moral principles, to put it in Harman's terms) recedes. In other words, the possibility of a reduction seriously challenges the form of debunking argument relying on an appeal to parsimony. Examples will help clarify this.

Consider trying to explain an avalanche. One explanation—call it the "vernacular"—speaks of melting snow and warm weather. Another—call it the "molecular"—speaks only of molecules of hydrogen and oxygen, thermal energy, and so forth. Suppose our evidence confirms the molecular hypothesis. Would this count as a disconfirmation of the vernacular hypothesis? Would we say, "There is no need to posit *snow* as part of the explanation of the avalanche, for to do so would be ontologically profligate"? Clearly not. Our complete molecular explanation of the avalanche may not mention *snow* at all—it may mention only H_2O arranged in a crystalline lattice—but snow just *is* H_2O arranged in a crystalline lattice. There is a reductive relation between the two.

Similarly, we can explain Hitler's decisions without mentioning *depravity* at all. Delving into his past, we might locate character-forming forces that gave him certain beliefs and attitudes, traumatic events that prompted neurotic outlooks, situations that influenced him to respond in unusual ways, and so on—all of which ultimately led him to think that genocide was a reasonable course of action. But it can be responded that some of these psychological traits described in nonmoral terms just *are* depravity. The moral property of depravity is not some ontological extra, to be excluded by considerations of parsimony; it is there, implicitly, in the accepted psychological explanation.

Similarly again, we might be able to explain why the moral faculty evolved without making reference to any moral facts, but rather by presenting an adaptive-link account couched in terms of moral thinking strengthening our ancestors' motivation to cooperate and so on But it can be responded that moral facts are implicitly present in the adaptive-link hypothesis, in virtue of standing in identity or supervenience relations to the items explicitly mentioned (such as the tendency to act cooperatively). Again: the moral properties need not be thought of as ontological extras to be excluded by considerations of parsimony.[5]

Harman captures the point succinctly when he writes, "Even if assumptions about moral facts do not directly help explain observations, it may be that moral facts can be reduced to other sorts of facts and that assumptions about these facts do help explain observations" (1977: 13). So the answers to the questions "Are moral truths needed to explain our moral judgments?" and "Are moral truths needed to explain anything at all?" may both be "No"—and one may, moreover,

[5] We can now see what is misleading about Ruse's séance analogy. The grieving mother's belief ("Johnny talked to me from beyond the grave") may be explained entirely in naturalistic terms, but it is wholly implausible that one might take what would be necessary to render this belief *true* (i.e., Johnny's ghostly post-mortem existence) and locate identity relations between these truth-makers and the psychological items appealed to in the naturalistic explanation.

endorse a strong principle of parsimony that obligates us to banish non-explanatory items from our worldview—but for all this, moral facts may survive. The metaethical view that allows moral facts to survive—that maintains that identity or supervenience relations hold between moral properties and the entities appealed to in the endorsed explanatory hypotheses—is moral naturalism.[6]

We started out this section investigating whether moral nativism might undermine moral naturalism. An argument to this effect was outlined, but it turns out that the availability of a viable moral naturalism would render this argument flawed, and therefore this particular debunking strategy appears to be question-begging. On the other hand, moral nativism provides a degree of support for the moral error theory insofar as it offers an explanation of the source of the massive mistake of which the error theorist accuses ordinary thinking. Of course, the moral naturalist will deny that there is any massive mistake requiring explanation; the point is, though, that error theorists who can provide an explanation of the massive mistake they discern in ordinary thinking are certainly in a dialectically stronger position than error theorists who shrugs their shoulders when asked for an explanation. In this respect, at least, moral nativism is a friendly result for the error theorist.

Moral naturalism, we have seen, has the potential to debunk a certain form of debunking argument. But might one go further and claim that moral nativism actually provides support for moral naturalism? I'll address some preliminaries in the following section, before taking up this question in the section after that.

4. Debunking a debunking argument

In responding to moral debunking arguments, several critics have suggested that, even accepting the nativist genealogy, moral judgments might, more or less, track the truth. At this point, we should note that there are two different notions of "truth-tracking" in play: an evolutionary notion and an epistemological notion. The former refers to what a psychological faculty is *supposed to do* (evolutionarily); the latter is often taken to refer to a covariation between a belief and the fact that it represents.[7] Suppose an evolved faculty has the function of producing judgments of the form "X is P." These representational states might covary robustly with Y's being Q, but we would not on that account say that the faculty tracks the truth or has the function of doing so. Whether the faculty tracks the truth depends on whether the judgments covary with those fact(s) that they represent—in this case, X's being P. Whether the faculty *has the function of* tracking that truth depends on whether success at truth-tracking explains the emergence and persistence (and thus the very existence) of the faculty.

[6] Harman himself endorses a kind of naturalistic reductionism, though of a relativistic nature. He concludes that "there is empirical evidence that there are (relational) moral facts" (1977: 132).

[7] In fact, I think epistemological truth-tracking is quite difficult to spell out, and the covariation analysis runs into difficulties when beliefs concern necessary truths and necessary falsehoods. See Joyce (2016) for discussion.

Another way of putting this point is using Elliott Sober's (1984) useful distinction between *selection of* a trait and a trait's being *selected for*. The latter indicates that the trait is the target of selection, in that the nature of the trait plays a causal role in the selective process. The former, by contrast, indicates that the trait is a byproduct of the selective process. (Suppose that colored marbles are dropped through a sieve that allows only the small ones to pass, and suppose that all and only the small marbles happen to be red. There has been selection *of* redness but not selection *for* redness; the sieve selects for smallness.) If a faculty for moral judgment is the product of natural selection, it is possible that, even if its truth-tracking quality was not selected for, there has nevertheless been selection of a truth-tracking trait.

Given these considerations, there are several ways in which critics of debunking arguments may proceed. They may accept that the evolved moral faculty (assuming there is one) is not evolutionarily truth-tracking, but claim that it may be, nonetheless, epistemically truth-tracking. Perhaps the moral facts are causally connected to the natural facts invoked by the moral nativist. Perhaps there is a naturalistic reduction between the two. Perhaps there is a "third factor" that explains both the natural facts in question and the moral facts. Alternatively, the critic may maintain that the moral faculty is (at least in part) truth-tracking in the evolutionary sense. The differences among these options are subtle and, in some cases, potentially problematic; it is not my goal here to tease them apart. But it is reasonable to observe that, despite differences, these critical responses form a family of strategies.

Kevin Brosnan (2011), for example, suggests the possibility that cooperation with others is morally good. The evolutionary process would explain both why we believe that cooperation with others is morally good (because doing so enhances the tendency to cooperate in an adaptive manner, say) and why cooperation is in fact good (because it tends to promote well-being, say). David Enoch (2010) presents a structurally similar argument. He speculates that survival or reproductive success is morally good, and that Darwinian forces have shaped our moral beliefs such that they often concern actions and events that promote survival and reproductive success. Thus, even if the *truth* of our ancestors' beliefs does not figure in the account of why they were adaptive, nevertheless they *were* (sometimes and nonaccidentally) true. Erik Wielenberg advocates another such argument, supposing that natural selection has provided humans with beliefs concerning individuals being surrounded by "a kind of moral barrier that it is…illegitimate for others to cross" (2010: 444-445). Such a belief might well have been adaptive in various ways. Moreover, the very cognitive capacities that make forming such a belief possible also guarantee (or at least probabilify) that one has such a "moral barrier," thus ensuring the belief's truth.

All the aforementioned philosophers aim to find a place for moral facts within the naturalistic nativist hypothesis, though most of them are more concerned with outlining the formal relation between moral and natural facts than they are with making the relation plausible. One diagnosis of why someone might think that a mere sketch might suffice to defeat a debunking argument is a failure to distinguish between debunking arguments that aim to establish an error theory and debunking arguments that aim to establish the more modest conclusion that our moral judgments lack justification. (Thus far, this chapter has been concerned exclusively with the former. The reason for this is that the thesis that moral judgments lack justification is consistent with either the truth or the falsity of moral naturalism, and thus is peripheral to the focus of this chapter.) If one aims to establish an error theory—that no moral judgments are true—then an opponent's establishing the mere possibility that moral facts implicitly reside in the nativist genealogical account is pertinent. But if one is arguing that moral nativism undermines the epistemic status of our moral judgments, then an opponent's establishing the mere possibility of moral facts implicitly residing in the nativist genealogy misses the point. Analogy: If I deny that life on Mars ever existed, then your convincing me of the possibility that it once existed in underground aquifers should give me pause. By contrast, if my claim is that we have no grounds for believing one way or the other whether life on Mars existed, then your convincing me only of the *possibility* that it once existed in underground aquifers doesn't budge my claim. I may be quite aware of this possibility; my contention is that the evidence is insufficient to grant the belief in Martian life warrant.

It is worth noting at this point that though the previous section explored the possibility of establishing an error theory on the basis of moral nativism, this is by no means the typical debunking strategy. Street is certainly no error theorist, and nor is Harman. I myself have argued for an error theory (Joyce 2001), but on *a priori* metaethical grounds, not via an evolutionary debunking argument. The debunking argument that I have offered (Joyce 2006) has as its conclusion that no moral judgments are justified.[8] Ruse's metaethical position is open to interpretation; though earlier I quoted passages that seem to provide a bridging premise to an error-theoretic conclusion, Ruse never (to my knowledge) explicitly endorses that view. Generally, he expresses himself as (like Street) opposed to moral realism, not opposed to moral facts *per se*. If, then, a critic provides only a sketch of moral naturalism because he or she thinks that this is all that is necessary to refute the error-theoretic debunking argument, it is not at all clear who among actual philosophers is the target.

Another possible diagnosis of why a critic might think that a mere sketch of moral naturalism suffices to refute the debunking argument—even one with an epistemic conclusion—is that she interprets the would-be debunker as claiming that whatever the nature of moral facts, our beliefs concerning them would lack justification. Wielenberg, for example, attributes to the epistemic debunker the position that even if moral facts exist, nativism shows that we would not know of them. He thus grants himself license to offer only a sketch of the nature of moral facts (mentioned previously)—virtually stipulating their existence and nature—and then proceeds to show how knowledge of these facts would be consistent with moral nativism. But this misrepresents the epistemic debunking argument. To recycle an analogy: Suppose one's claim that we do not know one way or the other whether life on Mars ever existed meets with the following response:

[8] I once made the ill-thought-through decision to label the thesis that moral judgments lack justification a version of "error theory" (Joyce 2006: 223), but have since recanted this (Joyce 2013a; 2016).

> Imagine that Martians live in large underground cities because their leaders want to stay hidden from human view, but suppose also that a small minority of Martians desire to secretly reveal the truth to humans. These rebels leave clues for our little robotic vehicles to find: rocks moved overnight, boulders that look a bit artificial, odd-shaped shadows, and so forth—that is, things that we actually do occasionally observe on the Martian surface. If the facts are as just described, then it can be said that those actual humans who believe in Martian life on the basis of weird-shaped rocks and so forth form their beliefs via a reliable process: these clues connect their beliefs about Martian life with actual Martian life. Therefore, it is possible that we (or at least a discerning few of us) know that there is life on Mars.

This response seems entirely wrongheaded. If one claims that we lack moral knowledge, one's opponent is not free to stipulate moral facts (and the nature of our connection with them) as he or she wishes and then declare that the skeptical claim is mistaken.

In short, merely pointing out how some form of moral naturalism *might* block a debunking argument does not show that any form of moral naturalism *does* block a debunking argument. In order to actually debunk a debunking argument, a particular theory of moral naturalism has to be carefully articulated and defended from criticism. Possibilities have to be converted into plausibilities.

5. The case for moral nativism supporting moral naturalism

One of the more developed versions of evolutionary moral naturalism currently available is that put forward by Kim Sterelny and Ben Fraser (forthcoming). Sterelny and Fraser argue that the evolutionary function of moral cognition is, in part, to track moral facts. More particularly, they argue that moral cognition (when functioning properly) is sensitive to "facts about profitable forms of cooperation, about social arrangements and cognitive dispositions positively and negatively relevant to the stable exploitation of those opportunities" (forthcoming). "Given a social and physical environment, and a set of interacting agents with their opinions and motives," they write, "there will be facts about whether their current norms are efficient means to stable and profitable cooperation" (forthcoming). These facts are the naturalistic reductive base of moral facts; they are what human moral cognitions are designed to pick out. This allows that one culture's moral system may be closer to the truth than another's; in principle, a culture's moral views (assuming homogeneity) could be entirely true. These moral judgments, moreover, will be true in virtue of the obtaining of certain naturalistic facts—complex and epistemically hard-to-access facts, to be sure, but certainly facts that fall under the purview of the natural sciences.

Sterelny and Fraser invoke the "Canberra Plan" to support their evolutionary naturalism (see Lewis 1970; Jackson and Pettit 1995; Jackson 1997). This method (popular among some prominent philosophers connected to the Australian National University in Canberra) consists of reconstructing the folk platitudes surrounding a concept—a reconstruction that may involve systematization and reflective equilibrium—the results of which can then be assessed from the perspective of our best scientific worldview. They suggest, furthermore, that the folk platitudes in question need not all be items of propositional knowledge, but may include practical skills. Thus, that the folk can successfully *use* a set of concepts for negotiating the world stands in favor of those concepts being vindicated by the Canberra Plan, as opposed to being debunked. Ancient astronomy, for example, for all its horrendous errors, was to some extent responsive to celestial facts in a manner that "guided navigation, calendar construction and time keeping" (Sterelny and Fraser forthcoming), and thus this methodology partially vindicates its conceptual scheme; we are not error theorists about the ancient discourse of *stars*, *the moon*, and so on. The concept *witch*, by contrast, does not pass:

> Even if those persecuted [as witches] were an identifiable subgroup…discrimination did not leverage adaptive behavior, even by the lights of the witch-burners. It did not prevent crop failures or other misfortunes…. So nothing in the world remotely corresponds to the witch-identifying maxims; nor did witch representation leverage adaptive behaviour. (Sterelny and Fraser forthcoming)

An application of the Canberra Plan, Sterelny and Fraser think, will underwrite the reductive relation they propose between moral concepts and the cooperation-oriented naturalistic properties in question, thus offering a partial vindication. They recognize that the world may not supply a property that perfectly satisfies the best systematization of folk moral platitudes, but they think that, on balance, the kind of naturalistic properties they have in mind are the best deservers of this honor. The folk moral concepts may contain errors, but if they also contain truths and, importantly, if they sufficiently promote adaptive behaviors, then they may pass the Canberra test. This is where the nativist hypothesis may provide specific support for moral naturalism, since nativism implies that moral cognition was, relative to ancestral environments, broadly adaptive.

That Sterelny and Fraser present a nuanced and noteworthy form of evolutionary moral naturalism is not in doubt; what is less clear is whether they offer anything to counter the standard kinds of objection that are typically leveled at moral naturalism. If their aim is to make a move against a presupposed backdrop of sympathy with moral naturalism, then this is no great criticism; but if their theory is supposed to be persuasive in a broader metaethical setting—one which includes interlocutors with grave misgivings about moral naturalism in general—then it is doubtful that they make very much headway. Of the several "grave misgivings" one might have about moral naturalism, its inability to adequately accommodate any notion of real normativity is the one on which I will focus in my closing discussion. (For further criticism of moral naturalism, see Mackie 1977; Horgan and Timmons 1991; Kelly 2004; Parfit 2011: pt. 6; Tropman 2012; Miller 2013.)

The moral wrongness of an action, it might be said, directs one against its performance; it tells one what to do. But natural properties appear to be inert in this respect. The fact that some action promotes stable forms of cooperation does not, in and of itself, guide one to perform it. The key phrase here is "in and of itself." Of course, if a person cares about promoting cooperation, then that an action has this property will be of practical import. But for the person who lacks this care, the fact that an action has this property may be an item of little interest; indeed, it looks like something that may be (*ceteris paribus*) legitimately ignored by him or her.

The view often promoted by reductive naturalists (explicitly or otherwise) is that a given property might be picked out by two different terms: a purely descriptive predicate (e.g., "…erodes stable cooperative practices") and a moral predicate (e.g., "…is wrong")—the latter of which may also have certain extra features, such as expressing to one's audience, by linguistic convention, one's disapproval. But in embracing such a view, the naturalist locates normativity in those who make moral judgments: in the force with which they utter their judgments, in the emotions they thereby express, in the persuasive intentions that accompany their utterances, and so forth. It may be objected, however, that a moral realism deserving of the name should locate normativity *in the object*. It is surely supposed to be the wrongness of stealing that should guide one not to do it, not the qualities of the utterance made by the person telling one not to steal (see Dancy 2006). Moral naturalists cannot accommodate this. The best they can manage is to locate a naturalistic property that most people, as a matter of contingent fact, care about, implying that most people, as a matter of contingent fact, will have reason to comply with moral imperatives and promote moral values.[9]

Many find this aspect of moral naturalism wholly inadequate, and it is not unreasonable to think that it represents the failure to accommodate a central folk moral platitude. The folk think that what Jack the Ripper did was *wrong* regardless of whether he cared about morality. Indeed, his not giving a fig wouldn't incline us to concede that morality is of no practical significance to him—that he can legitimately ignore it—but rather would encourage us to condemn him all the more vigorously. (And the fact the Jack went to his grave unpunished eliminates any complications that might arise concerning the punitive responses of his fellows, yet in no way softens our moral condemnation.) This failure to accommodate so central a platitude suggests that no set of naturalistic properties could *be* wrongness.

Sterelny and Fraser may counter that it is one of the great virtues of the Canberra Plan that concepts can survive the revelation that the folk are mistaken about certain elements of the concept. Perhaps the folk are simply misguided about the idea that wrongness in and of itself should guide practical choices; perhaps they are misguided in thinking that moral facts are relevant to agents irrespective of their desires or interests. But the flexibility of the Canberra Plan cuts both ways. Perhaps instead the folk are simply misguided in thinking that moral properties exist at all. After all, there is nothing inherent in this method that stipulates that only naturalistic and scientifically respectable platitudes count, or that they count for more; if the folk tacitly embrace platitudes that are inimical to moral naturalism, then so be it. One should expect to find many folk concepts for which a Canberra-Plan-type naturalization will fail, for there is no reason to think (and many reasons to deny) that folk concepts emerged and evolved with methodological naturalism as a background constraint. J.L. Mackie, for example, argues that the

[9] Note that the issue here does not, in the first instance, concern *motivation*. The claim is not that buried in folk thinking is motivation internalism: the thesis that moral judgments necessarily motivate. Sterelny and Fraser admit that according to their view, moral facts are such that "their power to motivate us is contingent" (forthcoming). This admission, however, does not speak to the complaint under discussion: that moral naturalism fails to accommodate any real normativity.

fairly unattractive kind of non-naturalism espoused by G.E. Moore (1903/1948) does a pretty good job of capturing the folk conception (Mackie 1977: 31-32). Nor, it is important to note, is there anything inherently conservative about the Canberra Plan; it is no more a tool for showing that things exist than it is a tool for showing that things do not exist; it should be considered no more friendly to naturalists and realists than it is to error theorists.

What of the supposition that moral cognition "leverages adaptive behavior"—does this lean the argument back in favor of the naturalist? This consideration, upon reflection, is less decisive than appearances suggest. For a start, it is not obvious that it divides examples cleanly in the intuitive manner. Sterelny and Fraser maintain that the concept *witch* failed pragmatically: burning "witches" didn't prevent crop failure or other misfortunes. But it is conceivable that the concept may have been useful in other ways: encouraging social homogeneity and reducing instances of defection, for instance. Certainly it is conceivable that the use of the concept could be adaptive (even if at the actual world it generally was not), which seems to imply that, by Sterelny and Fraser's lights, the concept fails only contingently. The concern is not, of course, that there are possible worlds at which witches exist (presumably there are); the concern is that there are fairly nearby possible worlds at which the use of the concept *witch* is all-things-considered quite useful—leading to the unsavory conclusion that witches thereby exist at those possible worlds.

Or consider another example: phlogiston. This obsolete concept, posited by 17th-century scientists to explain combustion and rusting, could be employed perfectly well in everyday contexts. A person could divide materials according to whether they were phlogisticated or not (wood = yes; water = no), could select only the former type to build the household fire, could seek out escaping phlogiston on cold winter evenings, could point to flames and say, "See the phlogiston escaping," and never be caught out, and so forth. Only in careful experimental conditions do the theory's flaws become apparent (e.g., measuring the mass of items before and after burning). This everyday usefulness of the concept *phlogiston*, however, is clearly not enough to outweigh its central erroneous platitude about combustion being a process of a stored substance released from flammable material. There is no phlogiston.

In the case of moral cognition, the "adaptive leverage" can be overstated. Certainly moral nativism implies that moral concepts must have pulled their weight in terms of enhancing reproductive fitness. But that was back in the Pleistocene. In the modern era of world wars and terrorism, the usefulness of morality can be questioned (see Moeller 2009; Garner 2010; Marks 2013). Of the hundreds of thousands who died in the trenches of the First World War, for example, how many would have been there had they not been susceptible to the storm of moralistic propaganda surrounding recruitment, or the sense of dutiful camaraderie felt toward their fellow soldiers? Would large-scale wars or terrorist violence even occur without the possibility of morality's being recruited to their service?

These are rhetorical questions, but let us suppose that we were to come to the conclusion that while moral cognition *was* adaptive for our ancestors, it *is*, all things considered, no longer useful. This draws attention to the possibility that the folk moral platitudes, considered only as they currently stand, might fail the Canberra test, whereas the folk moral platitudes of our prehistoric ancestors might pass. Such an odd conclusion could be avoided by emphasizing that the current concepts are the natural continuers of the past concepts, and therefore presumably inherit their referential success. Nevertheless, an uncomfortable *contingency* lingers. A concept that is put to good use in one environment may be futile in another not-so-different environment, and if this adaptive utility is what tips the scales between a Canberra-style vindication versus a debunking, then it looks like the wrong kind of criteria are being used to determine whether or not moral properties exist.

6. Conclusion

Whether one looks to moral nativism as a way of debunking morality or as a way of vindicating morality, the arguments discussed in this chapter tend to lead to the same place: to questions about the viability of moral naturalism. If one hoped, therefore, to use such an argument to verify or to refute moral naturalism, then one is likely to be disappointed. It is not unreasonable to claim, however, that it is the moral naturalist who bears the burden of proof to present a compelling theory. Yet even this somewhat weak declaration may be sufficient to power a debunking argument to an epistemological conclusion (rather than to an error-theoretic conclusion). If (1) moral nativism raises doubt about the truth of our moral judgments and (2) moral naturalism would dispel this doubt, but (3) there are serious doubts surrounding the viability of moral naturalism (whose resolution we await), then it seems sensible to conclude that our confidence in our moral judgments should be provisionally lowered. This is essentially the "more modest" conclusion encountered earlier: that our moral judgments lack justification.

This chapter has discussed the rather exciting idea that empirical findings can impact on metaethics: that philosophers can roll up their sleeves and get their hands dirty using *a posteriori* data to solve perennial problems. But ultimately the arguments examined lead back to the need to do metaethics the old-fashioned way. Sometimes there is nothing for it but to wash your hands, roll your sleeves back down, and sink back into the armchair.

Acknowledgments

This chapter has drawn freely from Joyce (2013a,b; 2016).

References

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

Butterworth, B. (1999). *What Counts? How Every Brain is Hardwired for Math.* New York: The Free Press.

Dahaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. Oxford: Oxford University Press.

Dancy, J. (2006). Nonnaturalism. In *Oxford Handbook of Ethical Theory*, edited by D. Copp, pp. 122-145. Oxford: Oxford University Press.

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.

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

Friedrich, J. (1996). On Seeing Oneself as Less Self-Serving than Others: The Ultimate Self-Serving Bias? *Teaching of Psychology* 23: 107-109.

Garner, R. (2010). Abolishing Morality. In *A World Without Values*, edited by R. Joyce and S. Kirchin, pp. 217-233. Dordrecht: Springer Press.

Harman, G. (1977). *The Nature of Morality: An Introduction to Ethics*. Oxford: Oxford University Press.

Harman, G. (1986). Moral Explanations of Natural Facts: Can Moral Claims be Tested against Moral Reality? *Southern Journal of Philosophy* 24(Suppl.): 57-68.

Horgan, T. and Timmons, M. (1991). New Wave Moral Realism Meets Moral Twin Earth. *Journal of Philosophical Research* 16: 447-465.

Jackson, F. (1997). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.

Jackson, F. and Pettit, P. (1995). Moral Functionalism and Moral Motivation. *Philosophical Quarterly* 45: 20-40.

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

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

Joyce, R. (2013a). Irrealism and the Genealogy of Morals. *Ratio* 26: 351-372.

Joyce, R. (2013b). Ethics and Evolution. In *The Blackwell Guide to Ethical Theory*, 2nd edn., edited by H. LaFollette and I. Persson, pp. 123-147. Oxford: Blackwell.

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

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

Kelly, D. (2014). Selective Debunking Arguments, Folk Psychology, and Empirical Moral Psychology. In *Advances in Experimental Moral Psychology: Affect, Character, and Commitments*, edited by J. Wright and H. Sarkissian, pp. 130-147. New York: Continuum.

Kelly, E. (2004). Against Naturalism in Ethics. In *Naturalism in Question*, edited by M. De Caro and D. MacArthur, pp. 259-274. Cambridge, MA: Harvard University Press.

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

Krebs, D. (2005). The Evolution of Morality. In *The Handbook of Evolutionary Psychology*, edited by D. Buss, pp. 747-771. New Jersey: John Wiley and Sons.

Lewis, D. (1970). How to Define Theoretical Terms. *Journal of Philosophy* 67: 427-446.

Machery, E. and Mallon, R. (2010). Evolution of Morality. In *The Moral Psychology Handbook*, edited by John M. Doris and the Moral Psychology Research Group, pp. 3-46. Oxford: Oxford University Press.

Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin Books.

Marks, J. (2013). *Ethics Without Morals: In Defence of Amorality*. New York: Routledge.

McKay, R. and Dennett, D. (2009). The Evolution of Misbelief. *Behavioral and Brain Sciences* 32: 493-510.

Mikhail, J. (2011). *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge: Cambridge University Press.

Miller, A. (2013). *Contemporary Metaethics: An Introduction*. Malden: Polity.

Miller, G. (2007). Sexual Selection for Moral Virtues. *Quarterly Review of Biology* 82: 97-121.

Moeller, H.-G. (2009). *The Moral Fool: A Case for Amorality*. New York: Columbia University Press.

Moore, G.E. (1903/1948). *Principia Ethica*. Cambridge: Cambridge University Press.

Nesse, R. (2007). Runaway Social Selection for Displays of Partner Value and Altruism. *Biological Theory* 2: 143-155.

Noë, R. (2001). Biological Markets: Partner Choice as the Driving Force behind the Evolution of Mutualisms. In *Economics in Nature: Social Dilemmas, Mate Choice, and Biological Markets*, edited by R. Noë, J. van Hoof, and P. Hammerstein, pp. 93-118. Cambridge: Cambridge University Press.

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

Pronin, E., Lin, D., and Ross, L. (2002). The Bias Blind Spot: Perceptions of Bias in Self versus Others. *Personality and Social Psychology Bulletin* 28: 369-381.

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

Ruse, M. (2006). Is Darwinian Metaethics Possible (And If It Is, Is It Well-Taken)? In *Evolutionary Ethics and Contemporary Biology*, edited by G. Boniolo and G. de Anna, pp. 13-26. Cambridge: Cambridge University Press.

Ruse, M. (2009). Evolution and Ethics: The Sociobiological Approach. In *Philosophy After Darwin*, edited by M. Ruse, pp. 489-511. Princeton: Princeton University Press.

Sober, E. (1984). *The Nature of Selection*. Cambridge, MA: MIT Press.

Sterelny, K. and Fraser, B. (Forthcoming). Evolution and Moral Realism. *British Journal for the Philosophy of Science.*

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

Street, S. (2012). Coming to Terms with Contingency: Humean Constructivism about Practical Reason. In *Constructivism in Practical Philosophy*, edited by J. Lenman and Y. Shemmer, pp. 40-59. Oxford: Oxford University Press.

Sturgeon, N. (1985). Moral Explanations. In *Morality, Reason and Truth*, edited by D. Copp and D. Zimmerman, pp. 49-78. Totowa: Rowman and Allanheld.

Taylor, S. and Brown, J. (1988). Illusion and Well-Being: A Social Psychological Perspective on Mental Health. *Psychological Bulletin* 103: 193-210.

Tropman, E. (2012). Can Cornell Moral Realism Adequately Account for Moral Knowledge? *Theoria* 78: 26-46.

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.