

*Evolution, truth-tracking, and moral skepticism*

Richard Joyce

Penultimate draft of paper appearing in R. Joyce, *Essays in Moral Skepticism* (OUP, 2016).

***Introduction***

Genealogical debunking arguments are as old as the hills. In the 4th century BC, Euhemerus sought to undermine religious beliefs by alleging facts about their origins: He suspected that divine and mythological beings were in fact historical persons whose qualities and exploits had become exaggerated and distorted with the passage of generations. Early Christian writers like Lactantius and Origen deployed euhemeristic arguments discriminately: against paganism and idolatry, but not (of course) against their own religion. Nietzsche, Marx, and Freud (to name but three intellectual heavyweights) all sought to disparage religion by presenting views on its sociological or psychological source.

How exactly are such arguments supposed to work? Nietzsche might have hoped that once Christians had seen their religion labeled “a slave morality”—once they had been persuaded that it was historically rooted in hatred dressed up as love—they would turn away from it blushing. But if such a rejection were to occur simply as a psychological response then it would not be a debunking *argument* at all. Freud seemed to acknowledge this when he wrote that once we recognize religion as a grand wish-fulfillment, “our attitude to the problem of religion will undergo a marked displacement” (Freud [1927] 1987: 215). That may well be so, but, again, if this is no more than a prediction of psychological causal consequences, then one is left wondering whether there is supposed to be an actual *refutation* of theism in Freud’s writings. Could one accept all that Freud says about religion as neurotic wish-fulfillment while reasonably maintaining theistic belief?

My interest here is in morality rather than religion. In recent years there has been a burgeoning of interest in genealogical arguments targeting morality, with Darwinian evolution being the genealogy appealed to most conspicuously as responsible for the debunking. The empirical relation between Darwinian evolution and human moral judgment is something about which the jury still deliberates. One hypothesis, called “moral nativism,” is that the human mind contains faculties dedicated to moral judgment and forged by natural selection for this task. A competing hypothesis is that moral judgment is a cultural achievement involving an array of psychological faculties that evolved for other purposes. Adjudicating these and alternative views is not the purpose of this paper. Rather, my goal is to examine whether the truth of the former hypothesis would undermine morality. In one sense, the details of the Darwinian hypothesis are redundant to this task, for, as we shall see, what really matters to a debunking argument is whether our moral judgments are the output of a faculty that, we have reason to suspect, does not track the moral facts. And this might be so even if moral nativism is false; for even if moral judgment is not a biological adaptation, it might nevertheless be produced by psychological faculties that fail to track the truth. For example, Freud’s wish-fulfillment theory of religion could be extended quite naturally (and not entirely implausibly) to moral systems, in which case the possibility of a debunking

argument would arise without evolution getting a look-in.<sup>1</sup> Nevertheless, I judge that there is good reason to continue to keep the evolutionary nativist hypothesis in clear sight, since (A) there is a decent chance that it is true, and (B) there is a solid case to be made that it implies that moral judgments are the output of a non-truth-tracking process. To this end, I will continue throughout to refer to *evolutionary debunking arguments* (EDAs), though I should like the reader to note that I have resisted a strong temptation to prefer the broader label *genealogical debunking arguments*.<sup>2</sup> One must bear in mind that the evolutionary hypothesis might well be substituted in an EDA by some other kind of genealogical hypothesis.

### *Types of debunking*

Before examining the logic of an evolutionary debunking argument, let us pause to wonder about the nature of the conclusion. What, exactly, is *debunking*? The answer is that it can be many things, though no doubt the associated arguments vary greatly in plausibility. Here are some potential conclusions to moral EDAs, followed by a quick survey of philosophers who argue for these theses.

1. All moral judgments are false.
2. All moral judgments are false insofar as they involve a claim to objectivity.
3. Certain normative moral theories (e.g., Kantianism)—but not all such theories—should be rejected.
4. All moral judgments lack justification.
5. All moral judgments lack justification and permanently so.

This is not an exhaustive list, but it suffices for my purposes. Both (1) and (2) undermine certain metaethical theories while vindicating others. The metaethical theory vindicated by (1) is the error theory (see Mackie 1977; Joyce 2001). (2), on the other hand, undermines any metaethical theory according to which there are objectively true moral claims, which (in many people's books) amounts to saying that (2) undermines moral realism. The claims made in (4) and (5), by contrast, are compatible with the possibility that moral facts exist and objectively so. Therefore any EDA that aims to establish (4) or (5) is not an attack on moral realism.

It is possible that (2) implies (1), but only by assuming a controversial bridging premise: namely, that all moral judgments necessarily include a claim to objectivity. One person who possibly argues in this manner is Michael Ruse (2006, 2009). He speculates that having a faculty that issues moral judgments was adaptive to our ancestors because those judgments strengthened their motivation to cooperate, and that this occurred precisely because those judgments were imbued with objectivity: "The Darwinian argues that morality simply does not work (from a biological perspective) unless we believe that it is objective" (1986: 253).

---

<sup>1</sup> In fact, Freud views morality either as the internalization of parental norms in order to resolve the Oedipal complex, or as a means of suppressing aggressive and sexual instincts in order to make possible a cooperative society.

<sup>2</sup> In part this is because "EDA" appears elsewhere in the literature, and I am loath to multiply abbreviations.

He goes on: “[M]orality is a collective illusion foisted upon us by our genes. Note, however, that the illusion lies not in the morality itself, but in its sense of objectivity” (ibid.). The last comment makes it seem as if Ruse is seeking to establish (2). But elsewhere he appears to endorse the bridging premise that would take him to (1): “Ethics is subjective, but its meaning is objective” (Ruse 2006: 22); “[W]hat I want to suggest is that ... the *meaning* of morality is that it is objective” (Ruse 2009: 507). If morality is necessarily objective, then establishing that there are no objective moral truths—as (2) claims—entails that there are no moral truths *tout court*—as (1) claims.

Of this strong EDA tentatively attributed to Ruse two things should be noted. First, it requires a highly contentious bridging premise, which itself cannot be established on genealogical grounds alone. Second, this attempt to establish (1) by EDA would be, to my knowledge, unique in the literature. Contrary to widespread opinion, establishing the error theoretic thesis that all moral judgments are false is simply not the evolutionary debunker’s typical goal. It is frustrating to read a commentator on my own debunking argument interpret the intended conclusion as “*all* moral judgements are false” (Mason 2010: 775).<sup>3</sup>

Sharon Street’s EDA seeks to establish (2) while explicitly resisting (1); she aims to debunk not morality but moral realism (Street 2006, 2008). The slide from (2) to (1) is avoided because Street allows the viability of moral *constructivism*, according to which objectivity is not an essential feature of moral concepts. Kevin Brosnan interprets Street as inferring “that our moral beliefs are probably false” (Brosnan 2011: 52), but this is incorrect; the conclusion of her EDA is that *moral realism* is probably false. The clause about “probability” arises because part of Street’s EDA is that the moral realist might claim that our evolved moral judgments match the objective moral facts without being causally connected to them, in which case (Street thinks) the realist is committed to “a fluke of luck that’s ... extremely unlikely” (12).

One finds EDAs with thesis (3) as their conclusion advocated in the work of Joshua Greene (2008) and Peter Singer (2005). Greene and Singer argue that an examination of the evolutionary basis of certain widespread moral intuitions casts their practical relevance into doubt. They go on to argue that certain normative theories—basically Kantian in flavor—draw their appeal largely from their conformity with these intuitions, and thus to undermine this class of intuitions is to undermine this class of theories. Whether they think that this reveals the theories to be unjustified, or false, or probably false, is not entirely clear; hence my preference for the vague phrasing of (3): “...should be rejected.” Greene allows that Kantian intuitions *might* be correct by “coincidence,” but (like Street) he thinks that, given their evolutionary history, it is “*unlikely* that they reflect any sort of deep moral truth” (2008: 71). In any case, neither Greene nor Singer is out to support the error theory; each thinks it likely that many moral judgments remain true, so long as these judgments can be

---

<sup>3</sup> Mason might perhaps be forgiven this infelicity on two grounds. First, I certainly have argued for (1) (see my 2001—where, for trifling reasons, I prefer “untrue” to “false”); I have not, however, tried to establish (1) using an EDA. Second, I did once suggest that the label “error theory” might be expanded so as to denote the disjunction of (1) and (4) (Joyce 2006: 223); and I then proceeded to attempt to establish (4) by an EDA and called it an “error theoretic” conclusion. I now think that this suggested expansion of the label was injudicious. The observation made earlier—that (4) is in fact compatible with realism—should suffice to show why. (Thanks to Hatha McDivitt for pointing this out to me.)

underwritten on consequentialist grounds. It is, furthermore, possible that they can maintain that many moral judgments remain *objectively* true, to the extent that consequentialism allows this conclusion.

Indeed, Guy Kahane (2011) argues that debunkers like Greene and Singer *must* presuppose a kind of objectivism, though I confess to finding his argument rather perplexing. Of EDAs and subjectivist/constructivist metaethical possibilities, Kahane writes:

if there is no attitude-independent truth for our attitudes to track, how could it make sense to worry whether these attitudes have their distal origins in a truth-tracking process? (2011: 112).

Perhaps I am missing something, but I find Kahane's rhetorical question easy to answer. Consider something about which we can all agree has constructivist status: the value of money. A given piece of paper is worth \$10 because and only because of some kind of collective decision to treat it as worth \$10 (construing the word "decision" very broadly); its value is not an objective matter. Suppose Fred is a newcomer to our country, and unsure about the respective values of the various pieces of metal and paper that we use as money; but he is also an idiot, and decides to form his beliefs on the matter on the basis of consulting tea leaves (a.k.a. tasseography). Clutching a piece of paper (which happens to be a \$10 note), Fred examines the tea leaves and decides that it's worth \$10. It is clear that Fred's belief, though true, has something wrong with it. We could, in fact, deploy a kind of genealogical debunking argument against the belief, based on the fact that the belief is the product of a process that doesn't track the truth. And yet we can all be sure (including Fred) that the value of money is not an objective phenomenon. Thus, *contra* Kahane, subjectivist/constructivist metaethical views may be just as subject to EDAs as objectivist views. What Kahane seems to overlook is that subjectivist/constructivist facts are *trackable* too, and thus beliefs and intuitions about such facts can be produced by processes that succeed or fail to track them.<sup>4</sup>

What I think one should say about Fred's belief is that it is (though true) *unjustified*. This brings us to thesis (4), which is the conclusion of the EDA that I have myself advocated against morality (Joyce 2006, chapter 6), and which will be the focus of the following discussion. Contributing to the difficulty of articulating and defending such an argument is the fact that epistemological justification is a disputed concept, and the argument may be more or less plausible according to one's epistemological leanings. The comments of the following brief paragraph, however, I think we can take as common ground.

Many types of things can be correctly spoken of as "justified" or "unjustified"—actions, emotions, plans, laws, etc.—but our current concern is epistemological justification, which applies paradigmatically (and probably solely) to *beliefs*. Epistemological justification is distinct from truth: A belief can be justified but false, or unjustified but true. Epistemological

---

<sup>4</sup> Consider, for example, Ronald Milo's contractarian constructivism (Milo 1995), according to which moral facts are determined by the choices of a hypothetical idealized group of rational contractors. It is not an objectivist theory, since the instantiation of moral facts depends on some "state of affairs [being made] the object of an intentional psychological state" (192). (Milo calls it a "stance dependent" theory.) But it is an "evidence independent" theory, in as much as the moral facts obtain independently of our having reasons or evidence for believing them. Given the latter quality, the moral facts are (according to Milo's theory) things that we can track well or poorly, employing sound or foolish methods.

justification is relative: One person's belief that  $p$  may be justified while another person's belief that  $p$  is not justified. A person's belief that  $p$  may be unjustified at time  $t$  but later gain justification; or justified at time  $t$  but later lose justification. Beyond these and a few other platitudes, a bewildering degree of disagreement reigns.

### ***Justification and truth-tracking***

A justified belief, if true, cannot be true by accident. What makes it justified is that it connects to the facts in the appropriate manner. (Even this second comment might be rejected by certain epistemological coherentists, but one cannot please everyone, especially if they insist on holding outlandish views.<sup>5</sup>) According to evidentialism, the justifying relation between a belief and the facts is a complicated affair that depends on the possession of evidence in favor of that belief—evidence being something that confers reasonableness or confirmation on a belief. According to the rival reliabilism, the justifying relation between a belief and the facts depends on the existence of a reliable mediating belief-formation process—reliability being a quality that probabilifies the truth of output beliefs. Both families of theories allow that *false* beliefs may have the characteristic that renders them justified.

In recent years, several critics of EDAs have favored reliabilism as a key part of their critiques (Carruthers & James 2008; Wielenberg 2010); thus, in order to meet these critics on their own ground, for the purposes of this paper I am willing to accept a reliabilist perspective. In the end it makes little difference, since what is really at issue is whether moral beliefs are the product of a truth-tracking process, and recognition of this fact appears to be common ground for both reliabilists and evidentialists.<sup>6</sup> Let us say that a belief-formation process is reliable in so far as it produces beliefs that track the truth. The intuition at the heart of truth-tracking is that beliefs may or may not be *sensitive* to the facts which they represent.

It is important to stress that we are not directly assessing beliefs; the claim is not that a belief is unjustified if it fails to track the truth—for any false belief trivially fails to track the truth, yet nobody wants the conclusion that all false beliefs are unjustified (see White 2010: 580). Rather, the claim is that a belief (whether true or false) is unjustified if it is the product of a belief-formation *process* that sufficiently fails to track the truth (Goldman 1979, 2008). Suppose that Fred forms the belief that  $p$  on the basis of tasseography, and that  $p$  happens to be false. The fact that this belief is insensitive to the truth does not render it unjustified; rather, what renders it unjustified is that tasseography is a process that is insensitive to the truth. This conclusion would remain even if the belief that  $p$  turned out, flukily, to be true.

---

<sup>5</sup> I made critical comments focused specifically on coherentist moral epistemology in Joyce 2006: 216-7, which I more-or-less stand by. A coherentism that aims to forge moral intuitions into a tidy package and declare them thereby justified is one whose flaws are so glaring as to require no further comment. By contrast, a truly *wide* reflective equilibrium, which takes into account empirical evidence and probabilities concerning the origin of those human moral intuitions, has more to recommend it; but the price of plausibility is the possibility of moral debunking. Here I second Singer's thoughts: "If the interpretation is truly wide enough to countenance the rejection of all our ordinary moral beliefs, then I have no objection to it" (Singer 2005: 347).

<sup>6</sup> The key difference between the views is that the evidentialist holds that the truth-probabilifying elements of the process must be accessible to the agent. The way I see it, the evidentialist can still maintain that in order to be justified, beliefs must track the truth; it is just that the evidentialist has a particular way of restricting what satisfies this criterion.

There is a natural temptation to understand this truth-sensitivity in terms of counterfactual covariation: (i) if  $p$ , then  $S$  believes  $p$ , and (ii) if not- $p$ , then  $S$  does not believe  $p$ . But this interpretation is problematic and is, in any case, optional. One problem for the counterfactual interpretation is to account for beliefs concerning necessary truths or necessary falsehoods, since in these cases the antecedents of (i) or (ii) may count as necessarily false, yielding counterpossible conditionals that are, at first blush, vacuously true. This has been thought to be a particular problem for EDAs concerning morality, for here, it may be argued, beliefs with non-contingent content abound. (See White 2010: 583; Enoch 2010: 433; Wielenberg 2010: 455-456.) The proposition “Promise-breaking is wrong” (perhaps with a “*ceteris paribus*” thrown in) is, some would claim, not merely true but *necessarily* so. This would problematize any attempt to assess the counterfactual “If it were not the case that promise-breaking is wrong, you’d still believe that it is.” Ruse stumbles into this problem when he writes “Given two worlds, identical except that one has an objective morality and the other does not, the humans therein would think and act in exactly the same ways” (1986: 254; see also Mackie 1977: 49). If some cogent views about moral supervenience hold—that two situations identical in all non-moral properties must be identical in all moral properties—then Ruse’s thought experiment is tricky at best.

Moves are afoot to make sense of non-vacuous counterpossibles (Restall 1997; Vander Laan 2004; Brogaard & Salerno 2013), and there does seem to be an intuition in favor of doing so being a desideratum. Compare the following two claims: “If 6 were a prime number we’d have some serious rethinking to do” and “If 6 were a prime number it would make no difference to mathematical calculations.” The former seems intuitively to have a claim to truth that the latter lacks. A promising solution is to appeal to impossible worlds, which, while sounding fishy, can be understood in innocuous terms. (I like Ed Mares’ description: “If we pay for ideology in the coin of ontology, then the doctrine of impossible worlds can be bought with loose change lying round the house of almost any possible worlds theorist” (Mares 1997: 525).)

However, even if such attempts to vindicate non-vacuous counterpossibles fail, there remains plenty of room for endeavoring to make sense of the idea of processes that succeed or fail to track necessary truths. Consider Fred again, who decides whether or not numbers are prime on the basis of tasseography. Let’s say that he forms the belief that 7 is prime. Intuitively, one still wants to say that tasseography fails to track the truth in this case—that it is in some sense *insensitive* to the truth—and one should be able to do so without having to consider propositions like “If 7 were not a prime number, then Fred would still believe it.” The crucial characteristic seems to be that mathematical facts about the primeness of numbers play *no explanatory role* in why the tea leaves formed a certain pattern in the cup, which is the sole factor in determining Fred’s belief. And this, surely, is something we can understand without needing to assess the truth of counterpossibles.

Here I am agreeing with Gilbert Harman’s claim that the counterfactual test of whether  $H$  explains  $E$  “is only appropriate in certain contexts” (1986: 63). In the case of moral facts explaining moral judgments, Harman says that rather than an assessment of counterfactuals, “what’s needed is some account of *how* the actual wrongness of [something] could help explain [someone’s] disapproval of it. And we have to be able to believe in this account. We cannot just make something up” (ibid.). Such an account is what’s lacking regarding the

relation between numbers being prime and the pattern of tea leaves in a cup.

[Aside: Imagine someone, Mary, being exposed to skewed “evidence” indicating that tasseography *is* an effective way of determining whether numbers are prime. Suppose Mary lived at a time (say, one thousand years ago) where believing in supernatural forces, etc., was not unreasonable (in the absence of better hypotheses), where all the epistemic authorities in her community unanimously endorsed tasseography, and so forth. If this story is fleshed out, there comes a point where many (myself included) become inclined to say that Mary’s belief that 7 is prime, though formed solely on the basis of tea leaves, is justified. But surely it remains true that 7’s being prime plays no explanatory role in her forming that belief, which seems to conflict with the previous suggestion. The reliabilist might respond that the process of forming beliefs on the basis of tasseography was not really the same process one thousand years ago as it is now. *Then* someone might consult epistemic authorities, weigh data, examine alternative hypotheses, etc., and reasonably come to the conclusion that tasseography is dependable. (I am, in any case, supposing this for the sake of argument.) *Now*, by contrast, an ordinarily-situated person can give credence to tasseography only by neglecting superior alternative hypotheses and rejecting a great deal of available data about the causal structure of the natural world. If one widens one’s attention from the narrow process of *tasseography* toward the broader process of employing techniques that do (or do not) involve consulting one’s epistemic community, that do (or do not) involve ignoring available alternative hypotheses, and so forth, then one can say that historical-Mary’s belief is the product of a reliable kind of process whereas contemporary-Fred’s belief (with the same content, formed on the basis of the same *narrow* process) is not. Roughly, we can say that Fred “should know better,” whereas we cannot say this of Mary. Of course, 7’s being prime still plays no role in explaining Mary’s belief that 7 is prime, but one might nevertheless maintain that Mary is employing *broad* belief formation practices which probabilify that her beliefs are (A) true, and (B) explicable (in part) by the relevant facts—whereas contemporary Fred is not. Should we, then, construe a person’s processes narrowly (according to which Mary’s belief is unjustified) or broadly (according to which Mary’s belief is justified)? The challenge of answering this question is, I take it, equivalent to the need for the reliabilist to answer the well-known *generality problem*, which I consider a serious obstacle. (See Conee & Feldman 1998.) The ecumenical conclusion that Mary’s belief is in one sense epistemically justified and in another sense epistemically unjustified is one from which I don’t recoil. (See Goldman 1988.)]

### ***Adaptation and truth-tracking***

In the context of evolutionary discussion, it is sometimes useful to speak of traits whose function it is to track the truth, which is to say that the trait’s meeting a certain threshold of success at matching the facts enhanced reproductive fitness among ancestors (relative to the competition) and thus explains why the trait emerged and persisted. This is intended to apply only to traits that involve *representational* states. Perspiration may covary with bodily temperature, but we don’t want to say that perspiration represents bodily temperature or anything else (at least, *I* don’t want to say that), and thus the question of whether perspiration “tracks the truth” does not arise. If, by contrast, an evolved psychological faculty produces

some species of *judgment*, then the prerequisite of producing representational states is satisfied, and one can then sensibly ask whether the function of the faculty is to track the truth. (I remain uncommitted on the extent of the domain of representational states.) The only “truth” with which we are concerned here is the one that is represented. Suppose an evolved faculty has the function of producing judgments of the form “X is P.” These representational states might covary robustly with (or be explained by) *X’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 (or are explained by) *those fact(s) that they represent*—in this case, X’s being P. And 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. (When Street talks of a “tracking account” of the evolution of a moral faculty—as opposed to what she calls an “adaptive link account”—she means the hypothesis according to which the faculty *has the evolutionary function* of truth-tracking. See Street 2006 and 2008: 210.)

Some concrete examples might help. Let us suppose that humans have been hard-wired by natural selection to perform certain basic arithmetical functions (see Butterworth 1999; Dohaene 1997). Simplifying things for the sake of illustration: let us suppose that the belief that  $2+3=5$  is an innate adaptation. Having this proposition immediately available at one’s mental fingertips, without calculation or hesitation, might well have been useful for our ancestors in a myriad of ways. But such arithmetical beliefs are useful *only if they are true*. By contrast, having the proposition “ $2+3=6$ ” immediately available at one’s mental fingertips is going to lead to all sorts of practical problems. The faculty that produces these innate beliefs, we must conclude, has the function of tracking the truth: it exists in virtue of producing accurate arithmetical representations. (To reiterate the point made earlier: One need not interpret this as requiring that we can make sense of the counterpossible “If  $2+3\neq 5$ , then...”; it suffices that the fact of  $2+3$  summing to 5 explains (somehow) the faculty’s tendency to give the output belief that  $2+3=5$ .<sup>7</sup>) Thus the belief that  $2+3=5$ —though (we are supposing) the product of a faculty that evolved in order to help our ancestors make more babies—is the output of a reliable process, and thus counts (*ceteris paribus*) as justified.

---

<sup>7</sup> Justin Clarke-Doane (2012) doubts these claims. His worries about the counterpossible are not unreasonable, but I have already said something about this. (I hope I have said enough to refute his claim that if the counterpossible is unintelligible, “the argument on behalf of Joyce et al. fails” (2012: 329).) Clarke-Doane’s worries about the explanatory claim, on the other hand, would, if cogent, do more damage to my argument; but I do not find them cogent. He claims that “for any mathematical hypothesis that we were selected to believe, H, there is a nonmathematical truth corresponding to H that captures the intuitive reason that belief in H was advantageous is plausible” (2012: 332). What he has in mind is that the usefulness of the ancestral belief that  $2+3=5$  (say) might be explained by reference to “(first-order) logical truths regarding objects in our environments (it is conceivable that they would also sometimes be mereological or impure set-theoretic truths regarding such objects)” (332). There is of course a substantive philosophical question of how mathematical facts can explain mathematical beliefs, into which I have no space to delve here. (The only clarification I would add is that this is distinct from the question of how mathematical facts can have *causal* effects.) In reference to the central point being made here, the issue is whether Clarke-Doane’s explanation of the adaptiveness of mathematical beliefs by reference to first-order logical truths would render those beliefs *false*. If not, then my claim that such an explanation presupposes that the beliefs are true may be maintained. If so, then I would happily move on to using a more straightforward and less contentious example to illustrate the point that some beliefs are useful only because they are true, regarding which I have little doubt that one could be quickly found.

Compare this with another example. Let us suppose that humans have been hard-wired by natural selection to systematically make unrealistically positive self-evaluations (see Taylor et al. 1988; Alicke 1985; McKenna et al. 1991). 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. 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. The faculty that produces these innate beliefs, we must conclude, does not have the function of tracking the truth: it exists not in virtue of producing accurate self-appraisals, but rather in virtue of producing self-appraisals that benefit the agent's physical and/or psychological well-being. Thus the belief "I am a better-than-average driver" (say) is not the product of a reliable process, and thus is unjustified.

The last conclusion is, of course, far too hasty, and requires qualification. A great many people really are better-than-average drivers, and some of them may well have solid ground for believing this of themselves. Perhaps a person has passed various practical tests at driving school, taken courses in driving, etc., and has repeatedly come top of the class. Such a person seems to have come by her belief via a process that *is* reliable and thus surely has, one might think, a justified belief on the matter. If, however, we take at face value the claim that the belief that she is better-than-average is *hard-wired*, such that she would carry on believing it come what may, then it follows that had the practical driving tests and so forth supplied her with ample evidence that she is in fact a dreadful driver, she *still* would have concluded that she is better-than-average. Bearing this in mind, her belief that she is better-than-average—even when true and accompanied by ample evidence testifying to this fact—does not look so justified after all.

Once this consequence is brought out, it would be reasonable to complain that the idea of beliefs "hard-wired" by natural selection, immune to all evidence or reason, seems artificial and unfamiliar. This may well be so. The point of the above examples is not to be plausible, but to illustrate a contrast. A more realistic portrayal of the trait would maintain that while natural selection may have given us a tendency to over-estimate ourselves in various ways—as a kind of knee-jerk default—we are not incapable of forming accurate beliefs about our abilities when presented with conspicuous and incontrovertible evidence.

This point illustrates the important difference between theses (4) and (5) from earlier: the difference between certain beliefs lacking justification and their lacking justification permanently. In the case of inflated self-evaluations, the thesis that seems correct to me is that when the belief is formed as a knee-jerk default, without reflection or proper sensitivity to the available evidence—when, that is, the belief is to be explained largely by reference to the arousal of an evolved *non-truth-tracking* doxastic faculty—then it lacks epistemic justification. But one is not necessarily stuck in that position. We are creatures with the capacity to bring other psychological faculties to bear on the matter—faculties that can track the truth in a reliable manner—and when these are employed properly, the same belief that was once unjustified may become justified. A person's initial default belief that she is a better-than-average driver may have been without justification, but sensitive consultation of appropriate evidence may provide the belief with its missing justification.

Given this, the strength of this kind of EDA is to establish a challenge, or a burden of proof. The beliefs that have been called into question *may* be rendered justified by the employment of appropriate epistemic mechanisms (they may be “undebunked”), but until that is accomplished they cannot be considered justified. A prominent target of this kind of EDA is, thus, the epistemic conservative, who holds that firmly held beliefs are “innocent until proven guilty.” It’s not that the proponent of the EDA need maintain that epistemic conservatism is mistaken; but rather declares that evidence that the beliefs in question emanate from a non-truth-tracking source serves as a defeater of any *prima facie* justification which they might have been accorded.<sup>8</sup>

### ***Morality and truth-tracking***

Here is not the place to present in any detail hypotheses concerning the evolution of the human faculty for moral judgment (see Alexander 1987; Joyce 2006; Mikhail 2011; Kitcher 2011). It suffices for our purposes to note that none of the examples of such hypotheses mention that this faculty served reproductive fitness via the production of *true* judgments. Most nativist hypotheses suggest that morality plays a vital role in enhancing social cohesion. Perhaps the adaptiveness of moral thinking lies in the fact that judging an uncooperative action to be *forbidden* might engage a more stalwart motivation to refrain—more stalwart even than thinking of that action as against one’s own interests. Yet the plausibility of this adaptational hypothesis seems unaffected by whether one is a moral realist or an error theorist—that is, unaffected by whether one thinks that these uncooperative actions *are* forbidden. According to this hypothesis, then, the evolutionary function of the moral faculty is not truth-tracking.

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 some kind of 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 (Zahavi 1977; Noë 2001). Thus, if one’s reproductive capacities depend on being selected as a partner in various cooperative ventures (hunting, raising a family, etc.), and those doing the selecting 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. Thus, making moral judgments in a sincere manner may be adaptive as a signaling device; yet, again, one might be convinced of this hypothesis while maintaining an error theoretic metaethical stance. There is no pressure to assume that the

---

<sup>8</sup> There are some substantive options here, depending on what the details are of one’s favored epistemological theory. One might say, as I have, (i) that *evidence of* the beliefs’ non-truth-tracking source removes the justification that they heretofore had been accorded; or alternatively one might say (ii) that *the fact of* the beliefs’ non-truth-tracking source removes the justification that they heretofore had been accorded. Regarding (i), one might say either (ia) that this knowledge renders previously justified beliefs unjustified, or (ib) that this knowledge reveals that the beliefs were never justified in the first place. In the case of (ii), I take it that only the latter option is viable: that the fact of a non-truth-tracking source implies that the beliefs were never justified in the first place.

moral judgments need be *true* in order for them to play this adaptive role.

If these hypotheses are to be believed, then a truth-tracking moral faculty was not selected for. But it does not follow that a truth-tracking moral faculty was not selected *tout court*. Elliott Sober established the useful distinction between a trait's being *selected* and a trait's being *selected for* (Sober 1984). 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 by-product of the selective process. As whales evolved, the lipid content in their blubber increased. The insulating quality of lipid-rich tissue is what was *selected for*; the exceptional flammable characteristics of the blubber, by contrast, were merely naturally *selected*.

With this distinction in hand, one can see that it remains entirely possible that an evolved moral faculty does in fact track the truth (i.e., that this has been selected), even though it does not have this as its evolutionary function (i.e., even though it has not been selected *for* this quality). Several critics of EDAs have pressed this point, in different ways, in making their case. Brosnan (2011) 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 (2009) 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 non-accidentally) true. Erik Wielenberg (2010) advocates another such argument, conjecturing 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" (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. (See also Skarsaune 2010.)

All such strategies can be seen as appealing to a "third factor" to explain the reliable connection between moral beliefs and the facts they represent. These views allow that moral beliefs did not evolve *in order* to track moral truths, but speculate that some aspect of the nativist genealogy raises the probability that the moral facts (those which the beliefs in question represent) obtain.

One logical possibility is that the evolutionary process that produced the moral faculty was somehow causally influenced by the existence of moral facts. Consider, by analogy, how Euhemerus sought to debunk the state religion. He claimed that on his travels to Crete he had stumbled upon an ancient tomb of a king named "Zeus," indicating that the stories of Zeus-the-divine were no more than exaggerations derived from the exploits of a historical figure. However, putting aside the obvious rejoinder that there might simply be two entities with the same name, one can imagine a religious believer claiming that *of course* Zeus-the-divine would create a kingly tomb bearing his own name, precisely to test the faith of potential doubters. If this hypothesis is given credence, then finding a tomb marked "King Zeus" is no evidence against the existence of Zeus-the-divine—not even if one has direct evidence that

the current conception of the god Zeus really does derive from inflated and half-forgotten stories about a real king. The religious believer simply claims that the god Zeus directly or indirectly causally established this evidence. In the same way, one might claim that the existence of moral facts somehow causally influenced the evolutionary genealogy. I mention this just for completeness; it is hard to imagine anything plausible being made of the idea. (See White 2010: 582-583.)

A more promising possibility is that the moral facts might be identical to, or supervene upon, the very facts described in the genealogical account. This is the strategy that all the aforementioned opponents of the moral EDA adopt. For example, the nativist hypothesis more-or-less explicitly mentions types of action that conduce to cooperation, and the anti-debunker can claim that whatever conduces to cooperation just *is* morally good. The evolutionary hypothesis might not explicitly invoke moral facts, but the moral facts might be implicitly presupposed by the hypothesis. In the same way, one might explain an alpine avalanche using terms like “snow” and “sunshine,” while not mentioning hydrogen atoms at all, but it’s not as if this demonstrates that hydrogen atoms played no role in the avalanche; an explanation couched in terms of “snow” implicitly concerns the activity of hydrogen atoms. The vernacular explanation might be more suitable to a given conversational context than the molecular explanation—or vice versa—but neither is superior in the sense of excluding the truth of the other.

On an earlier occasion I supplemented my version of the EDA with general metaethical arguments against moral naturalism, in order to cast doubt on this latter possibility (Joyce 2006, chapter 6). I don’t pretend that those arguments were comprehensive or presented a refutation, and in any case here is not the place to repeat them. Rather, I would like to draw attention to the logic of the dialectic, lest it be misunderstood. As I have stressed, the EDA that I favor represents a *challenge*; it says that the fan of morality has some work to do if justification is to be established or reinstated. Gestures toward a “third factor” don’t demonstrate some gaping hole in the EDA that has previously gone un-noticed, but rather represent attempts to meet the challenge that the EDA poses. But this also explains why I find these attempts so unsatisfying, for none of them seriously undertakes to argue for the connection between moral facts and the evolutionary process in any detail; they rather gesture—sometimes astonishingly vaguely—at a property that *might* suffice to establish the connection, as if showing the mere possibility should demolish the EDA. But the advocate of this kind of EDA was never under the illusion that establishing such a connection is *impossible*; the possibility that moral facts might find a place in the evolutionary genealogy was always acknowledged. What it comes down to is whether this mere possibility can be made *plausible*, and as far as I can see, not one of these debunkers of debunking has made a serious effort.

A slightly cruel comparison might be with those 19th-century Christian apologists who tried to explain away the growing evidence of the ancient age of the Earth which seemed to contradict the Biblical texts.<sup>9</sup> Someone who sensibly takes the fossil record as evidence of the venerable age of the Earth need not deny the *possibility* that it was all created by God 6,000

---

<sup>9</sup> E.g., Philip Gosse’s attempt to explain away all those mysterious Bible-debunking fossils, in his *Omphalos* (1857).

years ago (with fossils placed here to test our faith). Nor need such a person be disturbed by the coherent articulation of a hypothesis according to which dinosaur fossils and a young Earth are co-possible. All that really matters is whether that hypothesis has any plausibility.

By the same token, I am not much disturbed when someone says that *if* the moral facts were thus-and-so, then our ancestors' moral beliefs, though an adaptation, would have reliably tracked those facts. My response is: "OK, show me the theory; let us subject it to critical scrutiny." Wielenberg thinks he can help himself to a certain view of moral facts without pausing to make the view plausible, since he reads me as claiming that even if moral facts existed, *and whatever their nature*, the EDA would reveal our beliefs about them to be unjustified. Though I don't accuse Wielenberg of begging the question, I find his interpretation of the rules of engagement rather unfair. I certainly don't think that the EDA works against *any* construal of moral facts, even fantastic and gerrymandered ones. One might easily stipulate that moral goodness is whatever conduces to genetic replication, and then any evolutionary explanation of anything would of course implicitly involve moral facts. The onus, as I have emphasized, is on making this stipulation plausible. The role of the EDA is to place the burden of proof onto the shoulders of those who believe in justified moral belief. Perhaps Wielenberg's theory can be made to fly; perhaps it cannot. (Hint: I have a raft of objections, should the argument ever get that far.) If he and other opponents of the EDA see themselves as attempting to discharge this burden, then well and good—the EDA will have done its job and the discussion can continue. But if they see themselves as showing that the EDA never posed a challenge in the first place, then I protest.

Brosnan points out that "third factor" responses to the EDA challenge don't actually need to support a kind of naturalism according to which moral facts are identical to or supervene upon natural facts; all that is really needed is that the obtaining of the natural facts cited in the genealogy raises the probability of the obtaining of the moral facts (2011: 61). This is correct; but I disagree with Brosnan's diagnosis that this broadening of the requirement reveals that the challenge poses only "a very modest requirement" (*ibid.*). The broader requirement of *probabilification* may be logically weaker than the requirement of *identity or supervenience*, but it is very far from obvious to me that it is any easier to satisfy.

Something more needs to be said about how the challenge is expressed, for this depends on to whom one takes oneself to be speaking. Against the epistemic evidentialist, the EDA says (roughly) that until a plausible justifying account is presented, moral judgments must be considered unjustified. Against the epistemic reliabilist, by contrast, the EDA says (roughly) that until plausible evidence of a reliable process is presented, we *don't know* whether moral judgments are justified. On previous occasions I pictured my opponent principally as the evidentialist, and thus expressed the EDA in the former manner. The burden for the evidentialist is to provide a theory that *instates* or *reinstates* justification. But the reliabilist raises the possibility that our moral judgments might be the product of a reliable process (whether we know it or not) and thus that our moral judgments might have been justified all along and remain so. The burden for the reliabilist is to provide a persuasive account that establishes that moral judgments are indeed the product of such a reliable process.

This may seem like something of a climb-down for the proponent of the EDA, for he or she appears to have moved from asserting that moral judgments *are* unjustified to asserting that *we don't know* whether they're justified. But it's really not a substantive weakening at

all. The skeptical position doesn't change; it's just that how it is expressed needs to be tailored to one's interlocutor.<sup>10</sup> In other words, it might seem that to (1)-(5) from earlier we should add another possible debunking conclusion:

6. We don't know whether moral judgments lack justification,

but in fact this is better thought of as (4\*): that is, as (4) articulated for the reliabilist audience. That the proponent of the EDA hasn't really backed down from anything is evidenced by the fact that as soon as his or her interlocutor is again an evidentialist, then he or she will revert to asserting (4).

### ***Conclusion: Shifting the burden of proof***

Some philosophers will maintain that they can provide positive considerations indicating that many moral judgments connect to the facts in a manner that renders them justified. But even they cannot ignore the EDA, inasmuch as it represents the presentation of a new hypothesis about the place of moral judgments in the world (one, moreover, potentially with empirical backing), and thus any advocate of a truth-tracking hypothesis must either establish the superiority of his/her view over the new hypothesis or demonstrate that the two hypotheses can be jointly accommodated. Either way, any such advocate has some explaining to do.

Numerous other philosophers—and, I hazard to suggest, most of the folk—do not attempt to provide positive justifying considerations for morality, but nevertheless proclaim justification on the basis of some more or less clearly delineated principle of conservatism. It is to this position that the EDA does the most damage. The epistemic conservatist claims that no positive supporting considerations need be articulated in order for moral beliefs to be justified; rather, they receive justification in virtue of (A) the mere fact that we have these beliefs, and (B) the absence of defeating considerations. To my thinking, the provision of evidence that the moral faculty is the product of a non-truth-tracking process represents a significant defeating consideration. It is a *prima facie* defeater, since a persuasive articulation of one of the aforementioned strategies (e.g., that the moral facts are identical to certain natural facts) would defeat the defeater.

Conservatism in moral epistemology is rampant. Some explicitly embrace it (e.g., Huemer 2005; Lycan 1986), but even among those who eschew the label, conservatism is often quietly at work in the background. Roger White tries to nullify EDAs by expressing doubt that people typically take the fact that they believe something as crucial evidence for its truth (2010: 585), but while I accept that people do not typically *explicitly* do so, I am very far from convinced that this phenomenon is not extensive. A person will ordinarily take the fact that she believes there to be a cat in front of her as fairly important evidence that there is a cat. This is revealed by the fact that if the reliability of the belief is called into question—if,

---

<sup>10</sup> Here I use “skepticism” in its proper epistemological manner. Elsewhere (e.g., Joyce 2012) I have followed others in using the term to stand for the disjunction of the error theory and expressivism—a view that might be better classified as a kind of negative dogmatism.

say, evidence comes forth that her belief is the product of hypnosis—then a rational person will become considerably less confident that there is indeed a cat in front of her. It is not unreasonable for us to give some initial credence to our beliefs, especially if the practice of doing so has generally withstood the trials of everyday life, but the revelation of an unexpected alternative explanation for a token or type of belief alters the epistemic landscape. A person may then seek to reinstate confidence in her belief that there is a cat before her—by bringing forth other evidence, such a fur-balls and dead mice. But if it is revealed that the beliefs concerning all the “cat evidence” are also the product of hypnosis—such that not only does the belief “Here is a cat” have an alternative explanation, but so too does the belief “Here is a fur-ball,” and so forth—then reinstating confidence in the belief may be a trickier affair. Perhaps one might instead investigate the nature of the process of hypnosis—hoping to discover, say, that the hypnotist was the benign sort who instills in his subjects only *true* beliefs—but if no data is forthcoming on that count, then it will be of no avail.

This, I believe, is what nativism is likely to reveal about our moral judgments. First, there is no evidence that can be gathered concerning the obtaining of moral facts that is not itself mediated by the very moral judgments that are in question. We cannot “triangulate” on moral facts using faculties independently of our moral faculty; the very urge to try to do so is driven by intuitions that are themselves the output of the moral faculty. Second, there seems no ground for optimism that the evolutionary process that produced the moral faculty will turn out to be the benign sort that happens to track truths (even if it lacks this as its adaptive function). Of course, it *might* be—just as the hypnotist *might* have been benign—but epistemic justification (or, regarding reliabilism, our confidence in epistemic justification) requires more than a vague “might.”

The first observation is one that can be made with no particular reference to evolutionary matters. It is no great news that the epistemic status of moral judgments is suspect even before Darwinian considerations enter the conversation. But revelation of a non-truth-tracking function for the moral faculty contributes significantly to the second observation, because the expansion of the space of hypotheses forces those inclined to trust their moral beliefs to articulate a persuasive defense of their preferred hypothesis in the face of new competition. If our beliefs in general have tended to steer us right as we navigate the world, and if one considers moral beliefs as in the same boat, then it would not be unreasonable to have some optimism regarding the probable truth of moral beliefs. Moral nativism requires us to consider moral beliefs independently—as not in the same boat—as the output of a psychological faculty with a distinctive evolutionary trajectory. It provides us, moreover, with an explanation of how such a faculty might come to exist while generating systematically false beliefs; and, furthermore, an explanation for why such a body of false beliefs might seem so compelling and almost beyond question. It has not infrequently been claimed against the moral skeptic that one’s confidence in fundamental moral intuitions must be far more robust than one’s confidence in any obscure philosophical argument (see, e.g., Huemer 2005: pp.116-117). An EDA has the strength to overturn this comparison, by presenting evidence to account for those fundamental moral intuitions—which can account even for their persuasive felt quality—that is compatible with their falsehood (in the sense that even an error theorist can accept the evidence).

I started out this paper mentioning Nietzsche's, Marx's, and Freud's debunking efforts, pondering whether they should be taken as offering a *refutation* of theism. Could one accept all that Freud says about religion as neurotic wish-fulfillment, I asked, while reasonably maintaining theistic belief? And the answer is: "Not easily." Freud's analysis of the psychology of religion would (if plausible) enlarge the space of explanatory hypotheses, banishing the legitimacy of any kind of complacent optimism regarding theistic belief. The analysis would (if plausible) require the believer to earn the right to his or her belief, to articulate positive grounds for belief in the face of an alternative hypothesis (or show that the hypotheses are not true alternatives at all). A Darwinian analysis of moral belief should have the same effect of arousing suspicion about morality, to the extent that the burden lies on the shoulders of the anti-skeptic to articulate a persuasive defense.<sup>11</sup>

Victoria University of Wellington

#### References:

- Alexander R. 1987. *The Biology of Moral Systems*. Hawthorne, NY.: Aldine de Gruyter.
- Alicke, M. 1985. "Global self-evaluation as determined by the desirability and controllability of trait adjectives." *Journal of Personality and Social Psychology* 49: 1621-1630.
- Brogaard, B. & Salerno, J. 2013. "Remarks on counterpossibles." *Synthese* 190: 639-669.
- 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*. NY: The Free Press.
- Carruthers, P. & James, S. 2008. "Evolution and the possibility of moral realism." *Philosophy and Phenomenological Research* 77: 237-244.
- Clarke-Doane, J. 2012. "Morality and mathematics: The evolutionary challenge." *Ethics* 122: 313-340.
- Conee, E. & Feldman, R. 1998. "The generality problem for reliabilism." *Philosophical Studies* 89: 1-29.
- Dahaene, S. 1997. *The Number Sense: How the Mind Creates Mathematics*. 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.
- Freud, S. [1927] 1987. "The future of an illusion." In *Civilization, Society and Religion: The Pelican Freud Library Vol. 12*. J. Strachey (trans.). London: Penguin.
- Friedrich, J. 1996. "On seeing oneself as less self-serving than others: The ultimate self-

---

<sup>11</sup> Thanks to Hatha McDivitt for stimulating discussions on this topic.

- serving bias?" *Teaching of Psychology* 23: 107-109.
- Goldman, A. 1979. "What is justified belief?" In G. Pappas (ed.), *Justification and Knowledge*. Dordrecht: Reidel. 1-23.
- Goldman, A. 1988. "Strong and weak justification." *Philosophical Perspectives* 2: 51-69.
- Goldman, A. 2008. "Immediate justification and process reliabilism." In Q. Smith (ed.), *Epistemology: New Essays*. Oxford: Oxford University Press. 63-82.
- Greene, J. 2003. "From neural 'is' to moral 'ought': What are the moral implications of neuroscientific moral psychology?" *Nature Reviews Neuroscience* 4: 847-850.
- Harman, G. 1986. "Moral explanations of natural facts: Can moral claims be tested against moral reality?" *Southern Journal of Philosophy* 24, suppl. vol.: 57-68.
- Huemer, M. 2005. *Ethical Intuitionism*, Palgrave Macmillan.
- Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, R. 2006. *The Evolution of Morality*. Cambridge, MA.: MIT Press.
- Joyce, R. 2012. "Metaethical pluralism: How both moral naturalism and moral skepticism may be permissible positions." In S. Nuccetelli & G. Seay (eds.), *Ethical Naturalism: Current Debates*. Cambridge University Press.
- Kahane, G. 2011. "Evolutionary debunking arguments." *Noûs* 45: 103-125.
- Kitcher, P. 2011. *The Ethical Project*. Cambridge, MA.: Harvard University Press.
- Lycan, W. 1986. "Moral facts and moral knowledge." In *The Spindel Conference: The Southern Journal of Philosophy*, suppl. vol. 24: 79-94.
- Mackie, J. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.
- Mares, E. 1997. "Who's afraid of impossible worlds?" *Notre Dame Journal of Formal Logic* 38: 516-526.
- Mason, K. 2010. "Debunking arguments and the genealogy of religion and morality." *Philosophy Compass* 5: 770-778.
- McKenna, F, Stanier, R., & Lewis, C. 1991. "Factors underlying illusory self-assessment of driving skill in males and females." *Accident Analysis and Prevention* 23: 45-52.
- Mikhail, J. 2011. *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge: Cambridge University Press.
- Miller, G. 2007. "Sexual selection for moral virtues." *Quarterly Review of Biology* 82: 97-121.
- Milo, R. 1995. "Contractarian constructivism." *Journal of Philosophy* 92: 181-204.
- 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 R. Noë, J. van Hoof, & P. Hammerstein (eds.), *Economics in Nature: Social Dilemmas, Mate Choice, and Biological Markets*. Cambridge: Cambridge University Press. 93-118.
- Pronin, E., Lin, D., & Ross, L. 2002. "The bias blind spot: Perceptions of bias in self versus others." *Personality and Social Psychology Bulletin* 28: 369-81.
- Restall, G. 1997. "Ways things can't be." *Notre Dame Journal of Formal Logic* 38: 583-596.
- 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 G. Boniolo & G. de Anna (eds.), *Evolutionary Ethics and Contemporary Biology*.

- Cambridge: Cambridge University Press. 13-26.
- Ruse, M. 2009. "Evolution and ethics: The sociobiological approach." In M. Ruse (ed.), *Philosophy After Darwin*. Princeton, NJ: Princeton University Press. 489-511.
- Singer, P. 2005. "Ethics and intuitions." *Journal of Ethics* 9: 331-352.
- Skarsaune, K. 2011. "Darwin and moral realism: Survival of the fittest." *Philosophical Studies* 152: 229-243.
- Sober, E. 1984. *The Nature of Selection*. Cambridge, MA.: MIT Press.
- Street, S. 2006. "A Darwinian dilemma for realist theories of value." *Philosophical Studies* 127: 109-166.
- Street, S. 2008. "Reply to Copp: Naturalism, normativity, and the varieties of realism worth worrying about." *Philosophical Issues* 18: 207-228.
- Taylor, S. & Brown, J. 1988. "Illusion and well-being: A social psychological perspective on mental health." *Psychological Bulletin* 103:193-210.
- Vander Laan, D. 2004. "Counterpossibles and similarity." In F. Jackson & G. Priest (eds.), *Lewisian Themes: The Philosophy of David K. Lewis*. Oxford: Oxford University Press. 258-275.
- 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.
- Zahavi, A. 1977. "The cost of honesty: Further remarks on the handicap principle." *Journal of Theoretical Biology* 67: 603-605.