

Moral Naturalism and Evolutionary Debunking Arguments

Evolutionary Debunking Arguments (EDAs) seek to provide us with defeaters for our moral beliefs by drawing our attention to their evolutionary genealogy. Moral realism, very roughly, is the metaethical view that there are objective or mind-independent moral facts. EDAs are typically understood as a threat to the moral beliefs of moral realists.¹

Here is a schematic EDA:

Evolutionary Debunking Argument (EDA)

(EDA1) You are aware that there is an evolutionary genealogy of your moral beliefs.

(EDA2) If you are aware that there is an evolutionary genealogy of your moral beliefs, then you have a defeater for those beliefs.

(EDA3) Therefore, you have a defeater for those beliefs.

EDA is valid. Let us consider in more detail how the debunker defends EDA1 and EDA2, respectively.

EDA1 claims that you are aware of an evolutionary genealogy of our moral beliefs. Here is the gist of such a genealogy, synthesizing quite a bit.² The nearly universal human disposition to believe that cooperative behaviors are morally right and that uncooperative behaviors are morally wrong. Early hominids who possessed that disposition—and, perhaps, who lived in communities in which that disposition was relatively common—tended to survive and pass along their genes. Over many generations, standard evolutionary selected for and refined that disposition.

¹ Moral anti-realists, who believe that moral facts are at least partly determined by our moral beliefs and attitudes, are arguably exempt from the debunking threat. Indeed, Sharon Street's seminal EDA is framed as an argument *for* moral anti-realism. See Street 2006. Other debunkers who explicitly target moral realists include Joyce 2008; Bedke 2009 and 2014; Kahane 2011; Morton 2016. For a rare dissenting voice, see Berker 2014.

² The following evolutionary genealogy very loosely follows the debunking explanation in Joyce 2008.

The moral beliefs we hold today, then, are the causal products of an evolutionary process aimed entirely at maximizing reproductive fitness.

EDA2 claims that awareness of the evolutionary genealogy of our moral beliefs provides us with a defeater for those beliefs. An undercutting defeater—henceforth just “defeater”—for a belief that p is another belief that q that undermines one’s justification for believing that p without justifying one in believing that $\sim p$.³ How, exactly, does awareness of the evolutionary genealogy of your moral beliefs give you a defeater? There are at least three possible mechanisms.⁴

First, your new information about the evolutionary genealogy of your moral beliefs should lead you to conclude that your moral beliefs are *insensitive*—even if there were no objective moral facts corresponding to your moral beliefs, the same evolutionary forces and proximate natural causes would have ultimately caused you to have held the very same moral beliefs you in fact hold.⁵

Second, you should recognize that, given their evolutionary origins, it would be a massive coincidence if your moral beliefs turned out to be true. What are the changes that an evolutionary process aimed solely at producing fitness-maximizing traits also happened to steer our moral faculties in the direction of moral truth? The odds seem quite low.⁶

Third, you have learned that there is an “explanatory disconnect” between the moral facts, if they obtain, and your beliefs about them. After all, if the evolutionary facts alone suffice to explain why you hold the moral beliefs you do, then the moral facts, even if they exist, seem to play no non-redundant explanatory role in the genesis of those beliefs.⁷

³ See Pollock 1987.

⁴ These three mechanisms have received the most discussion in the literature..

⁵ See Joyce 2008; Kahane 2011; Clarke-Doane 2012; Baras and Clarke-Doane 2021

⁶ See Street 2006, p. 129; Bedke 2014; Berry 2020

⁷ In a recent series of papers, Daniel Z. Korman and Dustin Locke argued that such “explanatory disconnect” inferences are the fundamental sources of defeat in debunking arguments. See especially Korman & Locke 2020.

Moral naturalism (MN), very roughly, is the metaethical view that every objective moral fact *just is*—in some sense or other—some natural fact(s) (see §II). Non-naturalism is the view that objective moral facts are *sui generis* entities over and above the natural facts upon which they supervene.

According to the orthodoxy, EDAs pose a serious threat only to moral realists who are also non-naturalists. Moral naturalists, by contrast, can avoid the debunking threat with relative ease. For example, Richard Joyce writes:

If [...] moral naturalism is viable, then the apparent fact that human moral judgments can be explained without mentioning or presupposing moral facts will not have a debunking effect.⁸

Matthew Bedke agrees:

...[N]on-naturalism is a nice test case for seeing whether there is an epistemic *problem* [for moral realism] at all....I am skeptical that the sort of defeater at play for non-naturalists extends to other meta-normative views.⁹

Joyce and Bedke are both debunkers. Nevertheless, they both think that moral naturalists are well-positioned to resist EDAs. Indeed, proponents of EDAs sometimes frame those arguments as challenges to non-naturalist moral realism, in particular, rather than moral realism in general.¹⁰

Moral non-naturalist David Enoch expresses tentative support for the orthodoxy.¹¹ Another prominent non-naturalist, Russ Schafer-Landau, concurs with Enoch:

Ethical naturalists regard moral properties as natural properties, and moral facts as natural ones. So even if moral beliefs have a complete set of natural causes, this does not exclude the possibility that these beliefs also result from moral causes. On this line, some natural

⁸ Joyce 2007, p. 145.

⁹ Bedke 2009.

¹⁰ See Enoch 2014 and Bedke 2009.

¹¹ Enoch 2014.

facts are also moral facts, and some of these may well be causally responsible for our holding the moral beliefs we do.¹²

Unsurprisingly, the orthodoxy is quite popular among moral naturalists. For example, David Copp argues that the “society-centered theory,” his own version of moral naturalism, “shows...that the Darwinian challenge carries no risk for moral realism.”¹³ Indeed, the *SEP* entry on moral naturalism lists moral naturalism’s alleged immunity to defeat due to EDAs as one of the two main motivations for the view.¹⁴ So prominent evolutionary debunkers, moral naturalists, and even moral non-naturalists all number among the orthodoxy’s defenders.¹⁵

Here is the plan. First, I shall argue that the orthodoxy should be understood as the claim that a moral naturalist may use her moral-natural identity beliefs as “defeater-deflectors” to shield her moral beliefs from defeat due to EDAs (§II). Next, I shall introduce and defend a general independence constraint on epistemically legitimate defeater-deflectors (§III). I shall argue that the moral beliefs of most *a posteriori* moral naturalists violate this constraint (§IV). As a result, they are no better off with respect to EDAs than their non-naturalist counterparts. I will close by considering an essentialist epistemology of identity and reduction capable of evading debunking threats in a variety of philosophical domains (§V).

I. Identity Beliefs as Defeater-Deflectors

Consider the following pair of cases:

Main and Pain: You are relying on the voice directions from your car’s built-in GPS system to guide you through a large, unfamiliar city toward your destination at 123 Main St. “Your destination is 1 mile away,” the voice says. On that basis, you form the belief that 123 Main St. is 1 mile away. However, you then notice that the GPS is programmed for 123 *Pain* St. You

¹² Schafer-Landau 2012, p. 27

¹³ Copp 2008, p. 204.

¹⁴ Lutz 2018, Section I.2.

¹⁵ Although Street 2008, Section 7 is a prominent dissenting voice. See Toner 2011, Barkhausen 2016, and Bogardus 2016 for other opponents of the orthodoxy.

infer that you are 1 mile from 123 *Pain* St., and that your being proximity to that location is what caused your caused your GPS to say what it did. Therefore, so you conclude, your belief that you are 1 mile from 123 Main St. is causally explained by your being 1 mile from 123 Pain St.

Main and Pain 2.0: Everything is the same as in Main and Pain 1.0. This time, though, there is a trusted friend in the passenger seat. You have just realized that the GPS is programmed for 123 Pain St., and inferred that your proximity to Pain St. is causally responsible for your Main St. belief. Noticing your look of concern, your friend says: “Oh, don’t worry! I did a bit of research before we left. It turns out that ‘Main’ and ‘Pain’ are two different names for one and the same street—so 123 Main St. *just is* 123 Pain St.!” Based on your friend’s testimony, you form the belief that 123 Main St. = 123 Pain St, and decide that your belief that 123 Main St. is 1 mile away is in good standing after all.

Consider the original Main and Pain case first. Initially, your belief that you are 1 mile from 123 Main St.—which we can call your “Main St. belief”—enjoys *prima facie* justification. Plausibly, however, what you subsequently learn about the genealogy of your Main St. belief gives you a defeater for that belief.

For one thing, it seems to reveal that your Main St. belief is insensitive to the facts about how close you actually are to 123 Main St. After all, your GPS would have said the same thing regardless of whether 123 Main St. really is a mile away. Relatedly, it would seem to require a massively improbable coincidence for your Main St. belief to turn out to be true. What are the chances that your GPS took you to 123 Pain St. and, just by sheer luck, 123 Main St. also happened to be within a 1 mile radius?

Finally, it seems you have learned that there is no appropriate “explanatory connection” between your belief that 123 Main St. is 1 mile away, on the one hand, and the *fact* that 123 Main St. is 1 mile away, on the other. Even if that fact obtains, it seems to play no role in the causal process that produced your belief. Perhaps this recognition, all on its own, is what generates a defeater.

Now consider the second version of the Main/Pain case. As in the original case, in Main and Pain 2.0 your realization about the genealogy of your Main St. belief is poised to serve as a defeater for that belief. However, in this second case, you then gain another belief—your belief that Main is identical with Pain—that serves as a *defeater-deflector*. A belief that r is a defeater-deflector just in case, in virtue of holding the belief that r, one is prevented from gaining a defeater for one's belief that p.¹⁶

In Main/Pain 2.0, your identity belief—i.e. your belief that 123 Pain is identical with 123 Main—serves as a defeater-deflector. Specifically, that belief serves as what I have elsewhere called a *neutralizing* defeater-deflector. A belief that r is a neutralizing defeater-deflector just in case, in virtue of holding the belief that r, some other belief one holds, the belief that q, is prevented from serving as a defeater for one's belief that p. In other words, neutralizing defeater-deflectors undermine the defeating potential of a defeater belief *without* rationally requiring you to give up that defeater belief.

To see this, note that your identity belief prevents you from drawing any adverse inferences regarding the explanatory connection between the alleged fact that 123 Main is one mile away, on the one hand, and your *belief* that 123 Main is a mile away, on the other. You have just learned that 123 Pain St. is causally responsible for your Main belief. However, since you also think that 123 Main *just is* 123 Pain, you should conclude that 123 Main St.'s being one mile away *does* enter into the genealogy of your belief that 123 Main St. is a mile away.

Second, your awareness of the evolutionary explanation of your moral beliefs does not require you to infer that your moral beliefs are modally disconnected from the moral facts. To see this, note that identities hold of necessity—necessarily, if fact x is identical with fact y then,

¹⁶ See Plantinga 1993 and 2000, Moon 2017 and 2021, and Barker 2020.

necessarily, x obtains if and only if y obtains. Indeed, it is plausible that “reduction” relations other than identity are governed by similarly strong necessitation principles. For, in order to be truly reductive in nature, the obtaining of such relations should at the very least prohibit the reduced property from modally “floating free” of the property to which it is reduced. And something like the above necessitation principle is needed to secure that prohibition.

Consider the following counterfactual: if the moral facts were different, you still would have held the same moral beliefs. Does learning that there is a complete evolutionary explanation give you reason to accept counterfactual? Not if you believe that the moral facts are identical with some of the facts, such as the cooperativeness facts, cited in the evolutionary explanation.

For, given the necessity of identity, the nearest possible worlds in which the moral facts are different are also worlds in which different actions are cooperative. But if different sorts of acts had been cooperative, then evolutionary forces would have produced in us the predisposition to believe that *those* actions are morally right. In other words, worlds with different cooperativeness facts are also worlds in which we have different moral beliefs.

Therefore, the nearest worlds in which the moral facts are different are also worlds in which our moral beliefs are different. So you should deny that, if the moral facts were different, you would have held the same moral beliefs. On the contrary, your belief in the identity of moral facts with natural facts gives you reason to think your morally beliefs are appropriately modally sensitive to the moral facts.

II. The Orthodoxy

Moral naturalists endorse the following thesis:

Moral Reduction (MR) there are some natural facts, the Ns, such that *what it is* for there to be some objective moral facts, the Ms, *just is* for the Ns to obtain.

Note that the “what it is... just is for...” locution in MR is intentionally ambiguous. Different moral naturalists have different accounts of the relevant reductive or constitutive relation between the moral and the natural.¹⁷ For ease of exposition, I will assume a version of MR according to which the moral facts are *identical* with the natural facts. But nothing of substance hangs on this assumption—every argument I give below will work, *mutatis mutandis*, for non-identity versions of MR.

According to the orthodoxy, the moral naturalist’s belief in MR serves as a defeater-deflector for her. Specifically, the moral naturalist can agree with the antecedent of EDA2—that there is an evolutionary genealogy of her moral beliefs—while denying its consequent—that she thereby gains a defeater for those moral beliefs. The second Main/Pain case above provides an instructive analogy about *how* MR is supposed to deflect away the incoming defeater.

First, the moral naturalist’s reduction beliefs render it impermissible for her to draw the inference that her moral beliefs are insensitive. Although the natural facts in the evolutionary genealogy fully causally explain her moral beliefs, so she can reason, it is not the case that she would still have held those beliefs regardless of whether they are true. For, if the moral facts had been different or totally absent, then the natural facts in the evolutionary explanation would have been different as well. After all, given MR, the moral facts *just are* the natural facts!

Second, her reduction beliefs render it completely unmysterious why the moral facts just happened to align with her moral beliefs. For those moral beliefs were caused by the natural facts, which are *what it is* for the moral fact to obtain. It is no more a coincidence that her moral beliefs happened to be true than it is that your GPS, when programmed to Pain St., “happened” to take you to Main St. They’re one and the same!

¹⁷ Ex. Brink 1989 pp. 176–7 defends a version of MR according to which moral facts are *constituted by* natural facts without being identical with them. Also see Rosen 2017 for *grounding*-based and *essence*-based formulations of moral naturalism.

Third, the explanatory connection is secured. Since the natural facts in the evolutionary genealogy *do* causally explain her moral beliefs, and since she thinks the moral facts *just are* those natural facts, it follows that the moral facts *do* causally explain her moral beliefs. Explanatory connection restored!

In light of the above, I propose to understand the orthodoxy as follows:

The Orthodoxy: A moral naturalist can use her belief in MR-- or some relevantly similar precisification of MR—as a defeater-deflector, which prevents her recognition that there is an evolutionary explanation of her moral beliefs from giving her a defeater for those beliefs.

This formulation accounts for the idea that belief in moral naturalism gives one a *unique* reply to EDAs, i.e. a reply that is unavailable to non-naturalists. No non-naturalist believes MR or any precisification thereof. So, trivially, no non-naturalist can use her belief in MR as a defeater-deflector.

Unfortunately, the Orthodoxy is false. I shall argue that this disanalogy is also epistemically relevant. In the next section, I shall argue that defeater-deflectors must meet an independence constraint. Your belief about the identity of Main St. and Pain St. *does* meet that constraint, while the moral naturalist's belief in the identity of the moral and the natural facts does *not* meet that constraint.

III. An Independence Constraint on Defeater-Deflectors

Consider one more variation of the Main/Pain case:

Main and Pain 3.0: Everything is the same as in Main and Pain 1.0. Unlike in Main/Pain 2.0, however, you have no friend in the passenger seat. Instead, when you realize that your GPS is programmed for 123 Pain St., you engage in the following obviously illicit line of reasoning: “Ah, well, this is interesting! I know that 123 Main St. is just a mile away. Yet I also just learned that the fact that 123 *Pain* St. is just a mile away is responsible for my believing that 123 Main St. is just a mile away. Now, it would certainly be an improbable

coincidence if Main St. and Pain St. were different streets, and both just so happen to be within a one-mile radius of me. Yet I do know that I'm a mile from 123 Main St. So it must be that Main St. and Pain St. are one and the same street!"

Clearly, the line of reasoning you have just used to infer the identity of Pain and Main is illicit. Moreover, it would be even *more* illicit if you then went on to try to use your new identity belief as a defeater-deflector, blocking the otherwise defeating force of the new genealogical information you have gained about the origins of your Main St. belief.

Why? What explains the difference between the illicit defeater-deflector case that is Main/Pain 3.0 and the perfectly legitimate defeater-deflector case that is Main/Pain 2.0? The following principle explains the difference:

Independence: If one holds beliefs b_1 and b_2 , then b_2 may be a defeater-deflector for belief b_1 *only if* it is not the case that b_2 is epistemically dependent upon b_1 .¹⁸

Let us say that a belief b_2 is immediately epistemically dependent upon another belief b_1 just in case (i) b_1 is justified, (ii) b_2 is justified, and (iii) b_2 is justified in virtue of b_1 's being justified, and (iv) there is no other belief b_3 such that b_2 is justified in virtue of b_3 's being justified. And let us define epistemic dependence as the transitive closure of the immediate epistemic dependence relation.¹⁹

Independence explains the difference between Main/Pain 2.0 and Main/Pain 3.0. In Main/Pain 3.0, your belief in the identity of Main and Pain is epistemically based, at least in part, upon your belief that Main St. is 1 mile away. So Independence prohibits the use of that belief as

¹⁸ Independence principles have received plenty of discussion in the literature on the epistemology of disagreement. However, as far as I know, Moon 2017 and 2021 is the only extant discussion of independence principles for defeater-deflectors in the context of debunking arguments. Note that my Independence principle is weaker than those discussed in Moon 2017 and Moon 2021. On epistemic independence in the epistemology of disagreement, see Christenson 2010 and Lasonen-Aaarnio 2014.

¹⁹ See Audi 1997.

a defeater-deflector for purposes of protecting your belief that Main St. is 1 mile away from defeat. In Main/Pain 2.0, however, your identity belief does *not* depend for its justification upon your belief that 123 Main St. is a mile away. Rather, that belief is based entirely on your friend's testimony.

IV. The Independence Argument

I can now state my argument against the orthodoxy regarding moral naturalism and EDAs:²⁰

The Independence Argument

(IA1) Independence is true.

(IA2) If Independence is true, then no moral naturalist may use her belief in MR as a defeater-deflector.

(IA3) Therefore, no naturalist may use her identity beliefs as defeater-deflectors.

The argument is valid. I have already introduced and motivated Independence and, thus, IA1. So my main task is now to defend IA2, that Independence prohibits the moral naturalist from using her belief in MR as a defeater-deflector to protect her moral beliefs from defeat due to EDAs.

My defense of IA2 begins by noting that the moral naturalist's belief in MR is not a *basic* belief. Rather, that belief depends for its justification on some other belief(s) she holds. After all, moral naturalists offer *arguments* for their view. Let us consider the standard arguments for moral naturalism. What we shall find, I think, is that these arguments either explicitly or implicitly rely upon the assumption that *there are some moral facts*.

IV.1 Queerness

The first argument for MR turns on the alleged metaphysical "queerness" of irreducible moral facts.²¹ If there were such facts, they would be completely unlike facts of any other sort. In

²⁰ Street 2008 also claims that moral naturalists must base their moral-natural identity beliefs upon their moral beliefs. However, Street uses that claim for different purposes than I do. In particular, Street does not make the epistemic point I am making in this paper—namely, that the dependence of the moral naturalist's identity beliefs upon her moral beliefs prevents the former from serving as non-circular defeater-deflectors. See Street 2008, especially pp. 139—141.

²¹ Originally developed by J.L. Mackie as an argument for moral anti-realism. See Mackie 1977.

particular, they would be accompanied by a worldly quality of “to-be-done-ness”— they would be mind-independent, yet they would somehow make demands upon us, that *require* us to do this or that.

Here is the argument:

Queerness Argument (QA)

(QA1) There are objective moral facts.

(QA2) If there are objective moral facts, then either they are reducible to natural facts or else they are metaphysically queer.

(QA3) There are no metaphysically queer facts.

(QA4) Therefore, objective moral facts are reducible to natural facts.

QA is valid. QA4 is just another way of saying that MR is true.

However, notice that the argument’s first premise, QA1, says that there *are* objective moral facts. But that is exactly what E1 says. So, if the moral naturalist’s belief in MR is based upon QA, then her belief in MR is in part based upon E1. Notice that any reconstruction of QA that dispenses with QA1 will be unable to validly deliver QA4.

QA is not the only motivation for moral naturalism. Nor is it—these days at least—the most popular argument for moral naturalism. Nevertheless, the above discussion does illustrate something important about all the main arguments for moral naturalism—all such arguments, as we shall see, rely on the premise that there are objective moral facts.

IV.2 Parsimony

Consider a familiar maxim:

Ockham’s Razor: Do not multiply distinct entities without necessity!

According to the traditional reading, the term “distinct” means “numerically distinct.” So, according to the traditional reading, Ockham’s Razor prohibits the needless postulation of any

entity or entities that are numerically distinct from whatever entities you have already postulated. For clarity's sake, I will work with the traditional reading. Feel free to substitute your preferred reading in what follows.²²

The second *a posteriori* argument for MR relies upon the truth of Ockham's Razor. Here is that argument:

Parsimony Argument (PA)

- (PA1) There are objective moral facts.
- (PA2) If there are objective moral facts, then either MR is true *or* moral non-naturalism is true.
- (PA3) No theory that violates Ockham's Razor is true.
- (PA4) Moral non-naturalism violates Ockham's Razor.
- (PA5) Therefore, MR is true.

PA5, which is equivalent to MR, validly follows from PA1—PA4, taken together. The moral naturalist will defend PA4 by arguing that the moral non-naturalist's postulation of irreducible moral facts is *unnecessary*, especially for purposes of explaining the phenomena in need of explanation. PA3 appeals to the truth of Ockham's Razor and its reliability as a guide to theory-choice, while PA2 makes the trivially true claim that if indeed objective moral facts exist then they either are or are not reducible to the natural facts.

That leaves us with PA1 which, like QA1 above, is equivalent to E1. Moreover, as we saw in the case of QA, no revised version of the argument can validly deliver the unconditional conclusion that MR is true. The most a PA1-free version of the argument could deliver is the conditional claim that *if* there are objective moral facts *then* they are reducible to natural facts.

²² Contemporary philosophers have offered other readings of "distinct", resulting in various weakened versions of Ockham's Razor. David Lewis advocated a *qualitative* reading of Ockham's Razor, which prohibits only the needless multiplication of numerically distinct *types* of entities. Schaffer 2015 defends a *fundamentality* reading of Ockham's Razor, which prohibits only the needless multiplication of numerically distinct types of *fundamental* entities.

IV.3—The Supervenience Argument

Perhaps the most influential argument for moral naturalism is based upon a principle about the supervenience of the moral on the natural.²³ Here is a version of that principle:

Strong Supervenience (S): necessarily, for any moral fact M, there are some natural facts, the Ns, such that, necessarily, M obtains if and only if the Ns obtain.

The Supervenience Argument begins with the premise that S is true. The argument's second premise is that S's truth "calls out", "cries out", or "demands" explanation. In virtue of what, in other words, should the moral facts be so closely modally correlated with the natural facts? Surely there is an explanation.

Finally, the argument's third premise claims that the truth of MR provides the best explanation for the truth of S. If the moral facts *just are* the natural facts, so the thought goes, then the modal co-variation of the moral and the natural is grounded in the reducibility or identity of the moral to the natural. Just as Cicero co-varies modally with Tully *because* Cicero is identical with Tully, so too the moral facts co-vary modally with the natural facts because the moral facts *just are* the natural facts.

Here is the argument, which I am representing as an IBE argument:

(SA1) S is true.

(SA2) S's truth calls for explanation.

(SA3) MR is the best explanation of S's truth.

(SA4) Therefore, MR is true.

²³ See Blackburn 1971 for an early formulation of the argument. More recently, see Horgan 1993, McPherson 2012, and Väyrynen 2017.

Why does the moral naturalist think that S is true? Note that the moral naturalist cannot defend SA1 on the grounds that S is an *analytic* truth. For, if S were analytically true, then either S's truth would not call for explanation or else the best explanation of S's truth would be its analytic status. Either way, one of SA2 or SA3 would then be false and the argument would be unsound.

There are certainly those in the history of metaethics who have thought that S is analytically true. For current purposes, though, we can set them aside. I'll deal with a version of that view in the next subsection. Right now, all that matters is how those who base their belief in MR upon the Supervenience Argument can support S. For *those* defenders of moral naturalism, at least, S is synthetic rather than analytic.

If S is synthetic, then moral naturalists who endorse SA either base their belief in its truth upon further beliefs or not. If not, then it is either based upon perceptual experience or not. If not, then it is a synthetic *a priori* belief. Either way, as we shall see later on, a belief like this one is susceptible to a "revenge" EDA—

Suppose, as most moral naturalists do, that MR is based upon further beliefs. Here is the sort of argument that the moral naturalist presumably offers in support of S:

Inductive Argument

- (IA1) St. Francis's act of generosity was both morally right and fitness-maximizing, and Torrance's act of cruelty was both morally wrong and not fitness-maximizing, and Peter's act of kindness was both right and fitness maximizing, and....
- (IA2) Therefore, every act that is morally right is also fitness-maximizing and every act that is morally wrong does not maximize fitness (via inductive generalization)
- (IA3) The fact that IA2 is true calls for explanation.
- (IA4) S is the best explanation of IA2.
- (IA5) Therefore, S is true. (via IBE)

The inference from IA1 to IA2 is a universal generalization from the particular correlations listed at IA1. Then IA3 states that if every moral fact is indeed correlated with a particular type of natural fact, then that correlation calls for explanation. IA5 is then inferred from IA3 and IA4 via IBE.

Set aside your thoughts on the quality of this argument. If it succeeds, it does manage to establish S using premises that are at least arguably justifiable on *a posteriori* grounds. But the argument relies upon IA1, which is a long conjunction of particular claims about which acts are objectively morally right or wrong and which natural properties those actions have. As a result, anyone who believes IA1 is thereby committed to the existence of objective moral facts.

Therefore, any moral naturalist who relies upon the above argument to establish S is thereby basing her belief in S in part upon her belief that there are objective moral facts. Moreover, since the moral naturalist under consideration is using her belief in S as the first premise in an IBE argument for MR, she is thereby basing her belief in MR indirectly upon her belief in objective moral facts. She is, in other words, an existence baser.

IV.4—Metasemantic Argument

Consider the Kripke-Putnam account of natural kind terms like ‘water’, ‘gold’, etc. First, the reference of such a term is by whatever external thing is causally responsible for the internal states we associate with the term. For example, we associate the term ‘water’ with perceptual experiences as of a clear, odorless, tasteless, and thirst-quenching liquid. Since H₂O molecules are causally responsible for these perceptual experiences, the term ‘water’ refers to H₂O.²⁴

²⁴ Putnam 1975 and Kripke 1980.

Second, natural kind terms are *rigid designators*—they have the same extension or reference in every possible world in which they refer. For example, the term ‘water’ is a rigid designator. So, if ‘water’ actually refers to X, then ‘water’ refers to X in every possible world in which ‘water’ refers at all.

Third, co-referring rigid designators license a corresponding identity claim. That is, if ‘T1’ and ‘T2’ are both rigid designators, then if ‘T1’ and ‘T2’ both refer to b, then $\langle T1 = T2 \rangle$ is true. For example, ‘water’ and ‘H2O’ are co-referring rigid designators. And so we can conclude that $\langle \text{water} = \text{H2O} \rangle$ is a true identity claim.

Note that biconditionals associating the relevant natural kind with the associated cluster of internal states—ex. propositions like $\langle x \text{ is water iff } x \text{ tends to cause water-}y \text{ perceptual experiences} \rangle$ —are analytically truths. So those biconditionals are knowable *a priori*. However, we must engage in empirical inquiry to discover what external stuff is causally responsible for these internal states. As a result, our knowledge of the identity of natural kinds—ex. propositions like $\langle \text{water} = \text{H2O} \rangle$ —is *a posteriori*.

Some moral naturalists have extended Kripke-Putnam metasemantics to moral terms.²⁵ On this view, the term ‘right’ is a rigid designator. Moreover, there are some internal states, the *M-states*, which we closely associate with the term ‘right’—ex. the feeling approval toward an action, the feeling shame at having failed to do an action, etc. There is also some natural property N such that an action’s being N tends to cause the M-states. Finally, the identity of natural property N is discoverable *a posteriori*. For example, perhaps it is discoverable *a posteriori* that N is the property of being fitness-maximizing.

²⁵ See especially Boyd 1988, Jackson 1998, and Sterenly and Fraser 2013.

Consider the following story, inspired by Horgan and Timmons' famous moral twin earth thought experiment:²⁶

Painlings: We discover a distant planet inhabited by billions of sentient humanoids called 'Painlings,' each of which experiences regular 'pain-seizures', which are apparently random bouts of excruciating pain. To our horror, we begin to notice striking correlations between their pain-seizures and the behavior of human beings back on Earth—for each and every human being, there is a unique Painling who experiences an acute pain-seizure if and only if that human being does something we would tend to call 'morally right'—ex. giving to charity, driving a friend to the airport, saying a kind word, fighting for racial justice, etc. After years of careful study, scientists conclude that these human behaviors reliably *cause*—via causal mechanisms that are as-yet unknown—Painlings to experience their pain-seizers.

Painlings does not show that there are no analytic application conditions for moral terms. Rather, Painlings shows that the analytic application conditions of our moral terms is not exhausted by the functional role they play. In other words, such terms don't apply as long as certain causal roles are fulfilled. In addition, the actions exemplifying the natural properties that fulfill those causal roles must have the right *moral* qualities as well.

Moral terms, in other words, have the existence of moral properties and facts as part of their application conditions. This distinguishes them from familiar natural kind terms like 'water' and 'gold.' In these cases, it is analytic that the meeting of the right functional role is necessary and sufficient for the relevant natural kind term to apply. It is no part of the application condition of the term 'water' that it apply only if the stuff fulfilling the functional role *really is* water. For there's nothing more to being water than fulfilling that functional role!

By contrast, there is more to being morally right or wrong than simply whether or not you exemplify a natural property that fulfills the causal role of regulating the use of our moral terms. What other metasemantic conditions, in addition to the causal role condition, must an action meet

²⁶ Horgan and Timmons 1992.

in order to be the referent of a moral term, such as ‘right’? There is only one conceivable answer: it is a part of the meaning of the moral term ‘right’ that ‘right’ applies to an action *A* *only if A really is* right.

Here is how the above argument connects to the epistemology of moral reference. Consider an individual’s “reference” beliefs about what her moral terms do and do not refer to. Specifically, consider her belief that the term ‘right’ refers to *something or other*. What is this individual’s epistemic basis for this belief? What justifies her in believing that ‘right’ refers to something?

Presumably, her reference belief based in part upon her belief that it is part of the meaning of ‘right’ that it applies to any action with certain natural properties $N_1 \dots N_n$. Perhaps those natural properties include or are exhausted by causal-functional properties. Regardless, that belief alone is not enough to justify her in believing that ‘right’ refers. She must also believe that some action actually has natural properties $N_1 \dots N_n$.

The Painlings case illustrates that her reference belief is based upon two additional beliefs as well—her belief that it is part of the meaning of ‘right’ that it only applies to actions that really are right, and that at least one action really is morally right. As a result, this individual’s reference belief is based, at least in part, upon her belief that there is some action that is morally wrong. Of course, that is a *moral belief*. Therefore, this individual’s belief about the reference of her moral terms is based in part upon her moral beliefs.

More generally, our beliefs about the reference of our moral terms are based in part upon our first-order moral beliefs. The epistemology of moral reference is, in this respect, disanalogous to the epistemology of scientific and other natural kind terms. Perhaps this is a strike against the view that the semantics for moral kind terms is relevantly like the semantics for other natural kind

terms. Or perhaps this epistemological asymmetry does not point to an underlying metasemantic asymmetry.

I shall not try to adjudicate this question here. For our purposes, what matters is simply that our beliefs about the reference of moral terms are based upon our moral beliefs. For it follows that, given Independence, neither those reference beliefs nor any beliefs upon which they are based may be used as defeater-deflectors. Specifically, given Independence, it follows that no moral naturalist who bases her identity beliefs upon her beliefs about the metasemantic and referential properties of moral terms may permissibly use those identity beliefs as defeater-deflectors to protect her moral beliefs from the epistemic threat posed by debunking arguments.

V. Essentialist Reductionism

Existence basers about a given domain F are reductionists about the F-facts who base their reduction beliefs, at least in part, upon their existence beliefs about the F-facts, i.e. their belief that there are F-facts and/or their beliefs about which F-facts there are. Note that, just as one can be a reductionist about certain domains but not about other domains, one can also be an existence baser about some but not all domains.

For example, *moral* existence basers hold both moral existence beliefs and moral identity beliefs. Moreover, their moral identity beliefs are epistemically based, at least in part, upon their moral existence beliefs. As we have just seen, the vast majority of *a posteriori* moral naturalists are existence basers in the moral domain. Similarly, *object* existence basers hold both object existence beliefs and object identity beliefs, and they base the latter beliefs at least in part upon the former beliefs. And so on, *mutatis mutandis*, for other domains.

Independence prohibits the use of a given belief as a defeater-deflector for any belief(s) the former epistemically depends upon. Existence basers in a given domain F are those whose identity beliefs about the Fs are epistemically based upon her existence beliefs about the Fs. Therefore, existence basers in a given domain F may *not* use their identity beliefs about the Fs as defeater-deflectors for their existence beliefs about the Fs. Thus—assuming the reductionist has no other defeater-deflectors at the ready—the debunker has succeeded in undermining her justification for holding existence beliefs in the relevant domain.

Consider the following essentialist claims about putative kinds of entities:

(ES1) It is essential to moral rightness that there is some natural property N such that an action is morally right if and only if it is N.

(ES2) It is essential to composition that, for any xs, the xs compose a y if and only if the xs exist.

(ES3) *What it is* for something x to be a mathematical entity is for x to be some natural entity.

(ES4) It is essential to redness that something x is red if and only if x has some physical property P.

(ES5) It is essential to necessity that a fact is necessary if and only if it has some category property C.

ES1—ES5 are all statements about what lies in the *essence* or *nature* of entities of type F. Let an essence belief be a belief in the truth of some proposition relevantly like E1—E5, respectively.

The existence baser's essence beliefs, like her identity beliefs, are in part based upon her existence beliefs. Perhaps her reduction beliefs are based upon her essence beliefs, which are in turn based upon her existence beliefs. Maybe her essence beliefs are based upon her reduction beliefs, which are in turn based upon her existence beliefs. Or perhaps her essence beliefs and her reduction beliefs are independently inferred from her existence beliefs.

Serious essentialists, by contrast, do not base their essence beliefs—even in part—upon their existence beliefs. Instead, their essence beliefs are epistemically independent of their existence beliefs. Moreover, if a serious essentialist holds reduction beliefs about a given domain

F, then her reduction beliefs will be inferred from her essence beliefs, rather than her existence beliefs. For the serious essentialist, in other words, the question of what a thing's essence is—and, consequently, the question of what sort of thing it is identical with—is to be answered independently of the question of whether that thing exists.

Serious essentialism comes in different strengths and varieties. The details need not detain us here. The important point, for our purposes, is just that serious essentialists do *not* engage in existence basing—their identity and reduction beliefs, in other words, are not based upon their existence beliefs. Instead, they engage in essence basing—their identity and reduction beliefs are based upon their essence beliefs. Moreover, their essence beliefs are epistemically independent of their existence beliefs.

Essence basers *can*—unlike existence basers—use their identity and reduction beliefs as defeater-deflectors to protect their existence beliefs from defeat due to EDAs. Or, at least, an essence basing reductionist who uses her identity and reduction beliefs as defeater-deflectors does not—unlike her existence basing counterpart—thereby violate the Independence constraint.

One way to see this is to recall the main difference between the good and bad defeater-deflector cases from Section I. An existence baser who uses her identity beliefs as defeater-deflectors is more like the bad case and less like the good case. For the existence baser's reduction beliefs are in part based upon her existence beliefs, which are the very beliefs her reduction beliefs are supposed to protect from defeat.

Meanwhile an essential baser who uses her identity beliefs as defeater-deflectors is more like the good case and less like the bad case. For her identity beliefs are based exclusively upon her essence beliefs, which are in turn not even in part based upon her existence beliefs. So her

reduction beliefs are available for use as defeater-deflectors to protect her existence beliefs from defeat due to debunking.²⁷

²⁷ [Acknowledgements removed for blind review]

Bibliography

- Audi, Robert (1997). *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. Routledge.
- Barker, Jonathan (2020). Debunking Arguments and Metaphysical Laws. *Philosophical Studies* 177 (7):1829-1855.
- Barkhausen, Max (2016). Reductionist Moral Realism and the Contingency of Moral Evolution. *Ethics* 126 (3):662-689.
- Bedke, Matthew S. (2009). Intuitive non-naturalism meets cosmic coincidence. *Pacific Philosophical Quarterly* 90 (2):188-209.
- Bedke, Matthew S. (2014). No Coincidence? *Oxford Studies in Metaethics* 9:102-125.
- Berry, Sharon (2020). Coincidence Avoidance and Formulating the Access Problem. *Canadian Journal of Philosophy*.
- Blackburn, Simon (1971). Moral realism. In John Casey (ed.), *Morality and Moral Reasoning*. Methuen.
- Bogardus, Tomas (2016). Only All Naturalists Should Worry About Only One Evolutionary Debunking Argument. *Ethics* 126 (3):636-661.
- Boyd, Richard (1988). How to be a Moral Realist. In G. Sayre-McCord (ed.), *Essays on Moral Realism*. Cornell University Press. pp. 181-228.
- Brink, David Owen (1989). *Moral Realism and the Foundations of Ethics*. Cambridge University Press.
- Christensen, David (2010). Higher Order Evidence. *Philosophy and Phenomenological Research* 81 (1):185-215.
- Copp, David (2008). Darwinian skepticism about moral realism. *Philosophical Issues* 18.
- Horgan, Terence E. (1993). From supervenience to superdupervenience: Meeting the demands of a material world. *Mind* 102 (408):555-86.
- Horgan, Terence & Timmons, Mark (1992). Troubles on moral twin earth: Moral queerness revived. *Synthese* 92 (2):221 - 260.
- Jackson, Frank (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford University Press.

- Joyce, Richard (2005). *The Evolution of Morality*. Bradford.
- Lasonen-Aarnio, Maria (2014). Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research* 88 (2):314-345.
- Maddy, Penelope (1990). *Realism in Mathematics*. Oxford University Press.
- Moon, Andrew (2017). Debunking Morality: Lessons from the EAAN Literature. *Pacific Philosophical Quarterly* 98 (S1):208-226.
- Moon, Andrew (2021). Circular and question-begging responses to religious disagreement and debunking arguments. *Philosophical Studies* 178 (3):785-809.
- Kahane, Guy (2011). Evolutionary Debunking Arguments. *Noûs* 45 (1):103-125.
- Kripke, Saul A. (1980). *Naming and Necessity*. Harvard University Press.
- Mackie, John Leslie (1977). *Ethics: Inventing Right and Wrong*. Penguin Books.
- McPherson, Tristram (2012). Ethical Non-Naturalism and the Metaphysics of Supervenience. In *Oxford Studies in Metaethics Vol 7*. pp. 205.
- Morton, Justin (2016). A New Evolutionary Debunking Argument Against Moral Realism. *Journal of the American Philosophical Association* 2 (2):233-253.
- Plantinga, Alvin (1993). *Warrant and Proper Function*. Oxford University Press.
- Plantinga, Alvin (2000). *Warranted Christian Belief*. Oxford University Press USA.
- Pollock, John L. (1987). Defeasible Reasoning. *Cognitive Science* 11 (4):481-518.
- Putnam, Hillary (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science* 7:131-193.
- Rosen, Gideon (2010). Metaphysical Dependence: Grounding and Reduction. In Bob Hale & Aviv Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*. Oxford University Press. pp. 109-36.
- Rosen, Gideon (2017). Metaphysical Relations in Metaethics. In Tristram McPherson & David Plunkett (eds.), *The Routledge Handbook of Metaethics*. Routledge. pp. 151-169.
- Shafer-Landau, Russ (2012). Evolutionary Debunking, Moral Realism and Moral Knowledge. *Journal of Ethics and Social Philosophy* 7 (1):1-38.

- Schaffer, Jonathan (2015). What Not to Multiply Without Necessity. *Australasian Journal of Philosophy* 93 (4):644-664.
- Sterelny, Kim & Fraser, Ben (2016). Evolution and Moral Realism. *British Journal for the Philosophy of Science* 68 (4):981-1006.
- Street, Sharon (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies* 127 (1):109-166.
- Street, Sharon (2008). Reply to Copp: Naturalism, normativity, and the varieties of realism worth worrying about. *Philosophical Issues* 18 (1):207-228.
- Tahko, Tuomas E. & Lowe, E. J. (2020). Ontological Dependence. *Stanford Encyclopedia of Philosophy*.
- Toner, Christopher (2011). Evolution, naturalism, and the worthwhile: A critique of Richard Joyce's evolutionary debunking of morality. *Metaphilosophy* 42 (4):520-546.
- Väyrynen, Pekka (2017). The Supervenience Challenge to Non-Naturalism. In Tristram McPherson & David Plunkett (eds.), *The Routledge Handbook of Metaethics*. Routledge. pp. 170-84.
- Wilkins John S., & Griffiths Paul E., (2012). Evolutionary debunking arguments in three domains: Fact, value, and religion. In James Maclaurin Greg Dawes (ed.), *A New Science of Religion*. Routledge.