Evolutionary Debunking and Moral Relativism

Daniel Z. Korman and Dustin Locke

Penultimate Draft
Forthcoming in The Routledge Handbook of Philosophy of Relativism, Martin Kusch (ed.)

Abstract
Our aim here is to explore the prospects of a relativist response to moral debunking arguments. We begin by clarifying the relativist thesis under consideration, and we explain why relativists seem well-positioned to resist the arguments in a way that avoids the drawbacks of existing responses. We then show that appearances are deceiving. At bottom, the relativist response is no less question-begging than standard realists responses, and—when we turn our attention to the strongest formulation of the debunking argument—the virtues of relativism turn out to be vices.

1. A Neglected Response to the Moral Debunking Argument

The moral debunking argument aims to show that a proper appreciation of the source of our moral beliefs threatens to undermine those beliefs. In broad brush strokes, the argument runs as follows:

D1. Our moral beliefs are not explained by the moral facts and we realize this
D2. If so, then we’re rationally committed to withholding our moral beliefs
D3. So, we’re rationally committed to withholding our moral beliefs

D1 can be motivated in a variety of ways. One common strategy is to appeal to evolutionary considerations and, in particular, the selective pressures that shaped our moral beliefs. The belief that it’s good to feed one’s children, for instance, can be accounted for in terms of
the adaptive value of attitudes that motivate one to care for one’s children; and the
(putative) moral fact that it truly is good to feed one’s children has no role to play in this
evolutionary explanation. In what follows, we’ll be restricting our attention to this way of
developing the debunking argument, though much of what we say won’t hang on how D1 is
motivated. The idea behind D2 is then that, in light of the explanatory disconnect between
the moral facts and the factors influencing our moral beliefs, it would be an extraordinary
coincidence if those beliefs turned out to be true, the realization of which should convince
us to withhold belief.¹

Numerous responses to the debunking argument can be found in the literature, including
(but not limited to) the following three. According to the naturalist response, the moral
facts do explain our moral beliefs, insofar as they are identical to certain of the mundane
natural facts that explain our moral beliefs. According to the constructivist response, the
explanation runs in the opposite direction: the moral facts are as they are because we have
the moral beliefs that we do, which is why it’s no coincidence that we have accurate moral
beliefs. According to the third-factor response, while the moral facts neither explain nor are
explained by our moral beliefs, what secures the noncoincidental accuracy of our moral
beliefs is some further fact (a “third factor”) that is responsible both for the moral beliefs
and for the moral facts.²

Each response has its drawbacks. Naturalist reductions tend to have profoundly
counterintuitive implications regarding the first-order moral truths; constructivists seem
committed to saying that we can change the facts about what’s right and wrong simply by
changing our minds about what’s right and wrong; and third-factor theorists seem to
unabashedly beg the question against the debunker. (More on each of these below.)

¹ Arguments of this sort can be found in Ruse (1985), Joyce (2006: ch. 6), and Street (2006).
The arguments are often wielded with a restricted scope, targeting only realists. We focus
on the unqualified formulation, so as to shed light on why certain sorts of antirealists—in
particular, moral relativists—are supposed to be able to resist the argument.
² For representative defenses, see Copp (2008), Street (2006: §10), and Enoch (2010)
respectively.
Our aim here is to explore the prospects of a relativist response to the debunking argument. We begin by clarifying the relativist thesis under consideration (§2), and we explain why relativists seem well-positioned to resist the debunking argument in a way that avoids the drawbacks of existing responses (§3). We then show that appearances are deceiving. At bottom, the relativist response is no less question-begging than the third-factor response (§4), and—when we turn our attention to the strongest formulation of the debunking argument—the virtues of relativism turn out to be vices.

2. Relativism

Relativism, as we will understand it here, is the thesis that (i) no moral propositions are true simpliciter and (ii) for some moral proposition p and moral frameworks F and F', p is true relative to F and false relative to F'. A moral framework may be understood simply as a set of moral propositions, and a moral proposition is true relative to a given framework just in case that proposition is a member of that framework.

Since moral propositions are true only relative to a framework, beliefs can be accurate only relative to a framework. Unless otherwise specified, when we say that S's belief is accurate, this should be understood to mean that the content of S's belief is true relative to S's framework. What is it then for a given framework to be S's framework. In other words, what is it to have a given framework? Following the lead of Gilbert Harman (1996: ch. 1.3), we won't assume that S's framework is invariably going to be the set containing all and only the propositions S in fact believes, since that would make it impossible to have inaccurate moral beliefs. While there are a variety of ways one might fill in the details, for the sake of concreteness we'll assume that S's framework is the set of propositions that S would believe upon reaching reflective equilibrium.3

3 There also are other ways for relativists to fill in the details of what truth, frameworks, and accuracy amount to, as well as other fine points about the semantics of moral attributions (see, e.g., Brogaard 2008). We leave it to others to explore how different ways of
One might naturally assume that relativists are committed to saying that right actions are right and wrong actions are wrong because we have the moral frameworks we do—for instance, that one ought to feed one’s children because we have a framework relative to which it’s true that one ought to feed one’s children. In fact, and perhaps surprisingly, they are committed to no such thing. Propositions of the form \( p \) because \( q \) are themselves moral propositions, and as such can be assessed for truth only relative to a framework. What’s true relative to our ordinary moral framework(s), the relativist will say, is *that one ought to feed one’s children because they will die without our assistance*. The proposition *that one ought to feed one’s children because we have the framework we do*, by contrast, comes out false relative to the framework of any normal person (relativists included). Relativists can therefore happily deny that it is anyone’s having a given belief or framework that makes moral propositions true. As we shall see, herein lies one of the distinctive advantages of relativism.

Relativism is often taken to be motivated by its explanation of fundamental moral disagreement. So long as the disputants have different moral frameworks, the relativist can maintain that neither party is mistaken (simpliciter); each is saying something true relative to his or her own framework. At the same time, the relativist can affirm that the disagreement is genuine: the one party is denying the very thing that the other is affirming. Relativism also makes it easy to see how moral knowledge is possible. Having accurate moral beliefs is just a matter of believing propositions that are in one’s own framework. This isn’t automatic—since one’s framework consists of the propositions one developing a relativist thesis might affect what we say below and, in particular, the relativist treatment of the debunking argument.

---

5 Likewise, while relativists may say that what makes a moral proposition \( p \) true relative to \( F \) is that \( p \) is in \( F \), that is not to say that \( p \) is made true by anyone’s having that framework.
would believe under certain idealized conditions—but it does put accurate beliefs within reach, without requiring anything like responsiveness to some transcendent realm of moral facts.

3. The Relativist Response to the Debunking Argument

Relativists seem to have a straightforward way out of the debunking argument, by denying the second premise:

D2. If we realize that our moral beliefs are not explained by the moral facts, then we’re rationally committed to withholding our moral beliefs

The usual motivation for this premise, again, turns on the idea that, if indeed the moral facts aren’t explaining our moral beliefs, then it would be a massive coincidence if our moral beliefs turned out to be accurate. Relativists will deny that any coincidence is required. For as we just saw, by relativist lights one’s moral beliefs are guaranteed to be accurate so long as they’re aligned with one’s own framework. And while there is room for moral error, it certainly isn’t a coincidence if our beliefs are largely aligned with our respective frameworks. Accordingly, it’s no coincidence if our moral beliefs are largely accurate, and that is so regardless of the source of our moral beliefs, even if it’s “blind” evolutionary forces aimed only at keeping us and ours alive.

Now let us see how the relativist response avoids the drawbacks of the other responses. The naturalist resists D1 by identifying moral facts with the very natural facts that explain our moral beliefs. But such reductions are plagued by apparent counterexamples.8 (You don’t need our help to see the nasty consequences of, for instance, identifying that fact that something is the right thing to do with fact that such actions tend to promote human flourishing.) Since relativists are able to put moral truths within our epistemic reach without

8 Not to mention the open question argument (Moore 1903) and the moral twin earth argument (Horgan and Timmons 1991).
requiring moral facts to explain our beliefs, there is no need for them to identify moral facts with any such (causally efficacious) natural fact.

Constructivists maintain that the moral facts are as they are because we have the moral beliefs that we do. For instance, a certain sort of Humean constructivist might say that, for any given person S, what makes it true that S is required to Φ is that S believes that she herself is required to Φ. Since explanations support counterfactuals, the constructivist is thereby committed to such “repugnant counterfactuals” as *if I were to stop believing that I’m required to feed my children, I wouldn’t be required to feed my children.* The relativist, as we saw above, need say no such thing. For, according to the relativist, the fact that a given person believes the proposition that she is required to feed her children, or has a framework that includes that proposition, is no part of the explanation of why it’s true (relative to her framework) that she is required to do so. Since the relativist can reject the constructivist’s (counterfactual-supporting) explanations, she isn’t saddled with the repugnant counterfactuals.10

Finally, there are the third-factor responses according to which some natural fact explains both the moral beliefs and the associated moral facts. The fact that feeding one’s children promotes their survival, for instance, may be cited as the third factor that figures both in the evolutionary explanation of why we believe it’s good to feed them and the explanation of why it is good to feed them. Or the fact that altruistic acts promote social cohesion may be cited as the third factor that explains both why we believe that we are required to act altruistically and why we are required to act altruistically. These lines of reasoning straightforwardly beg the question against the debunker, by relying on the moral belief that promoting survival or social cohesion are good-making features, which is precisely the sort ____________________________

9 More nuanced versions of constructivism are committed to more nuanced, but no less repugnant, repugnant counterfactuals.

10 At most, the relativist is committed to saying: *if I didn’t believe p, p wouldn’t be true relative to the framework I would then have had.* This, however, is not an affirmation of the repugnant counterfactual; rather, it just reports the triviality that p isn’t true relative to frameworks it isn’t in.
of belief that the debunker is calling into question. The relativist, by contrast, doesn’t rely on any specific first-order moral beliefs in her response to the debunker. What’s doing the work for the relativist is just the metaethical assumption that having accurate moral beliefs (whatever they may be) is simply a matter of having beliefs that align with one’s own moral framework.\textsuperscript{11}

If indeed the relativist response to the debunking argument avoids all of the shortcomings of existing responses, this could potentially serve as an important secondary motivation for relativism, alongside its explanation of moral disagreement. On closer inspection, however, matters are not so simple, and relativism turns out to be ill-equipped to handle a fortified version of the debunking argument.

4. Is the Relativist Response Question-Begging and Does it Matter?

The relativist response to the debunking argument, as we just observed, does not directly invoke any of the first-order moral beliefs that the debunker means to be calling into question. But the relativist may still be accused of begging the question indirectly. In particular, one might object that the rationale for embracing relativism relies on first-order moral beliefs.

To see this, consider the motivations for relativism canvassed above. One was that relativism yields the best explanation of fundamental moral disagreement, for instance when one person thinks it’s wrong to kill someone for dishonoring one’s family, whereas another thinks it’s the right thing to do. The relativist is able to explain how her belief that honor killings are wrong can be accurate without thereby privileging her own belief. After all, the proponent of honor killings also has an accurate belief (true relative to his framework).

Here, though, the relativist is assuming that among the things to be explained is the accuracy of the belief that honor killings are wrong. This is, she is assuming that honor

\textsuperscript{11} Cf. Street (2006: 163-164) on whether constructivism is question-begging.
killings are wrong. In assuming this, she is indirectly relying on her first-order moral beliefs. Without that assumption, the relativist’s explanation of moral disagreement is no improvement on the error theorist’s. The error theorist can give precisely the same explanation of why the disputants disagree, in terms of their having substantially different moral frameworks, but will say that both are mistaken, since nothing is right or wrong. We have reason to prefer the relativist’s account of moral disagreement to the error theorist’s only if we take for granted that our moral beliefs are accurate and that this is something a theory of moral disagreement ought to account for.

Likewise for the contention that relativism yields the best response to the debunking argument, insofar as it is able to explain (quite easily) the striking correlation between adaptive moral beliefs and accurate moral beliefs. In assuming that there is such a correlation to be explained, the relativist is taking for granted that adaptive moral beliefs—for instance that it is good to feed one’s children—are indeed accurate. So the rationale again ultimately takes for granted the accuracy of her first-order moral beliefs.

The relativist does help herself to her first-order moral beliefs (albeit indirectly), and is therefore no less open to the charge of question-begging than the third-factor theorists considered above. That said, we don’t think that there is anything so bad about relying on beliefs that one’s interlocutor has called into question. The reason, in short, is that the mere fact that someone has presented you with an argument against some belief of yours is not by itself enough to remove whatever entitlement you have to rely on that belief, and in particular doesn’t prevent you from relying on that belief when assessing whether you ought to accept the premises of that argument. Otherwise, anyone could be driven to global skepticism by the following inane argument: “Everything in the Harry Potter stories is true, and every proposition you believe is contradicted by something in the Harry Potter stories.”

5. The Explanatory Constraint

12 See our (forthcoming: §4) for more on charges of begging the question.
There is, however, a potentially more serious problem for relativists who help themselves to their moral beliefs in the face of the debunking argument. The problem isn’t that they beg the question, but rather that they concede that moral facts have no role to play in explaining our moral beliefs.

To see why such explanatory concessions are problematic, imagine that, after consulting your magic 8-ball toy, you form the belief that your crush likes you back. Now you learn, much to your disappointment, that facts about who has a crush on who play no role in determining the 8-ball’s outputs, and hence no role in determining what you believe on the basis of those outputs. Clearly, you should then withhold belief about whether your crush likes you back.

Extrapolating from this, the realization that you have some beliefs for reasons having nothing to do with the range of facts those beliefs purport to be about rationally compels you to abandon those beliefs. Getting the explanatory constraint exactly right is a delicate matter, but the following will do for our purposes here:

**The Explanatory Constraint:** If p is about domain D, and S believes that her belief that p isn't explained by D-facts, then S is thereby rationally committed to withholding belief that p.\(^{13}\)

The relativist response runs afoul of the explanatory constraint. After all, the relativist concedes that the moral facts have no role to play in explaining why she has the moral beliefs that she does, but persists in relying on those very beliefs in reasoning her way to relativism.

\(^{13}\) This formulation of the constraint has the flexibility to accommodate beliefs about the future. Your belief that the sun will rise tomorrow isn’t explained by the fact that it will, but it is explained by facts about sunrises (viz. past sunrises), thereby satisfying the constraint. Still, we are not entirely satisfied with this formulation. For some alternative formulations, see Setiya’s K (2013: 96), Locke’s CD (2014: 232), McCain’s EF (2014: §4.4 and 6.4), Lutz’s EAD (2017: §2), and Korman’s EC5 (2019: §8).
The explanatory constraint also has ramifications for our understanding of the debunking argument. Thus far, following the literature, we have motivated D2 in terms of coincidence: the realization that the moral facts don’t explain moral beliefs undermines those beliefs by way of revealing that moral accuracy would at best be coincidental. The explanatory constraint eliminates the middleman by directly entailing D2. To our mind the strongest defense of the debunking argument is one that makes no reference to coincidence and that reasons directly from an explanatory constraint to D2.

Relativists who wish to resist this fortified version of the argument must therefore reject the explanatory constraint. But they must do so in a way that does justice to the insight that explanatory concessions have the power to undermine one’s beliefs (as in the 8-ball case). We will explore two strategies that the relativist might pursue.

6. Loosening the Explanatory Constraint

The first strategy that relativists might pursue is to amend the explanatory constraint, permitting the explanatory arrows to run in either direction:

**The Bidirectional Explanatory Constraint.** If p is about domain D, and S believes that her belief that p isn’t explained by and doesn’t explain D-facts, then S is thereby rationally committed to withholding belief that p.

There is some reason to believe that the constraint needs to be loosened in this way. Imagine a dictatorship in which, if the dictator so much as believes that something is illegal, it is thereby illegal. Suppose the dictator one day forms the belief that it’s illegal to wear orange on Sundays. The fact that it’s illegal doesn’t explain his belief that it is. But he’s certainly rational in believing that it’s illegal. Plausibly, that’s because there’s an explanatory connection running in the other direction: his believing it’s illegal makes it illegal.
We’re not entirely convinced that the explanatory constraint needs loosening in order to handle this case. The dictator presumably believes it’s now illegal to wear orange partly on the basis of his belief that the laws are determined by his beliefs. So long as this latter belief is explained by the fact that his beliefs determine the laws (which is itself a fact about the laws), the dictator’s law beliefs satisfy the original explanatory constraint.

But let’s just grant that the explanatory constraint does need to be loosened in this way. Still, shifting to the bidirectional constraint helps the relativist only if she is willing to affirm that our beliefs explain the moral facts. But that would then undermine one of the main attractions of the relativist response to the debunking argument, namely that it permits one to avoid the drawbacks of constructivism. For as soon as the relativist affirms that moral truths are made true by our beliefs—that is, as soon as she embraces constructivism—she will be saddled with the constructivist’s repugnant counterfactuals (see §3).

Additionally, relativism and constructivism make for strange bedfellows. For suppose that the relativist does go in for constructivism. It will then be true relative to her framework that (e.g.) we are required to feed our children because we believe we’re required to feed our children. But, relative to the framework of any other normal person, it’s false that that’s why we are required to feed our children (see §2). This leaves us with the uncomfortable result that the truth of constructivist relativism would not be enough to save the masses from debunking. Rather, those exposed to the debunking argument remain rationally required to withhold moral belief until they accept constructivism. None of the other responses to the argument has this odd result. For instance, on a nonrelativist constructivist view—on which moral propositions are made true (simpliciter) by our adoption of frameworks—one doesn’t have to accept constructivism in order for one’s beliefs to be noncoincidentally accurate; it’s enough just that constructivism is correct.

7. The Modal Constraint
Relativists might instead insist that the explanatory constraint misdiagnoses what’s going on in the 8-ball case. It’s not the explanatory concession per se that undermines the beliefs, they might say. Rather, it’s that explanatory concessions often commit one to conceding that one could easily have ended up with inaccurate beliefs, and it’s this modal concession that does the defeating. Applied to the 8-ball case, the idea is that, when you realize that your crush beliefs aren’t explained by the crush facts, you realize that, in forming beliefs on the basis of the 8-ball’s outputs, you could easily have ended up with inaccurate crush beliefs. And it is this modal concession—that this could easily have happened—that rationally commits you to withholding belief about whether your crush likes you back.

This diagnosis of the 8-ball case opens up the possibility that explanatory concessions don’t always defeat. Rather, they defeat when, only when, and because they rationally commit one to believing that one could easily have been mistaken (or, in other words, that the belief is not “safe”). More precisely:

**The Modal Constraint.** If p is about domain D, S’s rational commitment to believing that [her belief that p is not explained by some D-facts] defeats her belief that p when, only when, and because she is thereby rationally committed to believing that she could easily have had an inaccurate belief about whether p.

If true, the modal constraint creates a toe-handle for relativists, for morality is arguably one of the domains where the explanatory concession doesn’t rationally commit one to the modal concession. After all, the relativist will reason, our moral beliefs are more or less bound to be accurate, regardless of what’s explaining them. For even had we had different beliefs, those beliefs would still have been true relative to the frameworks we would then have had. In this case, the explanatory concession does not force a modal concession.

Unfortunately for the relativist, the modal constraint is false. To see this, consider the following case (adapted from Locke 2014). Suppose Jack believes P:

**P.** Protons cause streaks of type S in cloud chambers.
Jack believes P, but not because he has received the training of an ordinary physics student. Rather, it’s because he asked some Martians (who had previously convinced him of their superior intellect) what causes those streaks, and they replied: protons. Later, however, Jack learns that the Martians told him that it’s protons that cause those streaks, not because they themselves had done any physics, but simply because they liked the sound of the word ‘proton’. You may even suppose, if you like, that there is some deep law of Martian psychology that makes them like the sound of ‘proton’, and so it could not easily have happened that the Martians told Jack that such streaks were caused by anything else.

Upon learning all of this, it is clearly irrational for Jack to stand by his belief that P. But suppose that, rather than abandoning the belief, Jack attempts to assure himself that it is still in good standing by reasoning as follows:

Yes, my belief that it’s protons that cause streaks like that is not explained by the facts about protons. Still, given what I have just learned about Martian psychology, I could not easily have formed a different belief about whether protons cause streaks like that. Moreover, protons do cause streaks like that in cloud chambers, and—since this interaction is surely underwritten by natural laws—it could not easily have failed to be the case that protons cause such streaks. Putting the pieces together: I could not easily have been mistaken about whether such streaks are caused by protons.

Clearly the reasoning is illicit. And the obvious diagnosis of why it is illicit is that the reasoning relies on the belief that P (see the italics), which has already been undermined by his explanatory concession that he believes that those streaks are caused by protons for reasons having nothing to do with whether they in fact are caused by protons—just as our explanatory constraint would have it. The modal constraint, by contrast, leaves us with no explanation of why Jack’s reliance on P is illicit. After all, just as in other cases of testimony from an otherwise reliable source, Jack is initially entitled to the testimonial belief that P.\(^\text{14}\)

---

\(^{14}\) Those who doubt that we have a default entitlement to such testimonial beliefs may instead suppose that Jack was brainwashed to have an intuition or disposition to believe that
And if the modal constraint is correct, then Jack’s entitlement to P is undermined only once he is driven to the conclusion that his belief could easily have been false. Thus, by the lights of the modal constraint, there should be nothing illicit about relying on his not-yet-impugned belief in P when assessing whether P satisfies this modal condition.

In short: The Jack case shows that when explanatory concessions defeat, it isn’t always by way of committing one to a modal concession. So the modal constraint is mistaken, and the relativist is again without any viable replacement for the explanatory constraint.

8. Relativism Naturalized?

We have been assuming that the relativist will grant that moral beliefs are not explained by moral facts. There is, however, a different route open to the relativist, which is to follow the naturalist in identifying moral facts with some of the natural facts that explain our moral beliefs. Exploring this route requires taking a closer looks at what “moral facts” would amount to in a relativist setting and how the reference to moral facts in the explanatory constraint is to be understood.

Relativists will of course regard moral facts as entities that obtain only relative to a framework—lest it end up being an absolute fact that it’s good to feed one’s children or that honor killings are wrong. For concreteness, let’s assume they say that a moral fact that p obtains relative to S if and only if p is a member of S’s framework. Relativists may then insist upon understanding the explanatory constraint as follows:

**The Relativistic Explanatory Constraint.** If p is about domain D, and S believes that her belief that p isn’t explained by D-facts that obtain relative to S’s framework, then S is thereby rationally committed to withholding belief that p.

protons cause those streaks, or to have experiences that (richly) present those streaks as caused by protons.
To show that this explanatory constraint is satisfied, the relativist will need to find natural facts that explain our moral beliefs with which to identify the moral facts that obtain relative to our frameworks. For instance, she may identify the moral fact that it’s good to feed one’s children with certain psychological facts about us, for instance that we are strongly disposed to have positive attitudes towards feeding one’s children. Or she may identify it with the fact that feeding one’s children tends to promote human flourishing. Since such natural facts are indeed part of what explains why we believe that p, the relativist seems able to satisfy the relativistic explanatory constraint.

There are two potential problems with this strategy. First, the relativist response will now inherit the drawbacks of the naturalist response alluded to above, for instance that candidate reductions of moral facts are plagued by apparent counterexamples. Second, it would seem that the relativist cannot identify moral facts with natural facts—even natural facts about people’s evaluative dispositions. Suppose, for the sake of illustration, that the moral fact that honor killings are wrong which obtains relative to S is identical to the natural fact that S is strongly disposed to have a negative attitude towards honor killings. The trouble is that, by relativist lights, the former fact obtains relative to some frameworks but not others. So, by Leibniz’s Law, the latter fact also must obtain relative to some frameworks but not others. Plainly though, the latter fact, if it obtains at all, obtains simpliciter (or relative to all frameworks). So, on the envisaged relativist conception of moral facts, the fact that honor killings are wrong cannot be identified with this or any other natural, framework-invariant fact.

9. Conclusion

We have been exploring a relativist response to the debunking argument. At first glance, the relativist response seemed to enjoy a number of advantages over existing responses. It seems able to secure the noncoincidental accuracy of our moral beliefs without commitment to repugnant counterfactuals, without directly begging the question against the debunker, and without controversial reductions of the moral to the natural.
However, when the debunking argument is properly understood—as underwritten directly by an explanatory constraint—the advantages of the relativist response vanish. Like third-factor theorists, relativists rely on their antecedent moral beliefs (albeit indirectly) in the face of the debunking argument, which in turn are jeopardized by their explanatory concessions. They are able to accommodate a loosened version of the explanatory constraint only by collapsing into constructivism and inheriting all of its problems. And relativism, by its very nature, prevents them from accommodating the explanatory constraint by embracing a naturalistic reduction of the moral to the natural.

Acknowledgments
We would like to thank Charity Anderson, Sinan Dogramaci, Kenny Easwaran, Justin Fisher, Amy Flowerree, William Roche, Josh Thurow, and Joel Velasco for valuable discussion.

References


