13 Debunking Arguments in Metaethics and Metaphysics

Daniel Z. Korman

13.1 INTRODUCTION

A central project of cognitive science is to explain, at a variety of different levels, certain types of mental representations and the cognitive processes responsible for them. What is the evolutionary purpose (if any) of the representations? What sorts of information are the processes operating on to yield the representations? How are the processes and representations implemented in the brain?[1]

In some cases, answers to these questions presuppose that the representations are more or less accurate. The explanation of mental states that represent regions of space as empty or matter-filled is bound to cite the fact that successfully navigating the world involves accurately tracking which regions are and aren't empty, and explanations for why in a particular instance a region is represented as empty will typically cite the fact that it is empty. Likewise for representations of the relative distances of filled regions, and (one might naturally think) for visual representations of three-dimensional midsized objects.

In other cases, it seems quite possible to answer the above questions without presupposing the accuracy of the representations. Cognitive scientists have developed various accounts of the production and adaptive value of theistic belief that in no way presuppose the existence of deities. Color perception can likely be explained in terms of the adaptive value of easily detecting ripe fruit or young foliage, by means of detecting wavelengths in the visible spectrum, and without presupposing that the phenomenal colors represented are actually out there on the surfaces of objects. The processes underwriting our moral beliefs and intuitions can presumably be explained in terms of the adaptive value of detecting certain benefits and dangers to

---

[1] These correspond roughly to Marr's 1982 three levels of analysis. Needless to say, not all traits will have an evolutionary purpose or adaptive value. Some, for instance, are byproducts ("spandrels") of adaptations; see Gould and Lewontin 1979 for discussion.

337

ourselves and our kin, and without any reference to a realm of moral facts that the processes are supposed to be tracking.[2]

Suppose that a range of representations can be explained without reference to the putative objects, features, or facts represented. Does the availability of such explanations *debunk* those representations, undermining their justificatory status? This is an epistemological question, one whose answer can be informed by, but not directly read off of, findings in cognitive science. There has been some discussion in metaphysics of such explanations, and there have been attempts to wield them with debunking intent. But such discussions tend to be carried out in isolation from the extensive and nuanced literature on explanatory challenges and debunking arguments in metaethics. My aim here is to forge some connections, drawing lessons from the moral-debunking literature for the metaphysical-debunking literature and vice versa.

As a case study, I'll focus on material-object metaphysics and our beliefs about which midsized objects there are, though lessons learned generalize to explanatory challenges that arise in connection with such other metaphysical domains as color, causation, time, modality, and persistence.[3] In sections 13.3–13.4, I examine and combat the idea that, because our beliefs about midsized objects can be given an "evolutionary vindication," evolutionary debunking arguments won't arise for these beliefs. In section 13.5, I defend an alternative strategy for vindicating them, which involves bootstrapping from the very beliefs targeted by the debunking arguments. In section 13.6, I explain why debunking arguments framed in terms of the insensitivity of targeted beliefs are ineffective against bootstrappers. Finally, in sections 13.7–13.8, I develop and defend an explanatory constraint on bootstrapping, and I identify some existing responses to the debunking arguments that run afoul of the constraint.

13.2 THE DEBUNKING ARGUMENTS

You turn the corner and see some kids setting fire to a cat.[4] You immediately form the belief that they're doing something wrong. You feel inclined to believe this because it looks to you or seems to you intuitively that they are doing something wrong. I'll use the term "moral reactions" to cover this whole range of mental states: the experiences, intuitions, felt inclinations, and beliefs.

---

[2] See Leech and Visala 2011 on religious belief. See Sumner and Mollon 2000 on the evolution of color vision, and see Schaffer 2016, sec. 2, on the prospects of using the cognitive science of color vision to debunk color beliefs. Much more on moral belief later in the chapter.

[3] See, e.g., Goldman 1992, chaps. 2–3; Stroud 2000; Rea 2002; Chalmers 2006, sec. 6; Paul 2010, 2016; Pautz 2011, sec. 3; Leslie 2013; Gow 2014, sec. 6; Baron et al. 2015; Benovsky 2015; Crisp 2016; Schaffer 2016; and Baron 2017.

[4] The example is from Harman 1977, 4.

Moral-debunking arguments attempt to discredit moral reactions by exposing an explanatory disconnect between the reactions and the associated facts. Here, in its simplest form, is a moral-debunking argument:

(M1) Your moral reactions are not explained by the moral facts nor are the moral facts determined by your moral reactions.

(M2) If so, then you are not justified in retaining your moral beliefs.

(M3) So, you are not justified in retaining your moral beliefs.

Some may want to see the argument restricted in one way or another, so that it targets only realists, or only cognitivists, or only nonnaturalists, or only naturalists.[5] But there is no danger in sticking with this simple version of the argument, so long as we bear in mind that some philosophers—perhaps because they are antirealists, or noncognitivists, or naturalists, or nonnaturalists—may be in a position to wield the argument against their intended targets without opening themselves up to a *tu quoque*. The simple argument also nicely structures the moral-debunking debate, separating nonskeptics into two categories: those who deny M1, affirming the indicated sort of explanatory connection between the facts and the beliefs, and those who deny M2, rejecting the need for any such connection.

There are different routes to M1, and one common strategy involves pointing to an evolutionary explanation of our moral reactions.[6] Those actions that we represent as right or obligatory are ones that tend to keep us and ours alive; representing them as right or obligatory better motivates us to perform them; and these moral reactions now prevail precisely because they enhanced our ancestors' reproductive success. Mutatis mutandis for actions represented as wrong or forbidden.

The idea behind M2 is that if moral facts aren't explaining our moral reactions (or vice versa), then it would take a massive coincidence to end up with beliefs that line up with the facts. Since we have no grounds for believing that such a coincidence occurred, we shouldn't believe that it did, in which case we should not go on thinking that our beliefs line up with the facts.

There is also going to be an evolutionary explanation of the perceptual and cognitive mechanisms that convert the "blooming buzzing confusion" in the retinal image into representations of three-dimensional midsized objects. Accordingly, we ought to consider the prospects of a parallel debunking argument against our "object

---

[5] See Street 2006; Kitcher 2005, 175–176; Joyce 2006, chap. 6; and Bogardus 2016 respectively. See Bedke 2009, 200–201 and Locke 2014, 224 n. 8 on restricting the argument to those who have actually encountered the relevant explanatory challenges; cf. White 2010, 575, on undermining and blocking debunkers.

[6] Evolutionary moral-debunking arguments have been advanced in one form or another by Ruse 1986, chap. 3; Lillehammer 2003; Kitcher 2005, sec. 4; Joyce 2006; Street 2006; Wilkins and Griffiths 2012, sec. 2; Fraser 2014; and Braddock 2016, 2017.

reactions": beliefs and associated experiences and intuitions concerning which midsized objects there are.

(O1) Your object reactions are not explained by the object facts, nor are the object facts determined by your object reactions.

(O2) If so, then you are not justified in retaining your object beliefs.

(O3) So, you are not justified in retaining your object beliefs.

What was said above in defense of M1 seems to apply equally to O1: our object reactions, like our moral reactions, can be explained by the adaptive value of an ability to represent certain kinds of midsized objects. And like M2, O2 can be motivated by the observation that once you acknowledge the explanatory disconnect, you'd have to admit that it would be a massive coincidence if our beliefs lined up with the facts.

In metaethics, there is near consensus that the O-argument is a complete non-starter. In metaphysics, by contrast, arguments along these lines are deployed constantly against common-sense judgments about how the world divides up into midsized objects.[7] I will argue that, while parties to the object debates have been somewhat overzealous about the prospects of such arguments, they are right to take them seriously. But first let us see why the O-argument is summarily dismissed in the literature on moral debunking.

13.3 THE EVOLUTIONARY VINDICATION

As indicated at the outset, the evolutionary explanation of a range of representations sometimes will and sometimes won't presuppose the accuracy of those representations. Accordingly, in order to get from an evolutionary explanation to the crucial first premise of the debunking argument—that there is no explanatory connection between the reactions and the associated facts—one needs a premise to the effect that the facts do not enter into the evolutionary explanation. To wit:

(M*) The adaptive value of your moral reactions is not explained by the moral facts.

(O*) The adaptive value of your object reactions is not explained by the object facts.

---

[7] As with the moral arguments, such arguments are sometimes advanced, not with the aim of establishing O3, but rather in support some departure from common sense (e.g., antirealism or mereological universalism) that would enable one to resist the skeptical conclusion. For arguments in the vicinity, see Heller 1990, 41–42; Goldman 1992, 45–46; Merricks 2001, 72–76; Hudson 2001, sec. 3.8; Sider 2001, 156–57; Rea 2002; Hawthorne 2006, vi and 109; Benovsky 2015, sec. 2; Sattig 2015, 25–26; Osborne 2016; Schaffer 2016, sec. 3.2; and Rose and Schaffer 2017. For some discussion of object-debunking arguments within cognitive science, see Hoffman et al. 2015 and Hummel 2015.

And herein, metaethicists are wont to claim, lies the difference between the two arguments: M* is plausible and O* is implausible.[8]

Why is M* supposed to be plausible? Following Street (2006, sec. 6), let's compare two accounts of the adaptive value of our moral reactions: the adaptive link account (ALA) and the tracking account (TA):

(ALA) Representing actions $a_1 \ldots a_n$ as morally right (or wrong) enhances reproductive success because representing $a_1 \ldots a_n$ as such motivates us to perform (avoid) them and performing them enhances (inhibits) reproductive success.

(TA) Representing actions $a_1 \ldots a_n$ as morally right (or wrong) enhances reproductive success because $a_1 \ldots a_n$ really are morally right (wrong).

ALA seems preferable to TA for a number of reasons. First, it is more parsimonious than TA, insofar as the latter invokes a domain of moral truths. Second, ALA is illuminating: it makes good sense of why beings with lives like ours—bearing relatively few offspring and needing to cooperate with neighbors—would have the particular kinds of moral reactions that we do. Third, TA is obscure: it is entirely unclear how, for instance, the putative fact that protecting our children *really is* obligatory is supposed to explain, even in part, why believing it's obligatory helps keep us and ours alive. For these reasons, Street suggests, we should reject TA and accept M*.

Why is O* supposed to be implausible? Because it is hard to see how the adaptive value of our object reactions could be explained without supposing them to be accurate. It's adaptive to represent the world in terms of tigers and cliffs, the idea goes, only because there are tigers and cliffs. As Street (2006, 160–61) puts the point:

In order to explain why it proved advantageous to form judgements about the presence of fires, predators, and cliffs, one will need to posit in one's best explanation that there *were indeed* fires, predators, and cliffs, which it proved quite useful to be aware of, given that one could be burned by them, eaten by them, or could plummet over them.

It is not as if there is some more parsimonious, tiger-free explanation of the adaptive value of tiger experiences! Thus, the idea goes, we should reject O*. And since the object facts do ultimately get in on the explanatory action, insofar as they explain why our object reactions are adaptive, O1 of the object-debunking argument is false. The full evolutionary story vindicates our object reactions.

---

[8] See Gibbard 2003, 253–58; Lillehammer 2003, 570–71; Joyce 2006, 182; Street 2006, 130 and 160–61; 2008, 217; Bedke 2009, 202; Enoch 2010, 436; Schafer 2010, 472; White 2010, 587; Wielenberg 2010, 441 n. 4; Dreier 2012, 277; Wilkins and Griffiths 2012, 134–39; Fitzpatrick 2015, 885; and Vavova 2015, 108; for a notable exception, see Clarke-Doane 2015, sec. 4. Alternatively, one might point to differences in *proximate* explanations: midsized objects can cause object reactions, whereas moral objects (properties, facts) are causally inert. See my 2014, sec. 5; 2015, chap. 7.4, on the inadequacy of this sort of response.

Or so it would seem. As we are about to see, there is more to O*—and less to the putative difference between the two debunking arguments—than meets the eye.

13.4 NO EVOLUTIONARY VINDICATION

The envisaged evolutionary vindication of object reactions turns on the idea that we need to assume the accuracy of our object reactions in order to explain their adaptive value. But that's just a mistake. In order to explain the adaptive value of our object reactions, it's enough to suppose that we are reliably tracking what nourishes us, endangers us, or can otherwise feature in adaptive or maladaptive interactions. But reliably tracking such things is entirely compatible with certain ways of misrepresenting which midsized objects there are. In particular, misrepresenting the world of midsized objects will still be adaptive so long as we *reliably* misrepresent it.[9]

To see what I have in mind, let's return briefly to what was supposed to be a par-adigm case of representations whose adaptive value presupposes their accuracy: our representations of regions as empty or matter-filled. Not so fast. Even the regions we visually represent as empty are typically filled by gaseous matter. So we misrepresent those regions. Nevertheless, the misrepresentations reliably track the difference that matters to successful navigation, namely, which are and aren't filled by nongaseous matter. They reliably misrepresent such regions as empty, and they are adaptive de-spite the harmless misrepresentation.[10]

The case for moral debunking is also usefully framed in terms of reliable misrepre-sentation. The error theorist will say that our moral reactions—which divide actions into the right and the wrong and the obligatory and the forbidden—misrepresent those actions but are nevertheless tracking something important. Those represented as right or obligatory tend to promote reproductive success. Those represented as wrong or forbidden tend to inhibit reproductive success. The moral reactions can be adaptive, despite misrepresenting, precisely because of what they reliably track. Indeed, the error theorist might add, misrepresenting actions by imbuing them with moral features is superior (evolutionarily) to accurately representing them merely as beneficial or detrimental to reproductive success, since the morally charged representations better motivate us to choose the beneficial actions.[11]

Now back to objects. Let's suppose (for the moment) that our own object reactions are generally accurate, and let's contemplate two different ways in which a creature might have erroneous object reactions. First, it might misrepresent the regions of space filled by midsized objects as containing no solid matter at all. Second, it might

---

[9] I borrow this useful notion from Mendelovici 2013, 2016.

[10] Perhaps you'll think that the sort of "emptiness" we represent the regions as having is compatible with their being gas-filled—just as a refrigerator can be empty, in a sense, despite containing shelves and drawers. Fair enough. My point can instead be made with hypothetical beings whose emptiness representations are more demanding.

[11] Cf. Joyce 2005 on the utility of moral fictions.

misrepresent how a situation divides up into objects, but without misrepresenting how solid matter is distributed. For instance, there could be creatures—let's call them "troglodytes"—who, upon encountering a situation containing an ordinary dog and tree, believe that there is a dog and a trunk and some branches, but do not believe that there is any further object composed of the trunk and the branches. Let's imagine that they don't even experience the region the tree occupies as being filled by a single object, despite experiencing the region as being filled with solid matter. Odder still, they represent the scattered region jointly occupied by the dog and the trunk as containing a single object. It seems to them that they see a *trog*: a partly wooden and partly furry object composed of the dog and trunk. This would be a misrepresentation; there are no trogs.[12]

Importantly, we don't have to imagine the troglodytes to be under any misconceptions about how stuff moves around. They know that the branches are firmly attached to the trunk and don't move (much) independently from one another, and that the dog and trunk aren't attached and won't move together. When there's a particularly vicious dog, they run only from the dog, not from the tree. Or, as they might put it, it's only the furry part of the trog—not its wooden part—that you have to worry about. Just like us, these troglodytes are tracking the presence, changes, and movements of matter that are relevant to their survival and reproductive success. They reliably misrepresent the world of midsized objects.

Imagining such ways of misrepresenting the object facts points to easily overlooked ways in which our own object reactions may be inaccurate. Perhaps we're wrong and the troglodytes are right: there are trogs but no trees. Or perhaps we're both wrong: there are only the atoms arranged dogwise and atoms arranged treewise, but no midsized objects composed of the atoms.[13]

With all this in mind, let us return to the question of whether it is possible to explain the adaptive value of our object reactions without supposing those reactions to be accurate. Of course it is. When we seem to see a tree falling towards us, we step out of the way and successfully avoid a dangerous situation. To explain why seeming to see a tree is adaptive in such dangerous situations, there is no need to suppose that they contain falling trees, as opposed to just some branches and the wooden part of a trog, or just some atoms arranged treewise. What's important for survival and reproductive success is that we respond appropriately to different distributions of matter—pursuing those that tend to nourish us, avoiding those that tend to harm us, and anticipating the behaviors of each. What this requires is that we have representations that reliably track the dangerous and nourishing stuff, and

---

[12] Or so I shall assume throughout. See my 2015 for a defense of a trogless ontology, and see in particular chapter 7.3 for an argument that postulating trogs and other such extraordinary objects won't help with the debunking arguments.

[13] I am taking for granted that the atoms arranged treewise are not themselves (identical to) a tree. See Merricks 2001, chap. 1 and my 2015, sec. 3.2.2, on the complaint that the very idea that there are atoms arranged dogwise but no dogs is somehow incoherent.

it is neither here nor there whether they accurately represent how the dangerous stuff being tracked divides up into objects. Reliable misrepresentation is up to the explanatory task.

Not only is the assumption of accuracy not *needed* to explain the adaptive value of our object reactions, it's hard to see what an assumption of accuracy *could* add to an explanation of the adaptive value of our object reactions. To see this, compare the following analogues of Street's adaptive link account and tracking account:

(ALA_O) Perceptually and conceptually dividing perceived distributions of matter into midsized objects of kinds $k_1 \ldots k_n$ enhances reproductive success because doing so better enables us to track and anticipate the movements of matter that are relevant to our reproductive success.

(TA_O) Perceptually and conceptually dividing perceived distributions of matter into midsized objects of kinds $k_1 \ldots k_n$ enhances reproductive success because the matter really does divide into objects of kinds $k_1 \ldots k_n$, whose movements we are thereby able to track and anticipate.

It would seem that all of the reasons from section 13.3 for preferring ALA to TA are equally reasons for preferring ALA_O to TA_O. First, TA_O looks to be less parsimonious: it is committed to the existence of midsized objects, whereas ALA_O incurs no such commitment.[14] Second, ALA_O makes good sense of why beings with lives and environments like ours—which give us no reason to track dogs and trunks as one—would wind up with the particular sortal concepts we do, *tree* and not *trog*.[15] Third, TA_O is obscure: it is entirely unclear how the putative fact that matter really does divide up into trees and the like is supposed to help explain why tracking matter under the kind *tree* is selectively advantageous.

One way to see this last point is to imagine discovering that some of our ancestors *were* troglodytes, but proved to be less fit than those of our ancestors who divided up the world into trees and other such familiar kinds.[16] What could explain why they were less fit? One promising explanation is that successfully tracking matter under sortals like *trog* rather than *tree* required more cognitive resources or slowed processing. But in that case, it's not the inaccuracy of the object reactions that explains why they were less fit but rather their cognitive limitations. It's hard even to make sense of the idea that dividing the successfully tracked matter into trogs is less

---

[14] While there is room for doubt about whether postulating composites in addition to simples renders a theory less parsimonious (cf. Schaffer 2015), requiring the composites to do critical explanatory work does seem to reduce the simplicity of the theory. In any event, ALA_O is no *less* parsimonious than TA_O. Thanks to Elanor Taylor for helpful discussion.

[15] Though see Hirsch 1993, chap. 5, on some difficulties for explaining why strange ways of dividing the world would in principle be maladaptive.

[16] Another way to see it is to imagine that *we* have it wrong and that in fact there are trogs but no trees. In that case, would troglodytes thereby enjoy a selective advantage? It's hard to see why.

adaptive or less efficient than dividing successfully tracked matter into trees because there really are trees but no trogs.

For these reasons, the evolutionary vindication of object reactions is unconvincing. One does not need to assume that object reactions are accurate in order to explain their adaptive value. The adaptive value of our object reactions can be explained—and, it would seem, explained just as well (if not better)—on the assumption that they reliably misrepresent how distributions of matter divide up into midsized objects.

13.5 BOOTSTRAPPING TO VINDICATION

Our object reactions cannot be vindicated on the grounds that their evolutionary explanations make indispensable reference to their accuracy. They don't. There is, however, a more promising way of vindicating our object reactions.

Even when some phenomenon can be explained without reference to a certain range of entities or facts, we may still sometimes have good reason to prefer a less parsimonious explanation that does make reference to them. Suppose, for instance, that I observe the creation of a xion in a particle accelerator. This is typically the result of a single yion colliding with a single zion, though it could in principle result from two yions colliding with a zion and overdetermining the creation of the xion. Which explanation should I prefer? If all I observe is the effect, the newly created xion, it would be unreasonable to prefer the latter explanation. But if I actually observed the two yions simultaneously colliding with the zion, then given my total evidence I should prefer the second explanation. This is so even though the second is a less parsimonious explanation of the appearance of the xion considered in isolation.

Similarly, my object reactions can either be explained simply by reference to the atoms arranged midsized-objectwise or, less parsimoniously, by both the atoms and the midsized objects they compose. Which explanation should I prefer? If my observations include only the effects, the object-representing mental states, then it would be unreasonable to adopt the latter explanation. But this isn't all that I have observed. I have observed the midsized objects themselves. I have seen dogs and trees and even a few tigers, and I can see that they have volumes and masses and everything else it takes to be causally interacting with things. So, given my total evidence—which includes the noninferential, direct perceptual evidence that there are dogs and trees—I should prefer the second explanation, on which my object reactions are caused by midsized objects. This is so even though this is a less parsimonious explanation of my object reactions (the mental states) considered in isolation.

There is, to be sure, a glaring disanalogy between the two cases. In the xion case, I presumably have independent evidence of the reliability of whatever equipment is informing me about the number and behavior of the yions. But when it comes

to my object reactions, I have no independent evidence of their reliability, no independent means of confirming that there really are trees out there—not just trogs and branches, or just atoms arranged treewise—answering to my tree experiences. And yet, despite lacking any independent evidence for the reliability of my object reactions, here I am reasoning from those reactions to the conclusion that midsized objects are causing my experiences.

To put a label on it, let's say that one *bootstraps* from the deliverances of some source of information when one reasons from the deliverances of that source to a vindication of that source, while having no independent evidence of its reliability or of the accuracy of those particular deliverances. What makes the object-reaction reasoning different from xion-creation reasoning is the bootstrapping.

There is something unsettling about bootstrapping. It reeks of "question-begging." And yet bootstrapping is rampant in the moral-debunking literature. Moral realists almost invariably respond to debunking arguments by reasoning from their antecedent moral beliefs—the very beliefs that the debunker is calling into question—to the conclusion that evolution is more or less bound to lead us to accurate moral beliefs. For instance, one can find Robert Nozick and David Enoch advancing something like the following response: It's no accident that we evolved to believe that protecting our children is good and obligatory, since such moral beliefs motivate us to perform actions that promote inclusive fitness. Moreover, protecting our children *is* good and obligatory (there's the bootstrapping). So, it's no accident that we have accurate moral beliefs.[17]

Indeed, consider what Street herself has to say on behalf of her antirealist response to the moral-debunking arguments:

[O]f course it is no coincidence that there is such a striking overlap between the content of evaluative truths and the content that natural selection would have tended to push us toward. Of course it's no coincidence that, say, breaking one's bones *is* bad and that's also exactly what evolutionary theory would have predicted we think. But whereas the realist is forced to offer the scientifically unacceptable tracking explanation of this overlap, the antirealist is able to give a very different account. (2006, 153)

In other words, the antirealist thesis that moral reactions determine the moral facts best explains the striking overlap between adaptive moral beliefs and accurate moral beliefs. What, though, is her reason for thinking that they overlap? Evidently it's that believing bone-breaking to be bad is adaptive *and it's true that it's bad*; believing we have obligations to our children is adaptive *and it's true that we do*; and so forth. What reason is there, though, to believe that it's true that bone-breaking is bad or that we have obligations to our children? Street is silent on the matter, presumably

---

[17] See Nozick 1981, 342–48, and Enoch 2010, secs. 3–5; 2011, chap. 7. Other bootstrappers include Copp 2008, Schafer 2010, Wielenberg 2010, Skarsaune 2011, and Clarke-Doane 2015, 2016.

because she takes herself to be stating the obvious—the obvious deliverances of her moral reactions. She is bootstrapping from these reactions to an antirealist vindication of her moral reactions.

How can metaethicists be so complacent about bootstrapping? The short answer (which is all that we will have space for here) is that, as incredibly tempting as it may be to dismiss bootstrapping as question-begging, one of the lessons of contemporary epistemology is that we must resist temptation on pain of sweeping skeptical results.[18] For instance, one might complain that the bootstrapping is question-begging because the bootstrapper is relying on beliefs that the debunker does not agree are true. But, if we really are prohibited from relying on beliefs that one's skeptical interlocutors find suspect, then merely *encountering* a skeptic who insists that none of your beliefs are justified would be enough to render you unable to escape their arguments, however bad those arguments may be. So we had better not accept any such evidence neutrality constraint.

Alternatively, one might complain that the bootstrapping is question-begging because the bootstrapper relies on her moral or perceptual faculties without any independent evidence of their reliability. But requiring that one always have independent evidence of the reliability of a source of information before relying on it leads straight to a fairly obvious regress. Each source would have to be validated by some further source, which would have to be validated by another source, ad infinitum. Without infinitely many distinct sources to draw on, the chain of sources must either form a loop, with some source ultimately validating itself, or the chain must come to an end, with an unvalidated source ultimately validating all the others. If relying on self-validating and unvalidated sources is forbidden—as it would be if we require all sources to be independently validated—then we are stuck with the intolerable skeptical conclusion that *no* sources can be relied upon. Thus, on pain of global skepticism, one must allow that there are some basic sources to which we are simply entitled to help ourselves, even without independent evidence of their accuracy. And, once one recognizes the need for this, it is hard to see what would be wrong with believing what follows deductively, inductively, or abductively from the deliverances of those basic sources—including when what follows is some vindication of the sources themselves (e.g., that they are reliable).

I obviously don't expect to have won many over by these brief remarks in defense of bootstrapping. Still, I'm right about this, and I will proceed on the assumption that *if* there's something wrong with bootstrapping vindications of object reactions and moral reactions, it's not merely that they involve bootstrapping.[19] That said,

---

[18] For the long answer, see Nozick 1981, 197–98; Alston 1986; Bonjour 1998, sec. 5.5; Huemer 2001, sec. 8.2; Van Cleve 2003; Bergmann 2004; Pryor 2004; Feldman 2005, 95; Williamson 2007, sec. 7.3; Kelly 2008; Vogel 2008; Titelbaum 2010; and Alexander 2011 on the skeptical ramifications of universal demands for evidence neutrality, independent evidence, and/or noncircularity.

[19] Those familiar with this literature might wonder whether one can embrace bootstrapping as I have without also licensing Roxanne's track-record reasoning from Vogel 2000. The answer is yes; see Bergmann 2004 and Vogel's more bootstrapping-friendly 2008 for two promising approaches.

whatever entitlement we have to trust these sources, and whatever justification we have for beliefs formed on their basis, is surely defeasible. The remainder of the paper is devoted to examining what might be able to defeat these beliefs. We'll begin with the suggestion that the beliefs are defeated once they are shown to be *insensitive* to the facts in question.

13.6 SENSITIVITY
13.6.1 Merricks's Overdetermination Argument

Trenton Merricks advances an overdetermination argument against midsized inanimate objects.[20] Any phenomenon that we might ordinarily point to as being caused by midsized inanimates is already being caused by the atoms that putatively compose it. Accordingly, the idea goes, we should deny that the midsized inanimates exist, on pain of implausibly and unparsimoniously taking such phenomena to be systematically overdetermined by midsized objects and pluralities of atoms.

Merricks anticipates the bootstrapping response advanced in the previous section, that I am entitled to believe in midsized objects because I can see them and see that they have everything it takes to enter into causal relations. In response, Merricks argues that any perceptual reasons we have for believing in midsized objects are undermined as soon as we acknowledge that such objects would be mere overdeterminers.

How? "[E]ven if there were no statues," he says, "our perceptions of 'each statue' would be the same, caused by atoms arranged statuewise working in concert" (2003, 739). And this realization is meant to act as a defeater for our beliefs:

[O]nce a perceiver realizes that her perceptual experiences of statues would be the same whether or not statues existed, it seems clear that those experiences no longer justify her belief in statues (in addition to atoms arranged statuewise). It seems clear that this realization "defeats" any justification that might have come from seeing statues. (2003, 739–40)

Extrapolating somewhat, we get the following argument:

(S1) You would have believed that there are statues even if it were false that there are statues, and you realize this.

(S2) If you realize that you would have believed p even if p were false, then you are not justified in believing p.

(S3) So, you are not justified in believing that there are statues.[21]

---

[20] See Merricks 2001, chap. 3; 2003. See his 2001, chap. 4, on why some composites are spared from elimination.

[21] Merricks never *explicitly* endorses anything as general as S2. It may be that the argument would be more faithfully rendered as an argument from analogy, turning on the premise that there is no

S1 is meant to be supported by the observation that statues (if they exist) are mere overdeterminers. Since the presence and activities of the atoms arranged statuewise is already enough to produce the experience as of a statue, the absence of a statue composed of them would make no difference to whether you end up having such an experience.

S2 says that recognizing that a belief is insensitive—that you would have had it even if it were false—undermines the belief.[22] In support of the claim that recognized insensitivity is a defeater, Merricks presents the following case:

You gaze upon the Emerald City. Its buildings appear to be green. You are then informed that your glasses have green lenses. Thus you learn that the buildings would appear green to you even if they were some other color. And so you are no longer justified in believing that the buildings are green. (2003, 739)

Recognized insensitivity, the idea goes, explains why you lose your justification upon learning about the colored lenses. And if this is what undermines justification here, it plausibly undermines justification wherever else it rears its head.[23]

In what follows, I'll grant S2 for the sake of the argument, and focus just on S1.[24] But before turning to S1, it will be illuminating to consider an episode from the history of metaethics in which sensitivity also makes a notable appearance. Curiously, in this episode the roles are reversed: sensitivity is wielded *by* the bootstrapper *against* the debunker.

13.6.2 The Harman-Sturgeon Debate

Well before the recent surge of interest in evolutionary moral-debunking arguments, there was the Harman/Sturgeon debate. Gilbert Harman (1977, chap. 1) maintained that moral reactions can be explained simply by appeal to mundane natural facts.

epistemically significant difference between belief in statues and your color beliefs in Emerald City (more on Emerald City shortly). But this premise is questionable for the same sorts of reasons given subsequently against S1. We have good reason to think that our color beliefs in Emerald City are insensitive (given our previous encounters with tinted lenses), whereas we don't have good reason to think that our statue beliefs are insensitive (as I argue later). That would seem to be an epistemically significant difference.

[22] The premise really ought to include a restriction to methods, to handle well-known counterexamples (see Nozick 1981, 179–85). I ignore this complication since what I say below applies equally to more complicated formulations of the sensitivity principle.

[23] Merricks also offers a second argument in support of what I am calling S2; see my 2015, chap. 10, for discussion of his 2001, 72–76, argument from strange fusions. See Benovsky 2015 and Osborne 2016, 216, for similar attempts to impugn object beliefs on grounds of insensitivity.

[24] See my 2015, sec. 7.2.2, for criticism of S2. See Vogel 2012 for a deluge of arguments that insensitivity does not preclude knowledge, arguments that can be repurposed to show that recognized insensitivity does not preclude justification.

This threatens to make moral facts entirely irrelevant to the explanation of our moral reactions, which in turn threatens to show that our moral reactions cannot serve as evidence for moral beliefs formed on their basis.

In reply, Nicholas Sturgeon (1988) suggested a way of checking whether the moral facts are or aren't explanatorily relevant. Explanations tend to support counterfactuals: if A explains B, then had it not been the case that A it would not have been the case that B.[25] Accordingly, to get to the bottom of whether moral facts explain moral reactions, we should ask ourselves whether we would have had the same moral reactions had the moral facts been different—or, in other words, whether our reactions are sensitive to the facts. If the reactions are sensitive, that's reason to think that the moral facts are explaining the reactions; if they're insensitive, that's reason to think that the moral facts aren't explaining the reactions.

Following Sturgeon (1988, 246–48), let's apply the test to Harman's own case. You turn the corner, see the kids setting fire to the cat, and immediately judge them to be doing something wrong. To check whether the fact that they are doing something wrong explains your judgment, we need to assess the following conditional:

(CF) You would have judged that they were engaged in wrongdoing even had they not been engaged in wrongdoing.

If CF is true, then the fact that they're engaged in wrongdoing fails the test for explanatory relevance. If CF is false, then that fact passes the test for explanatory relevance.

We know how to assess conditionals like CF. You imagine a situation in which the kids aren't doing something wrong but that is otherwise as similar as possible to how things actually are. That is, you imagine the "closest world" in which they aren't doing something wrong. Then you check whether in that situation you judge them to be doing something wrong.[26] What, then, is the closest world in which they're not doing something wrong?

Error theorists and realists will predictably reach different conclusions when they ask themselves this question. The error theorist thinks that nothing has any moral properties, in which case setting fire to the cat isn't even actually wrong. Thus, by her lights, the closest no-wrongdoing world is the actual world, a world in which they're setting fire to a cat. Since that world (a.k.a. this world) is one in which you do judge them to be doing something wrong, CF comes out true.

Now for the moral realist. What, by her lights, is the closest world in which the kids aren't engaged in any wrongdoing? Not one in which they are doing the same as what they actually are doing. We'd have to go to pretty far-out worlds to find ones in which the kids are setting fire to a cat and yet (by realist lights) doing no wrong—maybe worlds where cats are unaffected by fire, or where setting fire to the

[25] See Sturgeon 1988, 245 n. 19, on some notable exceptions.
[26] Cf. Stalnaker 1968 and Lewis 1973.

cat will foil some terrorist plot. Rather, the closest worlds in which they're doing nothing wrong are ones in which they're "engaged in some robust but innocent fun" (Sturgeon 1988, 247), like arm-wrestling or racing their bikes. Since we're not absolute curmudgeons—liable to think the kids are up to no good no matter what they're doing—we don't in these close-by worlds judge that they're engaged in wrongdoing. CF thus comes out false by the moral realist's lights, and the moral facts pass the test for being relevant to our moral reactions. So, moral realists should conclude that moral facts do explain our moral reactions.

Is this line of reasoning somehow illicit? True, the moral realist is taking for granted the accuracy of her moral reactions in the face of the explanatory challenge, without offering any independent evidence of their reliability. But, as Sturgeon emphasizes (1988, 237–38), there is nothing dialectically illicit about that. Harman's concern is that even if the moral facts are as the moral realist takes them to be, they are explanatorily irrelevant. Harman is therefore inviting us to start from the assumption that the moral facts are as we think they are and see if he is right about what follows. Nor, I would add, is there anything *epistemically* illicit about the reasoning. As we saw in section 13.5, it had better at least sometimes be legitimate to believe the deliverances of a source of information even when one has no independent evidence of its reliability. Some auxiliary argument would be needed to establish that our moral faculties are not among those sources for which we enjoy this default entitlement.

Here then, as I see it, is the upshot of the Harman/Sturgeon debate. Sturgeon is right. Counterfactuals like CF cannot be assessed without first taking a stand on what the moral facts actually are. And just as you are entitled to rely on your moral reactions in deciding whether to call the cops on the kids, you are entitled to rely on your moral reactions in assessing Harman's concern that moral facts don't explain our moral reactions. Drawing on such reactions, those of us who were realists to begin with ought to think that CF is false and that the facts do explain the reactions.

13.6.3 The Sensitivity of Object Beliefs

Now back to Merricks's argument, which similarly turns on assessing our reactions to CF-like conditionals. Recall S1:

(S1) You would have believed that there are statues even if it were false that there are statues, and you realize this.

Should we accept S1? Certainly a statue eliminativist, who believes only in the atoms arranged statuewise, will think it's true of you. After all, the eliminativist thinks the closest world in which it's false that there are statues is the actual world—in which case, since you (I assume) do actually believe in statues, S1 comes out true.

What, though, should the statue realist think is the closest world in which there are no statues? Not one in which there are atoms arranged statuewise. We'd have to go to pretty far-out worlds to find ones in which there are atoms arranged statuewise but no statues—perhaps all the way out to impossible worlds. Rather, the closest worlds in which there are no statues are perfectly normal statueless worlds in which there are no atoms arranged statuewise either. Since we're not weirdly fanatical about statues—liable to think there are statues when there aren't even atoms arranged statuewise around—we'll judge in such close-by worlds that there are no statues. S1 comes out false by the statue realist's lights.

Just like the moral realist in section 13.6.2, the statue realist is within her dialectical and epistemic rights to rely on her object reactions when reasoning her way to the falsity of S1. Merricks and other eliminativists are of course welcome to accept S1, but noneliminativists shouldn't be convinced by his argument.

Some, I suspect, will complain that Sturgeon and I have missed the point. If you vary the moral facts *while holding fixed* what the kids are doing, they'll say, our moral beliefs would be the same; if you vary the midsized object facts *while holding fixed* how all the atoms are arranged, they'll say, our object beliefs would be the same. Maybe so.[27] But it's hard to see what of interest is supposed to follow from these counterfactuals. To derive any interesting epistemic consequences from them, we'd have to replace S2 with something like the following:

(S2′) If you realize that you would have believed that p even if (i) p were false and (ii) all non-p facts were exactly the same, then you are not justified in believing p.

But this is going to undermine beliefs that moral skeptics and object eliminativists never meant to challenge. You look at a reliable thermometer—or, if you insist, atoms arranged thermometerwise—that reads 87 degrees, and you're justified in believing that it's 87 degrees. But you would have believed this even if (i) it weren't 87 degrees and (ii) the thermometer still read 87 degrees. So S2′ wrongly classifies your belief as unjustified.[28]

13.7 LIMITS ON BOOTSTRAPPING
13.7.1 Explanatory Concessions

We have a default entitlement to believe and reason from the deliverances of our perceptual and moral faculties. The entitlement is of course defeasible. We have just been assessing one candidate defeater—insensitivity—but we did not find any convincing reason to agree that the targeted reactions are insensitive.

---

[27] Though I have my doubts; see my 2015, 98.
[28] Cf. Sturgeon 1988, 250–53. I leave it as an exercise for the reader to try to find a sensitivity constraint that undermines moral beliefs and object beliefs without overgeneralizing in this way.

I turn now to another putative defeater: the discovery that our moral reactions and object reactions bear no explanatory connection to their subject matter. Plausibly, discovering, or even just conceding (rightly or wrongly), that one is forming one's beliefs about some domain D for reasons having entirely nothing to do with the D-facts is poised to serve as a defeater for any justification one might otherwise have for one's D-beliefs.[29] This suggests an explanatory constraint on justification: one is justified in believing p only if one believes that it is at least in part because of p that one believes p. (We'll see a more adequate formulation of the constraint in section 13.8.) This is why it would be troubling if Merricks were right that our object beliefs are insensitive: if indeed they are, then that would suggest that it's not *because* there are statues that we believe in statues, which in turn undermines our entitlement to such beliefs. This is why it is troubling to find out that you are wearing green lenses in Emerald City: your entitlement to your color beliefs is defeated once (and because) you become convinced that it's because the lenses are green, and not because the buildings are green, that you believe the buildings to be green. Ultimately, it is the explanatory concessions that defeat the beliefs.

Crucially, Sturgeon *does* accept that moral facts play a role in explaining our moral reactions. Indeed, he invokes sensitivity precisely for the purpose of establishing the explanatory relevance of moral facts. So he doesn't run afoul of this defeater. But forget about Sturgeon, and consider Shmurgeon. Shmurgeon concedes that there is no explanatory connection between his moral reactions and the moral facts—maybe because he can't see how moral features could have any kind of causal efficacy, or because he thinks that citing moral facts would be objectionably redundant or unparsimonious. Shmurgeon concedes that his belief that the kids are engaged in wrongdoing is not even partly explained by the action's in fact being wrong. Nevertheless, Shmurgeon stands by his sensitivity beliefs. Like Sturgeon, he bootstraps from his moral reactions to the conclusion that CF is false: the closest no-wrongdoing worlds are ones in which the kids are racing their bikes.

This, I submit, is epistemically impermissible. Once you concede that your moral faculties render the verdicts they do about the moral facts for reasons having nothing at all to do with what the moral facts actually are, you lose any entitlement you had for relying on those faculties. Shmurgeon accepts that it is not because the actions are wrong that he believes or perceives or intuits them to be wrong. He accepts that which moral features things have has nothing to do with why he makes the judgments he does. And once he makes this explanatory concession he thereby loses his entitlement to the moral judgments, and bootstrapping from them is no longer an option.

The lesson here is that sensitivity derives whatever significance it has from its connection to explanation. Sensitivity is a mark of an explanatory connection. The

[29] Cf. Benacerraf 1973; Sayre-McCord 1988, secs. 5–6; Field 1989, 25–30; Gibbard 2003, chap. 13; Locke 2014; and Korman 2015, chap. 7.3.

absence of the former is some evidence for the absence of the latter, and discovering the absence of the explanatory connection is what does the defeating. Accordingly, once one has made the explanatory concession, it is neither here nor there that one can reason one's way from the disconnected beliefs to sensitivity. For whatever justification those beliefs once had has already been undermined by the explanatory concession.

13.7.2 Flouting the Explanatory Constraint

The above observation about explanatory connections and their relation to sensitivity has ramifications for a certain style of response to debunking arguments found both in the objects literature and in the moral literature. I'll start with the former.

In his recent book, Thomas Hofweber (2016, chap. 7.3) offers a response to Merricks's argument that is somewhat different from the one I have advanced. Whereas I maintain that midsized objects and their atomic parts overdetermine our object reactions, Hofweber suggests that we can stand by our object beliefs "even if Merricks is right about the simples doing all the causal work" (195):

How our beliefs are causally produced in us is not the crucial question for epistemology. Whether they track [i.e., are sensitive to[30]] the facts is relevant for this, and the answer is that they do, or so we have reason to think. (194)

And what reason do we have for thinking that our object beliefs track the object facts? Our object reactions, to which we are defeasibly entitled, and from which we can bootstrap to the sensitivity of the reactions:

The counterfactual situation relevant to evaluate the counterfactual conditional about my cup not being there [viz. that if the cup didn't exist, then I wouldn't believe it exists] isn't one where the simples are still there, but they somehow don't compose a cup, but rather one where both the simples and the cup are gone. But then nothing would cause me to believe in a cup in front of me. (194)

On a natural reading, then, Hofweber is saying that our object beliefs are sensitive to the object facts even though the facts don't causally explain the beliefs (they do no "causal work").[31]

---

[30] Hofweber doesn't explicitly define tracking in terms of sensitivity, but he does informally gloss the former in terms of the latter: "My belief that there is a cup in front of me does track the cup in the sense that if the cup didn't exist then I wouldn't have this belief" (2016, 194).

[31] Since I wrote this chapter, Hofweber has informed me that this is a misconstrual. He does say that the object facts do no *causal work* but doesn't thereby mean to deny that they *causally explain* the object reactions (a distinction I'd be interested to see spelled out in greater detail). I've chosen to leave

So understood, Hofweber is a real-life Shmurgeon, bootstrapping from his object reactions to a vindication of those reactions, all the while granting that the sensitivity is not underwritten by an explanatory connection between the reactions and the facts. Thus, he runs afoul of our explanatory constraint, and thereby undermines his entitlement to the object reactions from which he bootstraps. It is neither here nor there that he can reason his way from the discredited object reactions to sensitivity. *Pace* Hofweber, how the beliefs are produced *is* the crucial question, and the answer needs to be: by the associated facts.[32]

The explanatory constraint also has ramifications for the moral-debunking literature. There, one of the dominant strategies—what Dustin Locke (2014) dubs "the minimalist response"—has been to grant M1 and instead deny M2. The strategy is to grant that one's moral reactions are not explained by the moral facts, but show how our moral reactions *could* still turn out to be sensitive, or safe, or in some other way nonaccidentally accurate. Do we have reason to think that this not only could happen, but has happened? Yes, the idea goes, for we can bootstrap from our moral reactions to the conclusion that they are nonaccidentally accurate (recall the Nozick/Enoch reasoning from section 13.5).[33]

What minimalists are presumably thinking is that explanatory connections have epistemic import only insofar as they are decent indicators of nonaccidental accuracy. Justin Clarke-Doane (2015, sec. 6), for instance, insists that if—using our antecedent moral beliefs—we can assure ourselves that our moral beliefs are "bound to be true," it couldn't possibly matter whether one accepts an explanatory connection. But if the explanatory constraint is correct, then this sort of bootstrapping cannot even get off the ground; the explanatory concession undermines our entitlement to the moral reactions before the bootstrapping begins. Indeed, I think the minimalist has things exactly the wrong way round: once we concede that there is no explanatory connection between our moral beliefs and the moral facts, it couldn't possibly matter that we can use the discredited, disconnected beliefs to assure ourselves that the beliefs are bound to be true.[34]

the criticism in the text, with apologies to Hofweber, since I suspect some will still be tempted by the strategy I understood him to be advancing—and I didn't want to try readers' patience by changing it to a critique of Shmofweber.

[32] See my 2014, secs. 3–4; 2015, chap. 7, on other attempts to block the object-debunking arguments while granting that there is no explanatory connection.

[33] Minimalist strategies have been advanced or defended by Nozick 1981, 342–48; Dworkin 1996, 117–26; Huemer 2005, 218–19; Enoch 2010, secs. 3–5; 2011, sec. 7.4; Schafer 2010; White 2010, 588–89; Wielenberg 2010, secs. 4–8; 2014, chap. 4; 2016, sec. 3; Brosnan 2011, 60–63; Parfit 2011, 532–33; Skarsaune 2011, sec. 3; Berker 2014; Clarke-Doane 2015, secs. 4–6; 2016, secs. 2–4; Talbott 2015; Moon 2017; and Schechter 2018, sec. 6. Some on this list are more explicit (or aware) than others about the need for bootstrapping.

[34] See Korman and Locke (forthcoming) for more on minimalist responses to moral-debunking arguments.

13.8 WHAT IS THE EXPLANATORY CONSTRAINT?

What we have seen is that debunking arguments may be resisted by bootstrapping from antecedent beliefs, so long as bootstrappers respect the explanatory constraint on their entitlement to those beliefs. But what exactly is the explanatory constraint? Let's start simple and work our way to a viable formulation.[35] Impatient readers can jump down to EC5 to see where we are headed.

We'll begin with EC1:

(EC1) S is justified in believing p only if: [S believes p] because p.

In other words: only if the fact that p at least partly explains S's believing p. This is obviously too demanding. An unwitting brain in a vat is justified in believing that it has hands even though its belief that it has hands is not explained by its having hands.

This leads us to EC2:

(EC2) S is justified in believing p only if: S believes that [[S believes p] because p].

Unlike EC1, EC2 does not imply that the brain in a vat is unjustified, since the brain at least *thinks* that its beliefs are explained by the associated facts. But EC2 cannot be right as it stands because it cannot accommodate inductive beliefs. Take, for instance, beliefs about the future. The sun's rising tomorrow clearly doesn't explain your belief that it will, and you realize this. Nevertheless, your belief that it will rise tomorrow is surely justified. Other counterexamples to EC2 include cases where the belief and the associated fact have a common cause. I believe that the email I sent you is in your inbox, not because it is in your inbox, but because I clicked *send*.

The fix is to allow beliefs to be justified even when they're not believed to be explained by the associated facts, so long as they're prima facie justified by beliefs that themselves are believed to be explained by the associated facts:

(EC3) S is justified in believing p only if: *either* (i) S believes that [[S believes p] because p], *or* (ii) S's belief that p is based on beliefs $Bp_1 \ldots Bp_n$ that prima facie justify S in believing that p and S believes that [[S believes $p_1 \ldots p_n$] because $p_1 \ldots p_n$].

Your belief that the sun will rise tomorrow is based on your beliefs that it rose today and yesterday and the day before that; those beliefs do prima facie justify your belief that it will rise tomorrow; and you *do* think that its rising today explains your belief that it rose today, and likewise for the rest of the beliefs in your inductive base. So

---

[35] For some alternative formulations of the constraint, see Locke 2014, 232; McCain 2014, sec. 4.4 and 6.4; Lutz 2018, sec. 2; Schechter 2018, sec. 3; and Korman and Locke forthcoming, sec. 8.

the second disjunct is satisfied. Similarly, my belief that my email is in your inbox is prima facie justified by my belief that I clicked *send*, and I do think that my belief that I clicked *send* is explained by my clicking *send*.

EC3 makes room for justified inductive beliefs. But it threatens to overintellectualize justification (as does EC2). EC3 will entail that a young child—old enough to have beliefs about apples but too young to have beliefs about beliefs—cannot be justified in believing that there is an apple present. This can be handled by weakening the constraint so that what's required isn't belief but rather not withholding belief:

(EC4) S is justified in believing p only if: *either* (i) S doesn't withhold belief that [[S believes p] because p], *or* (ii) S's belief that p is based on beliefs $Bp_1 \ldots Bp_n$ that prima facie justify S in believing that p and S doesn't withhold belief that [[S believes $p_1 \ldots p_n$] because $p_1 \ldots p_n$].

Withholding belief is something more deliberate than failing to believe.[36] It involves entertaining a proposition and then opting not to believe it, either by disbelieving it or by remaining agnostic. A young child who, lacking the requisite concepts, cannot entertain propositions about what she does or doesn't believe won't count as withholding belief in those propositions. Counterexample averted.

EC4 permits justified inductive beliefs without depriving brains in vats and young children of justified perceptual beliefs. But it is still open to counterexamples like the following. Upon naming my dog "Emmylou," it is not her being named "Emmylou" that explains my belief that her name is "Emmylou." Rather, it's my taking her name to be "Emmylou" that explains why she is named "Emmylou." Since I recognize that the explanation doesn't run from the fact to the belief, EC4 entails that I'm not justified in believing her name to be "Emmylou." But obviously I am so justified.

The fix is to allow explanations to run in either direction: world to mind or mind to world. It's fine to acknowledge that my belief that p isn't explained by p so long as I think that my believing p explains why it's true that p. Let's say that the belief that p is *e-connected* to p iff either the belief that p is explained by the fact that p or vice versa. Now we can say:

(EC5) S is justified in believing p only if: *either* (i) S doesn't withhold belief that [S's belief that p is e-connected to p], *or* (ii) S's belief that p is based on beliefs $Bp_1 \ldots Bp_n$ that prima facie justify S in believing p and S doesn't withhold belief that [S's beliefs that $p_1 \ldots p_n$ are e-connected to $p_1 \ldots p_n$].

---

[36] See Bergmann 2005, 420–21 and Locke 2014, 223, on withholding.

I do believe that my belief that my dog's name is "Emmylou" is e-connected to her being named "Emmylou," since I take the former to explain the latter. So EC5 won't entail that the belief is unjustified.[37] This revision also makes room for antirealist treatments of the debunking arguments, denying O1 or M1 on the grounds that our reactions determine the associated facts.[38]

These refinements won't affect my earlier objections to Hofweber. Hofweber's strategy (as I interpreted it) involves withholding belief as to whether I believe there is a cup before me because there is a cup before me. And he certainly doesn't think that the explanation runs in the other direction, that there is a cup before me because I believe there to be one. So condition (i) of EC5 is not satisfied. What about condition (ii)? The only reasons Hofweber cites for his cup beliefs are his cup experiences. Trivially, then, his cup beliefs are not based on any beliefs that he takes to be e-connected to the cup facts. So condition (ii) is not satisfied either.

One might suggest that there is still a problem with EC5, insofar as it cannot accommodate foundational a priori beliefs about necessary truths. On pain of antirealism or skepticism about mathematics, EC5 looks to require us to accept (or at least not withhold belief) that mathematical facts explain our mathematical beliefs. Some may find this intolerable. But one's modus tollens, as they say, is another's modus ponens. I say that realists cannot stand by the belief that 1 + 1 is 2 once they concede that they believe this for reasons having nothing whatsoever to do with what the sum of 1 and 1 actually is. Benacerraf (1973) and Field (1989) are right that this is an unstable epistemic position. To be sure, I agree that it's unclear how mathematical facts could explain our beliefs. But, unless you're ready to embrace mathematical skepticism or antirealism, you had better be prepared to accept that they somehow manage to do so.[39]

13.9 CONCLUSION

Here is what I hope to have shown. Even what would seem to be some of our most secure beliefs, our perceptual beliefs about midsized objects, are not safe from debunking by cognitive science and evolutionary psychology. But even when cognitive scientists or evolutionary psychologists are able to explain some range of beliefs without reference to the associated range of facts, we needn't accept that these explanations are the full story. For we are often entitled to bootstrap to an expanded

---

[37] Actually, it's probably more complicated than that. It's my *intention* that her name be "Emmylou" that explains both why her name is "Emmylou" and why I believe that her name is "Emmylou." Accordingly, it's the second clause of EC5 that handles this case: my belief that I intend for her name to be "Emmylou" is explained by my having that intention; I believe that this is so; and my belief that I intended that this be her name, together with the belief that she is mine to name (which I take to be e-connected to the fact that she is mine to name), prima facie justify my belief that this is her name.

[38] Cf. Goldman 1992, chaps. 2–3, for subjectivist "demotions" of objects, color, and modality in response to challenges from cognitive science.

[39] See Bengson 2015 for one proposal.

explanation that does cite the associated facts. That said, one's bootstrapping had better culminate in embracing some such explanatory connection, for conceding that there is no connection would undermine one's entitlement to bootstrap.

I certainly don't take myself to have resolved the dispute between the debunker and the bootstrapper. In truth, I think that we have only just scratched the surface. Among other things, I think that there are further constraints on legitimate bootstrapping than simply allowing some explanatory connection between your beliefs and associated facts. Elsewhere, I have argued that the explanatory connection one takes to hold between beliefs and the facts must be "nondeviant," and that bootstrappers have their work cut out for them showing how a causal connection between midsized objects and object beliefs could be nondeviant.[40] If I am right about this, then bootstrappers cannot dismiss the object-debunking arguments simply on the grounds that midsized objects are poised to cause our object beliefs. And it is worth considering whether the threat of deviance also arises for responses to the moral-debunking arguments according to which moral properties are proximate or supervenient causes of moral beliefs,[41] which is something I hope to explore in future work.

ACKNOWLEDGMENTS

REFERENCES

Alexander, D. J. (2011). In defense of epistemic circularity. *Acta Analytica* 26: 223–41.

Alston, W. P. (1986). Epistemic circularity. *Philosophy and Phenomenological Research* 47: 1–30.

Baron, S. (2017). Feel the flow. *Synthese* 194: 609–30.

Baron, S., Cusbert, J., Farr, M., Kon, M., and Miller, K. (2015). Temporal experience, temporal passage and the cognitive sciences. *Philosophy Compass* 10: 560–71.

Bedke, M. S. (2009). Intuitive non-naturalism meets cosmic coincidence. *Pacific Philosophical Quarterly* 90: 188–209.

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy* 70: 661–79.

Bengson, J. (2015). Grasping the third realm. *Oxford Studies in Epistemology* 5: 1–38.

Benovsky, J. (2015). From experience to metaphysics. *Noûs* 49: 684–97.

Bergmann, M. (2004). Epistemic circularity: Malignant and benign. *Philosophy and Phenomenological Research* 69: 709–27.

---

[40] See my 2014, secs. 5–6 and 2015, sec. 7.4.

[41] Cf. Shafer-Landau 2012, 27, and Mogensen 2015.

Bergmann, M. (2005). Defeaters and higher-level requirements. *Philosophy Quarterly* 55: 419–36.

Berker, S. (2014). Does evolutionary psychology show that normativity is mind-dependent? In J. D'Arms and D. Jacobson, eds., *Moral Psychology and Human Agency*. Oxford: Oxford University Press, 215–52.

Bogardus, T. (2016). Only all naturalists should worry about only one evolutionary debunking argument. *Ethics* 126: 636–61.

BonJour, L. (1998). *In Defense of Pure Reason*. Cambridge: Cambridge University Press.

Braddock, M. (2016). Evolutionary debunking: Can moral realists explain the reliability of our moral judgments? *Philosophical Psychology* 29: 844–57.

Braddock, M. (2017). Debunking arguments from insensitivity. *International Journal for the Study of Skepticism* 7 (2): 91–113.

Brosnan, K. (2011). Do the evolutionary origins of our moral beliefs undermine moral knowledge? *Biology and Philosophy* 26: 51–64.

Chalmers, D. (2006). Perception and the fall from Eden. In T. S. Gendler and J. Hawthorne, eds., *Perceptual Experience*. Oxford: Oxford University Press, 49–125.

Clarke-Doane, J. (2015). Justification and explanation in mathematics and morality. *Oxford Studies in Metaethics* 10: 80–103.

Clarke-Doane, J. (2016). Debunking and dispensability. In N. Sinclair and U. Leibowitz, eds., *Explanation in Ethics and Mathematics*. Oxford: Oxford University Press, 23–36.

Copp, D. (2008). Darwinian skepticism about moral realism. *Philosophical Issues* 18: 186–206.

Crisp, T. M. (2016). On naturalistic metaphysics. In K. J. Clark, ed., *The Blackwell Companion to Naturalism*. Malden, MA: Wiley Blackwell, 61–74.

Dreier, J. (2012). Quasi-realism and the problem of unexplained coincidence. *Analytic Philosophy* 53: 269–87.

Dworkin, R. (1996). Objectivity and truth: You'd better believe it. *Philosophy and Public Affairs* 25: 87–139.

Enoch, D. (2010). The epistemological challenge to metanormative realism. *Philosophical Studies* 148: 413–38.

Enoch, D. (2011). *Taking Morality Seriously*. Oxford: Oxford University Press.

Feldman, R. (2005). Respecting the evidence. *Philosophical Perspectives* 19: 95–119.

Field, H. (1989). *Realism, Mathematics, and Modality*. Oxford: Blackwell.

Fitzpatrick, W. J. (2015). Debunking evolutionary debunking of ethical realism. *Philosophical Studies* 172: 883–904.

Fraser, B. (2014). Evolutionary debunking arguments and the reliability of moral cognition. *Philosophical Studies* 168: 457–73.

Gibbard, A. (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Goldman, A. (1992). *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, MA: MIT Press.

Gould, S. J., and Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm. *Proceedings of the Royal Society of London* 205: 581–98.

Gow, L. (2014). Colour. *Philosophy Compass* 9: 803–13.

Harman, G. (1977). *The Nature of Morality*. New York: Oxford University Press.

Hawthorne, J. (2006). *Metaphysical Essays*. Oxford: Oxford University Press.

Heller, M. (1990). *The Ontology of Physical Objects: Four-Dimensional Hunks of Matter*. New York: Cambridge University Press.

Hirsch, E. (1993). *Dividing Reality*. Oxford: Oxford University Press.

Hoffman, D. D., Singh, M., and Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin Review* 22: 1480–506.

Hofweber, T. (2016). *Ontology and the Ambitions of Metaphysics*. Oxford: Oxford University Press.

Hudson, H. (2001). *A Materialist Metaphysics of the Human Person*. Ithaca, NY: Cornell University Press.

Huemer, M. (2001). *Skepticism and the Veil of Perception*. Lanham, MD: Rowman & Littlefield.

Huemer, M. (2005). *Ethical Intuitionism*. New York: Palgrave.

Hummel, J. E. (2015). Is interface theory a special case of Q-morphs or full-blown solipsism? *Psychonomic Bulletin and Review* (invited commentary). http://www.psychonomic. org/news/297868/Is-Interface-Theory-a-Special-Case-of-Q-morphs-or-Full-blown-Solipsism.htm.

Joyce, R. (2005). Moral fictionalism. In M. E. Kalderon, ed., *Fictionalism in Metaphysics*. Oxford: Oxford University Press, 287–313.

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

Kelly, T. (2008). Common sense as evidence: Against revisionary ontology and skepticism. *Midwest Studies in Philosophy* 32: 53–78.

Kitcher, P. (2005). Biology and ethics. In D. Copp, ed., *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press, 163–85.

Korman, D. Z. (2014). Debunking perceptual beliefs about ordinary objects. *Philosophers' Imprint* 14: 1–21.

Korman, D. Z. (2015). *Objects: Nothing out of the Ordinary*. Oxford: Oxford University Press.

Korman, D. Z., and Locke, D. (forthcoming). Against Minimalist Responses to Moral Debunking Arguments. *Oxford Studies in Metaethics*.

Leech, D., and Visala, A. (2011). Naturalistic explanation for religious belief. *Philosophy Compass* 6: 552–63.

Leslie, S.-J. (2013). Essence and natural kinds: When science meets preschooler intuition. *Oxford Studies in Epistemology* 4: 108–66.

Lewis, D. (1973). *Counterfactuals*. Malden, MA: Blackwell.

Lillehammer, H. (2003). Debunking morality: Evolutionary naturalism and moral error theory. *Biology and Philosophy* 18: 567–81.

Locke, D. (2014). Darwinian normative skepticism. In M. Bergmann and P. Kain, eds., *Challenges to Moral and Religious Belief*. Oxford: Oxford University Press, 220–36.

Lutz, M. (2018). What Makes Evolution a Defeater? *Erkenntnis* 83: 1105–26.

Marr, D. (1982). *Vision*. Cambridge, MA: MIT Press.

McCain, K. (2014). *Evidentialism and Epistemic Justification*. New York: Routledge.

Mendelovici, A. (2013). Reliable misrepresentation and tracking theories of mental representation. *Philosophical Studies* 165: 421–43.

Mendelovici, A. (2016). Why tracking theories should allow for clean cases of reliable misrepresentation. *Disputatio* 8: 57–92.

Merricks, T. (2001). *Objects and Persons*. Oxford: Oxford University Press.

Merricks, T. (2003). Replies. *Philosophy and Phenomenological Research* 67: 727–44.

Mogensen, A. L. (2015). Evolutionary debunking arguments and the proximate/ultimate distinction. *Analysis* 75: 196–203.

Moon, A. (2017). Debunking morality: Lessons from the EAAN literature. *Pacific Philosophical Quarterly*. 98 (S1): 208–26.

Nozick, R. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Osborne, R. C. (2016). Debunking rationalist defenses of common-sense ontology. *Review of Philosophy and Psychology* 7: 197–221.

Parfit, D. (2011). *On What Matters*. Vol. 2. Oxford: Oxford University Press.

Paul, L. A. (2010). Temporal experience. *Journal of Philosophy* 107: 333–59.

Paul, L. A. (2016). Experience, metaphysics, and cognitive science. In J. Sytsma and W. Buckwalter, eds., *A Companion to Experimental Philosophy*. Malden, MA: Wiley Blackwell, 419–33.

Pautz, A. (2011). Can disjunctivists explain our access to the sensible world? *Philosophical Issues* 21: 384–433.

Pryor, J. (2004). What's wrong with Moore's argument. *Philosophical Perspectives* 14: 349–78.

Rea, M. C. (2002). *World without Design*. Oxford: Oxford University Press.

Rose, D., and Schaffer, J. (2017). Folk mereology is teleological. *Noûs* 51(2): 238–70.

Ruse, M. (1986). *Taking Darwin Seriously*. Amherst, NY: Prometheus.

Sattig, T. (2015). *The Double Lives of Objects*. Oxford: Oxford University Press.

Sayre-McCord, G. (1988). Moral theory and explanatory impotence. *Midwest Studies in Philosophy* 12: 433–57. Reprinted in his *Essays on Moral Realism*.

Schafer, K. (2010). Evolution and normative scepticism. *Australasian Journal of Philosophy* 88: 471–88.

Schaffer, J. (2015). What not to multiply without necessity. *Australasian Journal of Philosophy* 93: 644–64.

Schaffer, J. (2016). Cognitive science and metaphysics: Partners in debunking. In B. P. McLaughlin and H. Kornblith, eds., *Goldman and His Critics*. Malden, MA: Wiley Blackwell, 337–65.

Schechter, J. (2018). Explanatory challenges in metaethics. In T. McPherson and D. Plunkett, eds., *Routledge Handbook of Metaethics*. New York: Routledge, 443–59.

Shafer-Landau, R. (2012). Evolutionary debunking, moral realism and moral knowledge. *Journal of Ethics and Social Philosophy* 7: 1–37.

Sider, T. (2001). *Four-Dimensionalism*. Oxford: Clarendon Press.

Skarsaune, K. (2011). Darwin and moral realism. *Philosophical Studies* 152: 229–43.

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher, ed., *Studies in Logical Theory*. Oxford: Blackwell, 98–112.

Street, S. (2006). A Darwinian dilemma for realist theories of value. *Philosophical Studies* 127: 109–66.

Street, S. (2008). Reply to Copp: Naturalism, normativity, and the varieties of realism worth worrying about. *Philosophical Issues* 18: 207–28.

Stroud, B. (2000). Evolution and the necessities of thought. In *Meaning, Understanding, and Practice*. Oxford: Oxford University Press, 52–67.

Sturgeon, N. (1988). Moral explanations. In G. Sayre-McCord, ed., *Essays on Moral Realism*. Ithaca, NY: Cornell University Press, 229–55.

Sumner, P., and Mollon, J. D. (2000). Catarrhine photopigments are optimized for detecting targets against a foliage background. *Journal of Experimental Biology* 203: 1963–86.

Talbott, W. J. (2015). How could a "blind" evolutionary process have made human moral beliefs sensitive to strongly universal, objective moral standards? *Biology and Philosophy* 30: 691–708.

Titelbaum, M. G. (2010). Tell me you love me: Bootstrapping, externalism, and no-lose epistemology. *Philosophical Studies* 149: 119–34.

Van Cleve, J. (2003). Is knowledge easy—or impossible? Externalism as the only alternative to skepticism. In S. Luper, ed., *The Skeptics*. Burlington, VT: Ashgate, 45–59.

Vavova, K. (2015). Evolutionary debunking of moral realism. *Philosophy Compass* 10: 104–16.

Vogel, J. (2000). Reliabilism leveled. *Journal of Philosophy* 97: 602–23.

Vogel, J. (2008). Epistemic bootstrapping. *Journal of Philosophy* 105: 518–39.

Vogel, J. (2012). The enduring trouble with tracking. In K. Becker and T. Black, eds., *The Sensitivity Principle in Epistemology*. Cambridge: Cambridge University Press, 122–51.

White, R. (2010). You just believe that because … *Philosophical Perspectives* 24: 573–615.

Wielenberg, E. J. (2010). On the evolutionary debunking of morality. *Ethics* 120: 441–64.

Wielenberg, E. J. (2014). *Robust Ethics*. Oxford: Oxford University Press.

Wielenberg, E. J. (2016). Ethics and evolutionary theory. *Analysis* 76: 502–15.

Wilkins, J. S., and Griffiths, P. E. (2012). Evolutionary debunking arguments in three domains. In G. W. Dawes and J. Maclaurin, eds., *A New Science of Religion*. New York: Routledge, 133–46.

Williamson, T. (2007). *The Philosophy of Philosophy*. Malden, MA: Blackwell.