CHAPTER 41

....

THE EVOLUTIONARY CHALLENGE TO KNOWING MORAL REASONS

....

TERENCE CUNEO[*]

41.1 INTRODUCTION

....

MANY philosophers find themselves attracted to the following three claims:

(1) Moral non-naturalism: there are robust non-natural moral facts, including reasons to act in certain ways.
(2) Moral knowledge: we have some moral knowledge.
(3) Evolutionary influence: evolutionary forces have not only determined which cognitive faculties we have but also heavily shaped their workings, including the workings of the moral faculty, which is simply that ability (or cluster of abilities) by which we form moral judgments.

Let me begin by offering a word of explanation regarding each of these claims.

Philosophers do not agree about what a robust non-natural moral fact would be. In what follows, I'll not attempt to resolve this issue, assuming only that, were such facts to exist, they would have four distinguishing characteristics.[1] First, many of these facts would be strongly mind-independent in the sense that they exist, but not in virtue of our valuing or desiring (or being such that we would value or desire) non-moral features of the world. In this regard, they differ from the fact *that Yo-Yo-Ma is*

[1] For discussion, see Copp (2007), Cuneo (2007a), Cuneo and Shafer-Landau (2014), and Shafer-Landau (2003). Defenders of non-naturalism include Enoch (2011), Fine (2002), FitzPatrick (2008, 2011), Hampton (1998), Parfit (2011), Scanlon (1998), Shafer-Landau (2003, 2006), and Wedgwood (2007).

an admired cellist, which does depend on the attitudes of admiration that people take toward him. Second, these facts would not play (or be reducible to any facts that play) explanatory roles in the usual sciences, such as physics, biology, or chemistry. Third, and relatedly, these facts would be causally inert, not (in any direct sense) entering into the causal flow of nature, as they would be abstract entities. And, fourth, these moral facts would be (or would necessarily provide) reasons of various sorts. Some of these reasons would be categorical in the sense that they apply to agents regardless of what desires, goals, or social allegiances they may have. Indeed, in the remainder of this discussion, when I speak of moral facts, I will have in mind facts that are or generate moral reasons.

As for moral knowledge, in what follows, I propose to remain relatively non-committal regarding what it is, assuming only the following. First, states of moral knowledge would be constituted by moral judgments or beliefs. And, second, moral judgments would be candidates for moral knowledge only if they have the proper sort of etiology: they could not be the product of luck, accident, or the like, but must be "well-formed" in the sense of being reliably formed or the product of cognitive faculties that are working well in a congenial environment. Moral knowledge probably requires more than this but, I'll assume, it would involve at least satisfying this "well-formedness" condition.

As for the evolutionary influence on our cognitive faculties, I will assume that, according to the picture that evolutionary biology bequeaths us, natural selection has rewarded whatever cognitive traits that caused our Pleistocene ancestors to maximize the relative representation of their genes in the gene pool over generations, often doing so by promoting their survival. In many cases, I'll also assume, we have excellent reason to believe that natural selection has given us cognitive capacities that track independent truths about the world with reasonable accuracy, such as truths about mid-sized material objects in our visual field.

These introductory comments having been made, we can see that the three claims stated above yield a puzzle, which I will call *the evolutionary debunker's puzzle*—or *the debunker's puzzle*, for short.[2] According to this puzzle, moral knowledge requires that some of our moral judgments be well-formed, which is what claim (2) tells us. But it is difficult to see how these judgments could be well-formed if the forces of natural selection have deeply influenced the workings of our capacity for forming moral judgments, which is what claim (3) says. These forces, after all, are geared, not toward producing true beliefs but toward maximizing the relative representation of the genes of ancestral humans in the gene pool over generations. In certain cases, as already noted, it is plausible to hold that this resulted in cognitive capacities that accurately track independent facts in our environment, such as ordinary mid-sized material objects in our visual field. But if there were non-natural moral facts, it is very difficult to see how the forces of

[2] Variations of the puzzle can be found in Greene (2008), Joyce (2006), Kitcher (2005, 2011), Locke (2014), Ruse (1998: ch. 6), and Street (2006, 2008a, 2016).

natural selection would have influenced the workings of our moral faculty in such a way that we reliably grasp them.

The difficulty is described in different ways. Some proponents of the puzzle hold that since, as a conceptual matter, there could be any number of moral systems, we have little reason to hold that evolution would have "pushed" us toward the correct one. Other proponents maintain that since we can offer a complete non-moral genealogy for all our moral beliefs by appeal to evolutionary factors, it would be incredible if our moral beliefs happened to match the moral facts. However that may be, proponents of the puzzle agree that, since non-natural moral facts are not the sort of thing that could enter into causal relations and, hence, be tracked by our moral faculty, there seems to be little prospect of offering any kind of account according to which, surprisingly enough, our evolutionary heritage put us in a good position to track them. Something, it appears, has to give.

The responses to this puzzle fall into three broad categories.[3] Some, such as Alvin Plantinga and Thomas Nagel, have proposed that we reject claim (3), at least when it is understood to incorporate the assumption that the forces of natural section select, in the first instance, not for true beliefs but for traits that maximize the relative representation of genes in the gene pool. These philosophers, then, reject the standard evolutionary picture, as it's usually understood.[4] Other philosophers, such as Richard Joyce and Jonas Olson, propose that we reject claim (2), giving up the idea that there is any credible account of how we would gain moral knowledge if there were moral facts.[5] These philosophers embrace moral skepticism. Finally, other philosophers, such as David Copp and Peter Railton, recommend that we reject claim (1), embracing metaethical naturalism according to which moral facts are ordinary natural facts that our moral judgments could track by bearing causal relations to them.[6] These philosophers reject moral non-naturalism in favor of moral naturalism.

In principle, one of these responses to the debunker's puzzle might be the best available. But none is congenial to moral non-naturalists, who would like to affirm all three claims. Is there, though, any prospect of affirming all three claims stated above? Those who accept non-naturalism believe so; they maintain that their view has the resources to address the evolutionary debunker's puzzle. My primary project in this chapter is to explore the non-naturalists' response, canvassing some of the major moves made by its advocates.

Before diving into our topic, let me note that attempts to address the debunker's puzzle fall along a continuum. Those at one end of the continuum—what we might call *shallow responses*—attempt to show that the debunker's puzzle does not pose a genuine threat to non-naturalism, since it does not give us sufficient reason to believe that, given

---

[3] FitzPatrick (2014b), Shafer-Landau (2012), and Wielenberg (2010) distinguish different versions of the puzzle.

[4] See Nagel (2012), Plantinga (1993, 2002), and Plantinga and Tooley (2008).

[5] See Joyce (2001, 2006) and Olson (2011, 2014).

[6] See Copp (2008) and Railton (2010).

---

the truth of moral non-naturalism, it is unlikely that we would arrive at well-formed moral judgments if the workings of our cognitive faculties were shaped by evolutionary forces. Those at the other end of the continuum—what I'll term *deep responses*—go a step further. They not only contend that the puzzle does not pose a genuine threat to non-naturalism, but also formulate and defend positive proposals as to how we could reliably grasp moral facts were they to exist. In what follows, I am interested in exploring views that lie at that end of the continuum occupied by the "deeper responses." That is, I wish to investigate more ambitious—and, hence, controversial—proposals about how we might arrive at moral knowledge if moral non-naturalism were true. The payoff of doing so might be considerable. For if such proposals could be satisfactorily defended, then non-naturalists will have made significant progress toward addressing an important challenge—perhaps the most important challenge—to their view.

41.2 AN INITIAL PROPOSAL: THIRD-FACTOR VIEWS

In chapter 7 of his book *Taking Morality Seriously* (2011), David Enoch notes that philosophers have long worried about whether non-naturalists can combine their view with a satisfactory moral epistemology. The worries these philosophers raise take different forms. Some critics wonder how we could gain epistemic access to or get moral facts in mind if non-naturalism were true. Others worry that we could not arrive at justified moral beliefs were non-naturalism correct. Still others maintain that, given non-naturalism, it is very difficult to see how we could gain moral knowledge. Enoch contends that these concerns do not capture the real epistemological challenge to non-naturalism. The real challenge that faces non-naturalists, according to Enoch, is to explain how our moral beliefs could be *reliably correlated* with the moral facts.

Non-naturalists are, after all, committed to the claim that many of our moral beliefs are true, accurately representing the moral facts. But if non-naturalism were correct, it cannot be that our moral judgments are causally or constitutively responsible for the moral facts. For, as we saw earlier, non-naturalists hold that the moral facts are supposed to be robustly mind-independent; they do not depend on our moral judgments. Nor can non-naturalists, who believe that moral facts are causally inert, hold that the moral facts cause our moral beliefs.[7] Nor, finally, does there seem to be some other explanation of the correlation that is available to non-naturalists. It appears, then, that non-naturalist realists are committed to there being an unexplained striking correlation between our moral beliefs, on the one hand, and the moral facts, on the other. Indeed, if Enoch is right, the debunker's puzzle is simply an instance of this general worry about reliable

---

[7] Not all non-naturalists agree. For dissenters, see Oddie (2005) and Wedgwood (2007), as well as Cuneo (2006) and Shafer-Landau (2003: 27–8).

correlations. What the puzzle does, in effect, is to articulate the problem in an especially acute form. Given that the pressures of natural selection have deeply shaped the workings of our cognitive faculties but not in such a way that they were "aimed" at reliably yielding true moral beliefs, it would appear to be a miracle were we reliably to arrive at such beliefs. Of course there are people who do not mind accepting miracles. Nearly all party to this debate, however, maintain that good metaethical theories should not countenance miracles of this magnitude.

Suppose, for the time being, that Enoch's diagnosis is correct: the debunker's puzzle is a vivid example of the epistemological challenge to non-naturalist realism. Enoch maintains that if non-naturalist realists can address this puzzle, then they probably do not need to worry about the other ways of formulating the epistemological challenge mentioned earlier. For if they can satisfactorily address the debunker's puzzle, Enoch claims, then non-naturalists will have the resources to explain such things as how we could have epistemic access to the moral facts and how we could arrive at moral knowledge (Enoch 2011: 163). In a moment, I'll indicate why I believe that these claims are mistaken. In the meanwhile, it is worth noting that there is an initial line of response to the epistemological challenge available to non-naturalist realists at which Enoch gestures (see Enoch 2011: 167). The initial response is to note that even if non-naturalists cannot solve the puzzle, this would not imply that we should reject the view. For we can see that before we even canvass various attempts to address the debunker's puzzle, everybody's got problems when it comes to moral epistemology.

To see the point, consider some of the main rivals to non-naturalism. Begin with what is arguably the most radical rival, namely, the moral error theory. Advocates of the error theory, such as Richard Joyce and Jonas Olson, believe that there are no moral facts of any sort. These philosophers do not even attempt to explain how it could be that we reliably track moral reasons. Their view, after all, implies that there is nothing to explain, as we have no moral knowledge. Is this any reason to believe that, when it comes to the epistemology of moral belief, the error theory enjoys an advantage over its rivals? Not obviously. No one thinks, for example, that when it comes to the epistemology of perception, radical solipsism is superior to rival views because it relieves itself of having to explain knowledge of the external world. Better, we think, to have a position according to which we have such knowledge even if that position struggles to explain how we gain it. For it appears that we can know seemingly obvious facts, such as the fact (as I now write) *that there is a coffee cup in front of me*. Arguably, something similar is true of moral knowledge. After all, were the error theory true, we could not know apparently obvious moral truths, such as *that it is wrong to engage in recreational slaughter*. That, however, is not a virtue of the position.

Consider another rival to non-naturalism, the sophisticated expressivism defended by philosophers such as Allan Gibbard (1990, 2003) and Simon Blackburn (1993, 1998, 2005). According to these philosophers, moral judgments express, not attitudes that have moral representational content, which purport to represent the moral facts, but states of commendation or condemnation directed toward various features of non-moral reality. The sophisticated expressivists' position, like the error theory, implies that

our putative ability to reliably track reasons calls for no explanation. For, in their view, while we can *say* that there are moral facts in some minimal or deflationary sense, these moral facts would not be the sort of thing that can be reliably tracked.[8] As Blackburn puts the point in one place, deflated moral facts are simply not the sort of thing that "answer to" our moral judgments (1999: 216). When it comes to the epistemology of moral belief, is this better than a position according to which our moral judgments can track the moral facts, even if we have yet successfully to explain the relation? Not obviously. It is not clearly preferable to accept a position according to which moral judgments could not track the moral facts (or according to which there is nothing to say about the tracking relation) as opposed to one according to which there is, in principle, a substantive story to be told about the tracking relation although we have yet to articulate it.

Now consider a third view, which is more similar to non-naturalism than the two views just considered—namely, the idealized constructivism defended by philosophers such as Roderick Firth (1952) and Michael Smith (1994). These philosophers hold that there are moral reasons that are fixed by the attitudes of idealized agents. Under one prominent version of this view, you have a moral reason to act in a given way just in case and because an idealized version of yourself would want you to want to act in that way.[9] Unlike error theorists and sophisticated expressivists, idealized constructivists maintain that we can and sometimes do reliably track moral reasons. Is this view better placed than non-naturalist realists to explain how we could track these reasons? Not obviously. It is not at all evident that our present culture, time, and evolutionary history have placed us in a favorable position to grasp what idealized agents would care about. There is, after all, no idealized constructivist analogue to Moses on Mt Sinai who reveals to us to the moral truth!

Turn, finally, to non-idealized versions of constructivism, such as that developed by Sharon Street (who perhaps non-coincidentally has forcefully pressed the debunker's puzzle). Advocates of this position tell us that it is an agent's taking something to be a reason that accounts, in part, for it being a reason for that agent. What needs to be added is that an agent's taking something to be a reason generates a reason for that agent only if it withstands scrutiny in the sense of being "fully in line" with her other normative judgments. Since an agent's other normative judgments might not in fact cohere with what she takes to be reasons, Street acknowledges that agents "can be mistaken about what those reasons are" (2008b: 224). Unlike Michael Smith's idealized constructivist view, constructivism of this variety, Street writes, "is strongly Humean in that it accepts that practical reason as such commits us to no particular substantive conclusions about our reasons; depending on one's starting set of values, one could in principle have a reason for anything" (p. 244).

Does a view such as this account for how we could reliably track moral reasons? Perhaps in the first-personal case it does. Perhaps the view explains how I could reliably

---

[8] It is not easy to know what this deflationary sense comes to. For attempts to puzzle through the issue, see Cuneo (2007b: ch. 7; 2008, 2013).

[9] Smith (1994) develops a version of this position.

track the reasons I have, since I know what I take to be reasons and have a good sense of whether these takings are fully in line with my other normative commitments. But it is not at all apparent that non-idealized constructivism explains how we know what reasons others have.[10] To know that, we'd need to have extensive knowledge of what they take to be reasons and whether what they take to be reasons is fully in line with their other normative judgments. Do we have such knowledge? It is not at all apparent that we do.[11] It might also be worth noting that, when engaging with moral naturalists such as David Copp, Street herself makes it evident that she's interested in how we could know, not what morality or prudence favors, but what we have reason to do on the whole. It is knowledge of *decisive* reasons that moral realists are supposed to have difficulty explaining. But it is not at all apparent that non-idealized constructivism explains how we could reliably track reasons of this sort, especially in the third-person case. This is because, in part, it is not apparent what it is for an agent to have a decisive reason according to this view.[12]

In sum, when one considers a sampling of the major options in metaethics, it appears that no one gets a free pass when it comes to moral epistemology. Nearly everyone has bullets to bite or puzzles to solve. So, even if non-naturalists have little to say about how to explain the striking correlation between our moral beliefs and the moral facts which lies at the heart of the debunker's puzzle, it does not follow that, when it comes to the epistemology of moral beliefs, their view is the worst of a bad bunch.

Still, it must be admitted that the line of reply just voiced is what I earlier called a "shallow" response to the debunker's puzzle, since it sheds no light on how there could be a reliable correlation between our moral beliefs and the moral facts. To say it again, it merely implies that, when it comes to moral epistemology, the non-naturalists' worries are not clearly more pressing than those of their rivals. Non-naturalists would like to do better. Enoch thinks they can.

Central to Enoch's reply to the debunker's puzzle are two observations. First, the correlation between our moral judgments and the moral facts that needs to be explained is not that striking. No realist holds that competent moral agents have reliable moral beliefs across the board. Moral issues are simply too difficult and complex for that. What needs to be explained, then, is a correlation between a certain subset of moral beliefs and the moral facts—the moral facts perhaps being ones that are general and plainly evident, such as the facts *that it is wrong to engage in recreational slaughter* or *that it is wrong to torture others for mere pleasure*.[13]

Second, given a starting point of normative beliefs that are not too far off, there are presumably some reasoning mechanisms that can get us "increasingly closer to the truth by eliminating inconsistencies, increasing overall coherence, eliminating arbitrary distinctions,

[10] I owe this observation to Selim Berker.
[11] If knowledge of one's own reasons sometimes depends on our knowledge of others' reasons, this last point has the implication that we may not be capable of knowing some of our own reasons.
[12] On this topic, see Schroeder (2007).
[13] Some years ago, Boyd (1988: 201) made a similar observation.

drawing analogies, ruling out initially justified beliefs whose justificatory status has been defeated," and so on (Enoch 2011: 166). The activity of engaging in reflective equilibrium, which incorporates many of these reasoning procedures, is one such mechanism.[14]

It might be, for the reasons just cited, that the challenge facing non-naturalists is not as daunting as some have claimed. Even so, the challenge appears to be serious enough. Non-naturalists must hold, after all, that our moral judgments do not determine the moral facts and that the moral facts do not cause our moral judgments. But if this is so, what options are left for non-naturalists to explain the reliable correlation? In Enoch's view, non-naturalists should look for a "third-factor" explanation. For sometimes, Enoch writes, the "explanation of a correlation between [. . .] two factors A and B is in terms of a third factor, C, that is (roughly speaking) responsible both for A-facts and for B-facts" (2011: 167). For example, consider so-called pre-established harmony views about the mind defended by early modern philosophers such as Leibniz. These views imply that our beliefs and intentions do not themselves cause bodily movements. Rather, a third factor, God, ensures that our body moves in certain ways when we have certain beliefs and intentions. The question is what, in the case of our moral beliefs, such a third factor could be.

Enoch proposes the following: suppose we assume that our survival or reproductive success (or whatever evolution "aims" at) is good—by and large better than the alternative. (The sense of "good" here is not supposed to be moral.) If this assumption is correct, and it is true that the forces of natural selection "aim" at survival or reproductive success, then it will be the case that the forces of natural selection will have "pushed" us (or ancestral humans) to form normative judgments that are at least somewhat in line with the normative truths. "The causal influence of selective forces," Enoch writes,

only directly "pushes" us in the direction of having *evolutionarily beneficial* beliefs, not necessarily true ones. But here as elsewhere, the two may be systematically related. For we are the kind of creatures whose actions seem to be closely related to their normative beliefs about how they should act, or how it would be good to act, or what consequences it would be good to bring about [. . .] And this completes the explanatory story needed here: survival (or whatever) is good; so behaving in ways that promote it is (pro-tanto) good; but one efficient way of pushing us in the direction of acting in those ways is by pushing us to believe that it is good to act in those ways. And in fact, as we have just seen, it is good so to act. So the normative beliefs this mechanism pushes us to have will tend to be good. (Enoch 2011: 169)

If we grant this, Enoch maintains, then the rest of an explanation begins to fall into place. For

the fact that survival is good is plausibly related in coherence relations to many (though perhaps not all) other normative truths, like that pain is pro-tanto bad, that

[14] DePaul (1993) offers a sustained treatment and defense of reflective equilibrium in the moral domain.

some close relationships are good to have, etc. Relations of this nature between that survival is good and many other normative truths [. . .] explain the correlation between the normative truths and our normative beliefs [. . .] (Enoch 2011: 169)[15]

Enoch enters several caveats about his proposal. To begin with, this putative explanation of the reliable correlation is supposed to be a "just so" story. It does not purport to actually explain the correlation; rather the point is that it *could*, for all we reasonably believe, explain the correlation. In principle, we could appeal to other ways of developing the third-factor strategy.[16] In addition, much more work would have to be done to explain how the third-factor strategy would get us from mere evaluative beliefs, such as the belief that survival is good, to specifically moral ones. Furthermore, Enoch is willing to concede that, although he holds that the third-factor strategy does indeed enable non-naturalists to make progress toward solving the debunker's puzzle, it countenances a miracle of a sort. For it might be true that had the causal forces shaping our intellectual and other normative faculties been very different—had they "aimed" at things that are of no value at all or that are of disvalue—then "we would have been systematically mistaken in our normative beliefs" (Enoch 2011: 173). So, we are, Enoch says, lucky to have evolved in the environment that we did. But this sort of luck, Enoch maintains, seems relatively innocuous. In some sense, when it comes to *any* of our basic epistemic faculties that are in fact reasonably reliable, we are lucky that they have been shaped in such a way as to be reasonably reliable. If that is right, while non-naturalists might have to admit that there is a sense in which moral knowledge is a miracle, there is no special miracle in the normative case.

41.3 TWO WORRIES

According to Enoch's proposal, the third factor that explains the reliable correlation between our moral judgments and the moral facts is the complex fact that survival (or the maximization of the relative representation of genes in the gene pool over generations) (i) is what evolution "aims" at and (ii) is good.[17] What should we make of this proposal? The knee-jerk response is to pronounce it question-begging. In replying to the debunker's puzzle—so it might be said—one cannot simply help oneself to the truth of

[15] Elsewhere Enoch writes: "Selective forces have causally shaped our normative beliefs; that survival is good is (non-causally but closely) related to many normative truths; and so that survival (or whatever the evolutionary "aim" is) is good explains the correlation between our normative beliefs and the normative truths" (pp. 169–70).

[16] Enoch cites Parfit (2001), Skarsune (2011), and Wielenberg (2010) as developing similar strategies.

[17] Recall that, according to Enoch's official gloss, a third-factor explanation is one that explains "a correlation between [. . .]two factors A and B is in terms of a third factor, C, that is (roughly speaking) responsible both for A-facts and for B-facts" (2011: 117). Strictly speaking, Enoch's own proposal does not conform to this gloss, for the fact *that survival is the "aim" of evolution* does not explain the moral facts.

substantive normative beliefs, such as the belief that our survival (or whatever evolution "aims" at) is good.

It is not, however, so easy to make the charge of question-begging stick.[18] The advocate of the debunker's puzzle cannot, after all, simply assume at the outset of the argument that our common-sense normative beliefs—such as the belief that our survival is good—are *already* in doubt. If that assumption were made, the jig would be up for non-naturalists at the very outset of the argument. Rather, the charge must be that the puzzle provides non-naturalist realists with a defeater, a reason to *stop believing* what they may have default entitlement to believe. If this is right, the issue is whether the puzzle does indeed provide a defeater for accepting propositions such as that our survival is good.

In a series of papers, William FitzPatrick has defended a non-naturalist metaethical view very close to Enoch's (see e.g. FitzPatrick 2008, 2011). Still, FitzPatrick maintains that non-naturalists should not embrace Enoch's third-factor solution to the debunker's puzzle. For, when understood in a certain way, FitzPatrick argues, the debunker's puzzle does in fact provide a defeater for our commonsensical normative beliefs.

To see FitzPatrick's point, bring to mind the third claim that composes our original puzzle, namely:

(3) Evolutionary influence: evolutionary forces have not only determined which cognitive faculties we have but also heavily shaped their workings, including the workings of that faculty (of cluster of faculties) by which we form moral judgments.

FitzPatrick notes that there are different ways to understand the claim that evolutionary influences have "heavily shaped" the workings of our moral faculty. One way is what he calls the:

Extreme Explanatory Claim (EEC): the content of our moral beliefs *across the board*:
(i) reflects deep shaping by evolutionary forces that operated on the moral belief-forming dispositions of ancestral humans, according to principles insensitive to the truth of the content of moral beliefs; and
(ii) does *not* reflect any independent influence from developed forms of moral reflection *guided by independent moral facts as such*, through a grasp of their grounds as such. (FitzPatrick 2014b: 247)

(EEC) incorporates claims about both the scope and the depth of evolutionary influence. According to (EEC), the scope of evolutionary influence is wide: no moral beliefs have escaped it; the entire fund of moral beliefs has been affected. The depth of the influence, moreover, goes very deep: other independent influences of belief formation, revision, and maintenance, such as practices of moral training and reflection that involve the grasp of moral facts, have not counteracted these evolutionary

[18] As both Locke (2014) and Schafer (2010) note.

forces; to the contrary, the influence of these forces have saturated the entire fund of moral beliefs.

Suppose, for argument's sake, that the debunker's puzzle relies on (EEC). FitzPatrick's claim is that, if it does, it is difficult to see how third-factor explanations could be an adequate response to it. For when we run the response, we must trust certain normative judgments (or seemings), such as the judgment that our survival is good and that it is wrong to engage in recreational slaughter. These judgments may enjoy a default entitlement, an innocent-until-proven guilty status. But once we grant (EEC), FitzPatrick argues, this entitlement is defeated. For (EEC) implies that we cannot trust any of our moral judgments, including those on which the third-factor reply relies. If FitzPatrick is right about this, then an adequate reply to the debunker's puzzle—at least when it is understood to incorporate (EEC)—cannot simply appeal to a third-factor strategy. For any way of running the reply will have to rely on moral judgments that the puzzle calls into question.

It might be that a position such as Enoch's has the resources to respond to this first concern, although marshalling these resources may involve going beyond simply appealing to aspects of the third-factor strategy itself. There is, however, another, equally serious worry regarding the third-factory strategy that deserves attention. Recall that, in Enoch's view, the debunker's puzzle is best understood as drawing our attention to an apparently inexplicable correlation between our moral beliefs, on the one hand, and moral facts or truths, on the other. The correlation is that very often when we accept a moral proposition p, it is true that p; and very often when we do not accept a moral proposition p—such as when we reject p—it is indeed false that p (Enoch 2011: 159). Explain this correlation, Enoch maintains, and other versions of the epistemological challenges—ones that concern how we gain epistemic access to moral facts or gain moral knowledge—appear "redundant" (p. 163).

This last claim, however, is open to question. To see why, let me introduce a distinction between two ways in which our beliefs can be reliable to which epistemologists sometimes appeal. One way in which beliefs of a certain range can be reliable is when a sufficiently large portion of them is true. Call this *generic reliability*. Another way in which beliefs of a certain range can be reliable is when a sufficiently large portion of our beliefs are true because they are sensitive to features of the world that make them true (or are sensitive to features of the world that ground these truths)—where what makes these beliefs true corresponds to their contents. Call this *source reliability*, since the reliability is a matter of a believer's being sensitive to the source (or the grounds of the source) of their truth.[19]

Generic reliability, as many have pointed out, is compatible with a sufficiently large portion of our beliefs being true by accident. To borrow an example from Hartry Field, it is possible that by some lucky accident a sufficiently large portion of your beliefs about

---

[19] For present purposes, I aim to remain non-committal regarding that in which sensitivity consists, but I do not assume that it necessarily involves being in a position to be aware of the source of a belief.

a remote Nepalese village that you've never seen (or have much information about) are true (Enoch 2011: 158–9). Generic reliability is also compatible with our beliefs being true not because of luck but because of some systematic orchestrated correlation that has nothing to with our beliefs being sensitive to how the world is. Conceivably, for example, something like Descartes's Evil Genius could program you in such a way that you experience a series of hallucinations that reliably correlate with what is actually happening in your environment. In this case, a sufficiently large portion of your beliefs about your surroundings would be true not by accident but because of the systematic orchestrated correlation enacted by the Evil Genius. Source reliability, by contrast, is compatible with neither of these scenarios. If a sufficiently large portion of a range of your beliefs is true because it exhibits source reliability, then that is because it is sensitive to the source of their truth (or the grounds of these truths), such as the external objects in your surroundings.

Suppose, for argument's sake, we grant that a third-factor explanation such as Enoch's is successful in this sense: it explains why a sufficiently large range of our moral beliefs is reliably correlated with the moral facts. If so, third-factor explanations can account for why a range of our moral beliefs exhibits generic reliability. But the fact that these beliefs exhibit generic reliability does not imply that they exhibit source reliability. Indeed, something stronger can be said: if these beliefs exhibit generic reliability for the reasons Enoch offers, it appears that they do *not* exhibit source reliability. This is because, according to the third-factor proposal, the fact that a range of our beliefs is generically reliable is due to a systematic orchestrated correlation that has nothing to do with our (or our ancestors') being sensitive to normative reality. The orchestrator in this case, to be sure, is not a Cartesian Evil Genius. It is, rather, the forces of natural selection, aided by the activity of reflective equilibrium in which we engage in some epistemic "house-cleaning" by doing such things as eliminating inconsistencies in our belief system. Arguably, though, if it were true that our moral beliefs are produced in this way, then adverting to this truth would not be adequate to respond to the debunker's puzzle. For the debunker's puzzle is supposed to cast doubt on whether our moral beliefs are *well-formed*. It is highly plausible to hold, however, that for our moral beliefs to be well-formed, they must be sensitive to moral reality (or the grounds thereof). Imagine, for example, you correctly believe that engaging in recreational slaughter is wrong. Your belief would not be well-formed if it had nothing to do with the fact that recreational slaughter is wrong (aside from being about it). For it to be well-formed, your belief must have something to do with the fact that such slaughter is wrong, such as by incorporating a grasp of the grounds of its wrongness.

Let me summarize: third-factor explanations attempt to account for the reliable correlation between a range of our moral beliefs and the moral facts. They do not do so by maintaining that our moral beliefs determine the moral facts or that the moral facts cause our moral beliefs. Instead, they appeal to some third factor—in Enoch's case, the fact that our survival is good and that it is what the forces of natural selection "aim" at—to account for the reliable correlation between a range of our moral beliefs and the moral

facts. A view such as this faces two problems: first, when understood in a certain way, the debunker's puzzle generates a defeater for the normative beliefs to which advocates of the third-factor strategy appeal when defending their view. And second, the third-factor approach does not yield the conclusion, which is required to solve the debunker's puzzle, that our moral beliefs are well-formed. In fact, if this second concern is on target, it follows that third-factor strategies do not render other versions of the epistemological challenge redundant, as Enoch claims. For they offer no explanation whatsoever for how we could grasp the moral facts that would render our moral judgments true.

41.4 ADDRESSING THE FIRST CONCERN

If the argumentation in the last section is cogent, non-naturalists are free to employ third-factor strategies when replying to the debunker's puzzle. But if they do, they will need to supplement these strategies in such a way that they address the two concerns we just canvassed. In this section, let's explore some of the resources available to non-naturalists.

Let us begin with the first concern articulated in the last section. In his discussion of the debunker's puzzle, FitzPatrick stresses that nearly everyone party to the discussion of whether our moral judgments are well-formed accepts that evolutionary forces have influenced the workings of our cognitive faculties, including the moral faculty. The crucial question is *to what extent* evolutionary forces have influenced the workings of the moral faculty, and whether its influence has been counteracted by other influences, such as reliable moral reflection.

Here are two possibilities. On the one hand, the debunker's puzzle may rely on a very strong claim, such as the:

Extreme Explanatory Claim (EEC): the content of our moral beliefs *across the board*:

(i) reflects deep shaping by evolutionary forces that operated on the moral belief-forming dispositions of ancestral humans, according to principles insensitive to the truth of the content of moral beliefs; and

(ii) does *not* reflect any independent influence from developed forms of moral reflection *guided by independent moral facts as such*, through a grasp of their grounds as such.

On the other, it may rely on a significantly weaker claim, such as:

Modest Explanatory Claim (MEC): some familiar but suspect moral beliefs, such as those that concern racial purity, rigid gender roles, and clan loyalty, plausibly reflect only evolutionary influences unguided by the moral facts (and conditioned by cultural influences operating equally independently of the moral facts), which are

thus unlikely to be reliable and hence (given our knowledge of this) unjustified. (Cf. FitzPatrick 2014a.)

We have seen that if the debunker's puzzle relies on the first claim, (EEC), then it is difficult to see how, when responding to this puzzle, non-naturalists could avail themselves of the third-factor strategy, at least when it is taken as a self-standing response. But if the debunker's puzzle relies on only the second claim, (MEC), then it is easy to see that there is no deep puzzle for non-naturalists to solve. For while (MEC) concedes that some of our moral beliefs are suspect, it is compatible with other of our moral beliefs being reflective and informed, the result of our having grasped the moral facts and employed reliable strategies of refining our beliefs, such as reflective equilibrium.

Thus described, (MEC) is indeed a modest thesis, since it makes no sweeping claims about the scope and degree of evolutionary influence on our moral beliefs. But, as FitzPatrick points out, it is modest in another important sense, since it requires rather little of evolution. For, while non-naturalists must hold that the moral faculty yields a range of well-formed moral beliefs, they needn't defend the further thesis that evolution furnished this faculty. All that's needed is for evolution to have given us

the basic raw materials—reflective, intellectual, and emotional potentialities—necessary for us to develop reliable moral belief-forming dispositions ourselves, by developing those potentialities through the right forms of experience, training and reflection in rich cultural contexts, in such a way as to come reliably to track moral truths through gaining understanding. This is directly analogous to our cultural development in other domains of reliable capacities to track truths about non-linear algebra or quantum non-locality or metaphysical modality, none of which played any more role in the evolution of Pleistocene human cognitive capacities than moral truths did. (FitzPatrick 2014b: 245)

Let us suppose, then, that there are two different ways to understand the debunker's puzzle, one which depends on (EEC), another which relies on (MEC). The question that faces non-naturalists is which of these two claims they should accept.

Well, note that the relevant sciences, such as evolutionary biology, will not settle the issue. They do not tell us or imply that (EEC) is more likely to be true than (MEC). How, after all, would the relevant sciences, such as evolutionary biology, give us good reason to believe that the workings of the moral faculty do *not* reflect any independent influence from developed forms of moral reflection guided by independent moral facts, through a grasp of them or their grounds? Nor, for that matter, can proponents of the debunker's puzzle simply assume that (EEC) is true, for that would be to assume what needs to be established. What we need at this point in the dialectic is a compelling argument that non-naturalists should accept (EEC) rather than (MEC). While there might be such an argument, it is worth noting that, at this point in the discussion, advocates of

the debunker's puzzle have not produced it—the main reason being that they have not distinguished (EEC) from (MEC), as FitzPatrick does.

41.5 Addressing the Second Concern

The conclusion we reached in the last section is that the debunker's puzzle has force only when it incorporates an extremely ambitious claim, namely, (EEC), for which the debunkers offer no direct argument. Suppose that this conclusion is correct. Would it entirely drain the puzzle of its force? Perhaps not. When presented with the argument in the last section, advocates of the debunker's puzzle could retrench, offering an indirect argument for the claim that we should accept (EEC) rather than (MEC). While any such argument would probably have multiple strands, an initially attractive way to make a case for (EEC) would be to focus on its second clause, which states that the content of our moral beliefs:

> (ii) does *not* reflect any independent influence from developed forms of moral reflection *guided by independent moral facts as such*, through a grasp of their grounds as such.

Call this *the non-sensitivity thesis*. If an argument could be made for this thesis, then it could be the sort of consideration that could tip the balance in favor of accepting (EEC) rather than (MEC), thereby revitalizing the debunker's puzzle.

It is, however, one thing to identify a strategy for revitalizing the debunker's puzzle; it is another to provide the arguments to execute the strategy. Is there a promising way to argue for the non-sensitivity thesis? One promising approach would be to pick up the second concern raised earlier regarding Enoch's third-factor strategy. Recall that, according to this concern, to establish that our moral beliefs are well-formed, these beliefs must be not merely reliably correlated with but also sensitive to the moral facts. (MEC), we have noted, is compatible with our moral beliefs being sensitive to the moral facts. But it does not imply that they are or explain how they could be so. Debunkers could attempt to exploit this point, arguing that non-naturalists have nothing helpful to say about how our moral beliefs could be sensitive to the moral facts. The problem, these philosophers might claim, is that it is hard to imagine what the link could be between our moral beliefs and the moral facts such that the former are sensitive to the latter. Third-factor strategies, we have seen, do not seem to help; they give no indication of what the link could be. Moreover, non-naturalists cannot maintain—at least without risking collapsing their view into a version of naturalism—that the link in question is a causal one. If this is so, the debunkers might continue, the non-sensitivity thesis might have a lot going for it, enough to give us reason to accept (EEC) rather than (MEC).

Were advocates of the debunker's puzzle to pursue this strategy, it would yield a pair of interesting results. In the first place, it would highlight the fact that the debunker's puzzle

is no more a self-standing objection to non-naturalism than the third-factor reply is a self-standing response to the debunker's puzzle. To get off the ground, the puzzle, like the third-factor response, needs to be supplemented by additional arguments that bear much of the dialectical weight. In the second place, this strategy would shift the discussion of whether non-naturalists have a viable moral epistemology back to familiar concerns about how we could gain epistemic access to moral facts that lie outside the causal order. These concerns are familiar because they are the same that face everyone who accepts the existence and knowledge of abstracta such as properties, relations, numbers, sets, and the like, which are causally inert. If this is so, the place to look for a response to this challenge, it would seem, would be to accounts of how we grasp the abstract realm.

41.6 Addressing Non-Sensitivity

As it happens, philosophers such as John Bengson (2015) have developed such accounts. The proposal that Bengson advances, it is worth emphasizing, is not presented as the sober truth about how we grasp members of the abstract realm, such as non-natural moral facts. Instead, Bengson presents it as a promising model for how we might do so. In what remains, it will prove fruitful, I believe, to investigate this model in more detail. The reason is that by availing themselves of something like Bengson's model, non-naturalists might be better positioned to offer a deeper reply to the debunker's puzzle, one that goes some distance toward explaining why they should reject the non-sensitivity thesis. Admittedly, a deeper reply of this sort may not be dialectically persuasive; it might not rationally convince anyone skeptical of non-naturalism. But, if all goes well, it would illustrate that non-naturalists have more to say about the issue of non-sensitivity than many of their critics assume.

Let us remind ourselves of the challenge that faces non-naturalists. Non-naturalists hold that in order for our moral beliefs to be well-formed, they must be not only reliably correlated with but also sensitive to a range of the moral facts. The sensitivity in question cannot be a brute, inexplicable fact. Nor can it be a function of our moral beliefs constituting the moral facts. Nor can it be a matter of the moral facts causing our moral beliefs. If none of these answers is available, it is natural to wonder what the link could be.

Bengson suggests that this natural question has a natural answer, which is that our mental states bear some type of non-causal relation to abstracta that explains how we could apprehend them or get them in mind. To help us see that we might be able to bear such a relation to abstract objects of thought, Bengson asks us to consider a thought experiment. Imagine a person, Trip, who has never before encountered the colors red, orange, or blue. Nor has he ever encountered elliptical, circular, or hexagonal shapes. One evening, Trip experiences a vivid hallucination. When undergoing this experience, Trip

"sees" three items: a red ellipse, an orange circle, and a blue hexagon. On the basis of this experience, Trip forms these beliefs:

The color of the first item (the ellipse) resembles the color of the second item (the circle) more than the third item (the hexagon).

And:

The shape of the first item (the ellipse) resembles the shape of the second item (the circle) more than the shape of the third item (the hexagon).

It is plausible to believe that Trip now knows these things. Suppose, however, that a broadly realist account of colors and shapes is true and colors and shapes are mind-independent entities. And suppose—as it seems plausible to do—that Trip has had a perceptual experience of features—namely, shapes, colors, and resemblance relations—to which his experience bears no causal connections (after all, he has simply had an hallucinatory experience).[20] How could it be that, on the basis of his hallucination, which bears no causal connection to colors and shapes, he knows facts about colors and shapes? To explain this, Bengson suggests, we need to identify a non-causal relation between Trip's hallucinatory experience and the relevant facts about colors and shapes that explains how his experience can be non-accidentally correct with respect to those facts, providing him with knowledge of them.

Bengson's proposal is that Trip's mental state is not causally but rather *constitutively* related to that which it is about; his experience is constitutively tied to the colors and shapes, which is why it can serve as a source of knowledge about them. If this is right, the constitution relation is the non-causal relation that Trip's mental state bears to what it is about; it explains how he can know colors and shapes and the relations that they bear to one another.

As it is usually understood, the constitution relation has a number of formal characteristics, such as being not only non-causal but also an irreflexive, asymmetric ontological dependence relation. To illustrate: consider the relationship that a chunk of glass bears to a vase. If that glass constitutes the vase, the relationship is non-causal: the glass does not cause the vase to come into being. It is irreflexive: the vase is constituted by the piece of glass but the vase is not constituted by itself. It is also asymmetric: the vase is constituted by the piece of glass but the piece of glass is not constituted by the vase. Moreover, the glass is not one of the vase's parts. Nor does the vase contain the piece of glass. Nor does the vase merely modally co-vary with the piece of glass. Nor, finally, are the two identical. Rather, the vase constitutively depends on the piece of glass.

While hardly uncontroversial, appeal to the constitution relation is common in philosophy. Philosophers frequently appeal to the relation when puzzling through how it is

[20] Bengson (2015: 16–17) defends this claim, which is fairly controversial.

that pieces of glass are related to vases, bodies to persons, events to properties, and games to norms. When they do, they say such things as: the piece of glass constitutes the vase, a person is constituted by her body, the event of a person's shouting is constituted by that person and the property of shouting (at a time), and the game of chess is constituted by the norm *that the bishop moves any number of vacant squares in any and only a diagonal direction.*

We are, however, interested not simply in the properties of the constitution relation but also how it explains various phenomena. To that end, take a case in which we appeal to the constitution relation in the context of an explanation. For example, consider the statement that:

The vase is fragile because it is constituted by a piece of glass.

This, on the face of things, is a *constitutive explanation* of the fragility of the vase. The fragility of the vase is explained in terms of the fragility of the piece of glass of which the vase is constituted.

With examples such as this in mind, Bengson offers the following hypothesis:

When a's having F ensures b's having G, it may be possible to explain the fact that b is G by citing its constitution: b is G because it is *constituted* by a (which is F). (Bengson 2015: 22)[21]

So, to employ the example with which we've been working, if a piece of glass has the property of *being fragile* and a vase is constituted by this piece of glass, then the vase has the property of *being fragile* because it is constituted by the glass, which is fragile. In this case, the hypothesis seems to yield the right result.

Let us now return to Trip. As other philosophers have noted, it is plausible to suppose that Trip's hallucinatory experience is not caused but constituted by the relevant colors and shapes (see Hawthorne and Kovakovitch 2006). If this supposition is correct, then it can be used to explain the ability of Trip's hallucinatory experience to serve as a source of knowledge about those colors and shapes. The explanation would be that Trip's hallucinatory experience is non-accidentally correct, and able to serve as a source of knowledge about the relevant colors and shapes because it is partially constituted by those colors and shapes. If a constitutive explanation of this sort is on target, it pinpoints what makes it the case that Trip is in a position to know what he does.

Can we extend this basic account to other types of knowledge? In Bengson's view, we can. Consider ordinary empirical knowledge, such as the state of my knowing *that there is a coffee cup in front of me.* When I know this fact via perception, we can say that this fact constitutes my perceptual state. My perceptual state has a chunk of the world as a constituent, namely, the fact *that there is a coffee cup in front of me.* Or, to come closer

[21] Bengson (2015: 21) glosses the notion of ensurance thus: a's having F *ensures* b's having G iff: necessarily, if a is F and a constitutes b, then b is G.

to our target, take the case in which I know some putative mind-independent abstract fact such as *that the number two is the smallest prime*. If I know this fact, I do not perceive it but apprehend or intuit it. If the model with which we are working is correct, when I know this fact via apprehension or intuition, this fact constitutes my state of apprehension or intuition. This state of apprehension or intuition also has a chunk of the world as a constituent, albeit an abstract chunk of the world. Of course all these claims are controversial. But it is worth stressing that this approach has the virtue offering a unified account of knowledge states: hallucinatory, perceptual, and intuitive knowledge all have the same fundamental structure.

Let us take a final step. In principle, moral knowledge might come in different varieties, and concern different types of moral fact. But some such knowledge, presumably, concerns general moral truths or facts. So, consider a case in which an agent knows the general non-natural moral fact *that it is wrong to engage in recreational slaughter*. If an agent knows such a fact, then presumably it is also via apprehension or intuition, since it is abstract. According to the constitutive explanation model, this fact constitutes the state of moral apprehension or intuition, which, when all goes well, is itself a constituent of a state of moral knowledge. If it does, then non-naturalists have at their disposal a model for understanding moral knowledge. Moral knowledge is just like hallucinatory, empirical, and intuitive knowledge: it is not caused but constituted by the facts that it is about—in this case, the moral facts. When you know a moral fact, according to this model, you are grasping the fact itself; it is part of your mental state.

There is a great deal more to say about the constitutive explanation model that Bengson offers. (I have, for example, said nothing about how to understand non-veridical hallucinatory, perceptual, or intuitive states.) And it would be wishful thinking to hold that simply presenting this model is likely to quell misgivings about non-naturalism's epistemological commitments. Still, when philosophical issues are complex and difficult—as is moral epistemology—small steps, such as introducing promising models for understanding what the link might be between moral judgment and moral reality, can be helpful. And, in this case, I believe there is reason to hold that such a step has been taken.

41.7 Conclusion

Let us take stock. If the debunker's puzzle has force against moral non-naturalism, then it must incorporate a very ambitious claim about the extent to which evolutionary forces have operated on the workings of the moral faculty. We have noted that advocates of the debunker's puzzle have not offered arguments for this ambitious claim. Non-naturalists, then, are well within their rights to reject it in favor of more modest claims, which maintain that evolution's influence is not nearly as extensive as the debunkers claim and allows for this influence to be counteracted by states of moral understanding wherein we grasp the moral facts. To breathe new life into their objection, debunkers

could defend what I have called the non-sensitivity thesis. According to this thesis, moral beliefs do not reflect any independent influence from developed forms of moral reflection guided by independent moral facts. Defenders of the debunker's puzzle might argue that we have excellent reason to accept this thesis—and, hence, ambitious claims about evolution's influence—since non-naturalists have nothing illuminating to say about how we might grasp the moral facts. I have suggested that this is not so. By appealing to constitutive explanations, non-naturalists do have helpful things to say about what the link might be between our moral judgments and moral reality. If this is correct, non-naturalists have at hand a promising way to develop a deeper response to the debunker's puzzle.

References

Bengson, John (2015). "Grasping the Third Realm." *Oxford Studies in Epistemology* 5: 1–38.
Blackburn, Simon (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.
Blackburn, Simon (1998). *Ruling Passions*. Oxford: Oxford University Press.
Blackburn, Simon (1999). "Is Objective Moral Justification Possible on a Quasi-Realist Foundation?" *Inquiry* 42: 213–28.
Blackburn, Simon (2005). *Truth: A Guide*. Oxford: Oxford University Press.
Boyd, Richard (1988). "How to Be a Moral Realist." In Geoffrey Sayre-McCord (ed.), *Essays on Moral Realism*, 181–228. Ithaca, NY: Cornell University Press.
Copp, David (2007). "Why Naturalism?" In *Morality in a Natural World*, 33–55. Cambridge: Cambridge University Press.
Copp, David (2008). "Darwinian Skepticism about Moral Realism." *Philosophical Issues* 18: 184–204.
Cuneo, Terence (2006). "Moral Facts as Configuring Causes." *Pacific Philosophical Quarterly* 87: 141–62.
Cuneo, Terence (2007a). "Recent Faces of Moral Non-naturalism." *Philosophy Compass* 2: 1–32.
Cuneo, Terence (2007b). *The Normative Web*. Oxford: Oxford University Press.
Cuneo, Terence (2008). "Moral Realism, Quasi-Realism, and Skepticism." In John Greco (ed.), *The Oxford Handbook of Skepticism*, 176–99. Oxford: Oxford University Press.
Cuneo, Terence (2013). "Properties for Nothing, Facts for Free? Expressivism's Deflationary Gambit." *Oxford Studies in Metaethics* 9: 223–51.
Cuneo, Terence, and Russ Shafer-Landau (2014). "The Moral Fixed Points: New Directions for Moral Non-naturalism." *Philosophical Studies* 171: 339–443.
DePaul, Michael (1993). *Balance and Refinement*. New York: Routledge.
Enoch, David (2011). *Taking Morality Seriously*. Oxford: Oxford University Press.
Fine, Kit (2002). "Varieties of Necessity." In John Hawthorne and Tamar Gendler (eds), *Conceivability and Possibility*, 253–82. Oxford: Oxford University Press.
Firth, Roderick (1952). "Ethical Absolutism and the Ideal Observer." *Philosophy and Phenomenological Research* 12: 317–45.
FitzPatrick, William (2008). "Robust Ethical Realism, Non-Naturalism, and Normativity." *Oxford Studies in Metaethics* 3: 159–206.
FitzPatrick, William (2011). "Ethical Non-Naturalism and Normative Properties." In Michael Brady (ed.), *New Waves in Metaethics*, 7–35. Basingstoke: Palgrave Macmillan.