

# PART I

## Evolutionary Debunking Arguments



## 2

# Debunking and Dispensability

*Justin Clarke-Doane*

In his précis of a recent book, Richard Joyce writes, ‘My contention . . . is that . . . any epistemological benefit-of-the-doubt that might have been extended to moral beliefs . . . will be neutralized by the availability of an empirically confirmed moral genealogy that nowhere . . . presupposes their truth’ (2008 p. 216). Such reasoning—falling under the heading ‘Genealogical Debunking Arguments’—is now commonplace. But *how* might ‘the availability of an empirically confirmed moral genealogy that nowhere . . . presupposes’ the truth of our moral beliefs ‘neutralize’ whatever ‘epistemological benefit-of-the-doubt that might have been extended to’ them? In this chapter I argue that there appears to be no satisfactory answer to this question. The problem is quite general, applying to all arguments with the structure of Genealogical Debunking Arguments aimed at realism about a domain meeting two conditions. The Benacerraf–Field Challenge for mathematical realism affords an important special case.

### 2.1 Justifying and Undermining

Let *Harman’s Thesis* be the view that moral truths, realistically construed, are not implied by the best explanation of any of our ‘observations’.<sup>1</sup> An observation in Harman’s sense is an ‘immediate judgment made in response to the situation without any conscious reasoning’ (1977 p. 208), where a judgment is a mental event rather than a propositional content. Occurrent beliefs of all sorts may qualify. What is the epistemological upshot of this thesis, if it is true?

<sup>1</sup> Realism about an area, D, is roughly the view that D-sentences should be interpreted literally, and that some atomic or existentially quantified ones are true counterfactually, constitutively, and causally independent of anyone’s believing them to be. For a detailed explication of “D-realism,” see Clarke-Doane (2012 : Sec. 1).

Harman suggested one upshot. If the contents of our moral beliefs are not implied by the best explanation of any of our observations, then we cannot *empirically justify* them.<sup>2</sup> He writes: ‘Observation plays a part in science it does not appear to play in ethics, because scientific principles can be justified ultimately by their role in explaining observations... by their explanatory role... [M]oral principles cannot be justified in the same way’ (1977 p. 10). Harman’s point is evidently that there is no ‘Indispensability Argument’ for ‘moral principles’. Suppose that one claims to endorse the best (scientific) explanation of our observations of a kind, but refrains from believing our mathematical theories. Then *prima facie* one manifests incoherent beliefs. Any rigorous formulation of the former would be framed in mathematical language. Absent an argument that that language is redundant, or can be interpreted at other than face-value, if one endorses the former, then one ought (epistemically) to endorse the latter.<sup>3</sup> If these considerations afford the only *dialectically effective* argument for mathematical realism, as some philosophers allege, then perhaps Harman’s Thesis implies that there is no dialectically effective argument for moral realism.<sup>4</sup>

Does anything else follow? The key idea behind ‘Genealogical Debunking Arguments’ is that (knowledge of) Harman’s Thesis also *undermines* our moral beliefs.<sup>5</sup> Again, Joyce writes: ‘My contention... is that... any epistemological benefit-of-the-doubt that might have been extended to moral beliefs... will be neutralized by the availability of an empirically confirmed moral genealogy that nowhere... presupposes their truth’ (2008 p. 216).<sup>6</sup> Harman’s Thesis implies that there is a ‘moral genealogy [explanation of our moral observations] that nowhere... presupposes’ the truth of our moral beliefs. If no moral truths are implied by the best explanation of any of our observations, then, in particular, the truth of our moral beliefs is not implied by the best explanation of our moral

<sup>2</sup> I define ‘empirical justification’ at the end of this section.

<sup>3</sup> See Quine (1948) and Putnam (1971).

<sup>4</sup> For the claim that these considerations afford the only dialectically effective argument for mathematical realism see Field (1980: Preface, Sec. VII).

<sup>5</sup> Or it ‘undermines them, realistically construed’. The quoted phrase can mislead, however. Our moral beliefs have whatever contents they have. If realists are right about those contents, then Harman’s Thesis seems to undermine our moral beliefs simpliciter. If they are wrong, then it does not seem to undermine them, simpliciter (I will assume). Undermining is not literally relative to a construal, as debunkers sometimes seem to suggest. (More exactly still: Harman’s Thesis undermines our *non-logical* moral beliefs, realistically construed. Debunkers ought not deny that our belief that either Hitler was wicked or it is not the case that Hitler was wicked remains justified. I will ignore this complication in what follows.)

<sup>6</sup> While Joyce does not use the word ‘undermine’ in this quotation, he makes it clear elsewhere that this is what he intends. See, for instance, his (2006, 181). Note that ‘undermine’ is always shorthand for ‘all-things-considered undermine’ in what follows. Similarly for ‘undercut’ and ‘defeat’.

observations. Hence, if ‘the availability of an empirically confirmed moral genealogy that nowhere . . . presupposes’ the truth of our moral beliefs undermines our moral beliefs, then a fortiori, presumably, so does Harman’s Thesis. Of course, if we thought that Quine (1951 Sec. VI) was right that our belief that *p* is justified only if *p* is implied by the best explanation of (some of) our observations, then, rather than undermining our moral beliefs, Harman’s Thesis would give us reason to believe that they were never justified to begin with.<sup>7</sup> But debunkers cannot just be appealing to Quine’s constraint on justification. First, few accept Quine’s constraint and debunkers nowhere argue for it. Second, debunkers’ premise only concerns what is implied by the best explanation of our moral observations, so fails to imply that our moral beliefs do not satisfy Quine’s constraint. Finally, debunkers’ primary targets are ‘non-naturalist moral realists’.<sup>8</sup> These realists explicitly *accept* Harman’s Thesis. To argue against them on the basis of Quine’s constraint would be question-begging.

Debunkers must be claiming that *assuming that our moral beliefs are (defeasibly) non-empirically justified*, (knowledge of) Harman’s Thesis undermines them—where our belief that *p* is ‘non-empirically’ justified just in case our belief that *p* is justified, and *p* is not implied by the best explanation of any of our observations.<sup>9</sup> If this were so, then Harman’s Thesis would not just have dialectical significance (or show that our moral beliefs fail to meet Quine’s constraint on justification). Harman’s Thesis would give us reason to give them up (even if they were antecedently justified). The difference is like the one between an argument which shows that we believers in hands cannot convince a sceptic of our view, and an argument which shows that we believers in hands ought to stop being believers.

As an underminer, Harman’s Thesis contrasts with a rebutter. Whereas a rebutter would give us ‘direct’ reason to believe that our moral beliefs are false, an underminer merely gives us reason to no longer believe their contents, without giving us direct reason to believe that those contents are false (Pollock, 38–9).<sup>10</sup> The key question is: *how* could Harman’s Thesis undermine our moral beliefs? Genealogical Debunking Arguments are plausible only to the extent that there appears to be a satisfactory answer to this question.

<sup>7</sup> Quine and Harman do not mean exactly the same thing by ‘observation’. I ignore this complication.

<sup>8</sup> E.g., Joyce (2006 Ch. 6) and Street (2006).

<sup>9</sup> I will not continue to add the qualification ‘knowledge of’ in what follows, but this is always intended.

<sup>10</sup> The distinction between rebutting and undermining (undercutting) defeaters is less than crystal clear. This should not matter for what follows.

## 2.2 Sensitivity

Perhaps the most common answer to the key question is this. Harman's Thesis undermines our moral beliefs by giving us reason to doubt that *had the contents of our moral beliefs been false, we would not have believed them*. For convenience, I will say that our moral beliefs are *sensitive* if the italicized counterfactual is true.<sup>11</sup> Joyce sometimes suggests this answer. He writes:

Suppose that the actual world contains real categorical requirements—the kind that would be necessary to render moral discourse true. In such a world humans will be disposed to make moral judgments . . . for natural selection will make it so. Now imagine instead that the actual world contains no such requirements at all—nothing to make moral discourse true. In such a world, humans will still be disposed to make these judgments . . . for natural selection will make it so . . . [D]oes the truth of moral judgments . . . play a role in their usefulness? . . . I believe the answer is 'No'. (2001 p. 163)

Other debunkers make similar remarks. Ruse writes:

You would believe what you do about right and wrong, irrespective of whether or not a 'true' right and wrong existed! The Darwinian claims that his/her theory gives an entire analysis of our moral sentiments. Nothing more is needed. Given two worlds, identical except that one has an objective morality and the other does not, the humans therein would think and act in exactly the same ways. (1986 p. 254)

Similarly, Walter Sinnott-Armstrong writes, '[t]he evolutionary explanations [of our moral beliefs] work even if there are no moral facts at all' (2006 p. 46).

Since Harman's Thesis is not supposed to 'rebut' our moral beliefs, it must give us reason to doubt that if they are true, then they are sensitive, if it gives us reason to doubt that they are sensitive. But there is an obvious problem with the suggestion that Harman's Thesis does this. The contents of our explanatorily basic moral beliefs—our beliefs which purport to state the conditions under which a moral property is instantiated—are widely supposed to be metaphysically necessary, if true, and beliefs in metaphysically necessary truths are vacuously sensitive on a standard semantics.<sup>12</sup> Meanwhile, our non-basic moral beliefs seem to be sensitive, if true, even if the explanatorily basic moral truths are metaphysically contingent. Had *A* not been *M*, where *A* is a particular person, action, or event, and *M* is a moral property, then *A* would have been different in non-moral

<sup>11</sup> While related, the present notion of sensitivity is not the same as that of Nozick (1981). As with Nozick's notion, the present one must plausibly be relativized to a method of belief formation. Note that one way that the contents of our moral beliefs might have been false is by there being no atomic or existentially quantified such truths at all.

<sup>12</sup> And Harman's Thesis does not seem to give us reason to believe that they are contingent.

respects—since even if there are metaphysically possible worlds in which the explanatorily basic moral truths are different and *A* is not *M*, these worlds are presumably more ‘distant’ from the actual world than worlds in which the explanatorily basic moral truths are the same and *A* is not *M*. Since Harman’s Thesis does not seem to give us reason to doubt this reasoning, it does not give us reason to doubt that our moral beliefs are sensitive—on a standard semantics, at least.

Perhaps Harman’s Thesis gives us reason to doubt that our explanatorily basic moral beliefs are sensitive on a non-standard semantics incorporating ‘impossible worlds’ (our non-explanatorily basic moral beliefs will be sensitive on such a semantics if they are sensitive on a standard one)? Even if it did, it is hard to see how knowledge of this could undermine all of our moral beliefs.<sup>13</sup> For virtually *any* supervenient property, *F*, it appears that had—per impossible—the contents of our explanatorily basic *F*-beliefs been false, we still would have believed them. In particular, it appears that had atoms arranged chair-wise failed to compose a chair, we still would have believed that they did.<sup>14</sup> It might be thought that debunkers could accept this consequence. It is one thing to believe that atoms arranged chair-wise compose a chair. It is another to believe, e.g., that we are sitting in one. Arguably, only contingent beliefs like the latter are sacrosanct. But either our belief that we are sitting in chair is undermined if our belief that particles arranged chair-wise is, or not. If it is, then skepticism about explanatorily basic truths engenders ordinary skepticism. Since this is untenable, the apparent insensitivity of explanatorily basic moral beliefs does not even undermine them. If it is not, then Harman’s Thesis fails to undermine our contingent moral beliefs equally. For example, it fails to undermine our belief, discussed in Harman (1977 p. 8), that the children who are pouring gasoline on a cat are doing something wrong (even if it does undermine our belief which purports to state the conditions under which something is wrong). Either way, Harman’s Thesis does not undermine our contingent moral beliefs by giving us reason to doubt that our explanatorily basic ones are sensitive—even on an extended semantics.

Debunkers’ mistake is to assume that whether our moral beliefs are appropriately sensitive depends on whether their truth is implied by their best

<sup>13</sup> For reasons to think that Harman’s Thesis does not even give us reason to believe that our explanatorily basic moral beliefs are insensitive on an extended semantics incorporating ‘impossible worlds’, see Clarke-Doane [2015].

<sup>14</sup> More generally, it seems that, for virtually any of our beliefs in necessary truths, had—per impossible—their contents been false, we still would have believed them. In the mathematical case, it seems that “we would have had exactly the same mathematical...beliefs even if the mathematical...truths were different... [Field 2005, 81].”

explanation. If the explanatorily basic moral truths are metaphysically necessary, then our corresponding beliefs are vacuously sensitive, if true, on a standard semantics, even if Harman's Thesis is true. On the other hand, not even our non-moral explanatorily basic beliefs are sensitive on an extended semantics incorporating 'impossible worlds'. Either way, Harman's Thesis does not seem to give us reason to doubt that our moral beliefs are sensitive in any sense which could plausibly undermine all of them. To be sure, these considerations do not show that our moral beliefs *are* sensitive in some such a sense. What they show is that Harman's Thesis gives us no reason to believe that they are not.

So much for the first answer to the key question.

### 2.3 Safety

Another answer to the key question is that Harman's Thesis undermines our moral beliefs by giving us reason to believe that *we might have easily had false explanatorily basic moral beliefs* (it does not seem to give us reason to believe that we might have easily had false non-basic moral beliefs for reasons alluded to in 2.2). For convenience, I will say that our moral beliefs are safe if the italicized sentence is false.<sup>15</sup> Again, since Harman's Thesis does not 'rebut' our explanatorily basic moral beliefs, it gives us reason to doubt that if they are true, then they are safe, if it gives us reason to doubt that they are safe. Moreover, since Harman's Thesis does not give us reason to believe that *the contents of* our explanatorily basic moral beliefs might have easily been false, it gives us reason to believe that we might have easily had different explanatorily basic *moral beliefs*. As Braddock, Meogensen, and Sinnott-Armstrong (2012) write,

[D]ifferent instantiations of the process of cultural group selection have produced divergent normative systems, which nonetheless solve the same design-problem: namely, that of getting human societies to function as adaptive corporate units. In this way, one and the same process type may, through its various instantiations, easily result in divergent moral systems.<sup>16</sup>

<sup>15</sup> For my purposes, it will suffice to understand this condition in the most demanding way: moral beliefs fail to be safe in this sense if we could have easily had even one false moral belief. (Again, while the present notion of safety is related to notions prevalent in the literature on knowledge, it is not identical. For a similar approach to necessary truths in particular see Pritchard (2009). As with the condition discussed in Pritchard (2009), the present one must plausibly be relativized to a method of belief formation.)

<sup>16</sup> Sometimes debunkers seem to suggest that Harman's Thesis defeats our moral beliefs by merely giving us reason to believe that our explanatorily basic moral beliefs might have been false (even if not easily). For example, Ruse writes, "Had evolution taken us down another path, we might well think moral that which we now find horrific, and conversely. This is not a conclusion acceptable



However, *debunking arguments themselves* demonstrate that this cannot be right. Consider the most austere interpretation of Street's proposal that 'among our most deeply and widely held judgments, we observe many . . . with exactly the sort of content one would expect if the content of our evaluative judgments had been heavily influenced by selective pressures' (2006 p. 116). Suppose that we were evolutionarily 'bound' to have the explanatorily basic moral beliefs that we have, 'for reasons that have nothing to do with their truth.'<sup>17</sup> Then we could not have easily had different such beliefs. But, then, given that our moral beliefs are actually true, and that the explanatorily basic moral truths could not have been different, we could not have easily had *false* explanatorily basic moral beliefs. If our abductive methodology is also 'safe'—something debunkers, qua scientific realists, would presumably not deny—then we could not have easily had false moral beliefs generally. Since, again, Harman's Thesis does not seem to give us reason to doubt this reasoning, it cannot undermine our moral beliefs by giving us reason to doubt that they are safe.<sup>18</sup>

Of course, it is questionable whether we were evolutionarily bound to have the explanatorily basic moral beliefs that we do have. This is especially so in light of apparently pervasive moral disagreement. As before, the considerations above do not show that our moral beliefs *are* safe. (Even if it did show that we could not have easily had different moral beliefs, it would not show that those beliefs are true.) What they show is that Harman's Thesis does not give us reason to doubt that they are safe. Again, debunkers' mistake is to assume that whether our moral beliefs are safe depends on whether their truth is implied by their best explanation.

## 2.4 Explaining Reliability

If Harman's Thesis cannot undermine our moral beliefs by giving us reason to doubt that they are both sensitive and safe, is there some other way that it could

to the [moral realist].” (1986 p. 254) But the claim that our moral beliefs might have been false, had our moral beliefs been different, is a trivial consequence of moral realism plus the assumption that our moral beliefs are true. It says that the moral truths do not counterfactually depend on our moral beliefs. No one should think that this defeats those beliefs. Braddock, Meogensen, and Sinnott-Armstrong (2012) is an online discussion of Clarke-Duane's essay on PEA Soup: <<http://peasoup.typepad.com/peasoup/2012/03/ethics-discussions-at-pea-soup-justin-clarke-doanes-morality-and-mathematics-the-evolutionary-challe-1.html>>.

<sup>17</sup> The quoted phrase is badly in need of explication, as will be clear in Section 2.4.

<sup>18</sup> Note that I am not saying that Street is committed to holding that we could not have easily had different explanatorily basic moral beliefs. I am saying that the view that we could not have is consistent with, and on some presentations, even suggested by, Harman's Thesis. So, Harman's Thesis certainly does not seem to give us reason to believe that we could have.

undermine them? It might be thought that there is a natural way that it could. Harman's Thesis could give us reason to believe that the reliability of our moral beliefs would be an *inexplicable coincidence*. With respect to normative realists generally, Street writes:

[T]he realist must hold that an astonishing [inexplicable] coincidence took place—claiming that as a matter of sheer luck, evolutionary pressures affected our evaluative attitudes in such a way that they just happened to land on . . . the true normative views . . . [T]o explain why human beings tend to make the normative judgments that we do, we do not need to suppose that these judgments are *true*. (2008 pp. 208–9)

Street's suggestion is apparently that if the truth of our normative beliefs is not implied by their best explanation, then their reliability would be 'inexplicable' in a sense which is undermining (if they are construed realistically). The background principle is familiar from Field. Field claims that 'our belief in a theory [is] undermined if . . . [it] requires that it would be a huge coincidence if what we believed about its subject matter were correct.' (2005 p. 77). Unlike Street, however, Field does not seem to believe that in order to 'explain the reliability' of our beliefs, in the relevant sense, it is necessary to show that their truth is implied by their best explanation.<sup>19</sup>

What is the relevant sense of 'explain the reliability'? It must be such that the apparent impossibility of explaining the reliability of our beliefs *undermines* them. This seems to rule out a causal interpretation. Field himself notes that to require a causal link between our beliefs and their subject matter would be to revert to Benacerraf (1973), which 'almost no one believes anymore' (1989 p. 26).<sup>20</sup> It also seems to rule out a probabilistic interpretation of 'explain the reliability', contra Street (Manuscript). Whether the epistemic probability that our moral beliefs are true is high is just what is in dispute. But it does not appear impossible to show that the *objective* probability that our moral beliefs are true is high. For any explanatorily basic moral truth  $p$ , presumably  $\text{Pr}(p) = 1$ , given that such truths would be metaphysically necessary.<sup>21</sup> Also, as we have seen, it may be that  $\text{Pr}(\text{we believe that } p) \approx 1$ . But then  $\text{Pr}(p \ \& \ \text{we believe that } p) \approx 1$ , by the

<sup>19</sup> For example, in his (2005 p. 78) Field suggests that Balaguer (1995) relevantly explains the reliability of our mathematical beliefs, and Balaguer's view does not show that the truth of our mathematical beliefs is implied by their best explanation.

<sup>20</sup> Field is actually referring to the causal theory of knowledge. But the relevance to a Benacerrafian interpretation of 'explain the reliability' is clear.

<sup>21</sup> If one says that only truths which are necessary in an even stronger sense—e.g., 'conceptually necessary'—should get probability 1, then, e.g., our belief that atoms arranged chair-wise compose a chair will seem to have equal claim to be objectively improbable. See Clarke-Doane [2014, Sec. 3].

probability calculus. Since (p & we believe that p) implies (our belief that p is true), it may be that  $\Pr(\text{our belief that p is true}) \approx 1$ .<sup>22</sup>

Moreover, in light of what was argued in Sections 2.2 and 2.3, the relevant sense of ‘explain the reliability’ must be such that the apparent impossibility of explaining the reliability of our moral beliefs *does not give us reason to doubt that they are both sensitive and safe*.<sup>23</sup> There is such a sense of ‘explain the reliability’ only if the following principle is false.

*Modal Security*: If information, E, undermines all of our beliefs of a kind, D, then it does so by giving us reason to doubt that our D-beliefs are both sensitive and safe.<sup>24</sup>

The key idea behind Modal Security is that there is no such thing as a ‘non-modal underminer’. (Obviously, there is such a thing as a non-modal defeater—namely, a rebutter.<sup>25</sup>) If there were, then Harman’s Thesis could undermine our moral beliefs ‘immediately’. It could undermine them, but not by giving us reason to doubt that they are modally secure. The ‘by’ is needed, since anyone should agree that if E undermines our D-beliefs, then E gives us reason to doubt that those beliefs could not have easily been false. If our D-beliefs are actually false, then they could have easily been false.

Why would anyone believe Modal Security? Because it is hard to see why we should give up beliefs in light of information that neither tells ‘directly’ against their contents, nor against the ‘security’ of their truth. Consider a potential counterexample. Suppose that E is evidence that a false theory of justification is true according to which our D-beliefs are not justified. It might be thought that E could undermine all of our D-beliefs, but not by giving us some reason to doubt their sensitivity or safety. But, on inspection, this seems bizarre. Suppose that D includes only propositions for which we have excellent evidence, and E is

<sup>22</sup> Of course, this argument assumes that our explanatorily basic moral beliefs are (actually) true and that their contents are necessary. But debunkers themselves point out that these assumptions are unobjectionable in the present context. See Gibbard (2003 §13), Schechter (2010), and Street (Manuscript) for discussion of these issues, as well as Balaguer (1995) and Field (1989 p. 26) and the quotation from this page in §5, for a related discussion. For more on the justifiability of this assumption, see Clarke-Doane (Forthcoming A Sec. 4.4) and (Forthcoming B Sec. 2).

<sup>23</sup> With respect to metaphysically possible worlds, that is. I will not consistently add this qualification in what follows, for reasons discussed in Section 2.2. In particular, I will not add the qualification to the statement of ‘Modal Security’ below. (Also, strictly speaking, the sentence should read that Harman’s Thesis does not give us ‘direct’ reason to doubt that our moral beliefs are both sensitive and safe for reasons discussed in the following paragraph.)

<sup>24</sup> As before, “all of our beliefs of a kind, D” is shorthand for all our non-logical D-beliefs.

<sup>25</sup> In my [2015], I formulate Modal Security so as to cover both undercutting and rebutting defeaters. But I now think that this is problematic, for reasons to which I will allude presently.

evidence for the view that a belief is justified only if it is infallible. Perhaps we are students in a Philosophy 101 class, for example, and E is an apparently strong argument for the view that justification requires infallibility. To give up our D-beliefs on the basis of E—when E is neither ‘rebutting’ nor ‘direct’ reason to doubt that our D-beliefs are sensitive and safe—seems to be to give up those beliefs ‘for the wrong kind of reason’ (Barnett [Manuscript]).

A more worrying objection to Modal Security concerns necessary truths that we could not have easily failed to believe.

Suppose that a machine enumerates sentences, deeming them validities or invalidities. Independent investigation has confirmed its outputs prior to the last five. The last five outputs are ‘validity’. We defer to the machine’s last five outputs only, and have no prior metalogical beliefs. Today a trusted source tells us that the machine was ‘stuck’ in the last five instances. Call this evidence E. Then E does not seem to give us ‘rebutting’ reason to believe that the last five outputs are invalidities. Nor does it seem to give us ‘undermining reason to believe that it is not the case that had, as a matter of metaphysical possibility, the last five outputs—those sentences—been invalidities, the machine would have called them invalidities. (E does not seem to give us reason to believe that there is a metaphysically possible world in which those sentences are invalidities, and the negation in question can only be true with respect to metaphysically possible worlds if there is.) Nor does E seem to give us undermining reason to believe that the machine could have easily called the last five sentences invalidities. That is the point of calling it ‘stuck’. But E does seem to undermine all of our metalogical beliefs. Does not this show that E may undermine all of our beliefs of a kind without giving us reason to doubt that they are both sensitive and safe?<sup>26</sup>

It does not. E may not be evidence that, for any metalogical proposition *that we believe*, p, we could have easily had a false belief as to whether p. It does not follow that E is not evidence that we could have easily had a false metalogical belief. E is evidence that, even if the machine had considered an invalidity last, it still would have called the sentence a validity. But it is not just this fact which seems undermining. If we know that worlds in which it considers an invalidity last are ‘distant’, then it is hard to see how evidence that, had we been in one, we would have had false metalogical beliefs, could undermine our metalogical beliefs. It must apparently be added that we know that such worlds are ‘similar’ to the actual one. But if this is added, then E is evidence that we could have easily had different metalogical beliefs, had the machine considered different sentences

<sup>26</sup> Thanks to David James Barnett for pressing me with something like this example.

last, and hence, given the necessity of the metalogical truths, that we could have easily had false metalogical beliefs—i.e., that those beliefs are not safe.<sup>27</sup>

This response depends on the assumption that not just any grouping of beliefs counts as a ‘kind’. If we could let  $D$  be  $\{[b] = \{x: x = b\}\}$ , for some belief formed by the machine,  $b$ , then  $E$  might undermine all of our  $D$ -beliefs despite giving us no reason to believe that we might have easily had false  $D$ -beliefs. Intuitively, however, metalogical beliefs, like moral beliefs, are kinds, while ‘the last metalogical belief formed by the machine’ is not. If there is no principled argument for this, however, then Modal Security may not get off the ground.<sup>28</sup>

If we could explicate ‘stuck’ in such a way that learning that the machine was stuck in the last five instances undermine all of our metalogical beliefs, but has no modal implications at all, then the example above would indeed be a counter-example to Modal Security. But it is hard to see how we could do this. We might say that the machine is ‘stuck’ in that it is not ‘detecting’, ‘tracking’, or ‘sensitive to’ the metalogical truths—it is not generating its outputs ‘because’ they are true. But what do these locutions mean? They do not mean that the truth of the machine’s outputs is not implied by their best explanation. Had we imagined instead a machine that outputs only logical truths themselves, then, trivially, the machine would output such truths ‘because’ they were true, since every logical truth is a consequence of every explanation at all. We might explicate ‘stuck’ in terms of hyperintensional ideology like constitution or ground. But why exactly should give up beliefs which we learn are not ‘constituted by’ or ‘grounded in’ their truth?

Whether Modal Security is true is uncertain. But if it is, then Harman’s Thesis cannot give us reason to doubt that it is possible to ‘explain the reliability’ of our moral beliefs in any sense which is such that the apparent impossibility of explaining their reliability could undermine them. More generally, if Modal Security is true, then there is no satisfactory answer to the question that began this chapter. Harman’s Thesis cannot undermine our moral beliefs, because it does not give us reason to doubt that those beliefs are both sensitive and safe.

<sup>27</sup> This assumes that we believe of at least one sentence that it is invalid. But it is hard to see how we could believe of any sentence that it is valid while failing to believe of any other that it is invalid. If we believe that  $p$  is valid, then presumably we believe that  $\sim p$  is invalid.

<sup>28</sup> The problem is similar to the ‘generality problem’ for process reliabilism. See, e.g., Connie & Feldman (1998). Another interpretation of Modal Security would take kinds to be individuated by methods of belief. Modal Security would then say that if  $E$  epistemically obligates us to give up all of our beliefs formed by method,  $M$ , then it does so by giving us some reason to believe that our beliefs formed by that method are not both sensitive and safe (i.e., it gives us reason to believe that had the contents of those beliefs been false, we still have believed them or that we could have easily had false beliefs formed via method  $M$ ). Given natural assumptions, this formulation is strictly stronger than Modal Security as I have interpreted it.

## 2.5 Broader Relevance

The problem with Genealogical Debunking Arguments is actually very general. It arises for any argument which purports to undermine our beliefs from an area, D, satisfying two conditions.

1. The (explanatorily basic) D-truths would be metaphysically necessary, if true at all.
2. There is a plausible explanation of our having the D-beliefs that we have which shows that we could not have easily had different ones.

Consider, for sake of illustration, what has become the canonical formulation of the so-called ‘Benacerraf–Field Challenge’ for mathematical realism. Field writes:

We start out by assuming the existence of mathematical entities that obey the standard mathematical theories; we grant also that there may be positive reasons for believing in those entities. . . . But Benacerraf’s challenge . . . is to . . . explain how our beliefs about these remote entities can so well reflect the facts about them. . . . [I]f it appears in principle impossible to explain this, then that tends to *undermine* the belief in mathematical entities, *despite* whatever reason we might have for believing in them. (1989 p. 26, italics in the original)

This argument has the dialectical structure of debunkers’. But whereas debunkers claim that learning that the truth of our moral beliefs is not implied by their best explanation undermines those beliefs, Field claims that learning that the reliability of our mathematical beliefs would be ‘inexplicable’ undermines them. Street’s argument can be understood as an application of Field’s, under the assumption that explaining the reliability of our beliefs from an area, in the relevant sense, requires showing that their truth is implied by their best explanation.<sup>29</sup>

Like morality, mathematics arguably satisfies 1 and 2. With respect to 1, (pure) mathematical truths are paradigmatic metaphysical necessities. If our mathematical beliefs are true, then they are vacuously sensitive on a standard semantics.

<sup>29</sup> For more on the relationship between debunking arguments and the Benacerraf–Field Challenge, see Clarke-Doane (2015 and Forthcoming A). It is an interesting question whether the mathematical analogue of debunkers’ premise is true. In his (forthcoming), Joyce counters an argument from my (2012) that the contents of our mathematical beliefs are not implied by their best evolutionary explanation. He claims that it at most shows that those contents *realistically construed* are not implied by this. He concludes that he, unlike Street, is ‘free to maintain that an explanation of the usefulness of the ancestral belief that  $2+3=5$  in terms of first-order logic is nevertheless one that presupposes that our ancestors’ arithmetic beliefs were *true*’ (forthcoming p. 8 fn.7). But I do not know what this means. Surely Joyce does not claim that arithmetic truths *are* first-order logical truths. Which such truth is  $1 + 1 = 2$ ?

As for safety, our epistemically basic mathematical beliefs may be evolutionarily inevitable. If they are, then given the ‘safety’ of our abductive methodology and the necessity of their contents, our mathematical beliefs are safe. I am not endorsing this argument. The point is that Field does not seem to give us any reason to doubt it. He does not even pretend to give us ‘rebutting’ reason to believe that our mathematical beliefs are false, and he does not seem to give us any undermining reason to doubt that they are both sensitive and safe. If Modal Security is true, then Field’s premise cannot undermine our mathematical beliefs.

## 2.6 Conclusions

If Modal Security is true, then Genealogical Debunking Arguments are unsound. But while successfully defending Modal Security would suffice to deflate those arguments, successfully challenging it would not suffice to rehabilitate them. The question would remain: *how* does Harman’s Thesis undermine our moral beliefs? To answer simply, ‘by giving us reason to believe that the reliability of our moral beliefs would be an inexplicable coincidence’, merely invites a follow-up question: *In what sense* of ‘explain the reliability’ is it plausible *both* that it appears impossible to explain the reliability of our moral beliefs *and* that the apparent impossibility of explaining their reliability undermines them? An analogous question plagues the Benacerraf–Field Challenge. Absent a satisfactory answer to these questions, such arguments have little force.<sup>30</sup>

## References

- Balaguer, M. 1995. ‘A Platonist Epistemology’. *Synthese*. 103: 303–25.
- Barnett, David James. [Manuscript] ‘Higher-Order Evidence is the Wrong Kind of Reason.’
- Benacerraf, Paul. 1973. ‘Mathematical Truth’. *Journal of Philosophy* 70: 661–79.
- Clarke-Doane, J. 2012. ‘Morality and Mathematics: The Evolutionary Challenge’. *Ethics* 122: 313–40.
- Clarke-Doane, J. 2014. ‘Moral Epistemology: The Mathematics Analogy’. *Noûs* 48: 238–55.
- Clarke-Doane, J. [Forthcoming A] ‘Justification and Explanation in Mathematics and Morality’. In: Shafer-Landau, R. (ed). *Oxford Studies in Metaethics Volume 10*. Oxford: Oxford University Press, Ch. 4.

<sup>30</sup> Thanks to Tomas Bogardus, Jennifer McDonald, Alex Silk, Neil Sinclair, Folke Tersman, and an anonymous referee for helpful discussion. Special thanks to David James Barnett for extensive discussion of Modal Security and related principles.

- Clarke-Doane, J. [forthcoming B] 'What is the Benacerraf Problem?'. In: Pataut, F. (ed). *New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity*. Dordrecht: Springer.
- Connie, E. & Feldman, R. 1998. 'The Generality Problem for Reliabilism'. *Philosophical Studies* 89: 1–29.
- Darwin, C. 1871. *The Descent of Man*. New York: Appleton.
- Field, H. 1980. *Science without Numbers*. Princeton: Princeton University Press.
- Field, H. 1989. *Realism, Mathematics, and Modality*. Oxford: Blackwell.
- Field, H. 2005. 'Recent Debates about the A Priori'. In: Gendler, T. S. & Hawthorne, J. (eds). *Oxford Studies in Epistemology*. Oxford: Oxford University Press, pp. 69–88.
- Gibbard, A. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Harman, G. 1977. *The Nature of Morality*. New York: Oxford University Press.
- Joyce, R. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, R. 2006. *The Evolution of Morality*. Cambridge, Massachusetts: MIT Press.
- Joyce, R. 2008. 'Precis of Evolution of Morality'. *Philosophy and Phenomenological Research* 77: 213–18.
- Joyce, R. [Forthcoming] 'Evolution, Truth-Tracking, and Moral Skepticism'. In: Reichardt, B. (ed). *Problems of Goodness: New Essays on Metaethics*. Bernstein Verlag.
- Nozick, R. 1981. *Philosophical Explanations*. Oxford: Oxford University Press.
- Pollock, John. 1986. *Contemporary Theories of Knowledge*. Savage, MD: Rowman and Littlefield.
- Pritchard, D. 2009. 'Safety-Based Epistemology'. *Journal of Philosophical Research* 34: 33–45
- Putnam, H. 1971. *The Philosophy of Logic (Essays in Philosophy)*. New York: Harper and Row.
- Quine, W. V. O. 1948. 'On What There Is'. *Review of Metaphysics* 2: 21–38.
- Quine, W. V. O. 1951. 'Two Dogmas of Empiricism'. *Philosophical Review* 60: 20–43.
- Ruse, M. 1986. *Taking Darwin Seriously*. New York: Basil Blackwell.
- Schechter, Joshua. 2010. 'The Reliability Challenge and the Epistemology of Logic'. *Philosophical Perspectives*. 24: 437–64.
- Sinnott-Armstrong, W. 2006. *Moral Skepticisms*. Oxford: Oxford University Press.
- 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.
- Street, S. [Manuscript]. Objectivity and Truth: You'd Better Rethink It. Available at: <[https://files.nyu.edu/ss194/public/sharonstreet/Writing\\_files/Paper%2012%20for%20website%20-%20Objectivity%20and%20Truth%20-%20You'd%20Better%20Rethink%20It.pdf](https://files.nyu.edu/ss194/public/sharonstreet/Writing_files/Paper%2012%20for%20website%20-%20Objectivity%20and%20Truth%20-%20You'd%20Better%20Rethink%20It.pdf)>.
- Sturgeon, N. 1985. 'Moral Explanations'. In: Copp, D. & Zimmerman, D. (eds). *Morality, Reason, and Truth: New Essays on the Foundations of Ethics*. Totowa, NJ: Rowman and Allanheld, pp. 49–78.
- Sturgeon, N. 1986. 'Harman on Moral Explanations of Natural Facts'. *The Southern Journal of Philosophy (Supplement)* 24: 69–78.