

---

# EXPLAINING OUR MORAL RELIABILITY

BY

SINAN DOGRAMACI

---

**Abstract:** I critically examine an evolutionary debunking argument against moral realism. The key premise of the argument is that there is no adequate explanation of our moral reliability. I search for the strongest version of the argument; this involves exploring how 'adequate explanation' could be understood such that the key premise comes out true. Finally, I give a *reductio*: in the sense in which there is no adequate explanation of our moral reliability, there is equally no adequate explanation of our inductive reliability. Thus, the argument that would debunk our moral views would also, absurdly, debunk all inductive reasoning.

## 1. Introduction

My topic is a form of evolutionary debunking argument that concerns our ability to explain our own reliability. Advocates of this argument include Allan Gibbard (2003, ch. 13) and Sharon Street (2006, 2008, ms.). They argue that there is no explanation of the reliability of our moral beliefs. From this, they draw certain revisionary conclusions. Gibbard proposes that moral knowledge, if it is any kind of knowledge, is knowledge of an inferior grade; he suggests it is not knowledge 'in a strict sense', a sense in which he says perceptual beliefs are knowledge (pp. 256, 258). Street concludes that moral realism must be false and moral constructivism must be true, since, she argues, only constructivism can salvage our indispensable claim to moral knowledge (2006, sect.10; ms. sect.6). Moral realism is the pre-theoretically plausible view that there are known, mind-independent, logically atomic moral truths. ('Atomic' because e.g. knowing the negation of a necessarily false moral contradiction doesn't count.) I won't examine the positive details of constructivism here.

I'll consider different standards for what counts as explaining our reliability, and I'll argue that, to whatever extent we cannot explain our moral reliability, we likewise cannot explain the reliability of our inductive

inferences. Since it would be absurd to think we lack justification for our inductive inferences, it follows that no revisionary conclusions ought to be drawn for our moral beliefs either, and moral realism remains safe.

## 2. *An alleged requirement on rational belief*

Many philosophers seem to be attracted to a principle like this:

**Explainable Reliability Requirement on Rationality (ERROR):** for a belief-forming method of yours to yield rational beliefs, and thus for it to yield knowledge, you must *not* have all things considered reason to believe there *exists no adequate explanation* of the method's reliability.

**ERROR** doesn't say you need to have in hand the explanation of your method's reliability; you just need to *not* have reason to believe there's no adequate explanation out there.

By 'reliability' I mean that, by and large, the propositions you (actually) believe are (actually) true. The same meaning for 'reliability' is adopted by all the authors I quote in this section. (See Field, 1989, pp.230, 238; Schechter, 2010, pp.438, 441; Enoch, 2011, p.155; and Setiya, 2012, p.68, who says: '[Some] clarifications are worthwhile. First, being reliable is getting things mostly right'. In Street's (and others') usage, notice the constant talk of the 'correlation' or 'coincidence' between our beliefs and the facts, e.g. in the quote below.)

**ERROR** speaks of 'adequate' explanation. In the following sections, I'll explore various ways of drawing the line between adequate and inadequate explanations.

**ERROR** is the natural interpretation of the Benacerraf (1973) problem for mathematical knowledge, also called the Benacerraf-Field problem (Field, 1989, pp. 25–6, 230–9; 2005, sect. 5). In the 2005 paper, Field puts it like this:

I will not consider Benacerraf's own formulation – it relies on a causal theory of knowledge that simply seems inapplicable to a priori knowledge – but rather, will try to capture its general spirit. The key point, I think, is that our belief in a theory should be undermined if the theory requires that it would be a huge coincidence if what we believed about its subject matter were correct.

Field's next sentence (and the 1989 material) indicates that he is worried about *unexplained* coincidences:

But mathematical theories, taken at face value, postulate mathematical objects that are mind-independent and bear no causal or spatio-temporal relations to us, or any other kinds of relations to us that would **explain why our beliefs about them tend to be correct**; it seems hard to give any account of our beliefs about these mathematical objects that doesn't make the correctness of the beliefs a huge coincidence (2005, p. 77, my bold; see too the cited pages of Field, 1989, especially p. 233).

Sharon Street seems to rely on *ERROR* in her argument against moral realism. She writes:

Insofar as we regard our normative judgments as true, we must agree that there is a striking coincidence between (1) the normative judgments that are true, and (2) the normative judgments that causal forces led us to believe. This coincidence between the normative truth and what causes led us to believe is puzzling and demands some explanation (ms., p. 10).

Street also says ‘the challenge I am raising for normative realism bears major similarity to what has become known as the Benacerraf Problem’ (ms., fn. 33). (See also Street, 2006, p. 125; 2008, pp. 207, 211, 214.)

Joshua Schechter (2010, pp. 440, 444–6) also sympathizes with *ERROR*, writing:

In general, failing to possess an explanation of the reliability of a cognitive mechanism is not terribly worrisome. What would be worrisome is if we had compelling reason to suspect that there is no satisfying explanation to be had (p. 446).

Schechter cites Gibbard (fn. 29) as well as Benacerraf and Field.

Many moral realists also accept that *ERROR* poses a genuine requirement on rational belief, only they think the requirement can be met. Recent book length defenses of realism by David Enoch and Kieran Setiya include chapters dedicated to addressing the demand posed by *ERROR*. See Enoch, 2011, ch. 7, especially, pp. 159–160, and this passage:

[T]he need to explain the correlation between a specified class of truths and our relevant beliefs is quite global, since if we cannot explain this correlation in a given domain, we may conclude that no such correlation is likely to exist in that domain, and so that we are completely unreliable in our relevant beliefs, a conclusion that will defeat – and perhaps even undermine – any justification for the relevant beliefs, and therefore also knowledge in the relevant domain (p. 176).

And Setiya (2012, chapter 2) accepts this constraint that he draws from Benacerraf, Field, and Street:

If I know that a correlation of facts would be inexplicable, and I am not otherwise justified in accepting it, I should doubt that the correlation obtains (p. 73, italics omitted).

Setiya is discussing ‘the correlation of belief and fact’, saying ‘This correlation cries out for explanation’ (p. 66).

### 3. *An initial account of (in)adequacy: deep versus trivial vindication*

Gibbard’s own discussion (2003, ch. 13) introduces an important distinction. He draws a line between what he calls *deep* and *trivial* vindications. Deep and trivial vindications are two kinds of explanation of our reliability. He says it

is his 'suggestion that knowledge, in a strict sense, requires deep vindicability' (p. 258, see also p. 256). This, again, seems to be a version of **ERROR**, where the *adequate* explanation **ERROR** calls for must be a deep vindication.

I'm unsure what Gibbard means by the restriction to a strict sense of knowledge, and will ignore the restriction. If I can argue that *strict* knowledge does not require deep vindication, it should follow that neither does *ordinary* knowledge, since it's presumably weaker or equivalent.

What is the difference between deep and trivial vindication? Gibbard says this:

... it [deep vindication] should be more than a trivial vindication of the form, '*J*, and here's why we are the kinds of beings who would judge that *J*.' For any judgment we make, there is presumably a correct causal explanation of our making it (2003, p. 262).

While the notion of deep vindication is not independently defined (as Gibbard acknowledges: p. 263), its contrast, trivial vindication, appears to be clearly defined, as follows. A trivial vindication is given by simply doing two things: (i) you *use* one of your belief-forming methods to form some beliefs, that *X*, that *Y*, and that *Z*, and (ii) then you give a causal explanation of how you came to believe that *X*, that *Y*, and that *Z*.

Street, (ms., pp. 26–7), citing and crediting this same discussion of Gibbard's, considers exactly this model of trivial vindication and endorses the claim that knowledge requires a kind of explanation of reliability beyond this type of trivial vindication. See also Street, 2008, pp. 216–7, which also cites and credits the discussion by Gibbard.

It's true that any *interesting* condition on knowledge requires something more than trivial vindicability. As Gibbard notes, we can reasonably expect there to *always* be, for the reliability of any of our methods, the sort of argument he calls a trivial vindication. For, it's a contingent but undeniable fact that there is always some causal explanation of how we came to believe that *P*, and we can always bootstrap our way to the conclusion that this belief is accurate simply by saying, 'Not only did I come to believe that *P*, but also, lo and behold, *P*!'. So, while a trivial vindication is a kind of explanation of our reliability, it's indeed a trivial one. The interesting question is whether there is any stronger kind of explanation that knowledge requires, something that **ERROR** attempts to capture.

My aim is to object to an argument that Gibbard and Street make against moral realism, an argument they make partly on the basis of **ERROR**. My main claim is disjunctive: *either ERROR is false or there is an adequate explanation of our moral reliability*.

(Richard Joyce (2006, ch. 6) makes an evolutionary argument against moral realism, but I won't discuss that work here. Joyce is not explicit about whether he is applying the present model of argument, the sort inspired by Benacerraf and primarily reliant on **ERROR**.)

#### 4. *Is there an adequate evolutionary explanation of our moral reliability?*

Gibbard and Street claim that we can do no better than a trivial vindication when we try to explain our moral reliability. Here is Street (see also Street, 2008, pp. 214–7):

The general question we are asking in both the manifest surroundings case [perceptual knowledge] and the normative case [moral knowledge] is ‘Why think that the causes described by our best scientific explanations would have led us to the truth in this domain?’ In answer to this question, it is unsatisfactory to reply, ‘My judgments in this domain are true, and they’re also the ones that the causes described by our best scientific explanations led me to.’ Such a reply offers no *reason* for thinking that the causes led us to the truth; it merely reasserts that they did. So, for example, in the manifest surroundings case, it will not do to argue ‘There are objects *X*, *Y*, and *Z* in my manifest surroundings and that’s also what the causes led me to believe.’ Similarly, in the normative case, it will not do to argue ‘*X*, *Y*, and *Z* are valuable and that’s also what the causes led me to believe.’ The problem for normative realism is that in the normative case, empty replies of this sort are **all that can ever be given** (if one is a realist), whereas in the manifest surroundings case, a more informative reply is available (ms., pp. 27–8, bold added).

Gibbard and Street argue, while admitting the details are tricky, that we know the outlines of a plausible evolutionary explanation of how our *perceptual* belief-forming method is reliable, an explanation that satisfies the demands of **ERROR**. By contrast, they claim, there is no adequate evolutionary explanation of the reliability of our moral belief-forming method. Why do they think this? They think that when we try to sketch an evolutionary explanation, it collapses into a *trivial* vindication. (See Street, ms., pp. 27 ff., partly quoted above; and Gibbard, 2003, p. 265.)

Gibbard and Street believe there is no prospect of a *non-evolutionary* explanation of our moral reliability if moral realism is true (Gibbard, 2003, pp. 263, 267; Street, 2006, pp. 134–5). They don’t fully defend this claim, but I accept it and I think it is plausible enough to grant it here and focus on debating what follows. Gibbard and Street appear to want an explanation of how we *ended up* with a reliable method. (They want to answer what Schechter calls ‘the etiological question’ (Schechter, 2010, p. 444).) It’s plausible that an answer to this question needs to draw upon the details about whatever process is actually causally responsible for our ending up with the belief-forming methods we have (though this claim can be denied, e.g. by philosophers who argue that, *necessarily*, concept-users or language-users are reliable), and it’s plausible that, for us, that actual process is evolution. Thus, arguing that there is no adequate evolutionary explanation of our reliability, no evolutionary explanation of the sort **ERROR** demands for knowledge, Gibbard and Street conclude that moral realism leads to skepticism and so must be false.

I want to examine why Gibbard and Street think there is no adequate evolutionary explanation, no deep vindication, of our moral reliability. My strategy will be to consider various attempts to explain our moral reliability, and see whether it can be maintained that these all must be counted as trivial, counted as inadequate to the standards demanded by **ERROR**. I'll argue that, if these explanations are inadequate, then absurd skeptical implications follow for our other basic, non-moral belief-forming methods, in particular for induction.

To begin the examination then, consider the following explanation of our moral reliability. It would appear we have a perfectly adequate, non-trivial evolutionary explanation of our moral reliability in the following valid argument which, I claim, has plausible premises:

**The Evolutionary Explanation of My Moral Reliability:**

**Premise M1:** The moral belief-forming method of my ancestors, i.e. of historically successful reproducers, was fit for survival.

**Premise M2:** If my ancestors' moral belief-forming method was fit for survival, then it was reliable.

**Premise M3:** If my ancestors' moral belief-forming method was reliable, then so is mine.

**Conclusion MC:** Therefore, my moral belief-forming method is reliable.

(I don't assume a good explanation must be an argument. But an explanation can take the form of an argument, and the candidate explanations we'll be examining in this article all do.)

So why do Gibbard and Street think we don't have, in this argument, anything better than a trivial vindication? They dismiss this argument (or ones just like it) on the basis of two points. (1): They issue a challenge to the plausibility of premise M2; they ask that some justification be provided in support of M2. (See Gibbard, 2003, p. 265; Street, 2006, pp. 130–1.) And (2): They claim the only justification we can provide for M2, when that justification is explicitly made a part of the argument, makes the whole argument collapse into a trivial vindication. (Again, see Gibbard, 2003, p. 265, and see again the lengthy quote from Street above.) Let's consider both aspects of Gibbard and Street's dismissal of the argument as a trivial vindication.

(1) Must premise M2 be supported by some further premises we make explicit? Or can it just be a plausible assumption on its own? That is, can we decide it is justified not by other premises that we can explicitly cite, but rather is justified by default? This is one point where some critics may choose to push back or dig in their heels. Even if it is admittedly somewhat counter-intuitive to say M2 is justified by default, couldn't we still end up saying it is justified by default *if saying so would save moral realism*? People with

different basic intuitions won't have much to say to each other or argue over here. Later, at the article's end, I'll return to the issue of what's worth accepting or worth giving up in order to save moral realism.

(2) Gibbard and Street say the only way we can argue for M2 is by using our moral method. We must *make* moral claims, life-affirming moral claims, such as that you ought to take care of your kids. And then we need only notice that it is, obviously, fit to believe as much. Let me grant them, for the sake of argument, that this is the only way we can explicitly argue for M2.

So, the new, expanded argument that Gibbard and Street argue we're forced to offer, the one that allegedly collapses into trivial vindication, involves a pair of additional claims:

**Premise M2a:** If the moral belief-forming methods of my ancestors were fit for survival, then those methods delivered life-affirming moral beliefs.

**Premise M2b:** The life-affirming moral beliefs are reliable, i.e. a life-affirming morality is, at least by and large, true.

Now we face a problem, though: how is the result a trivial vindication? How does adding to the argument a life-affirming moral claim (M2b), and adding the observation that my fit ancestors accepted such views (M2a), result in a trivial vindication? The result still is not a trivial vindication as Gibbard and Street define trivial vindication. (See again the previous section. The resulting argument doesn't explicitly say why I believe we ought to affirm life. The argument doesn't even explicitly say what I believe. The only thing the argument explicitly says about my moral beliefs is that they're reliable.<sup>1</sup>)

It may seem like I'm too closely following the letter, and not the spirit, of what Gibbard and Street surely intended to fall under the standard of a trivial vindication, and of *inadequate* explanation of our reliability. Okay, so let's make an effort, then, to give a better characterization.

## 5. *An improved account of (in)adequacy*

The reasonable interpretation is that Gibbard and Street see M2b as trivializing the argument. A natural way of articulating their intended more general idea is something like this. As *ERROR* says, rationality (and thus knowledge) requires explainable reliability – but there's a constraint on what counts as an adequate explanation. I propose that the desired constraint here is that it's no adequate explanation of a method's reliability if your argument uses the method to supply the premises that *X*, that *Y*, and that *Z*, and the truth of *X & Y & Z* on its own entails the truth (or truth-preservation) of most of the propositions (or inferences) the method actually leads you to

believe (or draw). The idea here is that we want a kind of *non-circularity constraint*. Let me re-iterate the proposal once more, as an indented principle:

**Triviality Constraint** (*partial formulation – see below*): an argument does not adequately explain a method's reliability in the sense demanded by **ERROR** if:

- (i) the argument uses the method to supply the premises that *X*, that *Y*, and that *Z*; and,
- (ii) that *X & Y & Z* entails most of the propositions your method leads you to believe, or it entails most of the corresponding (material) conditionals for the inferences your method leads you to draw.

To illustrate, **Triviality Constraint** is not violated by M1 & M2 & M3; those premises entail that *I'm reliable* (MC), but don't entail the various particular moral propositions I believe. But, M2b all by itself violates the constraint, since the specific moral propositions that I do believe are all (well, more or less) life-affirming moral propositions, and M2b flatly says all this is true. Another illustration: while we are allowed to give something like an evolutionary explanation of our mathematical reliability, we are not allowed to give an explanation that itself states a handful of mathematical axioms that entail most of our mathematical beliefs.

As I stated **Triviality Constraint**, it speaks of 'most' of your beliefs or inferences. It's admittedly extremely vague as to what counts as 'most' of these, but the intended idea here just piggybacks on our intuitive understanding of reliability. I said at the start that I'm understanding a method to be reliable when, 'by and large', what the method leads you to believe is true (or, what you infer is truth-preserving); so, 'most' in the constraint just means that you're entailing enough to secure this 'by and large' correlation.

**Triviality Constraint** has some appeal. While circularity of some kind is inescapable when it comes to explaining the reliability of your most basic belief-forming methods, the constraint says it's no adequate explanation if you use a method to flat out state things that entail the truth (or truth-preservation) of most of the things it leads you to believe (or infer). Even if some kind of circularity must be permissible, such extremely shallow circularity seems bad.

A referee notes that the above arguments for conclusion MC (my moral belief-forming method is reliable) might still fail to meet other further constraints on what kind of explanation of my reliability is required in order for me to have moral knowledge. Perhaps there is a modal constraint on knowledge like the popular safety requirement, or perhaps there is a distinct kind of 'non-accidental reliability' constraint on knowledge. Since my focus



will be, with Gibbard and Street, just on the distinctive aspects of evolutionary debunking that have to do with epistemic circularity, I won't examine such constraints on knowledge in this article.<sup>2</sup>

Here is another virtue of interpreting Gibbard and Street's challenge in the form of *Triviality Constraint* (aside from the constraint's intrinsic appeal). Many critics have said Gibbard and Street (and Joyce) are holding moral knowledge to an absurd standard, a standard that perceptual knowledge cannot meet either. (See Schafer, 2010, pp. 476 ff.; White, 2011, sect. 4.2; Enoch, 2011, p. 175; Shafer-Landau, 2012, pp. 18, 22–3; Setiya, 2012, p. 79; Vavova, 2014, sec. 3.1; Berker, 2014, sect. 8.) The critics say that, since we must use our perceptual faculties to give an evolutionary explanation of the reliability of perception, the explanation of the reliability of our moral method should also be allowed to use that method. But, in Gibbard and Street's defense, there seems to be a significant asymmetry here between morality and perception, one that *Triviality Constraint* may capitalize on. In our argument above, M2b expresses a *very general* moral view, one that suffices for a very large core of our actual moral view, if not nearly the entirety of our moral view, to be true. Now, it would take a lengthy and scientifically well-informed book to adequately examine the contours of an evolutionary explanation of our perceptual reliability, but there does seem to be something intuitive to the idea that perception's evolutionary vindication can rely on a narrower application of our perceptual faculties. It might be that we can argue for the general reliability of perception by making an argument that, while it must use our perceptual method at some junctures, uses it to make claims that fall very far short of sufficing for the truth of most of the propositions my perceptual method actually leads me to believe. We might only need our perceptual method to supply some propositions about things like the archeological evidence that our ancestors were outfitted with excellent faculties for perceiving the environment in ways that made our ancestors highly fit for survival. Of course, to conclusively resolve the matter, we'd need a thorough understanding of perception's role in supporting the claim that our ancestors were fit, and in supporting the claim that if our ancestors were fit then their perceptual method was reliable. (For the record, I'm personally sympathetic even to the idea that we know *a priori* that fit perceptual methods are reliable; I think Gibbard (2003, pp. 255–6) shows how.)

So, it seems to me that *Triviality Constraint* does *not obviously* prevent our adequately explaining the reliability of perception (though the final verdict is not for armchair philosophy), but it obviously *does* prevent our adequately explaining our moral reliability in any way that includes M2b. If *Triviality Constraint* can help Gibbard and Street quiet their many armchair philosopher critics who accuse them of debunking perceptual beliefs along with moral beliefs, then Gibbard and Street should be very welcoming of the present proposal. So far, so good. Is *Triviality Constraint* otherwise attractive?

One problem with *Triviality Constraint* arises when we consider how we might explain the reliability of our *deductive* belief-forming methods, that is, our methods for inferring the logical consequences of what we believe. The problem is that the reliability of our deductive methods just consists in the truth of the corresponding (material) conditionals that link any premises and their logical consequences, and these conditionals will be logically necessary. Since those conditionals are logically necessary, they will be entailed by *any* proposed explanation of our deductive reliability (since they are vacuously entailed by anything at all).

But I think the problem here can be addressed by introducing an extra clause into *Triviality Constraint*. There's something intuitive to the idea that an explanation should fail *ERROR*'s standards if it says anything that is *trivializing*: an explanation should fail as soon as it makes any claim which turns the explanation from one that *didn't* into one that *does* entail the propositions your method leads you to believe, or entail the corresponding conditionals for the inferences your method leads you to draw. The following full formulation thus captures the intended intuitive idea, while not ruling out every possible explanation of the reliability of our deductive methods:

*Triviality Constraint (full formulation)*: an argument does *not* adequately explain a method's reliability in the sense demanded by *ERROR* if:

- (i) the argument uses the method to supply the premises that *X*, that *Y*, and that *Z*; *and*,
- (ii) that *X & Y & Z* entails most of the propositions your method leads you to believe, or it entails most of the corresponding (material) conditionals for the inferences your method leads you to draw; *and*,
- (iii) those believed propositions, or corresponding conditionals, are *not* entailed by the remainder of the explanation (whatever the remainder consists of, perhaps nothing).

To illustrate, consider Schechter's (2013) proposal to explain the reliability of our deductive method via an evolutionary explanation. While the initial partial formulation of *Triviality Constraint* would have automatically ruled out Schechter's strategy as inadequate, the full formulation allows that we may be able to explain the reliability of deduction evolutionarily. Of course, the full formulation still rules out the explanation of our moral reliability that uses M2b.

(A side note: having added clause (iii) to arrive at the full formulation, there will now be some intuitively trivial, inadequate explanations of reliability that aren't filtered out by *Triviality Constraint*; in particular *no* explanations of our deductive reliability will be counted as trivial, since all fail condition (iii). But, this is not a problem for our purposes. We needed to find a constraint that captured the spirit of Gibbard and Street's objections to M2b, and we have it. We don't need an exhaustive account of all trivial

explanations here. We only need this sufficient condition on trivial explanation.)

I do think *Triviality Constraint* has intuitive appeal, and I think it captures the spirit of trivially circular vindication as Gibbard and Street are interested in the notion, and as it stems from Benacerraf. In my view, this principle best characterizes the notion desired here.

So, to review, here is what seems to me the best reconstruction of the core of the case that Gibbard and Street wish to put together against moral realism. There are three key ingredients. (1) There is the endorsement of *ERROR*: if you have reason to believe there's no adequately explaining your reliability, you can't rationally maintain your beliefs. (2) There is the challenge to M2 in the initial evolutionary explanation, the valid argument consisting of just M1-M3: the debunkers say M2 is not plausible on its face, not justified by default, and they demand that M2b be added to the argument. And, (3) There is *Triviality Constraint*, which is applied to say that an argument containing M2b is not an adequate explanation, not the sort demanded by *ERROR*.

I want to argue that all this still fails to refute moral realism. I'll do this by considering, in the next section, how we might explain the reliability of induction. It's interesting that among the recent flurry of work on evolutionary debunking, there are discussions of morality, perception, mathematics, deductive logic and even modal metaphysics, but almost nobody discusses induction at all. I think considering induction can shed new light on the issues here.

(In the recent work on evolutionary debunking, I see only two places where induction or inference to the best explanation is discussed: Setiya, 2012, pp. 104–9 of ch. 3; and a brief paragraph in Wielenberg, 2010, p. 445, though Wielenberg points to a nice earlier discussion by Carruthers (1992, pp. 109–110 and ch. 12). For evolutionary issues concerning perception, see Schafer, 2010; White, 2011; Vavova, 2014. For evolutionary issues concerning mathematics, see Clarke-Doane, 2012; Schechter, 2010, p. 455; Field, 2005. For deductive logic, see Schechter, 2010, 2013; Field, 2005. For modality see Williamson, 2006, p. 136; FitzPatrick, 2015, sect. 1.)

## 6. *Is there an adequate evolutionary explanation of our inductive reliability?*

Our inductive belief-forming method is the method by which we form beliefs about unobserved contingencies. Or, oversimplifying a bit, it's our method for inferring how the future will be. Our beliefs about unobserved contingencies are, of course, typically inferred from observed contingencies; we predict the future based on the past.

I will assume the only hope for explaining our inductive reliability is an evolutionary explanation. (This is a simplifying assumption. Van Cleve's (1984) rule-circular argument for induction's reliability does not concern evolution. Though I won't take the space to elaborate, those familiar with the rule-circular explanation should see that the points I'll make still apply.) Here is how an attempted evolutionary explanation would seem to go:

**The Evolutionary Explanation of My Inductive Reliability:**

**Premise I1:** The inductive belief-forming method of my ancestors, i.e. of historically successful reproducers, was fit for survival.

**Premise I2:** If my ancestors' inductive belief-forming method was fit for survival, then it was reliable.

**Premise I3:** If my ancestors' inductive belief-forming method was reliable, then so is mine.

**Conclusion IC:** Therefore, my inductive belief-forming method is reliable.

In the moral case, Gibbard and Street challenged premise M2. They asked what justifies the claim that *fit* moral methods are *reliable* moral methods. Although that question can be answered, they said it can only be answered by using our moral method to make moral claims, to make the claim that we ought to promote life. And, this violated *Triviality Constraint* because our actual moral view is largely just the expression of life-affirmation.

Things are different with premise I2: if it were challenged, its justification would *not* plausibly require induction. Rather, it seems *a priori* clear that anyone who reliably predicted which berries will be poisonous was likelier to survive (and so on). But, note how much, or how little, is justifiable *a priori* here: it's *a priori* clear that anyone who won at the Darwinian competition knew how to avoid poisonous berries, at least long enough to reproduce. It's not *a priori* clear that *I* will succeed in avoiding poisonous berries long enough to reproduce. What's clear here is only that inductive methods that helped the survivors to survive were methods that were reliable for them. Premise I2 cannot be taken to say that this method *will be reliable for me*.

For my ancestors' reliability to be relevant to *my* reliability, we need something like premise I3, the premise that generalizes from reliability in my ancestors' time and environment to reliability more widely, including in my time and environment. What justifies this? Just as Gibbard and Street asked what justifies M2, it can be asked what justifies I3. It's presumably no less reasonable than asking, analogously, what justifies a prediction that the same species of berries that were poisonous to my ancestors will be poisonous to me.

Now, I certainly don't deny that premise I3 *is* perfectly justified. Premise I3 is extremely plausible. But, then again, I also think M2 is extremely plausible. M2 is plausible, and a challenge to its plausibility can be answered by using our moral belief-forming method. Likewise, I3 is plausible, and a challenge to its plausibility can be answered, though again, we answer it

by using our inductive method. (The need for a circular appeal to induction in order to defend the evolutionary explanation of induction's reliability against all challenges is also discussed by Carruthers (1992, ch. 12).) We need to use induction to deliver some additional explicit argumentation, such as:

**Premise I3a:** Wherever I go, whenever I go there, the universe will continue to resemble the past; in particular, the laws of nature will continue hold.

(Some may insist this is an *a priori* truth. *A priori* or not, it's a deliverance of our method for forming beliefs concerning unobserved contingencies. That's the method under examination.)

Now, of course, when we use induction to generate a premise like I3a, the premise that supports I3's generalization from our method's being reliable for my ancestors to its being reliable for me, we violate *Triviality Constraint*. We end up with something just like M2b, the premise that Gibbard and Street consider trivializing. Just like the explicit endorsement of a life-affirming morality, I3a explicitly says the world is a certain way, a way that entails that most of the inferences my inductive method actually leads me to make are truth-preserving (that is, I3a entails the corresponding material conditionals for these inferences). I3a states the reliability condition for induction, familiar from Hume, that nature will continue to remain appropriately uniform. Thus, if the evolutionary explanation of induction's reliability is challenged in the same way that Gibbard and Street challenge the explanation for morality, the prospects for a response are the same: we must use the method in question to make claims that suffice for the truth, or truth-preservation, of most of the method's actual prescriptions. Induction is no better off than morality.<sup>3</sup>

## 7. *What is worth conceding in order to save moral realism?*

I now return to an issue I said I'd return to. What if Gibbard and Street insisted that there is a difference between M2 and I3? What if they said M2 requires some explicit argument for its justification, while I3 does not (in particular, I3a need not be added before we have an adequate explanation of the reliability of induction)? Someone, of course, *can* insist that I3 is a plausible basic assumption, a premise justified by default.

The problem with this strategy, though, is that it undermines Gibbard and Street's demand for a non-default justification for M2. The argument consisting of just M1-M3 doesn't violate *Triviality Constraint*. There is only a violation if Gibbard and Street can rightly insist that the argument is not plausible until M2b is added as a supporting premise for M2. But, if it's fair to challenge M2 in this way, why can't I3 fairly be challenged as well?

Could Gibbard and Street defend the claim of an asymmetry between M2 and I3, saying: I3 is justified on its face, without our using induction to say anything in its support, while M2 is hardly justified on its face, and only appears plausible after we use our moral method to make moral claims, after M2b is added? In response, I could even *concede* that M2 is less plausible on its face, even far less, than I3 is. But, it seems, at least to me, that this difference in degree can be tolerated by the moral realist, as long as it's *more* plausible that M2 has default justification than it's plausible that moral realism is false. While philosophers can each weigh this matter for themselves, this seems to me to provide a line of defense that keeps moral realism safe.

Ultimately, I hope that I am clarifying the status of the arguments and counter-arguments available both to debunkers and realists. I myself do not think this debunking argument succeeds, but I believe that what we've put together here is a clear and compelling formulation of the debunking argument. In particular, I believe my formulation of *Triviality Constraint* may be the best that can be offered to the debunker who sees epistemic circularity at the root of the issue. For whatever my opinion is worth to my opponents, I suggest that the wisest strategy for the debunkers is to part company with me by insisting that I3 is justified without our having to cite any other premises to support it, while M2 requires a further premise, in particular M2b. I simply don't find this a plausible enough claim *to overturn moral realism*. But, the dispute here depends on how plausible one finds moral realism, and how implausible one finds the various revisionary proposals offered by anti-realists and skeptics.

## 8. Conclusion

Could we or should we, like Hume (or one caricature of him), take a skeptical attitude about *both* moral and inductive knowledge? No, I think that's just out of the question. So, my own conclusion is this: either **ERROR** is false, or *Triviality Constraint* is false, or M2 is justified by default. To say it again: either it's okay to think there is no explanation of our reliability, or one of the explanations considered here suffices, whether it be one with M2b or one without.<sup>4</sup>

Department of Philosophy,  
The University of Texas at Austin,

### NOTES

<sup>1</sup> A referee suggests an alternative approach. Perhaps it could be argued that M3, like M2, demands its own backing premises to be added to the full argument, and these backing premises then will entail a specification of what I believe: they will say that I believe (roughly) the same thing my ancestors believed, and the argument's other premises tell us that they believed a

life-affirming moral view. Perhaps then the resulting full argument could amount to (something much closer to) a trivial vindication even on Gibbard and Street's original definition. I don't have any big problem with this approach. The main point I make in this article would still go through if we took this course. But I think we can give a better, more focused diagnosis of the apparent source of triviality if we focus our attention, as I am going to, on M2b. The key issue I want to fight over is the way in which M2b contributes to the trivialization of the argument as a circular argument, and the approach I follow in the article allows us to focus on this most clearly.

<sup>2</sup> Clarke-Doane (forthcoming a,b) focuses on such modal issues. He does not focus on the issue of circularity at all. He interprets the Benacerraf problem as a demand that we be able to show (or at least believe it could be shown) that our beliefs have properties like those often labeled 'sensitivity' or 'safety' and which concern truth in non-actual worlds. I favor a very different view: my view is that a plausible demand for adequate explanation can at most demand that we show (or believe it can be shown), non-circularly, that our beliefs are true in the actual world.

Setiya (2012, ch. 2) focuses, as I do in this article, on issues concerning evolutionary debunking and epistemic circularity. However, in his chapter 3, Setiya goes on to defend another distinctive constraint on knowledge, namely that for a belief that  $p$  to be knowledge, it must be 'no accident' that the belief-forming method used by the subject is a reliable method (p. 96). Setiya argues that we should understand accidentality not in modal terms, but in terms of explanation. On his view, non-accidental reliability requires 'an explanatory connection: a connection between the reliability of [the subject's method]  $m$  and the fact that [subject]  $S$  is disposed to use it' (p. 97). Setiya might claim our argument for conclusion MC (my moral belief-forming method is reliable) doesn't provide the requisite explanatory connection.

Perhaps some modal constraint applies to our (ability to claim) *knowledge* of moral facts. (*Knowledge* is the title and the distinctive concern of Setiya's chapter 3.) Or perhaps knowledge has a non-accidentality condition. I'm not giving arguments against any of these views here. I'm just flagging that my concern in this article is with a non-circularity condition on rationality.

<sup>3</sup> A referee suggests an alternative strategy that the moral skeptic could pursue. The skeptic could argue that M2 itself cries out not only for justification but explanation, and that M2a and M2b can only justify without explaining, whereas I3a explains I3. The skeptic might argue this by alleging a violation of some constraint on explanation other than the sort of non-circularity constraint that has been my focus in this article (see the previous note). I myself am inclined to disagree with this skeptic's accusation; I think that M2a and M2b do offer some good explanation of M2; furthermore I think M1-M3 on their own offer some good explanation of MC. However, others may disagree with me and thus may see another strategy for the skeptic here.

<sup>4</sup> For help with this article, I thank: Justin Clarke-Doane, Alex Grossman, Karl Schafer, Miriam Schoenfield, David Sosa, and the anonymous referees.

#### REFERENCES

- Benacerraf, P. (1973). "Mathematical Knowledge," *The Journal of Philosophy* 70(19), pp. 661–679.
- Berker, S. (2014). "Does Evolutionary Psychology Show that Normativity is Mind-Dependent?," in J. D'Arms and D. Jacobson (eds) *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics*. Oxford: Oxford University Press, pp. 215–52.
- Carruthers, P. (1992). *Human Knowledge and Human Nature*. Oxford: Oxford University Press.
- Clarke-Doane, J. (2012). "Morality and Mathematics: The Evolutionary Challenge," *Ethics* 122 (2), pp. 313–340.
- Clarke-Doane, J. (forthcoming a). "Justification and Explanation in Math and Morality," in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics*, Volume 10. Oxford: Oxford University Press.

- Clarke-Doane, J. (forthcoming b). "What is the Benacerraf Problem?," in F. Pataut (ed.) *New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity*. Dordrecht: Springer.
- Enoch, D. (2011). *Taking Morality Seriously*. Oxford: Oxford University Press.
- Field, H. (1989). *Realism, Mathematics, and Modality*. Oxford: Blackwell.
- Field, H. (2005). "Recent Debates about the A Priori," in T. S. Gendler and J. Hawthorne (eds) *Oxford Studies in Epistemology*, 1. Oxford: Oxford University Press, pp. 69–88.
- FitzPatrick, W. (2015). "Debunking Evolutionary Debunking of Ethical Realism," *Philosophical Studies* 172(4), pp. 883–904.
- Gibbard, A. (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Joyce, R. (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.
- Schafer, K. (2010). "Evolution and Normative Scepticism," *Australasian Journal of Philosophy* 88(3), pp. 471–488.
- Schechter, J. (2010). "The Reliability Challenge and the Epistemology of Logic," in J. Hawthorne (ed.) *Philosophical Perspectives: Epistemology*, 24. Amsterdam: Elsevier, pp. 437–464.
- Schechter, J. (2013). "Could Evolution Explain Our Reliability about Logic?," in T. S. Gendler and J. Hawthorne (eds) *Oxford Studies in Epistemology*, 4. Oxford: Oxford University Press, pp. 214–239.
- Setiya, K. (2012). *Knowing Right from Wrong*. Oxford: Oxford University Press.
- Shafer-Landau, R. (2012). "Evolutionary Debunking, Moral Realism and Moral Knowledge," *Journal of Ethics and Social Philosophy* 7, pp. 1–38.
- Street, S. (2006). "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127, pp. 109–166.
- Street, S. (2008). "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About," *Philosophical Issues (supplement to Nous)* 18, pp. 207–228.
- Street, S. (ms.). 'Objectivity and Truth: You'd Better Rethink It' draft of July 2009 (the most recent draft, as of Spring 2014).
- Van Cleve, J. (1984). "Reliability, Justification, and the Problem of Induction," in P. French, T. E. Uehling, Jr. and H. K. Wettstein (eds) *Midwest Studies in Philosophy*, Volume IX. Minneapolis, MN: University of Minnesota Press, pp. 555–567.
- Vavova, K. (2014). "Debunking Evolutionary Debunking," in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics*, Volume 9. Oxford: Oxford University Press, pp. 76–101.
- White, R. (2011). "You Just Believe That Because...," in J. Hawthorne (ed.) *Philosophical Perspectives: Epistemology*, Volume 24. Amsterdam: Elsevier, pp. 573–615.
- Wielenberg, E. (2010). "On the Evolutionary Debunking of Morality," *Ethics* 120(3), pp. 441–464.
- Williamson, T. (2006). *The Philosophy of Philosophy*. Oxford: Blackwell.