

## **Explanatory Challenges in Metaethics\***

Joshua Schechter (Brown University)

### **1. Introduction**

There are several important arguments in metaethics that rely on explanatory considerations. Gilbert Harman has presented a challenge to the existence of moral facts that depends on the claim that the best explanation of our moral beliefs does not involve moral facts. The Reliability Challenge against moral realism depends on the claim that moral realism is incompatible with there being a satisfying explanation of our reliability about moral truths. The purpose of this chapter is to examine these and related arguments. In particular, this chapter will discuss four kinds of arguments – Harman’s Challenge, evolutionary debunking arguments, irrelevant influence arguments, and the Reliability Challenge – understood as arguments against moral realism. The main goals of this chapter are (i) to articulate the strongest version of these arguments; (ii) to present and assess the central epistemological principles underlying these arguments; and (iii) to determine what a realist would have to do to adequately respond to these arguments.

### **2. Robust Moral Realism**

Before we get to the explanatory arguments, it will be useful to first present a strong form of non-natural moral realism. I won’t defend this view here (though I am sympathetic to it). Rather, I present this view because it makes a natural target for the explanatory arguments discussed below. Versions of the arguments may have force against other metaethical views. But the issues

---

\* For Tristram McPherson and David Punkett (eds.) *Routledge Handbook of Metaethics*.

involved are delicate, and it is easier to get a sense of the crux of the arguments when they are targeted against a strong form of realism.

The version of moral realism that will serve as the target of the explanatory arguments has seven theses. The first thesis concerns the nature of moral language and thought:

(Cognitivism) Certain sentences and mental representations purport to represent moral facts. They are both meaningful and truth-apt (that is, capable of being true or false).

This thesis should be accepted by any moral realist.

The second thesis is the denial of moral error theory:

(Non-Error Theory) Some attributions of basic moral properties and relations are true.

Moral error theory is sometimes understood to be the claim that all moral sentences are false. But this is a mistake. If a moral sentence is false then its negation is true. Instead, moral error theory is better understood as the view that sentences that attribute basic moral properties and relations are false. For example, if wrongness is a basic moral property, then according to moral error theory, any sentence that attributes wrongness to an action will be false, but its negation – and the sentence “no action is wrong” – will be true.

The third thesis is a claim about the independence of the moral:

(Independence) The fundamental moral facts do not depend on us. In particular, they do not depend on facts about our minds, language, or social practices.

On any plausible realist view, some moral facts depend on facts about our minds, language, or social practices. For example, that my friends detest the taste of strawberry ice cream is part of why it would be better to bring a tub of vanilla to their party. That certain words and gestures have the meanings that they do – which itself depends on complex facts about our social practices – is part of why uttering certain sentences or making certain gestures is morally problematic. And so on. According to Independence, what are independent of us are the fundamental moral facts – those moral facts that (in combination with non-moral facts) determine the rest.

How should “independent” be understood here? The best way to understand this notion is in terms of a kind of explanation (Fine 2012). According to Independence, facts about our minds, language, and social practices do not constitutively explain the truth of the fundamental moral facts. In other words, the fundamental moral facts do not obtain in virtue of facts about our minds, language, or social practices.

Independence captures one sense in which morality may be claimed to be objective. The fourth thesis captures a different sense in which morality may be claimed to be objective. This is the sense on which not every practice of moral assessment is on a par.

(Non-Plenitude) Of the many possible coherent practices of moral assessment, only a few are correct. Not all practices are on a par.

This thesis requires some explanation. Consider a view of the metaphysics of morality on which there are a vast number of different moral properties. For instance, there are the properties of

rightness and wrongness. There are also the properties of rightness\* and wrongness\*, the properties of rightness\*\* and wrongness\*\*, and so forth. Indeed, suppose there are so many properties analogous to rightness and wrongness that every coherent practice of assessing actions as “right” or “wrong” fits one of these pairs of properties. Suppose, too, that none of the pairs of properties is metaphysically privileged over the rest. (This view is loosely analogous to plenitudinous Platonism about mathematics, according to which every consistent mathematical theory correctly describes some part of mathematical reality (Balaguer 1998).) For instance, on this view, if members of some society have a coherent practice of moral assessment that involves accepting the claim “causing gratuitous suffering is wrong, but not on Tuesdays”, this claim will be true when their word “wrong” is understood to stand for a different property than what our word “wrong” stands for. Such a plenitudinous view of morality is compatible with the first four theses. But it lacks an important kind of objectivity. The Non-Plenitude thesis is intended to rule out this kind of non-objective view.

The fifth thesis is the denial of moral naturalism.

(Non-Naturalism) Moral properties and relations are not natural properties and relations.

This is a metaphysical thesis about the nature of moral properties and relations. There are two conceptions of the natural that are relevant here. According to one conception, natural properties are the kinds of properties that play a role in the natural sciences (along with any properties that are reducible to or fully grounded in such properties). According to a second (and potentially broader) conception, natural properties are descriptive properties. (See “Metaphysical Relations in Metaethics” in this volume for one way to make this precise.) The difference between these

two conceptions will not be important in what follows. What will be important is the claim, plausible on either conception, that natural phenomena do not have non-natural causes. Given this claim, the fifth thesis entails that moral facts do not cause natural phenomena.

The final two theses concern the status of our moral beliefs:

(Justification) By and large, the moral claims we believe (upon reflection and discussion) are epistemically justified.

(Reliability) By and large, the moral claims we believe (upon reflection and discussion) are true. Or, at least, we do significantly better than chance would predict.

According to these two theses, we are not epistemologically hopeless about the moral. Many of our moral beliefs are justified – they are rational for us to have. Many of our moral beliefs are true. Even if our moral beliefs are imperfect, they are correct more often than they would be if they were generated at random.

These seven theses concern moral language, moral thought, moral metaphysics, and moral epistemology. Taken together, they characterize a strong form of moral realism.

Borrowing a term popularized by David Enoch, in what follows I will use “robust moral realism” to stand for this package of views (Enoch 2011). Robust moral realism will serve as the main target of the explanatory challenges discussed below.

### 3. Harman's Challenge

The first explanatory challenge I will discuss is due to Gilbert Harman. In *The Nature of Morality*, Harman presents a challenge to our belief in the existence of moral facts (Harman 1977: 3–10). Harman begins by noting a disanalogy between moral and scientific beliefs. Consider some scientific observation. To use Harman's example, consider a physicist observing a vapor trail in a cloud chamber and immediately coming to believe that there is a proton over there. Harman points out that the best explanation of the physicist's belief involves the existence of a proton. (Roughly: The proton interacted with the vapor in the chamber, causing there to be a vapor trail. Photons bounced off the vapor trail and interacted with photoreceptors in the physicist's eye, which led to a pattern of electric impulses in the physicist's optic nerve ... which led to the physicist's belief.) By contrast, consider some "moral observation". Again to use Harman's example, consider someone who observes children lighting a cat on fire and immediately comes to believe that what the children are doing is wrong. Harman suggests that the best explanation of the observer's belief does not involve the existence of moral facts, at least not as they are conceived of by the robust moral realist. The explanation of how the observer came to have the moral belief only seems to involve facts about the psychology (and culture and social background, etc.) of the observer.

Why does this (apparent) disanalogy pose some kind of problem for robust moral realism? There are a few different arguments that one could present that make use of Harman's disanalogy. For instance, one might claim that a belief in an independent truth doesn't count as a piece of knowledge unless the best explanation of how the thinker came to have the belief involves the truth in question. (Related principles concerning the nature of knowledge have been endorsed by several epistemologists (Goldman 1988: 22; Jenkins 2006: 139).) Given this

principle, one might argue that since the best explanation of our moral beliefs does not involve moral truths, either (i) the moral truths are not independent of us or (ii) our moral beliefs do not count as knowledge. (This argument is broadly analogous to Benacerraf's problem for Platonism about mathematics (Benacerraf 1973).) This argument runs into difficulty, however, since the proposed principle is false. One important kind of counterexample concerns knowledge of the future. By observing water in a pot on the stove, I can come to know that the water will boil. This is so even though the fact that the water will boil doesn't play a role in the explanation of my belief.

A second argument that could be presented goes as follows: Since moral facts don't play a role in the best explanation of our moral beliefs, it is mysterious how it could be that our moral beliefs are by and large true. This casts doubt on our reliability about morality, and provides us with reason to give up our moral beliefs. This argument is an important one, and will be discussed below. But it is not the challenge that Harman poses.

To understand Harman's Challenge, it is helpful to return to the case of protons. Why are we justified in believing that protons exist? A plausible answer is that the claim that protons exist is part of the best explanation of our observations of the world, including the observation of vapor trails in cloud chambers. Our belief that protons exist is justified because the claim that protons exist plays an important role in explaining our observations. The same does not seem to be true for our belief in moral facts – moral facts do not seem to play any role in explaining our observations. Thus, argues Harman, we have reason to worry that our belief that there are moral facts is unjustified.

This line of thought is tied to a general picture of epistemic justification. According to a popular view, many of our beliefs are justified on the basis of Inference to the Best Explanation

(Lipton 2004). Given some evidence, we are justified in coming to believe a hypothesis if it (i) does a sufficiently good job of explaining the evidence and (ii) does a significantly better job of explaining the evidence than alternatives. Candidate explanations are better than alternatives when they have various explanatory virtues – e.g., they are simpler, less ad hoc, have greater predictive power, yield greater understanding, are theoretically fruitful, and so on. Harman’s Challenge depends on something like a converse to Inference to the Best Explanation: If certain claims do not play a role in the best explanation of our observations, we are not justified in believing them.

Harman’s Challenge thus rests on a general epistemic principle. At a first stab, we can state the principle as follows: We are justified in believing that there exist objects of a certain kind only if the existence of those objects plays a role in the best explanation of the relevant kind of observations. As stated, this principle needs several refinements. One issue is that morality doesn’t obviously involve the existence of special moral objects. Rather, morality involves the attribution of moral properties and relations – the properties of rightness, wrongness, goodness, and badness, and the relations of is-morally-better-than and is-a-moral-reason-for, among others. So the epistemic principle needs to be expanded to apply to properties and relations, too.

A second refinement is that, presumably, for us to be justified in believing in the existence of objects, it is not necessary that the objects play a role in explaining our observations of those very objects. (And similarly for properties and relations.) It would suffice that they play a role in explaining something else, so long as what is explained is part of our total body of evidence. Moreover, if a claim is itself part of our total body of evidence, it doesn’t need to do any explanatory work for us to be justified in believing it. That it is part of our total evidence suffices.



A third refinement is that we may be justified in having some beliefs on a basis that doesn't involve Inference to the Best Explanation – for instance, we may be justified on the basis of perception or some kind of intuition, or perhaps because the beliefs in question are included in our epistemic starting point. (Harman himself endorses the claim that the beliefs included in one's epistemic starting point have a privileged epistemic status (Harman 1995: 189–93).) The epistemic principle at issue in Harman's Challenge is better understood as a principle governing the revision of belief. A belief in some claim loses justification when we acquire good reason to believe that the claim does not play an appropriate explanatory role.

Putting this all together, we get the following epistemic principle:

If we are justified in believing that the best explanation of our total body of evidence does not involve the existence of a certain kind of object or the exemplification of a certain property or relation (and our total body of evidence doesn't itself entail that the kind of object exists or that the property or relation is exemplified) then this defeats any initial justification we might have had to believe that the kind of object exists or the property or relation is exemplified.

This principle is a mouthful. But it captures a straightforward idea. For us to be justified in believing that there are moral facts – that is, facts that attribute basic moral properties – such facts had better “earn their keep”. They had better be part of our evidence or part of the picture of the world that does the best job of explaining our evidence.

Given this principle (perhaps with additional refinements), Harman's Challenge is generated by the following line of thought: Our direct evidence about the world does not include

moral facts. Moral facts are not part of the best explanation of our moral beliefs. Moral facts are not part of the best explanation of any other part of our evidence. So, by the epistemic principle, we are not justified in believing in the existence of moral facts.

As I've so far presented it, Harman's Challenge is a challenge to our justification for believing in the existence of moral facts. But it is better to view it as a challenge to robust moral realism taken as a package. This is because there are ways to avoid or answer the challenge by rejecting realist theses. It is easy to see that the challenge poses no difficulty for a moral error theorist or for someone who denies that moral beliefs are justified. What is more interesting is that some naturalists are also able to straightforwardly answer the challenge (Harman 1986). Consider a view according to which moral properties are reducible to natural properties that do explanatory work. For instance, consider a view on which the property of moral rightness just is the property of maximizing the satisfaction of preferences. Alternatively, consider a view on which the property of moral rightness just is the property of being what would be approved of by one's community if it were fully informed. On such a view, Harman's Challenge can easily be answered: The constituents of the property of moral rightness – e.g., preference, satisfaction, and the like – do important explanatory work for us. Since the property of rightness reduces to these constituents, it does not need to do any additional explanatory work to belong in our picture of the world. Such a reductive view of moral properties may raise further questions – for instance, why is it that this natural property (as opposed to some other natural property) plays such an important role in our thought and talk? And in answering these further questions, it may be helpful to show how the property of moral rightness plays a distinctive explanatory role. But to answer Harman's Challenge, no more is required.

This kind of response doesn't work for all naturalist views. Consider a naturalist view according to which moral properties are distinctive natural properties that are not reducible to other natural properties. Since moral facts are a new kind of natural fact, to directly answer Harman's Challenge, one would need to show how these natural facts earn their keep. Of course, a naturalist is in a better position than a robust realist to show this. This is for two reasons. First, adding new natural facts to one's picture of the world presumably requires less justification than adding new facts of a completely novel kind. Second, since the robust realist claims that moral facts are not natural facts, such a realist will presumably claim that moral facts cannot play a role in the causal explanation of natural phenomena. So they cannot earn their place in our picture of the world by providing causal explanations.

Harman's Challenge is therefore at its strongest targeted against robust moral realism. How might a robust realist respond to this challenge? One strategy is to find an explanatory role for moral facts to fill. Indeed, much of the debate over Harman's Challenge has taken this form, though typically on behalf of the non-reductive naturalist. For instance, Nicholas Sturgeon provides examples of cases in which we are happy to treat some moral fact as explaining some natural facts (Sturgeon 1988: 243–244). Sturgeon's examples include (i) the fact that Hitler was morally depraved helps to explain why Hitler did what he did (e.g., start a world war); and (ii) the fact that the institution of slavery was particularly bad in the 18<sup>th</sup> and 19<sup>th</sup> centuries in Europe and North America helps to explain why widespread opposition to slavery appeared then. Sturgeon argues that these examples provide an answer to Harman's Challenge.

Sturgeon is right that we are sometimes happy to treat moral facts as explaining natural facts. But his examples don't provide a good avenue of response for the robust realist. One difficulty is that, as we have seen, the robust realist claims that moral facts are not natural facts,

and so will resist the claim that moral facts can play a role in the causal explanation of natural phenomena. Instead, the robust realist will presumably account for Sturgeon's examples by claiming that what does the explanatory work in each of his cases is not a moral fact but whatever natural facts the moral fact is correlated with. Strictly speaking, it is not Hitler's depravity that explains his behavior but his racism, paranoia, and megalomania. It is not the badness of the institution of slavery that explains the widespread opposition to it, but the fact that it caused a great deal of suffering. On this view, when someone presents an apparent moral explanation of a natural fact, this is really just an elliptical way of pointing to a natural explanation of a natural fact.

A second issue with this strategy is that it is not enough to answer Harman's Challenge to present apparent examples of moral explanations of natural facts. Rather, it needs to be shown that the moral facts are indispensable for explaining the natural facts (Sayre-McCord 1988b). If we can explain the natural facts just as well without relying on moral facts, then the moral facts are not part of the best explanation of the natural facts – explanations that rely only on natural facts are more parsimonious. So Sturgeon's examples do not suffice to answer Harman's Challenge.

A different strategy for the robust realist to pursue is to reject the epistemic principle at the heart of Harman's Challenge. It is plausible that we are justified in believing that there are objects of a certain kind or that certain properties are exemplified only if the objects or properties do some important work for us. But one might think that this work need not be explanatory. There are other ways that a belief in moral facts, for example, could earn its keep.

The developments of this line of thought of which I am aware are all aimed at arguing for the existence of normative – and not strictly moral – facts. Belief in the existence of normative facts is then leveraged to help support belief in the existence of moral facts.

The most well-developed proposal of this kind is due to David Enoch (Enoch 2011: ch. 3). Enoch suggests that, just as we can be justified in believing in the existence of certain facts because they are indispensable to our best explanations, we can also be justified in believing in the existence of certain facts because they are indispensable to our deliberations. On Enoch's view, the explanatory project and the deliberative project – the project of deciding what to do – are on a par in the sense that thinkers are epistemically justified in holding those beliefs and employing those rules of inference that are indispensable for successfully carrying out either one of these projects (Enoch and Schechter 2008). Enoch further argues that belief in the existence of normative truths is indispensable to the deliberative project. Consider some difficult decision between two courses of action. For it to make sense to deliberate about which action to take, one has to assume that there is an answer to the question of what one should do. And one has to assume that this answer is independent of the deliberative process itself, and more generally, that it has the hallmarks of a robustly realist fact. This, Enoch argues, justifies us in believing that there are robustly realist normative truths.

Enoch's line of argument is appealing, but it faces a number of worries. One issue is that engaging in deliberation doesn't seem to require that one assume that there is an answer to the question of what one should do. Rather, it only seems to require that one not assume that there isn't an answer to that question. So the belief that there are normative facts may not be deliberatively indispensable. A second issue is that deliberation is a non-epistemic project, so indispensability to this project may not yield genuinely epistemic justification. (To avoid this

worry, Enoch could move to a different project – for instance, the explanatory project. Enoch might claim that it is indispensable to the project of explaining phenomena that we take there to be better and worse explanations, which is itself a normative fact. A version of this view has been proposed by Geoffrey Sayre-McCord (Sayre-McCord 1988b: 178–180).) A third worry is that epistemic justification is closely tied to truth, but being indispensable to deliberation is not. So deliberative indispensability cannot yield epistemic justification (McPherson and Plunkett 2015).

A final strategy available to the robust moral realist is to argue that some moral facts are part of our evidence, and so do not need to play an explanatory role for us to be justified in believing in them. Harman’s discussion seems to fit with a particular conception of what our evidence is – a subject’s evidence consists of claims about the subject’s empirical observations. But this is a contentious view of what our evidence consists of. One might argue that our evidence includes other claims, too, perhaps including simple logical, mathematical, and moral truths. If that’s right, then the moral truths that are part of our evidence don’t need to do any explanatory work in order for us for us to be justified in believing them. We may be justified in believing in the existence of those moral facts directly, and then gain further justification for believing in the existence of the moral facts that help to explain those moral facts, and so on. The nature of evidence is a vexed topic in epistemology (Kelly 2014), but this strikes me as a promising approach for a robust realist to take.

#### **4. Evolutionary Debunking Arguments**

The second kind of challenge I will discuss is evolutionary debunking arguments against morality. There have been many different presentations of evolutionary debunking arguments in

the literature (Ruse and Wilson 1986; Joyce 2006; Street 2006). Sometimes these arguments are focused on the moral. Other times they are focused on normativity more generally. Sometimes these arguments are used to argue for error theory. Other times they are used to argue for a mind-dependent (or society-dependent) view.

My discussion will take as its starting point the evolutionary debunking argument presented by Sharon Street (Street 2006). This is because Street's argument is among the most sophisticated in the literature, and because focusing on Street's discussion will be helpful for drawing out several lessons. Street's argument is aimed at realism about normativity in general. For simplicity, I will refocus it against realism about morality. Street also claims that her argument has force against naturalist as well as non-naturalist views. For simplicity, I'll primarily focus on non-naturalist realism.

In broad outline, Street's argument (the "Darwinian Dilemma") is as follows: Our cognitive mechanisms for making moral judgments are the products of evolution by natural selection. Indeed, "the forces of natural selection have had a tremendous influence on the content" of our moral beliefs (Street 2006: 113). According to robust moral realism, the fundamental moral truths are independent of us. However, there are good scientific reasons to think that our ancestors were not selected for believing independent moral truths (or for having the motivational tendencies that would lead to believing such truths). Rather, they were selected for having beliefs (or motivational tendencies) that would provide them with survival and reproductive advantages. There are many different possible collections of moral beliefs. It would therefore be an astonishing coincidence if our moral beliefs matched the independent moral truth. Thus, robust moral realism leads to the conclusion that it is highly likely that our moral beliefs are false. This is an unacceptable conclusion. Therefore, robust moral realism is false.

One thing to note about this argument is that it aims to show that there is an internal tension in the robust realist view. The realist claims that our moral beliefs are by and large true. If Street's line of thought is correct, there is also an argument from premises the realist should accept – a Darwinian account of the evolution of our moral capacities as well as Independence and other realist theses – to the conclusion that it is likely that our moral beliefs are false. This internal tension has epistemic force. It provides realists with reason to give up some part of their package of views. Street suggests that the best claim to give up is Independence, and that the realist should move to a mind-dependent view. But there are alternatives – one could instead give up the claim that our moral beliefs are by and large true, or the claim that moral sentences and mental representations purport to represent moral facts, or perhaps some other realist thesis.

A second thing to note is that Street's argument is not meant to be a conclusive argument against robust moral realism. The premises don't entail that robust moral realism is false. If the argument is successful, rather, it shows that there is strong reason to reject robust realism. This reason may, in principle, be outweighed by competing considerations.

How exactly should Street's argument be understood? The argument has several moving parts. It is not transparent what the crux of the argument is, or how best to characterize the epistemic principle underlying it. In what follows, I will discuss ways of trying to explicate Street's argument.

## **5. Irrelevant Influence Arguments**

A natural suggestion is to assimilate Street's argument to an "irrelevant influence" argument (Vavova 2016). This is an argument that relies on the claim that some of our beliefs were formed



in a way that reflects the impact of irrelevant causal influences. This, it is claimed, defeats our justification for those beliefs.

More carefully, the suggestion is to understand the crux of Street's argument as follows: Assuming robust moral realism, we have good reason to believe that our moral beliefs were formed in a way that reflects the significant impact of irrelevant factors – factors that are disconnected from the truth of the claims in question. (The irrelevant factors are the factors that were important in the evolution of our cognitive mechanisms for making moral judgments.) According to a general epistemic principle, if under certain background assumptions we have good reason to believe that our beliefs about a domain were formed in a way that reflects the significant impact of irrelevant factors, this puts pressure on the combination of our beliefs about the domain with the background assumptions. Thus, there is an internal tension in the robust moral realist's package of views.

The problem with this argument is that the "general epistemic principle" it relies upon has counterexamples. Many of our beliefs were formed in ways that reflect the significant impact of irrelevant factors. For example, the fact that I was bored earlier today led me to surf the Internet, which led me to read an article on the red-spotted newt, which led me to come to have several new beliefs about the red-spotted newt. But the fact that I was bored is disconnected from the truth of claims about newts. Similarly, the fact that I grew up in the United States led me to have many beliefs about what various words mean in the English language. But the fact that I grew up in the United States is disconnected from the truth of claims about the meaning of those words.

Perhaps these counterexamples can be avoided by somehow restricting the "general epistemic principle". For instance, one might note that the irrelevant influences in my examples

led me to acquire evidence about red-spotted newts and about the meanings of words in English. This suggests a potential fix: Irrelevant influences are not problematic when they lead to the acquisition of evidence. According to this proposal, learning about the significant impact of an irrelevant influence only defeats the justification of a belief if one is justified in believing that the irrelevant influence did not act via leading one to acquire new evidence that supports the belief.

The proposed fix, however, doesn't work. One problem is that learning about an irrelevant influence can defeat the justification of one of my beliefs even if the influence led me to acquire new evidence – for instance, I might find out that the influence led me to acquire evidence in a biased way or to misevaluate the new evidence. A second problem is that learning about an irrelevant influence can fail to defeat the justification of one of my beliefs even if the influence did not lead to me acquire new evidence – for instance, I might find out that the influence prompted me to better evaluate the evidence that I already possessed. Finally, there are areas of thought where talk of evidence seems inappropriate. For instance, it is plausible that whatever justifies my beliefs in basic logical, mathematical, and conceptual truths is not properly described as evidence. (What is my evidence for the claim that 2 is a number or that bachelors are unmarried?) If there is a distinction between problematic and unproblematic irrelevant influences in these areas of thought, this distinction cannot be drawn by appealing to the notion of evidence.

A different proposed fix is to claim that learning of the significant impact of irrelevant influences is problematic only when one has good independent reason to think that the influences have likely led one to have false (or unjustified) beliefs. More carefully, the proposed epistemic principle is as follows: If, under certain background assumptions, we have good independent reason to believe that our beliefs about a domain were formed in a way that makes it likely that

they are false (or unjustified), this puts pressure on the combination of our beliefs about the domain with the background assumptions. Here, “we have good independent reason” means something like “putting aside our beliefs and reasoning about the relevant domain, we have good reason”. Versions of this principle have been defended by David Christensen and Adam Elga in the context of peer disagreement arguments, and by Miriam Schoenfield and Katia Vavova in the context of irrelevant influence arguments (Christensen 2007; Elga 2007; Schoenfield 2014; Vavova 2016). To make use of this principle in an argument against robust moral realism, one would also have to argue that assuming robust realism, we have good independent reason to think that the evolutionary origin of our cognitive mechanisms for making moral judgments makes it likely that our moral beliefs are false.

This proposed fix faces difficulties, too. First, it is not clear exactly how to understand “good independent reason” – which beliefs and reasoning are we supposed to put aside, and what exactly is it to put these beliefs and reasoning aside? Second, the proposed epistemic principle has problematic consequences. Consider the case of perception. Putting aside our perceptual beliefs, we have good reason to think that perception is likely to yield false beliefs about our environment. After all, there are very many possible perceptual mechanisms that a creature could employ, and only a small proportion of them tend to yield true beliefs. (We do, of course, possess a psycho-physical explanation of how perception works and an evolutionary explanation of how we ended up with a reliable perceptual mechanism that explain how it is that our perceptual mechanism yields true beliefs about our environment. But these explanations are ultimately justified on perceptual grounds, and so presumably have to be put aside when applying the principle.) So, according to a minor generalization of the epistemic principle, there

is pressure to give up our perceptual beliefs or some anodyne background assumptions. That seems wrong.

What all of this suggests is that irrelevant influence arguments against robust moral realism are problematic. It is not clear how to state a correct general principle concerning the epistemic significance of irrelevant influences. (Indeed, I suspect there isn't such a principle.) Even if there is such a principle, it is not clear that it can be used to pose a difficulty for robust moral realism. If Street's argument is to have some force, we had better understand it in a different way.

## **6. The Reliability Challenge**

A better idea is to understand Street's argument – or the intuitive argument underlying it – not in terms of irrelevant influences, but in terms of explanation. What generates the intuitive problem for robust moral realism is that, assuming the truth of robust realism, there seems to be no satisfying explanation of how our moral beliefs are reliable (Enoch 2011: ch. 7). (This argument is analogous to the Benacerraf-Field argument against mathematical Platonism (Field 1989: 25–30). There are also analogous arguments concerning other areas of thought, including even logic (Schechter 2010).)

Put more carefully, the argument goes as follows: According to robust moral realism, the moral claims we believe are by and large true, at least given sufficient reflection and discussion (from Reliability). This is a striking fact, one that “cries out” for explanation. In particular, explanation is needed of how it is that we have cognitive mechanisms for making moral judgments that tend to yield true beliefs. Given robust moral realism, there is no satisfying explanation of our reliability that can be provided. That is because our moral beliefs do not

constitutively explain the moral facts (from Independence), the moral facts do not causally explain our moral beliefs (from Naturalism), and there is no third factor that somehow explains both our moral beliefs and the moral facts. According to a general epistemic principle, it is a cost of a theory if it treats a striking phenomenon within the scope of the theory as accidental or otherwise inexplicable. Thus, there is a tension in the robust moral realist's overall package of views.

This is a better explication of Street's argument. It correctly highlights the role that the "astonishing coincidence" that we are reliable about the moral plays in generating the problem (Bedke 2009). It also relies on a plausible epistemic principle – it is a cost of a theory if it treats a striking phenomenon within the scope of the theory as inexplicable. This general principle seems central to our practices of theory choice.

One important feature of this argument is that evolution turns out not to play a role in generating the challenge for robust moral realism. The problem for the robust realist does not stem from there being some causal story of how we came to have our cognitive mechanisms for making moral judgments. Rather, the problem stems from the fact that we seem to lack a satisfying explanation of how it is we have reliable cognitive mechanisms for making moral judgments. Or, more accurately, the problem stems from the fact that there seems to be principled reasons to think there cannot be a satisfying explanation of how we have reliable cognitive mechanisms for making moral judgments, at least assuming robust moral realism.

I think that the fact that evolution does not play a role in this argument is exactly right. Pointing to the evolution of our moral capacities is part of what makes Street's argument so arresting. But this is more of a psychological phenomenon than an epistemological one. The evolutionary story makes salient a worry the robust moral realist already should have had –

namely, that there is no satisfying explanation of how it is that we are reliable about morality. One way to see this point is to note that that if we did possess a satisfying explanation of our reliability about morality, the evolutionary story of how we came to have our cognitive mechanisms for making moral judgments would not seem at all worrisome. A second way to see the point is to note that the evolutionary story does not rule out any explanation of our reliability about morality compatible with robust moral realism that was not already ruled out by a broadly scientific view of the world. So the presence of the evolutionary story does not seem to put any pressure on robust moral realism that wasn't already there.

To be fair, evolution may have a role to play if the Reliability Challenge is targeted against moral naturalism. According to moral naturalism, moral facts are natural facts. So they can play a role in causing natural phenomena. At least in principle, we may be able to explain our reliability about the moral by claiming that the moral facts caused our moral beliefs. Street's evolutionary account might be seen to put pressure on this sort of approach. If Street's evolutionary account rules out (or renders less plausible) potential naturalist explanations of our reliability, it will contribute to generating the problem for moral naturalism. Of course, to see if this is really so, one would need to generate specific candidate naturalist explanations of our reliability and determine whether Street's evolutionary account rules them out (or renders them less plausible).

A second important feature of the argument is that it rests on explanatory considerations. To answer the Reliability Challenge, one does not need to provide a justification of our moral beliefs or of robust moral realism. One doesn't need to provide new evidence for these claims. Rather, what is needed is an explanation of how it is that we are reliable about the moral. In principle, this explanation could make use of our moral beliefs or robust moral realism. Since the

challenge is an explanatory rather than a justificatory one – and since the challenge is generated by an internal tension within the robust realist’s package of views – it is perfectly fine to make use of robust realist theses or substantive moral claims in answering the challenge.

How should we respond to the Reliability Challenge? One option is to reject robust moral realism. There are several ways in which this could be done. We could give up the claim that moral sentences and mental representations purport to represent moral facts. We could give up the claim that we are reliable about the moral. As Street recommends, we could give up the Independence thesis and instead claim that we are reliable because the moral truths are constitutively explained by our moral beliefs. Alternatively, we could move to a naturalist view and claim that we are reliable because our moral beliefs are caused by the moral truths. (Street presents arguments against such a response, most notably that if the naturalist explanation relies on our substantive views, it will be question-begging (2008: 215–217). As we’ve seen, this isn’t obviously correct.) Or, instead, we could move to a plenitudinous view of moral properties. If every coherent moral practice fits some set of moral properties, it is easy to explain our reliability – no matter our practice, we would be bound to get it right

Assuming we want to retain robust moral realism, how could we respond to the Reliability Challenge? One possible response is to reject the need to answer the challenge. The most sophisticated version of this response is due to Justin Clarke-Doane (Clarke-Doane 2015). He argues that information can provide a reason to give up beliefs about a domain only if it provides reason to doubt the safety or the sensitivity of those beliefs. (Beliefs about a domain count as safe just in case we could not easily have had false beliefs about the domain. Beliefs about a domain count as sensitive just in case had the truths about the domain been different, we would have had correspondingly different beliefs.) Clarke-Doane argues that arguments such as

the Reliability Challenge do not provide reason to doubt the safety or sensitivity of our moral beliefs, and so have no force. Clarke-Doane further argues that we have good reason to think that our fundamental moral beliefs are both safe and sensitive. If our fundamental moral beliefs are true, they are safe, since the evolutionary origins of our moral beliefs show that we could not easily have had different fundamental moral beliefs. If our fundamental moral beliefs are true, they are necessarily true, and so are automatically sensitive.

There are several problems with Clarke-Doane's response. One issue is that we might understand the Reliability Challenge as targeting the safety and sensitivity of our moral beliefs. If there is no explanation of how it is that we are reliable, this provides reason to think that we may not be reliable, which in turn provides reason to think that our beliefs are neither safe nor sensitive. A second issue is that there is reason to think that Clarke-Doane's epistemic principle is false. The principle makes the defense of necessarily true beliefs much too easy. There can be good arguments against beliefs that thinkers take to be both necessarily true and hard to avoid. If his principle were correct, these arguments could always easily be answered.

A second potential response to the Reliability Challenge is to claim that we are reliable about morality because our powers of rational reflection helped us to arrive at the moral truth. This proposal faces two difficulties. First, a version of the Reliability Challenge applies to our beliefs about what is rational. So this response may simply move the bump under the rug. Second, rational reflection requires a starting point – it starts with some initial beliefs and (hopefully) improves them. But if our initial moral beliefs are wildly unreliable, even careful rational reflection will be unlikely to yield true moral beliefs (Street 2006: 123–124).

A third proposal is to claim that what explains our reliability involves the nature of moral concepts (Cuneo and Shafer-Landau 2014: 424–8). On this view, if one accepts claims that are



far enough away from the genuine moral truths, one is no longer thinking or talking about morality, but about something else entirely. So there is a limit to how unreliable thinkers can be about the moral. There are three problems with this proposal. First, even if there is a limit to how unreliable someone can be about morality, one can still be pretty unreliable about the moral. So we still need an explanation of how we are as reliable as we (think) we are. Second, if the view is not to be a plenitudinous view, practices that are sufficiently far away from the genuine moral truths must either correspond to properties that are metaphysically less privileged than genuine moral properties or they must not correspond to properties at all. So we still would need an explanation of how it is that we ended up with a practice that corresponds to metaphysically privileged properties. Third, even if this proposal were to explain our reliability about the moral, it would not explain our reliability about what one ought to do in a more general sense (i.e., the “plain vanilla” ought). It is implausible that there are significant conceptual constraints built into our most general normative concepts, so the analogous strategy would not seem to work for explaining our reliability about beliefs involving these concepts.

In my view, the most plausible direct response to the Reliability Challenge is to present a “third-factor” view, according to which there is some factor that explains both the moral truths and our moral beliefs (Wielenberg 2010; Enoch 2011: 168–174; Sarksaune 2011). For example, according to a (toy) proposal due to Enoch, our ancestors were selected to behave in ways that promoted their survival. Natural selection did this in part by encouraging our ancestors to believe that it is good to behave in those ways. Moreover, survival is (typically) a good thing, and it is (typically) good to act in ways that promote survival. This is part of what explains our reliability about what’s good. On this view, facts about what promotes survival causally explain our moral beliefs and also constitutively explain the moral facts. One can challenge this specific

explanation. And one might worry that this view leaves something left to explain – for instance, that what natural selection aims at has a positive normative status. But this kind of view strikes me as a plausible avenue to pursue.

There is an indirect response to the Reliability Challenge also worth mentioning: Robust moral realism has “companions in guilt”. This is so in two different senses. First, other domains face analogous challenges – including logic, mathematics, modality, conceptual truth, and so on. It is implausible that we should give up realism about all of these domains. This suggests that there must be some way of responding to the challenge. Second, versions of the Reliability Challenge arise for alternatives to robust moral realism. For instance, naturalist realist views are subject to the challenge. More interestingly, mind-dependent views of morality may also be subject to the challenge. There are very many possible ways in which the moral facts could depend on facts about our minds. It is a striking fact that the way that the moral facts actually depend on facts about our minds (according to a mind-dependent view) is one where the moral truths turn out to by and large match our moral beliefs (as opposed to, for instance, their negations). This fact “cries out” for explanation, too. If this is right, then robust moral realism doesn’t face a distinctive problem.

Finally, it is worth noting that, even if it is successful, the Reliability Challenge does not conclusively show that robust moral realism is false. It only generates strong reason to reject robust moral realism. This reason may, in principle, be outweighed by competing considerations.

## **7. Conclusion**

Let’s take stock. If what I’ve argued here is correct, there are two pressing explanatory challenges facing robust moral realism. According to Harman’s Challenge, for us to be justified

in believing in the existence of moral facts, such facts had better be part of the picture of the world that does the best job of explaining our evidence. Given robust moral realism, it is difficult to see how this could be so. According to the Reliability Challenge, it is a striking fact that we are reliable about morality. Given robust moral realism, it is difficult to see how there could be an explanation of our reliability. This puts pressure on robust moral realism. (And, perhaps, analogous challenges put pressure on alternative metaethical views.)

There are plausible lines of response open to the robust moral realist. In response to Harman's Challenge, the robust realist might reject the central epistemic principle generating the challenge and claim that moral facts "earn their keep" in some way that does not involve explanation. Alternatively, the robust realist might claim that certain moral claims are part of our total body of evidence. In response to the Reliability Challenge, the robust realist might claim that some "third-factor" view is correct, according to which some factor both constitutively explains the moral truths and causally explains our moral beliefs. Alternatively, the robust realist might argue that the Reliability Challenge is a challenge facing all plausible views of morality, and so does not pose a problem specific to robust realism.

It is fair to say, though, that it has not been established that these challenges can be answered, either when targeted against robust moral realism or when targeted against alternative metaethical views. These explanatory challenges are still very much live.

### **Acknowledgements**

Thanks to Tristram McPherson and David Plunkett for their helpful comments and boundless patience. Thanks to Jamie Fritz for extremely useful comments. Thanks also to David Christensen and David Enoch for discussion of many of these issues. This chapter was written

while I was a visiting fellow at the Humanities Institute at the University of Connecticut. I am grateful to the University of Connecticut for its hospitality.

## References

- Balaguer, M. (1998) *Platonism and Anti-Platonism in Mathematics*, Oxford: Oxford University Press.
- Bedke, M. (2009) "Intuitive Non-Naturalism Meets Cosmic Coincidence," *Pacific Philosophical Quarterly* 90: 188–209.
- Benacerraf, P. (1973) "Mathematical Truth," *The Journal of Philosophy* 70: 661–79.
- Christensen, D. (2007) "Epistemology of Disagreement: The Good News," *Philosophical Review* 116: 187–217.
- Clarke-Doane, J. (2015) "Justification and Explanation in Mathematics and Morality," *Oxford Studies in Metaethics* 10: 80–103.
- Cuneo, T. and Shafer-Landau, R. (2014) "The Moral Fixed Points," *Philosophical Studies* 171: 399–443.
- Elga, A. (2007) "Reflection and Disagreement," *Noûs* 41: 478–502.
- Enoch, D. (2011) *Taking Morality Seriously*. Oxford: Oxford University Press.
- Enoch, D. and Schechter, J. (2008) "How Are Basic Belief-Forming Methods Justified?," *Philosophy and Phenomenological Research* 76: 547–79.
- Field, H. (1989) *Realism, Mathematics, and Modality*, Oxford: Basil Blackwell.
- Fine, K. (2012) "Guide to Ground," in F. Correia & B. Schnieder (eds.), *Metaphysical Grounding*, Cambridge: Cambridge University Press, pp. 37–80.
- Goldman, A. (1988) *Empirical Knowledge*, Berkeley: University of California Press.

- Harman, G. (1977) *The Nature of Morality*, Oxford: Oxford University Press.
- \_\_\_\_\_. (1986) "Moral Explanations of Natural Facts," *The Southern Journal of Philosophy* 24 supplement: 57–68.
- \_\_\_\_\_. (1995) "Rationality," in Smith, E. and Osherson, D. (eds.), *Thinking*, vol. 3, Cambridge, Mass.: MIT Press, pp. 175-211.
- Jenkins, C. (2006) "Knowledge and Explanation," *Canadian Journal of Philosophy* 36: 137–63.
- Joyce, R. (2006) *The Evolution of Morality*, Cambridge, Mass.: The MIT Press.
- Kelly, T. (2014) "Evidence," *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition). Available at: <http://plato.stanford.edu/archives/fall2014/entries/evidence/>.
- Lipton, P. (2004) *Inference to the Best Explanation*, second edition, New York: Routledge.
- McPherson, T. and Plunkett, D. (2015) "Deliberative Indispensability and Epistemic Justification," *Oxford Studies in Metaethics* 10: 104–133.
- Ruse, M. and Wilson, E. O. (1986) "Moral Philosophy as Applied Science," *Philosophy* 61: 173–192.
- Sarksaune, K. (2011) "Darwin and Moral Realism: Survival of the Fittest," *Philosophical Studies* 152: 229–43.
- Sayre-McCord, G. (ed.) (1988a) *Essays on Moral Realism*, Ithaca: Cornell University Press.
- \_\_\_\_\_. (1988b) "Moral Theory and Explanatory Impotence," in Sayre-McCord (1988a), pp. 256–281.
- Schechter, J. (2010) "The Reliability Challenge and the Epistemology of Logic," *Philosophical Perspectives* 24: 437–464.
- Schoenfield, M. (2014) "Permission to Believe: Why Permissivism Is True and What It Tells Us about Irrelevant Influences on Belief," *Noûs* 48: 193–218.

- Street, S. (2006) “A Darwinian Dilemma for Realist Theories of Value” *Philosophical Studies* 127:109–166.
- \_\_\_\_\_. (2008) “Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About,” *Philosophical Issues* 18: 207–28.
- Sturgeon, N. (1988) “Moral Explanations,” in Sayre-McCord (1988), pp. 229–255.
- Vavova, K. (2016) “Irrelevant Influences,” *Philosophy and Phenomenological Research*.
- Wielenberg, E. (2010) “On the Evolutionary Debunking of Morality,” *Ethics* 120: 441–464.

### **Further Reading**

- Berker, S. (2014) “Does Evolutionary Psychology Show That Normativity Is Mind-Dependent?,” in J. D’Arms and D. Jacobson (eds.) *Moral Psychology and Human Agency*, Oxford: Oxford University Press, pp. 215–52. (A careful discussion of Street’s evolutionary debunking argument.)
- Vavova, K. (2015) “Evolutionary Debunking of Moral Realism,” *Philosophy Compass* 10: 103–116. (An introduction to evolutionary debunking arguments, distinguishing them from other kinds of arguments.)
- White, R. (2010) “You Just Believe That Because...,” *Philosophical Perspectives* 24: 573-615. (A sensitive discussion of many different kinds of irrelevant influence arguments.)