

104 The Explanatory Redundancy Challenge to Moral Properties

THOMAS PÖLZLER

Department of Philosophy, University of Graz, Austria

ABSTRACT: Most philosophers assume that moral properties exist. One of the main challenges to this claim has been Gilbert Harman’s “explanatory redundancy challenge”: (1) we are justified in believing in the existence of a certain kind of entity only if these entities figure in the best explanations of observations that we make, and (2) moral properties do not figure in the best explanations of any observations that we make. In this chapter, I provide an overview of the explanatory redundancy challenge and critics’ main responses to it. Moreover, I tentatively assess these responses. It will turn out that there is little reason to reject the argument’s first premise. Critics should rather try to show that moral properties are in fact needed to best explain observations that we make (by showing that these properties reduce to or supervene upon explanatorily potent non-moral properties). However, success is not guaranteed with this latter strategy either.

Introduction

Some chapters in this volume attempt to establish some particular scientific hypotheses about morality. Others attempt to infer first-order ethical conclusions from such hypotheses. This chapter is different. The argument that will be addressed can work on the basis of various scientific hypotheses: hypotheses that pertain to the role of evolution, the role of emotions, etc. Moreover, its conclusion does not belong to first-order ethics but to metaethics (to theorizing about moral questions that does not purport to answer these questions but to explore their philosophical status and foundations). More specifically, the chapter will address the metaethical question of the existence of moral properties.

Philosophers disagree about which things exemplify which moral properties, for example, are morally good, bad, right, wrong, virtuous, vicious, etc. While some philosophers are utilitarians, others side with theories in the traditions of Kant or Aristotle; all of these theories are defended in numerous different versions; sometimes the existence of correct moral theories is denied altogether in favor of bottom-up approaches; while some philosophers think that greenhouse gas emissions should be distributed equally, others think that our share should (also) depend on the extent to which we benefitted from past emissions, and so on.

That said, for all these disagreements on all these levels, most philosophers agree on the following metaethical claim: there are at least some things—whatever these things might be—that are morally good, bad, right, wrong, virtuous, vicious, etc. In other words, moral properties exist (e.g., historically, Aristotle 2012; Hume 1983; Kant 1996).

In fact, so-called moral realists even go one step further. They argue that moral properties do not only exist but that they exist in an *objective sense*, i.e., whether things have these properties is independent of what anybody thinks about them. Suppose, for example, that it is wrong to set a cat on fire. Moral realists would then claim that setting a cat on fire would be
105 wrong even if you yourself believed that it is not wrong, even if most members of your culture believed that it is not wrong, or even if you yourself or most members of your culture did not respond with disapproval to setting a cat on fire (e.g., Brink 1989; Enoch 2011; Moore 1993; Plato 2010).¹

So again, most philosophers believe that there are moral properties, and many philosophers even believe that there are objective moral properties. Still, doubts about these properties have never been completely absent from philosophy either (e.g., Ayer 1952; Blackburn 2000; Gibbard 1990; Joyce 2001; Mackie 2011). Especially in recent decades, many of these doubts have been grounded in empirical evidence, including scientific evidence about morality. For example, critics of moral realism have argued that the high prevalence and persistence of moral disagreement might be best explained by, and hence support, the claim that there are no (objective) moral properties. Rather, being good just means being

¹ Some philosophers have argued that the moral realism/anti-realism debate is best interpreted as concerning not the metaphysical issue of the existence of objective moral properties (as proposed above), but, for example, the semantic issue of the truth-aptness and truth of moral sentences (Sayre-McCord 1988b), the epistemological issue of whether we have any moral knowledge (e.g., Horwich 1998), or several non-metaphysical metaethical issues combined (e.g., Boyd 1988). This is not the place to enter into debates about the relative strengths and weaknesses of these interpretations. Suffice it to say that the metaphysical interpretation assumed in this chapter has been historically dominant and still appears to be most popular and that there are a number of good reasons in its favor (e.g., Miller 2009).

considered good by the society you live in (e.g., Fraser and Hauser 2010) or does not refer to any existing property at all (e.g., Mackie 2011).

This chapter focuses on another influential empirical challenge to the claim that (objective) moral properties exist, namely on Gilbert Harman's "explanatory redundancy challenge" (henceforth ERC) (1977). Harman has pointed out that moral properties might not only be redundant when it comes to explaining moral disagreement (as proponents of the argument from moral disagreement maintain); these properties might not figure in the best explanation of *any* observations that we make. But if this is the case, the argument goes, then we are not justified in believing that these properties exist. We should deny the existence of rightness, wrongness, goodness, badness, etc., in any sense of these terms, both an objective and a subjective sense.

The ERC is still one of the most widely held arguments against the existence of moral properties (see, e.g., Blackburn 1984; Gibbard 1990; Joyce 2001, 2006; Leiter 2001; Mackie 2011; Ruse 1986, Ruse and Wilson 1986). Since it has first been proposed by Harman, it has sparked lively controversies from different philosophical perspectives. In light of this large and insightful body of literature, it would be presumptuous to try to establish anything close to a definitive judgment about the ERC in a brief chapter such as this. My ambition here is more modest: I will try to provide a helpful overview of the ERC and critics' main responses to it that may serve as an entry point for further reading. Moreover, I will suggest some very tentative assessments of these responses.

Section 1, "The Explanatory Redundancy Challenge," will introduce the ERC as it was proposed by Harman. Section 2, "Explanations and the Existence of Entities," will address critics' response that the challenge is based on a deficient criterion of the existence of properties, at least when it comes to morality. Section 3, "Moral and Non-moral Explanations," will address the response that contrary to the ERC, moral properties do figure in the best explanation of some observations that we make. In evaluating these two kinds of responses I will suggest that the ERC still has to be taken seriously and represents a plausible way in which scientific findings about morality might undermine the existence of (objective) moral properties.

The Explanatory Redundancy Challenge

Concerns about the explanatory role of moral properties were first and most famously raised by Gilbert Harman on the first pages of an introductory book to ethics (1977). Harman begins

his considerations by proposing a general test for the existence of entities. He suggests that if we want to figure out whether a certain kind of entities exists then we should ask ourselves the following question: Is the assumption that this kind of entities exists part of the best explanation of observations that we make?

It is often difficult to establish a positive answer to this question, more difficult than to establish a negative one. This particularly holds for cases in which the available empirical evidence underdetermines our selection of explanations (there are several explanations that are equally supported by the empirical evidence). But suppose the answer to the above question is yes, i.e., we do need to appeal to the entities to best explain observations. Then, according to Harman, we are justified in believing in these entities' existence. If, on the other hand, the answer is no (if the entities are "explanatorily redundant"), then we do not have reason to believe that they exist.

Thus, Harman's criterion for the existence of entities—henceforth referred to as the "explanatory criterion"²—looks as follows:

Explanatory Criterion: We are justified in believing in the existence of a certain kind of entities only if these entities figure in the best explanations of observations that we make.

Harman suggests that the explanatory criterion vindicates typical objects of scientific research. His main example involves a physicist in a cloud chamber who sees a vapor trail and judges that there is a proton ("There goes a proton", 1977: 6). The best explanation of this observation will appeal to a number of facts, including the presence of the vapor trail, the scientist's psychology, and his/her scientific theory. In any case, whatever these facts look like, Harman claims, the best explanation will typically also include reference to the actual existence of the proton. This ontological assumption is indispensable. It is indispensable in the sense that if we do not assume that the proton exists, then the physicist's observation cannot be explained in the best available way.

107 The ERC arises from the application of the explanatory criterion to morality. Again, Harman provides a helpful illustration (1977: 7). Suppose you go around a corner and find some children setting a cat on fire for the pure fun of it. Your immediate judgment might look something like this: "This is wrong!". To best explain this judgment we will again need to assume a number of facts, such as that the children are really setting the cat on fire and that you have a certain kind of "moral sensibility", i.e., that you tend to respond with certain kinds of moral judgments to certain stimuli. Still, there seems to be a crucial difference to the

² This label has been taken from Sayre-McCord (1988a).

scientific case. According to Harman, it might not be necessary to assume that setting the cat on fire *is in fact wrong*. In other words, in contrast to the proton, the property of wrongness does not appear to add anything to the explanation of the observation—it appears to be explanatorily redundant.

Let us assume that Harman is correct in that the property of wrongness does not add anything to the explanation of your moral observation that the children's action is wrong. Then, the explanatory criterion would entail that this observation does not give us any reason to believe in the property of wrongness. And in fact, this is exactly what Harman claims:

[...] you need to make assumptions about certain physical facts to explain the occurrence of observations that support a scientific theory, but you do not seem to need to make assumptions about any moral facts to explain the occurrence of the so-called moral observations. [...] The fact that an observation of an event was made at the time it was made is evidence not only about the observer but also about the physical facts. The fact that you made a particular moral observation when you did does not seem to be evidence about moral facts, only evidence about you and your moral sensibility. [...] there does not seem to be any way in which the actual rightness or wrongness of a given situation can have any effect on your perceptual apparatus. In this respect, ethics seems to differ from science. (Harman 1977: 6–8)

As this passage shows, Harman does not only suggest that the particular property of wrongness might be dispensable to explaining the particular moral observation that setting the cat on fire is wrong; he suggests that *all* moral properties (“any moral facts”) might be dispensable to explaining *all* moral observations (“the occurrence of the so-called moral observations”).

Suppose Harman got things right: no moral property is needed to explain any moral observation. Does this suffice to show that we lack reason to believe in moral properties? Critics of the ERC may still remain unconvinced—and for good reason. We do accept other kinds of entities that do not explain our particular observations *of them*. The prime example is mathematical facts.

108 Mathematical facts (such as $2 + 2 = 4$) are not causally efficacious, which means that they cannot stimulate our perceptual apparatus. There is hence no way in which these facts could explain our “mathematical observations” (such as the “observation”, if it makes sense to use this term, that $2 + 2 = 4$). Nevertheless, we believe that mathematical facts exist. This shows that the explanatory criterion is too strong and that the fact that moral properties are not needed to explain moral observations does not suffice to show that these properties do not exist.

Harman anticipated this objection. In response, he formulated the explanatory criterion in a way that allows entities to earn their existence not only by best explaining observations *of them* but also *of other phenomena* (as also reflected in the above formulation of the criterion). This enables him to say that mathematical facts meet the explanatory criterion. After all, while these facts do not explain “mathematical observations”, they do figure in our best explanations of physical phenomena.³

In contrast, things still do not seem to look better for morality. This is because moral properties do not only seem to be redundant in explaining *moral* observations but also in explaining any other—*non-moral*—kinds of observations (such as explanations about protons, mountains, revolutions, actions, etc.).

Since an observation is evidence for what best explains it, and since mathematics often figures in the explanations of scientific observations, there is indirect observational evidence for mathematics. There does not seem to be observational evidence, even indirectly, for basic moral principles. In explaining why certain observations have been made, we never seem to use purely moral assumptions. In this respect, then, ethics appears to differ not only from physics but also from mathematics. (Harman 1977: 10)

In summary, then, Harman’s ERC to moral properties looks as follows.

P1 We are justified in believing in the existence of a certain kind of entities only if these entities figure in the best explanations of our observations that we make (explanatory criterion).

P2 Moral properties do not figure in the best explanations of any observations that we make.

Ergo: We are not justified in believing in the existence of objective moral properties.

Harman’s challenge has sometimes been misunderstood, so let me add two clarifications. First, the ERC is not supposed to be a conclusive argument against moral properties (as suggested, e.g., by Koons 2000). Harman only meant to raise a *challenge*. As will be explained below, he himself (in contrast to some others) has suggested that this challenge can in fact be plausibly met. Second, some discussants have given the impression that Harman’s challenge only applies to moral realists, i.e., to those who think that moral properties are objective (Joyce 2006; Leiter 2001; Ruse 1986; Ruse and Wilson 1986). But the challenge actually applies to anybody who postulates the existence of rightness, wrongness, goodness,

³ As discussed above, the explanatory criterion is about “kinds of entities”. This is especially noteworthy in the present context. While mathematical facts as a category figure in the best explanation of physical phenomena, not each individual member of this category does so. There are some specific mathematical facts that have not (yet) figured in our best explanation of any phenomenon. The explanatory criterion of course cannot by itself give us reason to believe in the existence of these facts in particular. But it can prevent us from dismissing their existence on account of these specific mathematical facts being members of the category of mathematical facts.

badness, etc., even to those who think of these properties as subjective (including the subjectivist theory endorsed by Harman).

So what, then, to make of the ERC? Is it, like Harman believes, a challenge that can ultimately be overcome and just requires us to go this meta-ethical way rather than that? Or does it really radically sink the idea of there being moral properties? This is the question that we will turn to now. In the next section, I will address responses to the ERC's first premise: responses according to which the explanatory criterion or its application to morality is ill-founded and we can hence be justified in believing in the existence of entities even if these entities do not figure in the best explanations of observations that we make. Then, in the third section, I will move on to critics who, rejecting the ERC's second premise, claim that moral properties do in fact figure in the best explanations of observations that we make.

Explanations and the Existence of Entities

According to the first response to the ERC, we can be justified in believing in the existence of (moral) entities even if these entities do not figure in the best explanations of observations that we make. To properly assess this response, a distinction needs to be introduced between two claims about the nature of moral properties: the claim that these properties are natural (as argued by naturalists, e.g., by Brink 1989; Boyd 1988; Railton 1986) and the claim that these properties are non-natural (as argued by non-naturalists, e.g., by Enoch 2011; Huemer 2005; Shafer-Landau).

What is it for a property to be natural or non-natural? Unfortunately, this is not quite clear. G. E. Moore (1993) famously suggested that a moral property qualifies as natural if and only if it belongs to the subject matter of the natural sciences or psychology. But there are various problems with this definition.⁴ In my view, in the specific context of metaethics, the most appropriate way of distinguishing the natural from the non-natural is epistemological. In what follows, I will assume that for moral properties to be natural means for them to be such that we can only come to know about them empirically and that for moral properties to be non-natural means for them to be such that they can be known *a priori*, that is by pure

⁴ Among others, at any point in time the natural sciences and psychology will be both incomplete and involve errors, which means that, on Moore's definition, some properties will turn out non-natural even though they should be classified as natural and some properties will turn out natural even though they should be classified as non-natural. In light of these problems, contemporary proponents of science-based definitions often appeal to an ideal (complete and correct) rather than the actual version of the natural sciences (e.g., Brink 1989). However, this proposal faces serious challenges as well; most importantly, we just do not know what an ideal version of the natural sciences would look like.

thinking (the main proponent of this account of the natural/non-natural distinction is Copp 2003; see also Blackburn 2006; Shafer-Landau 2006).

110 From this definition, it might already be guessed that the ERC's first premise—the explanatory criterion—is particularly to the disliking of non-naturalists. If moral properties can be known *a priori* (or, on Moore's definition, cannot be the subject matter of the natural sciences or psychology), then applying this criterion to morality might not seem appropriate or fair in the first place. And indeed, non-naturalists (along with some naturalists) have put forward several arguments against the explanatory criterion and its application to morality (e.g., Huemer 2005; Shafer-Landau 2003). Here I will look at two of these arguments. I will tentatively suggest that these arguments are not convincing.

Proponents of the first and most prominent objection claim that requiring of non-natural moral properties that they figure in our best explanations of observations simply amounts to begging the question against them; it means starting from a premise that has the non-existence of these properties essentially built into it. In this vein, Thomas Nagel, for example, has complained about the ERC as follows:

Mackie meant that reasons play no role in causal explanations. But it begs the question to assume that this sort of explanatory necessity is the test of reality for values. The claim that certain reasons exist is a normative claim, not a claim about the best causal explanation of anything. To assume that only what has to be included in the best causal theory of the world is real is to assume that there are no irreducibly normative truths. (Nagel 1986: 144)

Nagel does not explain how exactly the explanatory criterion, in his view, begs the question against moral non-naturalism. However, the reasoning behind this objection does not seem too difficult to guess.

On Harman's criterion, the only source of evidence that can give us reason to believe in the existence of any entity is observations. By an observation Harman (1977) means an immediate judgment that is based on perception: one acquires information about the situation by seeing, hearing, smelling, tasting, or touching things. This means that by making observations the yardstick of existence, the explanatory criterion presupposes a general empiricist epistemology. It presupposes that existence beliefs can only be tested by experience and cannot be tested *a priori*, by pure thinking.

But clearly, this presupposition will strike any non-naturalist as unacceptable. Recall that on what I introduced as the most appropriate understanding of the natural/non-natural distinction, non-naturalism is the claim that whether a thing is right, wrong, good, bad, virtuous, vicious, etc., can be known *a priori*. So the explanatory principle indeed rules out

the existence of non-natural moral properties as a matter of definition. Moreover, even if
111 moral properties are defined in terms of science (e.g., according to Moore, as not being among
the subject matters of the natural sciences or psychology), the question-begging charge seems
to have some bite. After all, the natural sciences and psychology are endeavors that operate on
the basis of a perception-based (empiricist) methodology.

So is Nagel right that appealing to the explanatory criterion would be unfair to non-naturalists? Without further argument—and Nagel does not provide such an argument—I still think we should not accept this conclusion too readily. That the explanatory criterion involves a commitment to a general empiricist epistemology can only be said to be problematic to the extent to which we have reason to doubt such an epistemology. If, on the other hand, we had strong independent reasons to think that perception is our only or most reliable source of knowledge about the existence of entities, then it would seem misguided to object to the explanatory criterion on grounds of it “begging the question” against non-naturalism. The question to be asked would rather be: Given that this criterion is so well grounded and successful in other respects, why should it not deliver correct results with regard to morality? Why should this particular domain (as non-naturalists claim) be so different from all or most other domains?

This is of course not the place to settle general disputes between empiricism and rationalism or between naturalism and non-naturalism. Yet, if we move to such a more holistic assessment of Nagel’s charge and strive for overall coherence in our beliefs, then, in the absence of additional arguments, this charge does not look too powerful. There are in fact strong independent reasons that support the explanatory criterion. Reasoning along its lines has helped us to remove many entities from our ontology that today are widely regarded as non-existent, ranging from ghosts to witches to phlogiston. Moreover, the criterion’s general empiricist presupposition is supported by strong philosophical arguments in favor of empiricism as well.⁵

A second objection against the explanatory criterion’s application to morality, put forward by David Enoch (2011), can be understood as explaining in more detail how the criterion begs the question against non-naturalists. Enoch grants that moral properties (understood as objective irreducible non-natural properties) do not meet this criterion, i.e., they are not needed to explain any observations that we make. Still, he maintains that we are

⁵ The anonymous reviewer of this paper pointed out that while an empiricist epistemology might be plausible for some domains, it does not seem to be plausible for others. Again, mathematics serves as a prime example. I agree that this is a serious challenge that might ultimately lead to qualifications in the content or scope of the explanatory criterion. However, due to its entanglement with broader issues in epistemology, it does not make sense to pursue it here any further.

justified in believing in the existence of these properties. This is because, in his view, while the explanatory criterion is one potential source of justification for existence beliefs, it is not the only source of this kind.

Why is the explanatory criterion reasonable? Why are we justified in assuming that entities that are needed to best explain observations exist? Part of the answer, Enoch (2011) argues, is that such entities have turned out to be indispensable for the project of explaining the world. For example, our understanding of the world would become impoverished if we no longer assumed the existence of protons. But according to Enoch, it would be wrong to hold that explaining the world is the only project that can justify existence beliefs. There can be other “intrinsically indispensable” projects as well (projects that we have very strong reasons to participate in and that are essential to our lives).

One such project, in his view, is the deliberative project that aims at figuring out what we ought to do. As rational beings, we just cannot, and ought not, do without this kind of deliberation. Thus, in addition to the explanatory criterion, Enoch endorses a “deliberative criterion”.

Deliberative Criterion: We are justified in believing in the existence of a certain kind of entities if these entities are indispensable for deliberation.

Moral properties, Enoch (2011) argues, meet the deliberative criterion. In deliberating about what to do people assume that moral questions have single correct answers. For such answers to exist, there would have to be moral properties (understood as objective irreducible non-natural properties). Hence, these properties are indispensable to these deliberations, and we have reason to believe in their existence.

[...] irreducibly normative truths, though not explanatorily indispensable, are nevertheless deliberatively indispensable – they are, in other words, indispensable for the project of deliberating and deciding what to do – and that this kind of indispensability is just as respectable as the more familiar explanatory kind. Deliberative indispensability, I argue, justifies belief in normative facts, just like the explanatory indispensability of theoretical entities like electrons justifies belief in electrons. (Enoch 2011: 50)

Enoch’s argument is too sophisticated to be addressed satisfactorily in just a couple of paragraphs (for discussion see, e.g., Cline 2016; McPherson and Plunkett 2015; Worsnip 2016). Nevertheless, I would like to point to at least two potential problems. The first problem is that, contrary to Enoch, it may in fact not be the case that we assume that moral questions have a single correct answer when we deliberate. Thus, we have no reason to believe that we

assume the existence of (objective, irreducible, and non-natural) moral properties and that the deliberative criterion vindicates these properties. Let me explain this in more detail.

The claim that people assume that moral questions have a single correct answer is an empirical claim. Recent research in moral psychology casts doubt on this claim. For example, in a study by Sarkissian et al. (2011), the more the cultural distance between disagreeing parties was made salient to subjects, the less they believed that there is a single correct answer to moral questions; and Pölzler and Wright (2020) found that many more of their subjects believed that two contradicting answers to moral questions can be right than only one of them. If these findings should be corroborated by future research, then the deliberative principle could at best support the existence of some kind of subjective moral properties but not the objective properties that Enoch meant to vindicate.

However, things might be even worse. The design of most of the abovementioned psychological studies on folk metaethics is such that these studies cannot properly rule out that subjects do not have any robust or determinate view about whether moral questions have a single correct answer at all. There is at least some indication that this hypothesis is plausible. For instance, qualitative data collected by Bush and Moss (2020); Pölzler and Wright (2020); and Wright (2013) suggest that a fair number of subjects do not fully grasp what is at issue in questions about the objectivity of morality. If this alternative empirical hypothesis turned out true—the hypothesis that people do not have any robust or determinate views about whether moral questions have a single correct answer at all— then the deliberative principle could not even justify believing in subjective moral properties.⁶

In addition, and even more importantly, one can also reasonably doubt the validity of the deliberative criterion (Cline 2016; McPherson and Plunkett 2015; Worsnip 2016). It is easy to see why it would make sense to accept the constituents of best explanations into our ontology. These explanations are explicitly geared toward getting things about the world right, and many of them will be true. But why assume that the fact that an entity is necessary for *deliberation* justifies believing in the existence of this entity? Couldn't it be that in figuring out what to do people assume all sorts of things that actually do not exist; perhaps even *must* assume these things?

⁶ Enoch would probably dismiss the above-cited empirical studies as irrelevant, namely on grounds of their measuring people's explicit metaethical commitments—while what is relevant, according to him, are their implicit commitments, i.e., not what they say but what is presupposed by what they say and how they act (Enoch 2014). However, there is not yet any clear evidence that people's implicit metaethical commitments diverge from their explicit commitment. The only study that seems to have come close to measuring implicit commitments, on the contrary, found that these commitments are very much in line with what people actually say about their metaethical intuitions (Wagner et al. 2023).

This is the essence of a criticism that Brandan Cline has recently leveled against the deliberative criterion. Cline provides a plausible example of the potential insufficiency of the criterion. Following Searle, he speculates that the assumption that we have free will is needed for deliberation since if we do not think that we have different possibilities that are open to us, we cannot engage in rational decision-making. Yet, this deliberative indispensability of belief in free will does not seem to give us strong reason to believe that free will actually exists. It could still have to be treated as an illusion.

The conclusion that Cline draws is general. The deliberative criterion is not a legitimate independent means of settling disputes about the existence of entities. It may provide preliminary evidence (just as the fact that our perceptual experience commits us to the existence of some entities provides preliminary evidence of these entities' existence). But if we want to get to the bottom of things, we cannot but engage in explanatory considerations.

114 For example, only if free will is needed to best explain human action and phenomenology would we have strong reason to believe in it.

“ICD [Inference to the Commitments of Deliberation, i.e., the deliberative criterion]”, Cline summarizes, “even when successful, is in principle incapable of settling ontological disputes. This is why ICD is not fit to serve as an independent guide in our quest to work out what there is” (2016: 3248).

Moral versus Non-moral Explanations

In the previous section, I suggested that assuming the explanatory criterion and applying it to morality are legitimate theoretical moves. Yet, even if this is true, the ERC may of course still fail. For the explanatory criterion to cast doubt upon the existence of moral properties, it also needs to be the case that these properties do not figure in the best explanation of observations that we make. Many defenders of the existence of moral properties have argued that this claim can be shown to be wrong, i.e., that moral properties do in fact figure in the best explanation of observations (e.g., Brink 1989; Sturgeon 1984).

In investigating this response to the ERC, it will prove helpful to introduce labels for two distinct kinds of explanations: moral explanations and non-moral explanations (see Leiter 2001).

Moral explanation: Any explanation of an observation that assumes that moral properties exist.

Non-moral explanation: Any explanation of an observation that does not assume that moral properties exist.

Moral explanations are explanations that, in some sense or another, involve the assumption that moral properties exist. Often, the *explanandum* of such an explanation is a moral observation. For example, proponents of moral explanations claim that in Harman's cat example, the person observes that it is wrong to set the cat on fire because it is in fact wrong to set the cat on fire. In addition, some critics of the ERC have even proposed moral explanations for non-moral observations. Nicholas Sturgeon, for example, argues that the oppressiveness of slavery is part of the best explanation of why opposition to it arose at particular times and places.

115 An interesting historical question is why vigorous and reasonably widespread moral opposition to slavery arose for the first time in the eighteenth and nineteenth centuries, even though slavery was a very old institution; and why this opposition arose primarily in Britain, France, and in French and English speaking North America, even though slavery existed throughout the New World. There is a standard answer to this question. It is that chattel slavery in British and French America, and then in the United States, was much worse than previous forms of slavery, and much worse than slavery in Latin America. Equally good for my purpose would be the more limited thesis that explains the growth of antislavery sentiment in the United States, between the Revolution and the Civil War, in part by saying that slavery in the United States became a more oppressive institution during that time. The appeal in these standard explanations is straightforwardly to moral facts. (Sturgeon 1988: 245)⁷

Non-moral explanations, in contrast, do not appeal to moral properties. They explain relevant observations in completely non-moral terms, e.g., by referring only to physical properties of situations and psychological properties of the individuals who make observations.

The ERC does not depend on any *particular* non-moral explanation being the best explanation of moral or non-moral observations (see Joyce 2006). To illustrate, in one of the above quotations, we found Harman suggesting—very vaguely—that our observation that it is wrong to set the cat on fire may be partly explained by our “moral sensibility” (1977: 6). Suppose this explanation turned out to be false. Still, there are numerous other non-moral explanations that might be appealed to (many such explanations are discussed in this volume), and if any of these explanations comes out better than the moral explanation, then the ERC will still succeed.

In the context of this chapter, there is no need to figure out which non-moral explanation is best. Nevertheless, for the sake of illustration, let me sketch one kind of non-moral

⁷ Miller (2003: 144–145) quotes the same passage by Sturgeon.

explanation that has been widely discussed by empirical scientists. According to this explanation, the ultimate cause of humans' capacity to make moral observations is that these observations increased our ancestors' reproductive success by leading them to help others more readily (Joyce 2006). A "dual process" theory then explains the proximate nature of our moral capacity (Greene 2013). Often, moral observations are the product of automatic, unconscious, effortless emotional responses to situations (as stressed by Haidt 2012). In addition, there are also cases in which we come to make such observations by first going through some (conscious or unconscious) process of reasoning (as stressed by Mikhail 2011).

Suppose we settled on some non-moral explanation, such as the evolution/dual process explanation just described. How can we figure out whether this explanation is better than any non-moral explanation when it comes to observations such as "Setting the cat on fire is wrong" or "Opposition to slavery arose at particular times and places"?

116 For such a claim to be well grounded, we first need to think about what makes an explanation a good explanation, i.e., about what properties of explanations should count as "explanatory virtues" (Lipton 2004).⁸ There is some disagreement about the details of the matter. Most philosophers accept, however, that among the most important explanatory virtues are certain forms of coherence (being logically and probabilistically consistent with and best explaining/being best explained by other hypotheses), ontological simplicity (being committed to few kinds of entities), syntactic simplicity (involving few and non-complex hypotheses), and consilience (explaining many kinds of phenomena; Thagard 1978).⁹

The question to ask, then, is: Do we have good reason to assume that non-moral explanations are generally more coherent, ontologically simpler, syntactically simpler, and more consilient than moral explanations?

At first sight, this question seems to invite a positive answer (Leiter 2001). Plausible non-moral explanations, such as our evolutionary/dual process explanation, always involve commitments to fewer kinds of entities, because per definition they only refer to non-moral (and not also to moral) properties. Of course, non-moral explanations' number of hypotheses is higher, and the content of these hypotheses tends to be more complex. For example, the

⁸ One might worry that by appealing to explanatory virtues, proponents of the explanatory criterion give up on their general empiricist outlook. Which of a given number of explanations is best does not seem to be a purely empirical question any longer. Proponents of the explanatory criterion best reply that explanatory virtues can be themselves be identified with empirical properties, so that grasping and applying them becomes an empirical matter as well. Moreover, it seems perfectly consistent to deny the existence of "practical values" (such as moral properties), but still endorse the existence of "theoretical values" (such as explanatory virtues) in this empirical sense.

⁹ Needless to say, the above characterizations of coherence, ontological simplicity, syntactical simplicity, and consilience are very rough. For more detailed accounts, see Lipton (2004).

evolutionary/dual process explanation involves numerous complex sub-hypotheses, while on the face of it, the explanation that people observe things to be wrong or oppose things because these things are in fact wrong is very simple. But this deficiency is easily compensated for by much greater consilience.

One of the main problems with moral explanations is that they only seem to explain moral observations. Non-moral explanations, in contrast, also explain many non-moral kinds of observations. For example, in addition to explaining the phylogenetic development of the human moral capacity, evolutionary theory also contributes to explaining why humans have opposable thumbs, why giraffes have long necks, and why some animals can fly. Thus, Brian Leiter writes:

Moral properties [...] suffer from PEN [the problem of explanatory narrowness]: they are too neatly tailored to only one sort of explanandum— that which I am calling the moral phenomena—for us to think that moral properties are real (explanatory) properties. The comparison of MEs [moral explanations] with NEs [non-moral explanations] should make this obvious, but even without a comparison, moral facts, like our aforementioned spirit facts, seem to suffer from PEN. (Leiter 2001: 88)

117 Again, then, at first sight, the ERC's second premise may seem like an easy win. Moral properties indeed seem to be dispensable when it comes to explaining observations. But as acknowledged by Leiter, Harman, and most other discussants of ERC, things are in fact not quite as simple. This is because what appear to be non-moral explanations might in fact be claimed to refer to moral properties after all. In particular, these properties might be somehow implicit in (supposedly) non-moral explanations.

There are two ways in which moral properties might be implicit in what appeared to be non-moral explanations. First, as claimed by some naturalists, moral properties might be *reducible* to non-moral properties. This would mean that moral properties are identical to a subset of non-moral properties, in the way in which water may be said to be identical to H₂O (and thus reducible to molecules) (e.g., Railton 1986).¹⁰ Second, as accepted by almost all metaethicists, moral properties might *supervene* upon non-moral properties. This means that things can only differ in their moral properties if they differ in their non-moral properties: for example, in the way in which an image on your computer can only change if the pixels that constitute this image change.

¹⁰ For such an ontological (as opposed to semantic) understanding of the reduction relation see, e.g., Heathwood (2013). Needless to say, what is at issue here are “type-type” rather than “token-token” reductions. That is, the question is whether the kind of moral properties are identical to any kind of natural properties, and not whether particular moral properties are identical to particular natural properties.

To be sure, the mere possibility of reduction or supervenience relations does not get defenders of moral properties off the hook. For the ERC to be met, moral properties would not only have to reduce to or supervene upon non-moral properties in any general sense; they would have to reduce to or supervene on *those particular non-moral properties that are referred to by our best non-moral explanations*. For example, assuming the evolutionary/dual process explanation introduced above, moral properties such as goodness or rightness would have to reduce to or supervene upon properties such as increasing reproductive success or having positive emotional responses to a situation.

But this is of course exactly what many critics of ERC have claimed. Enoch, for example, who is also a proponent of this second response to the challenge, objects to evolutionary debunking arguments against moral realism by pointing out that “survival or reproductive success (or whatever else evolution ‘aims’ at) is at least somewhat good” (2010: 430): “Selective forces have shaped our normative judgments and beliefs, with the ‘aim’ of survival or reproductive success in mind (so to speak). But [...] these are by-and-large good aims—aims that normative truths recommend [...]” (Enoch 2010: 430) According to Jesse Prinz’s moral theory, moral properties are reducible to emotional responses: “[a]n action has the property of being morally right (wrong) just in case it causes feelings of approbation (disapprobation) in normal observers under certain conditions” (2006: 20).

Harman anticipates this potential response to the ERC. In his (1977) book he writes: “Even if assumptions about moral facts do not directly help explain observations, it may be that moral facts can be reduced to other sorts of facts and that assumptions about these facts do help explain observations” (1977: 13). In fact, Harman does not only anticipate this response, he even endorses it. In his view, while the ERC helps showing that rightness, wrongness, goodness, badness, etc., do not exist in any objective sense, these properties can at least ground a subjectivist theory of morality: roughly, for an action to be right means that it is subject to some kind of implicit agreement or understanding of the members of a group (1977, 1996).

To my mind, claiming that moral properties reduce to or supervene upon the non-moral properties appealed to in our best non-moral explanations is the most promising response to the ERC. But it is nevertheless far from being a done deal (see also Leiter 2001). I have already hinted at one main challenge for proponents of this response. They need to show that it is indeed plausible that moral properties reduce to or supervene upon those very non-moral properties that our best non-moral explanations appeal to (such as the properties of increasing reproductive success, or helping others, or having positive emotional responses to a situation).

This will be particularly challenging in the case of reductive accounts. Ever since Moore proposed his famous open-question argument, such accounts have been objected to in a number of ways (for overviews, see Lutz and Lenman 2018; Miller 2003).

What is more, ERC critics' reduction/supervenience response comes at significant explanatory costs, especially in terms of simplicity (Leiter 2001). First, non-moral explanations only commit us to non-moral properties and only involve a given number of hypotheses. If Enoch, Prinz, Harman, and others are right, however, then two things must be added to these explanations: first, a commitment to moral properties; and second, the hypothesis that particular moral properties reduce to or supervene upon particular non-moral properties that are appealed to in the explanations. This makes the new reduction/supervenience explanations less ontologically and syntactically simple than the original non-moral explanations and hence must be justified in terms of some corresponding explanatory gains. Put differently, proponents of the response need to show that if we would not add moral properties to the originally non-moral explanations then something would be lost (in the same way in which something would be lost by rejecting that water reduces to H₂O or that an image supervenes upon pixels).

This challenge to the reduction/supervenience response has provoked extensive discussions, including a famous back and forth between Harman and Nicolas Sturgeon, a proponent of non-reductive naturalism. Sturgeon provides examples of observations that, in his view, cannot be satisfactorily explained on purely non-moral grounds. Above I have already mentioned his slavery case. In addition, he also claimed, for example, that Hitler's depravity is part of the best explanation of why he started World War II and pursued the
119 extermination of Jews (1984). If we removed these moral properties (being depraved, being oppressive) from the picture, then, according to Sturgeon, in these cases we would end up with a worse explanation of the observations.

Is this claim plausible? Sturgeon (1984, 1988) seems to think that the inferiority of non-moral explanations is due to the fact that they simply do not fully explain all relevant observations. He proposes a counterfactual test for figuring out whether a kind of entity is relevant to an explanation. This relevance is said to obtain if, in the absence of the entity, the observed phenomenon would not have occurred. In his view, this counterfactual test delivers a positive verdict for moral properties. For example, going back to Harman's original case, if setting the cat on fire for the pure fun of it had not exemplified the property of moral wrongness, then we would not have observed this action to be morally wrong.

Assessing the appropriateness of this test and its application is a highly contested issue that would go beyond the scope of this chapter. Suffice it to say that, as far as I can see, the majority of metaethicists reject Sturgeon's arguments. They believe that they are biased in favor of the existence of moral properties (for a helpful discussion, see Miller 2003). To my mind, the appropriate conclusion to draw hence is that while the reduction/supervenience response to the ERC *might* turn out successful, at this point it has not yet been plausibly established. Further arguments in favor of the response would be needed from the side of critics of the ERC.

Conclusion

In this chapter, I have discussed one of the main challenges to the metaethical claim that some things are morally right, wrong, good, bad, virtuous, vicious, etc., namely Harman's ERC. This challenge is based on two premises: (1) we are justified in believing in the existence of a certain kind of entities only if these entities figure in the best explanations of observations that we make and (2) moral properties do not figure in the best explanations of any observations that we make.

I have suggested that there appears to be little reason to reject the first premise, i.e., the explanatory criterion. Neither does it need to beg the question against non-naturalism in any problematic way nor is it promising to complement it by an additional deliberative criterion (that could be used to vindicate moral properties). The hope of critics should rather rest on challenging the claim that moral properties are not needed to best explain observations that we make. In particular, it must be shown that these properties reduce to or supervene upon non-moral properties in our best explanations, and in a way that is both independently plausible and whose losses in terms of the explanatory virtue of simplicity are outweighed by corresponding explanatory gains.

120 As far as I can see, the reduction/supervenience response has not yet been plausibly established. The ERC, therefore, appears to be alive and well; it still deserves to be taken seriously as one of the main challenges to the existence of moral properties. Such a continued engagement with Harman's argument is also to be recommended for two other reasons. First, as has already been mentioned, the ERC is closely related to and sometimes even part of other important empirical arguments against (objective) moral properties, such as arguments from moral disagreement and evolutionary debunking arguments. And second, explanatory considerations along these lines have also been put forward *in favor* of these properties.

To elaborate on the latter of these points, if it is indeed the case that we are justified in believing in the existence of a certain kind of entities only if these entities figure in the best explanations of our observations that we make, then all it would take to establish the existence of moral properties is to show that these properties do figure in the best explanations of observations, i.e., to refute the ERC's second premise. This has indeed been argued by Sturgeon (1984, 1988) and other proponents of moral realism. Thus, by pursuing such a response to the ERC, defenders of moral properties might not only reject an argument against their view, they may also develop a powerful argument in its favor.

Much is at stake, then, metaethically speaking, in discussions about the explanatory role of moral properties. I hope that this chapter helped to clarify these discussions and can serve as a guide to more detailed investigations.¹¹

122 References

- Aristotle (2012). *Nicomachean Ethics*. Chicago, IL: The University of Chicago Press.
- Ayer, Alfred J. (1952). *Language, Truth and Logic*. New York: Dover.
- Blackburn, Simon (1984). *Spreading the Word*. Oxford: Clarendon Press.
- Blackburn, Simon (2000). *Ruling Passions: A Theory of Practical Reasoning*. Oxford: Oxford University Press.
- Blackburn, Simon (2006). Antirealist expressivism and quasi-realism. In: Copp, David (ed.): *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press, 146–162.
- Boyd, Richard (1988). How to be a moral realist. In: Sayre-McCord, Geoffrey (ed.): *Essays on Moral Realism*. Ithaca, NY: Cornell University Press, 181–228.
- Brink, David O. (1989). *Moral Realism and the Foundations of Ethics*. New York: Cambridge University Press.
- Bush, Lance S.; Moss, David (2020). Misunderstanding metaethics: Difficulties measuring folk objectivism and relativism. *Diametros* 17 (64), 6–21. <https://doi.org/10.33392/diam.1495>.
- Cline, Brendan (2016). Against deliberative indispensability as an independent guide to what there is. *Philosophical Studies* 173 (12), 3235–3254. <https://doi.org/10.1007/s11098-016-0661-z>
- Copp, David (2003). Why naturalism? *Ethical Theory and Moral Practice* 6 (2), 179–200. <https://doi.org/10.1023/A:1024420725408>

¹¹ For helpful comments on this chapter, I would like to thank the editors of the book, an anonymous reviewer, and Norbert Paulo.

- Enoch, David (2010). The epistemological challenge to metanormative realism: How best to understand it, and how to cope with it. *Philosophical Studies* 148 (3), 413–438. <https://doi.org/10.1007/s11098-009-9333-6>
- Enoch, David (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
- Enoch, David (2014). Why I am an objectivist about ethics (and why you are, too). In: Shafer-Landau, Russ (ed.): *The Ethical Life: Fundamental Readings in Ethics and Moral Problems*. New York: Oxford University Press, 192–205.
- Fraser, Ben; Hauser, Marc (2010). The Argument from disagreement and the role of cross-cultural empirical data. *Mind and Language* 25 (5), 541–560. <https://doi.org/10.1111/j.1468-0017.2010.01400.x>
- Gibbard, Allan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgement*. Cambridge, MA: Harvard University Press.
- Greene, Joshua (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin.
- Haidt, Jonathan (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. London: Penguin.
- Harman, Gilbert (1977). *The Nature of Morality: An Introduction to Ethics*. Oxford: Oxford University Press.
- Harman, Gilbert (1996). Moral relativism: In: Harman, Gilbert; Thompson, Judith J. (eds.): *Moral Relativism and Moral Objectivity*. Cambridge, MA: Blackwell, 1–64.
- Heathwood, Chris (2013). Reductionism in ethics. In: LaFollette, Hugh (ed.): *International Encyclopedia of Ethics*. Oxford: Wiley-Blackwell, 4459–4466.
- 123 Horwich, Paul (1998). *Truth*. Oxford: Clarendon Press.
- Huemer, Michael (2005). *Ethical Intuitionism*. Basingstoke: Palgrave MacMillan.
- Hume, David (1983). *An Enquiry Concerning the Principles of Morals*. Indianapolis: Hackett.
- Joyce, Richard (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, Richard (2006). *The Evolution of Morality*. Cambridge and London: The MIT Press.
- Kant, Immanuel (1996). *The Critique of Practical Reason*. Cambridge: Cambridge University Press.
- Koons, Jeremy R. (2000). Do normative facts need to explain. *Pacific Philosophical Quarterly* 81 (3), 246–272.
- Leiter, Brian (2001). Moral facts and best explanations. *Social Philosophy and Policy* 18 (2), 79. <https://doi.org/10.1017/S0265052500002910>
- Lipton Peter (2004). *Inference to the Best Explanation*. London: Routledge.

- Lutz, Matthew; Lenman, James (2018). Moral naturalism. In: Zalta, Edward N. (ed.): *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/naturalism-moral/>.
- Mackie, John L. (2011). *Ethics: Inventing Right and Wrong*. London: Penguin.
- McPherson, Tristram; Plunkett, David (2015). Deliberative indispensability and epistemic justification. In: Shafer-Landau, Russ (ed.): *Oxford Studies in Metaethics*, Vol. 10. New York: Oxford University Press, 104–133.
- Mikhail, John (2011). *Elements of Moral Cognition*. Cambridge: Cambridge University Press.
- Miller, Alexander (2003). *An Introduction to Contemporary Metaethics*. Cambridge: Polity Press.
- Miller, Christian B. (2009). The conditions of moral realism. *Journal of Philosophical Research* 34, 123–155. https://doi.org/10.5840/jpr_2009_5
- Moore, George E. (1993). *Principia Ethica*. Cambridge: Cambridge University Press.
- Nagel, Thomas (1986). *The View from Nowhere*. New York and Oxford: Oxford University Press.
- Plato ([380 B.C.] 2010). Euthyphro. In: Jowett, Benjamin (ed.): *Dialogues of Plato. Translated into English, with Analyses and Introduction*. Vol. 2. New York: Cambridge University Press, 293–318.
- Pölzler, Thomas; Wright, Jennifer C. (2020). Anti-realist pluralism: A new approach to folk metaethics. *Review of Philosophy and Psychology* 11 (1), 53–82. <https://doi.org/10.1007/s13164-019-00447-8>
- Prinz, Jesse J. (2006). The emotional basis of moral judgements. *Philosophical Explorations* 9 (1), 29–43. <https://doi.org/10.1080/13869790500492466>
- Railton, Peter (1986). Moral realism. *Philosophical Review* 95, 163–207. <https://doi.org/10.2307/2185589>
- Ruse, Michael (1986). *Taking Darwin Seriously. A Naturalistic Approach to Philosophy*. Oxford and New York: Blackwell.
- Ruse, Michael; Wilson, Edward O. (1986). Moral philosophy as applied science. *Philosophy* 61 (236), 173–192. <https://doi.org/10.1017/S0031819100021057>
- Sarkissian, Hagop; Park, John; Tien, David; Wright, Jennifer; Knobe, Joshua (2011). Folk moral relativism. *Mind and Language* 26 (4), 482–505. <https://doi.org/10.1111/j.1468-0017.2011.01428.x>
- Sayre-McCord, Geoffrey (1988a). Moral theory and explanatory impotence. In: Sayre-McCord, Geoffrey (ed.). *Essays on Moral Realism*. Ithaca, NY: Cornell University Press, 256–281.

- Sayre-McCord, Geoffrey (1988b). The many moral realisms. In: Sayre-McCord, Geoffrey (ed.): *Essays on Moral Realism*. Ithaca, NY: Cornell University Press, 1–26.
- Shafer-Landau, Russ (2003). *Moral Realism: A Defense*. Oxford: Oxford University Press.
- Shafer-Landau, Russ (2006). Ethics as philosophy: A defense of ethical nonnaturalism. In: Horgan, Terry; Timmons, Mark (eds.): *Metaethics after Moore*. New York: Oxford University Press, 209–232.
- Sturgeon, Nicholas L. (1984). Moral explanations. In: Copp, David; Zimmerman, David (eds.): *Morality, Reason and Truth*. Totowa: Rowman & Allanheld, 49–78.
- Sturgeon, Nicholas L. (1988). Moral explanations. In: Sayre-McCord, Geoffrey (ed.): *Essays on Moral Realism*. Ithaca, NY: Cornell University Press, 229–55.
- Thagard, Paul R. (1978). The best explanation: Criteria for theory choice. *Journal of Philosophy* 75 (2), 76–92. <https://doi.org/10.2307/2025686>
- Wagner, Johannes; Pölzler, Thomas; Wright, Jennifer C. (2023). Implicit metaethical intuitions: Validating and employing a new IAT procedure. *Review of Philosophy and Psychology* 14, 1–31. <https://doi.org/10.1007/s13164-021-00572-3>
- Worsnip, Alex (2016). Explanatory indispensability and deliberative indispensability: Against Enoch’s analogy. *Thought: A Journal of Philosophy* 5 (4), 226–235. <https://doi.org/10.1002/tht3.220>
- Wright, Jennifer C.; Grandjean, Piper T.; McWhite, Cullen B. (2013). The meta-ethical grounding of our moral beliefs: Evidence for meta-ethical pluralism. *Philosophical Psychology* 26 (3), 336–361. <https://doi.org/10.1080/09515089.2011.633751>