

BEST EXPLANATION AND SCIENTIFIC REALISM

YVONNE RALEY

I. INTRODUCTION

In his now famous 1977 essay, Harman argues against the existence of moral facts by claiming that they have no explanatory value. His argument starts with the assumption that both in science and in ethics, we invoke general principles to explain the observations we make. But, the argument continues, in the case of ethics, these principles are not confirmed. When it comes to science, “you need to make assumptions about certain physical facts to explain the occurrence of observations that support a scientific theory.”¹ Thus, the scientific theory is “tested against the world.”² This is not so in the case of moral observations. Rather, “it would seem that you need only make assumptions about the psychology or moral sensibility of the person making the moral observations.”³

To illustrate the scientific case, Harman discusses a scientist seeing a vapor trail in a cloud chamber:

[The physicist’s] observation confirms his theory, a theory that helps give meaning to the term “proton” as it occurs in his observational judgment. Such a confirmation rests on inferring an explanation. He can count his making the observation as confirming evidence for his theory only to the extent that it is reasonable to explain his making the observation by assuming that, not only is he in a certain psychological “set” given the theory he accepts and his belief about the experimental apparatus, but furthermore, there *really was* a proton going through the cloud chamber, causing the vapor trail, which he saw as a proton (emphasis mine).⁴

¹ Gilbert Harman, “Ethics and Observation,” *Essays in Moral Realism*, ed. Geoffrey Sayre-McCord (Ithaca, NY: Cornell UP, 1988) 119–124 at 121.

² *Ibid.*: 121.

³ *Ibid.*

⁴ *Ibid.*: 121–22.

The theory about protons is therefore confirmed because an explanation is inferred. But, more importantly, this explanation involves an ontological assumption. In particular, it involves the assumption that “there was a proton going through the cloud chamber, causing the vapor trail.”⁵ It is therefore *also* confirmed that protons exist.

The moral case, on the other hand, does *not* involve any ontological assumptions. To illustrate this, Harman provides the following case as an example: You see children setting a cat on fire, and “you immediately think: ‘that’s wrong.’”⁶ The observation that this is wrong (“observation” is here used in the sense of having the thought that this is wrong)⁷ for Harman, cannot be explained by the (supposed) existence of a moral fact (e.g., the moral fact that hurting a cat is wrong): “you do not need to make to make assumptions about any moral facts to explain the so-called observations I have been talking about.”⁸ Such observations are instead explained by the psychological states, or beliefs, of the person making the observation (e.g., that the person making the observation *believes* that hurting a cat is wrong). And if moral facts do not explain anything, then there are no such things.

Note that each of these arguments has two parts. The first part involves invoking a principle (e.g., “protons which pass through cloud chambers cause a vapor trail,” or “hurting a cat is wrong”) to explain an observation. The second part of the argument involves the inference to the existence, or non-existence, of the entities the explanatory principle (explanation, for short) posits. Given that the argument Harman makes relies on inference to the best explanation, this kind of argument will be called the “argument from best explanation,” or BE for short.

Since its inception, BE has generated a tremendous cottage industry, both in ethics and in the realist literature more generally.⁹ Probably the most famous use

⁵ Ibid: 122.

⁶ Ibid: 123.

⁷ Ibid.

⁸ Ibid: 121.

⁹ Here’s a (tiny) sampling: Bas C. van Fraassen, *The Scientific Image* (New York: Oxford UP, 1980); Nancy Cartwright, *How the Laws of Physics Lie* (New York: Oxford UP, 1983); Arthur Fine, *The Shaky Game: Einstein Realism and the Quantum Theory*, 2nd ed. (Chicago, IL: U of Chicago P, 1994); Nicholas L. Sturgeon, 1985, “Moral Explanations,” *Essays in Moral Realism*, ed. Geoffrey Sayre-McCord (Ithaca, NY: Cornell UP, 1988) 229–55; Warren Quinn, “Truth and Explanation in Ethics,” *Morality and Action*, ed. Warren Quinn (Cambridge: Cambridge UP, 1993) 109–33; Geoffrey Sayre-McCord (ed.), *Essays in Moral Realism* (Ithaca, NY: Cornell UP, 1988); Peter Lipton, *Inference to the Best Explanation*, 2nd ed. (New York: Routledge, 2004); and Barry Stroud, *The Quest for Reality: Subjectivism and the Metaphysics of Color* (New York: Oxford UP, 2000). A more general discussion of “best explanation” can be found in J. L. Mackie, *Problems from Locke* (New York: Oxford UP, 1976) and Thomas Nagel, *The View From Nowhere* (New York: Oxford UP, 1986). Sturgeon and Sayre-McCord specifically criticize Harman’s application of BE to the case of values.

of BE can be found in the defense of scientific realism, which will be the focus of this discussion. Scientific realism can be understood in two ways: as realism about scientific *theories* and as realism about the *entities* these theories posit.¹⁰ Realism about theories is the view that the scientific theories we hold true describe the world the way it is. Realism about entities is the view that the entities posited by the theories we hold true, exist. It can safely be said that most realists want to be both realists about theories and realists about entities (Harman would be an instance), but there are also philosophers who are just realists about entities (e.g., Hacking).^{11,12}

BE, as represented here, is ultimately an argument for the existence of the certain entities, and an argument against the existence of others: Protons are said to exist because they are posited by the explanations of our scientific observations. But when it comes to explaining observations in ethics, no entities need to be posited (one only needs to make assumptions about one's own psychology).¹³ And further examples can easily be generated: If phlogiston, or the ether, does not need

¹⁰ See, for instance, Ian Hacking, *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science* (Cambridge: Cambridge UP, 1983) 27. In Ian Hacking, "Experimentation and Scientific Realism," *The Philosophy of Science*, ed. Richard Boyd, Philip Gasper, and J. D. Trout (Cambridge, MA: MIT Press, 1991) 247–60, Hacking credits N. Cartwright with this distinction.

¹¹ See Hacking (1983): 28, 29.

¹² While it is logically possible to be just a realist about theories (and not about entities), this position is not well represented in the literature on the subject (there's one person who holds the view: Jody Azzouni, *Deflating Existential Consequence: A Case for Nominalism* [New York: Oxford UP, 2004]). In fact, with some exceptions (Hacking [1983], but also Michael Devitt, *Realism and Truth* [Princeton, NJ: Princeton UP, 1991]), the fact that these two types of realism can come apart is not made very explicit. Yet, the debate over scientific realism is largely a debate about whether or not the entities scientific theories posit (in particular, theoretical entities such as protons, electrons, and the like) exist. Correspondingly, one finds most definitions of scientific realism to concentrate on what one might call the "existence aspect" of scientific realism. For instance, Richard Boyd, "Confirmation, Semantics, and the Interpretation of Scientific Theories," *The Philosophy of Science*, ed. Richard Boyd, Philip Gasper, and J. D. Trout (Cambridge, MA: MIT Press, 1991) 3–35 describes the realist position like this: "According to realists, when a well-confirmed scientific theory *appears* to describe unobservable 'theoretical entities,'" it is almost always appropriate to think of its "theoretical terms" as *really* referring to real unobservable features of the world, *which exist independently of our theorizing about them*, and of which the theory is probably approximately true (p. 11—note that the italics are his). And Michael Devitt, "Scientific Realism," *The Oxford Handbook of Contemporary Analytic Philosophy*, ed. Frank Jackson and Michael Smith (New York: Oxford UP, 2007) 767–91, takes the following definition of realism to reflect the literature of the subject: "Most of the essential unobservables of well-established current scientific theories exist mind-independently." His definition ignores what might be called the "truth aspect" of scientific theories entirely.

¹³ According to BE, one cannot confirm the existence of any moral properties (e.g., the property of being morally wrong); therefore, it can be assumed that such properties do not exist.

to be posited to explain any observations, then there are no such things. If, on the other hand, theoretical entities such as electrons, neutrons, etc., do need to be posited to explain observations, then they do exist. Thus, it can be said that BE style arguments are employed as a *criterion*, or a necessary and sufficient condition, for what types of entities exist. And since along the way, certain scientific principles are confirmed as well, BE also exemplifies realism about theories.

This article explores how best to formulate the criterion that BE style arguments provide when they are employed to establish realism about theoretical entities. There seem to be only two options. The first and most natural option is that BE style arguments provide a metaphysical criterion, that is, a criterion for *what exists*. The second, and weaker, option is that they instead provide an epistemological criterion—a criterion for what one should *believe* exists.

Given that BE is seen as an argument for scientific realism, it seems that one can rule out from the start that it is an epistemic criterion. For, it is quite possible that despite the *belief* that the entities posited by explanatory principles exist, these entities are not real. After all, the existence of protons and electrons is in no way dependent upon human beings and their beliefs about the world. Thus, any beliefs about them are, strictly speaking, not relevant to whether or not such things are real.

On the other hand, if it turns out that BE style arguments cannot function as a metaphysical criterion—and the arguments in section II are intended to show exactly this—the only other option is that they represent an epistemic criterion instead. This option is explored in section III, and shown not to work either. This leaves the proponent of BE at an impasse, to put it mildly.

Even though BE style arguments do not yield a criterion for what exists, or even a criterion for what one should believe exists, the intuition that explanation has an ontological role is worth taking seriously. Therefore, its roots need to be investigated (section IV). Once the source of the intuition is revealed, however, it will become clear that when the chips are down, explanation drops out of the argument for scientific realism.

II. BEST EXPLANATION AS A METAPHYSICAL CRITERION

First, BE needs to be made more precise. This can be done by formulating it like so:

(*) If an entity (or a certain type of entity) is posited in the best explanation of an observation, then that entity is real.

Put like this, it looks as though the condition BE style arguments rely on is a merely sufficient condition. But a necessary condition is also needed. Construed

as a merely sufficient condition, BE really does not have any “ontological bite”: Protons, the scientific realist wants to say, exist, but it would stand to reason that he also wants to say that phlogiston and the ether do not exist. Certainly phlogiston and the ether have dropped out of scientific explanations, so it should follow that neither of them is real.

Scientific realism, properly understood, is not just a commitment to the existence of some entities, but also a commitment to the non-existence of others. It just seems wrong to think of scientific realism as a position that is *neutral* with regard to entities like the abovementioned phlogiston and ether. Therefore, BE style arguments, when they are used to justify this realism, must be understood as offering necessary and sufficient conditions for what exists.

How should this necessary and sufficient condition be formulated? The observant reader will already have noted that simply rewording (*) as a necessary and sufficient condition is problematic. Here’s how it would look:

An entity (or a certain type of entity) is real if and only if it is posited in the best explanation of an observation.

This formulation is absurd. It suggests that if an entity has never been posited in any explanation of observed events, the entity does not exist, which rings false. Even under the most liberal interpretation of “observe,” nobody has yet observed very distant planets (or any events on those planets), for instance, and probably nobody ever will (even though our explanatory principles are supposed to hold for the events on those very distant planets). But this does not license the conclusion that these planets, and the things on them, do not exist. So the term “observation” has to go.

What might it be replaced with? One term that is sometimes used in the formulation of BE is “phenomenon.” The way the term “phenomenon” is ordinarily used, a phenomenon does not have to be observed. The term thus has a wider scope than “observation.” Therefore, using it would get rid of the obvious anthropocentric slant of BE as formulated above. If “observation” is replaced with “phenomenon,” this results in the following new formulation of BE:

An entity (or a certain type of entity) exists if and only if it is posited in the best explanation of a phenomenon.

But is the criterion, as it is stated now, really any less anthropocentric? It does not seem so, for there is something unmistakably anthropocentric about the activity of explaining itself.

Kim’s (1987) observations about explanation show why:

Explaining is an epistemological activity, and “having” an explanation is, like knowing, an epistemological accomplishment. To be in need of an explanation is to be in an epistemologically imperfect state, and we look for an explanation in an attempt to remove that imperfection and thereby improve our epistemic situation [. . .]. Our explanations are part of what we know about the world.¹⁴

This makes sense. An explanation is a vehicle for gaining knowledge. If Luke does not know why his car does not start, and his mechanic explains to him that this is because the batteries are dead, then he has acquired a piece of knowledge. Similarly, having figured out the explanation for why there is a vapor trail in the cloud chamber, namely, that the vapor trail is caused by a proton, constitutes an epistemic advance.

The problem, then, is this: The existence of protons (or anything else, for that matter) does not, and should not, depend on whatever epistemic state we are currently in. It does not depend on whether or not we have already gained knowledge about their existence. Protons exist, whether or not we can explain any of the phenomena we observe as the result of their existence.

But if this is right, then it makes little sense to see BE as a metaphysical criterion. If explanations are epistemic, and not metaphysical, accomplishments, then BE style arguments provide epistemic, and not metaphysical, conclusions. So, perhaps BE is an epistemic criterion after all, namely, a criterion for what entities one should *believe* exist, rather than a criterion for what entities exist.

III. BEST EXPLANATION AS AN EPISTEMIC CRITERION

A bit more needs to be said to motivate this switch from interpreting BE style arguments as offering a metaphysical criterion to, instead, an epistemic criterion. This is because epistemological formulations, on the face of it, are too weak to capture the intuitions that lie behind BE. As Harman says, it seems that we are making assumptions *about the world* when we try to explain the physicist’s observation. But as an epistemic criterion, BE does not appear to make such assumptions.

On the other hand, if BE is understood as a metaphysical criterion, its assertions just seem too sweeping to be credible. After all, a criterion for what exists says that an entity (or type of entity) exists if and only if “—”. Whatever fills that blank is supposed to provide a characteristic that *every existing entity* has. And to say that *everything that actually exists* (regardless of our current state of information) has to have this one characteristic is a *very strong claim*. If there is to be a criterion for

¹⁴ Jaegwon Kim, “Explanatory Realism, Causal Realism, and Explanatory Exclusion,” *Explanation*, ed. David-Hillel Ruben (Oxford: Oxford UP, 1993) 228–45, at 228.

what exists, then this criterion has to describe a characteristic that entities can have independently of us, independently of what we believe, or hope, exists.¹⁵

Therefore, the idea that the entity (or the type of entity) has to be posited by an explanation—which is, as the previous section intimated, an *epistemic* notion—seems a bad place to start. Rather, if one is out hunting for a criterion for what exists, one might more naturally look at a characteristic like “is causally efficacious”¹⁶ or “is located in space and time” or “is concrete.” These are the sorts of characteristics that entities can have quite independently of us. And this provides some indication that perhaps BE at best is an epistemic criterion.

How might BE, as an epistemic criterion, be stated? It has already been shown that the term “observation” is problematic in a formulation of BE as a metaphysical criterion. The same applies when BE is formulated as an epistemic criterion. We obviously believe in entities (and types of entities) that have not been observed. Certain types of theoretical entities, like mesons or muons, would be an example. Any formulation of BE that employs the term “observation” must therefore be set aside.¹⁷

One can also exclude the possibility of stating BE as a merely sufficient condition, for reasons analogous to those for rejecting a metaphysical version of BE (when stated as a merely sufficient condition). As a sufficient condition, and using the term “phenomenon,” as opposed to “observation,” BE would read:

If an entity (or a type of entity) is posited in the best explanation of a phenomenon, then one should believe that the entity is real.

But this formulation is too weak as a criterion. In fact, it is even weaker than its metaphysical relative. Such a criterion provides reason to believe in some entities, but no reason whatsoever not to believe in others. According to this criterion we should believe in protons because they explain certain observations. But the criterion does not foreclose belief in the ether even though it does not explain anything. Yet we do not believe that the ether exists. Hence, what is needed is a necessary condition in addition to the sufficient one. Here, then, is how BE might look:

¹⁵ For most formulations of scientific realism, mind-independence appears to be a requirement. See, for example, Devitt (1991): 14–17, as well as Devitt (2007): 767–91.

¹⁶ See again Hacking (1983), who defends a causal criterion for what exists. Another proponent of a causal criterion is David Armstrong, *Nominalism and Realism: Universals and Scientific Realism*, Vol. I (Cambridge: Cambridge UP, 1978). The phrase “criterion for what exists” was coined by Jody Azzouni, “On ‘On What There Is,’” *Pacific Philosophical Quarterly* 79 (1998): 1–18.

¹⁷ The assumption here is that to claim that mesons and muons are observable stretches the term “observation” beyond its natural limits. However, as the considerations that follow show, this criterion will fail even if the term “observation” is retained in its formulation.

We should believe in the existence of an entity (or a type of entity) if and only if that entity (or entities of that type) is posited in the best explanation of a phenomenon.

Unfortunately, this formulation of BE creates new problems. For it is certainly legitimate to ask *why* one should believe in the existence of *all* of the entities posited by explanations. That this is an issue becomes especially evident when this criterion for what we should believe exists is contrasted with competitors. Here is an intuitive one: We should believe that an entity (or a type of entity) exists if and only if it is causally efficacious. Call that the “causal criterion.” The causal criterion is the epistemological analogue to the metaphysical causal criterion mentioned earlier.

The epistemic version of BE might force us to believe in numbers. For example, the fact that there are no seats left in the subway car might be explained by there being 60 people in the car but only 50 seats.¹⁸ But numbers, on most views, do not have causal powers. So, according to the causal criterion, we have reason to believe that numbers do not exist.¹⁹ Thus, the advocate of BE needs to provide some reason for thinking that his criterion—which is, *prima facie*, broader than the causal criterion—is correct.

A look at how philosophers (traditionally) justify the causal criterion shows how such an argument is made. It would have to be shown that the metaphysical analog of the epistemic criterion is correct. For example, if all and only causally efficacious entities *do* exist, then those are the only entities we *should* believe exist. Another way to put this is to say that *how the world is* provides justification for what we should believe the world to be like. So an argument that shows one or other metaphysical criterion to be correct must be provided first.²⁰

What this means for BE is easy to see. To justify BE as an epistemic criterion, an argument of the following form is needed: The metaphysical version of BE is correct; therefore, the epistemic version of BE is justified. But if this is the case, then it is an underlying metaphysical criterion that gives the epistemic criterion bite. And thus, a metaphysical version of BE is needed to show that BE—the epistemic criterion—is right. Unfortunately, as section II showed, a metaphysical criterion based on BE is not available.

¹⁸ Harman also argues that mathematical principles are needed in explained observations of physics; therefore, there is “indirect observational evidence for mathematics” (124). He does not explicitly say, however, that for this reason, numbers are real. See Harman (1988).

¹⁹ Any nominalist would defend this position.

²⁰ In fact, Azzouni (2004: 82–87) argues that epistemic claims about what we should believe exists can only be justified by “an excursion into *metaphysics*” (i.e., by resorting to a criterion for what exists).

IV. THE FACTICITY OF EXPLANATION

It is nevertheless hard to deny that there is something to the intuition that there is a connection between scientific realism and the posits of our explanations. So there must be another way to ground this intuition—one that does not require BE to be a *criterion*. One way to ground this intuition is by assuming that explanations must be true. The view that explanations must be true is called “the facticity of explanation.”²¹

The facticity of explanation can be argued for in two ways. The first argument is due to Azzouni (2000). According to Azzouni, the factive use of “explanation” corresponds to the way we ordinarily talk. Call this “the argument from ordinary language.” Azzouni says that

if you ask for an explanation of A, and I tell you something, B, and then add that B is false, you can respond: “B *would* explain A, if B were true.” If I have been using B as an explanation of A, and subsequently learn that it is false, I stop doing so; I can no longer offer it to others *as an explanation*.²²

If Azzouni is right, then “explanation” is never used in a way that is not factive, for it would invite precisely the sort of rejoinder that Azzouni envisions in the above exchange.

The second argument for facticity is provided by Kim (1987). If, as stated in section II, explanations were vehicles for gaining knowledge, it would stand to reason that explanations have to be true. As Kim observes: “Knowledge implies truth: we cannot know something that is not the case.”²³ Call Kim’s argument “the argument from knowledge.”

While the argument from ordinary language and the argument from knowledge provide a strong *prima facie* case for the facticity of explanation, this brief treatment of them is not decisive. But showing that explanation must indeed be factive is not the object here. For the purposes of this article, what is of interest is what would follow for BE if the assumption that explanations are factive were granted.

Return, once more, to Harman’s original example. According to Harman, the principles about protons and their properties (e.g., “protons cause vapor trails in

²¹ This term is coined by Jody Azzouni in his *Knowledge and Reference in Empirical Science* (New York: Routledge, 2000). Others hold the same view; see Michael Levin, “What Kind of Explanation Is Truth?” *Scientific Realism*, ed. Jarret Leplin (Berkeley: U of California P, 1984) 124–39; William Alston, *A Realist Conception of Truth* (Ithaca, NY: Cornell UP, 1996); and David-Hillel Ruben, *Explaining Explanation* (New York: Routledge, 1990). Ruben also thinks that all explanations are causal.

²² Azzouni (2000): 61.

²³ Kim (1987): 228.

cloud chambers”) explain the observations of the scientist. And if explanations are factive, then such principles are also *true*. Next, these principles posit the existence of certain entities (protons, in Harman’s example). And because these principles are also *true*, it can now be argued that these entities exist (if “protons cause vapor trails in cloud chambers” is true, and a vapor trail was observed, then there must *be* protons). In fact, when Harman employs the term “confirmation” in his reasoning, this is perhaps what he had in mind. To confirm scientific principles, on one (common) reading, is to provide evidence for their truth.

So far, so good. Unfortunately, this provides only a sufficient condition for what exists, and not a necessary one.²⁴ The foregoing does not—in fact, it cannot—be used to show that if an entity is not posited in an explanation, it does not exist. The reason is obvious: Facticity is not a characteristic unique to explanation. There are many true statements that are not explanations. If explanations commit us to the existence of the entities they posit *in virtue of being true*, then so do all other statements that are true.²⁵ And thus, there is nothing “ontologically” unique about explanation after all. Explanation seems to drop out of the argument.²⁶

V. CONCLUSION

So, is there no ontological role for explanation? This article has argued that there is not. There seem to be only two ways of understanding BE: Either BE style arguments offer a metaphysical criterion, or they offer an epistemic one. However, there are good reasons to think that both of these formulations fail, and a third option is not available. Furthermore, once it becomes clear that the underlying reason for the perceived connection between explanation and realism is due to the (hidden) assumption that explanations are factive, the ontological burden of explanation is shifted to truth instead.

But it should not be seen as such a bad thing that explanation has no ontological role. Explanations do a lot of other work for us. As Kim says, they are a vehicle for gaining knowledge, and if that is so, then they are also a vehicle for determining truth: If a statement counts as an explanation, then it has to be

²⁴ From this, in turn, one could then construct a sufficient condition for what we should believe exists.

²⁵ This, incidentally, does not mean that the view that the truth of a statement automatically commits us to all of the statement’s posits has to be endorsed (see Yvonne Raley, “Ontology, Commitment, and Quine’s Criterion,” *Philosophica Mathematica* [Forthcoming]). But it is an assumption that (1) is common and (2) seems to be made by the supporters of BE.

²⁶ If anything, then, it is the truth of a statement that could then be seen as a sufficient condition for the existence of the entities posited by such a statement. And this would also provide warrant for believing in the existence of such entities. But again, neither of these conditions involves the notion of explanation.

true. It follows that “elevating” a statement to the status of being explanatory is one way (though not the only way) of awarding that statement the status of “true.” The role of explanation in arguments for realism is thus much more indirect: It only helps establish the truth of certain scientific theories, or theory realism, for short.

Felician College