

Moral Theory and Explanatory Impotence

Geoffrey Sayre-McCord

1. Introduction

Among the most enduring and compelling worries about moral theory is that it is disastrously isolated from confirmation. The exact nature of this isolation has been subject to two interpretations. According to one, moral theory is totally insulated from observational consequences and is therefore in principle untestable. According to the other, moral theory enjoys the privilege of testability but suffers the embarrassment of failing all the tests. According to both, moral theory is in serious trouble.

After briefly defending moral theory against the charge of in principle untestability, I defend it against the charge of contingent but unmitigated failure. The worries about untestability are, I suggest, easily met. Yet the very ease with which they are met belies the significance of meeting them; all manner of unacceptable theories are testable. The interesting question is not whether moral theory is testable but whether moral theory *passes* the relevant tests. Recently, it has become popular to hold that a moral theory passes only if it is explanatorily potent; that is, only if it contributes to our best explanations of our experiences. The problem with moral theory, on this view, is that it apparently contributes not at all to such explanations.¹ Working out a plausible version of the demand for explana-

This essay first appeared in *Midwest Studies* (Morris: University of Minnesota Press, 1988), 12:433-57. Earlier versions were delivered at the 1985 Eastern Division meetings of the American Philosophical Association, at the Research Triangle Ethics Circle, and at the University of Wisconsin-Madison, University of Notre Dame, Virginia Polytechnic Institute, University of California-Irvine, Duke University, and University of California-San Diego. The essay has benefited considerably from exposure to these audiences, and especially from comments made by Kurt Baier, Douglas Butler, Joseph Camp, Jr., David Gauthier, Joan McCord, Warren Quinn, Michael Resnik, Jay Rosenberg, Robert Shaver, and Gregory Trianosky.

1. For example, see Gilbert Harman's *The Nature of Morality* (New York, 1977) [selections reprinted in this volume] and "Moral Explanations of Natural Facts—Can Moral Claims Be Tested

tory potency is surprisingly hard. Even so, once a plausible version is found, I argue, (some) moral theories will in fact satisfy it. Unfortunately, this too is less significant than it might seem, for any argument establishing the *explanatory* potency of moral theory still falls short of establishing its *justificatory* force. (My arguments are no exception.) And, as I will try to make clear, the pressing worries concerning moral theory center on its claim to justificatory force; its explanatory force is largely beside the point. So much the worse for moral theory, one might be inclined to say. If moral theory goes beyond explanation, it goes where the epistemically cautious should fear to tread. Those who demand explanatory potency, however, cannot afford the luxury of dismissing justificatory theory. Indeed, the demand for explanatory potency itself presupposes the legitimacy of justificatory theory, and this presupposition can be turned to the defense of moral theory's justificatory force. Or so I shall argue.

2. Observational insulation

Keeping in mind that observation is theory-laden, one way to put the charge of untestability is to say that moral theory appears not to be appropriately observation-laden; unlike scientific theories, moral theories seem forever insulated from observational implications.

This objection to moral theory emerges naturally from a variation on the empiricist verification principle. Of course, as a criterion of meaning, the verification principle has for good reason been all but abandoned. Still, taken as a criterion of justifiability, rather than as a criterion of meaning, the principle seems to impose a reasonable requirement: if there is no way to verify the claims of a proposed theory observationally, then there is no way to justify the theory (unless all its claims are analytic).² Even if moral claims are meaningful, then, they might nonetheless be impossible to justify.

In favor of thinking moral theory untestable is the apparently unbridgeable chasm dividing what is from what ought to be.³ After all, claims concerning moral obligations cannot be deduced from nonmoral claims ('ought', it is often said, cannot be derived from 'is'); which suggests (to some) that 'ought-claims' are not

against Moral Reality?" *Spindel Conference: Moral Realism*, *Southern Journal of Philosophy* 24 (1986), Supplement: pp. 57-68; J. L. Mackie, *Ethics: Inventing Right and Wrong* (New York, 1977) [selections reprinted in this volume]; Simon Blackburn, *Spreading the Word* (New York, 1984); Francis Snare, "The Empirical Bases of Moral Scepticism," *American Philosophical Quarterly* 21 (1984), 215-25; and David Zimmerman, "Moral Theory and Explanatory Necessity," in *Morality, Reason and Truth*, ed. David Copp and David Zimmerman (Totowa, N.J., 1985), pp. 79-103.

2. Although this change in emphasis, from meaning to justification, represents a natural development of the verification principle, it constitutes a significant change. With it comes the rejection of the verifiability principle as grounds for noncognitivism.

3. As Reichenbach notes: "Science tells us what is, but not what should be." *The Rise of Scientific Philosophy* (Berkeley, 1951), p. 287.

'is-claims'. Since observation is always of what is, we may have reason to suspect that observation is irrelevant to what ought to be.

This argument for the is/ought distinction is too strong, though. It mistakenly assumes that definitional reducibility is a prerequisite for putting what ought to be on an ontologically equivalent footing with what is. No matter what we know about the nonmoral facts of the case, the argument emphasizes, we cannot uncontroversially infer the moral facts. Moral assertions are not definitionally reducible to nonmoral assertions. Since nonmoral assertions report what is, and since moral claims are not reducible to these others, then moral claims must not report what is. So the argument goes.

Remarkably, by similar lines of reasoning we would be constrained to admit that the claims made in psychology are not claims about facts, for psychology, no less than morality, resists definitional reduction. No matter what we know of the nonpsychological facts of the case, we cannot uncontroversially infer the psychological facts. Psychological assertions are not definitionally reducible to nonpsychological assertions. Since nonpsychological assertions report what is, and since psychological claims are not definitionally reducible to these others, then (the argument would have it) psychological claims must not report what is. Consequently, if the argument offered in support of the is/ought distinction worked, we would find ourselves stuck with an *is/thought* distinction as well. Psychology, we would have to say, reports not what is but merely what is thought—which is silly.⁴

Yet even if we put aside the is/ought distinction, the claim that moral theory is not properly observation-laden still extracts admirable support from common sense. For if people or actions or states of affairs have a worth or a dignity or a rightness about them, this is something we seemingly cannot sense directly. And most moral theories recognize that we cannot by construing moral properties as not directly observable. The unobservability of moral properties cannot pose a special problem for moral theory's testability, however, since in *this* respect, moral theory is no different from those (obviously testable) scientific theories that postulate unobservable entities.

Moreover, on at least one standard construal of what counts as an observation, some moral claims will actually count as observation reports. Specifically, if one takes an observation to be any belief reached noninferentially as a direct result of perceptual experience, there is no reason to deny that there are moral observa-

4. Although it is true that what is thought to be is not always so (just as what ought to be is not always so), reports that something is thought to be (or that something ought to be), are still assertions concerning what is the case. Moral theory is as concerned with what is as is psychology. In making claims about what ought to be, moral theory is claiming that what ought to be *is* such and such. Moral theory characteristically makes assertions such as "Killing humans for entertainment *is* wrong"; "An action *is* made worse if it results in excruciating pain for others"; "The Ku Klux Klan is a morally corrupt organization."

tions. After all, just as we learn to report noninferentially the presence of chairs in response to sensory stimulation, we also learn to report noninferentially the presence of moral properties in response to sensory stimulation.

On this liberal view of observation, what counts as an observation depends solely on what opinions a person is trained to form immediately in response to sensory stimulation, and not on the the content of the opinions.⁵ Since such opinions are often heavily theory-laden and are often about the external world rather than about our experiences, the account avoids tying the notion of observation to the impossible ideal of theory neutrality or to the solipsistic reporting of the contents of sensory experience.

Of course, we may be too liberal here in allowing *any* opinion to count as an observation simply because it is reached directly as a result of perception. Surely, one is tempted to argue, we cannot observe what is not there, so that some opinions—no matter that they are directly reached as a result of perception—may fail to be observations because they report what does not exist. As a direct result of perception, I may believe I felt a friend's touch; but in the absence of her touch, my report seems most properly treated as an illusion, not an observation. Taking this into account, it is tempting to distinguish what are merely perceptually stimulated judgments from actual observations, thus reserving 'observation' for those perceptually stimulated judgments that are accurate.

If there were some observation-independent way to determine which judgments are accurate, we might legitimately dismiss a given class of purported observations (say, moral observations) on the grounds that they fail to report the facts accurately. Yet once the prospect of divining some set of basic (and indubitable) empirical statements is abandoned, so too must be the hope of establishing what things exist without appeal (at least indirectly) to observations. If some observations are needed to support the theories we then use to discredit other observations, we need some account of observation that allows us to isolate observations as such without assuming their accuracy has already been shown. Observations (in some ontologically noncommittal sense) will be needed to legitimize the theories we use to separate veridical from nonveridical observations. It is this ontologically noncommittal sense of 'observation' that may be characterized simply as any opinion reached as a direct result of perception; and it is in this sense of 'observation' that we must allow that there are moral observations. Once moral observations are allowed, the admission that moral theories can be tested against these moral observations will quickly follow. Just as we test our physical principles against observation, adjusting one or the other in search of a proper fit, so we can test our moral principles against (moral) observation, adjusting one or the

5. Paul Churchland defends this account in *Scientific Realism and the Plasticity of Mind* (Cambridge, Mass., 1979). See also Norwood Russell Hanson, *Patterns of Discovery* (Cambridge, Mass., 1958); and Wilfrid Sellars, *Science, Perception, and Reality* (London, 1963).

other in search of a proper fit. (Many have exploited the availability of this sort of observational testing and—unsatisfyingly—treated it as the sole criterion we have for the acceptability of theories.)⁶

So neither the is/ought distinction nor the unobservability of moral properties seems to support the charge of untestability. In fact, there is reason to think moral theory passes the testability requirement in the same way any respectable scientific theory does—even if moral properties count as unobservable. Of course, how scientific theories manage to pass the testability requirement is a notoriously complicated matter. As Duhem and Quine have emphasized, scientific theories do not pass the testability requirement by having each of their principles pass independently; many of the theoretical principles of science have no observational implications when considered in isolation. Observationally testable predictions may be derived from these scientific principles only when they are combined with appropriate background assumptions.⁷

In the same way, certain moral principles may not be testable in isolation. Nevertheless, when such principles are combined with appropriate background assumptions, they too will allow the derivation of observationally testable predictions. To test the view that an action is wrong if and only if there is some alternative action available that will bring about more happiness, we might combine it with the (plausible) assumption that punishing the innocent is wrong. From these two principles taken together, we get the testable prediction that there will never be a time when punishing the innocent brings more happiness than any other action that is available. Alternatively, consider Plato's contention that 'virtue pays'. If combined with some account of what virtue is and with the (non-Platonic) view that 'payment' is a matter of satisfying preferences, we get as a testable consequence the prediction that those who are virtuous (in whatever sense we settle on) will have more of their preferences satisfied than if they had not been virtuous. Or again, if a moral theory holds that a just state does not allow capital punishment, and if we assume some particular state is just, we get as a testable consequence the claim that this country does not allow capital punishment.

In each case our moral principles have observationally testable consequences when combined with appropriate background assumptions. Experience may show that punishing the innocent does sometimes increase happiness or that misery often accompanies virtue or that the state in question does allow capital punishment. Upon making such discoveries we must abandon (or amend) our moral principles or our background assumptions or the confidence we place in our

6. See, as examples, John Rawls, *A Theory of Justice* (Cambridge, Mass., 1971); Ronald Dworkin, *Taking Rights Seriously* (Cambridge, Mass., 1977); and Philip Pettit, *Judging Justice* (London, 1980).

7. See Pierre Duhem, *The Aim and Structure of Physical Theory* (Princeton, N.J., 1954); and W. V. O. Quine, "Two Dogmas of Empiricism," in *From a Logical Point of View* (Cambridge, Mass., 1964), pp. 20–46.

discoveries. Something has to give way.⁸ Of course, we can often make adjustments in our overall theory in order to save particular moral principles, just as we can adjust scientific theories in order to salvage particular scientific principles. In science and ethics, background assumptions serve as protective buffers between particular principles and observation. Yet those same assumptions also provide the crucial link that allows both moral and scientific theories to pass any reasonable testability requirement. If the testability requirement rules out relying on background assumptions it will condemn science as untestable. If it allows such assumptions, and so makes room for the testability of science, it will likewise certify moral theory as testable. Once—but only once—background hypotheses are allowed, both scientific and moral principles will prove testable. Hence, if moral theories are unjustified, it must be for reason other than that moral theories have no testable consequences.⁹

3. Explanatory impotence

Disturbingly, just as moral theory survives any reasonable standard for testability, so too do phlogiston theory, astrology, and even occult theories positing the existence of witches. Like moral theories, each of these theories (when combined with appropriate background assumptions) generates testable consequences, and each makes cognitively packed claims about the world. Yet given what we now know about the world, none of these theories has a claim on our allegiance. Although testable, they fail the test.

Quite reasonably, then, we might wonder whether moral theories likewise fail the empirical tests to which they may admittedly be subjected. Perhaps we ought to think of moral theories as failed theories—as theories betrayed by experience. Perhaps we ought to give up thinking there are moral facts for a moral theory to be about, just as we have abandoned thinking there is such a thing as phlogiston, just as we have abandoned the belief that the heavens control our destiny, and just as we have abandoned the idea that bound women who float are witches.

In our search for an understanding of the world, each of these theories seems to have been left in the dust; every phenomenon we might wish to explain by appeal to these theories can be explained better if they are put aside. Like phlogiston theory, astrology, and theories positing witches, moral theories appear explanatorily impotent.

8. See Morton White, *What Is and What Ought to Be Done* (Oxford, 1981); and Nicholas Sturgeon, "Moral Explanations," in *Morality, Reason, and Truth*, pp. 49–78, reprinted in this volume.

9. The last few paragraphs reiterate points made in my "Logical Positivism and the Demise of 'Moral Science,'" in *The Heritage of Logical Positivism*, ed. Nicholas Rescher, University of Pittsburgh Philosophy of Science Series (Lanham, Md., 1985), pp. 83–92.

The problem is that we need suppose neither that our particular moral judgments are accurate nor that our moral principles are true in order to explain why we make the judgments or accept the principles that we do. It seems we make the moral judgments we do because of the theories we happen to embrace, because of the society we live in, because of our individual temperaments, because of our feelings for others, but not because we have some special ability to detect moral facts, not because our moral judgments are accurate, and not because the moral theories we embrace are true. Given our training, temperament, and environment, we would make the moral judgments we do and advance the moral theories we do, regardless of the moral facts (and regardless of whether there are any).

To clarify the challenge facing moral theory, consider two situations (I take these from Gilbert Harman, who has done the most to advance the charge of explanatory impotence.)¹⁰ In one, a person goes around a corner, sees a gang of hoodlums setting a live cat on fire, and exclaims, "There's a bad action!" In the other, a person peers into a cloud chamber, sees a trail, and exclaims, "There's a proton!" In both cases, part of the explanation of why the report was made will appeal to the movements of physical objects and the effects these movements have initially on light and eventually on the observers' retinas. A more complete explanation would also have to make reference to the observers' psychological states as well as the background theories each accepted. Certainly the scientist would not make the report she did if she were asleep or, even if awake and attentive, if she did not accept a theory according to which vapor trails in cloud chambers evidence the presence of protons. Had she thought witches left such trails, she might have reported a witch in the chamber instead of a proton. Similarly, the moral judge would not have made the report he did if he were asleep or, even if awake and attentive, if he did not accept a view according to which burning live animals is wrong.¹¹ Had he thought cats the embodiments of evil, he might have reported the action as right instead of wrong.

Whatever explanations we give of the reports, one thing is striking: protons will form part of our best explanation of why the proton report was made; in contrast, moral properties seem not to form part of our best explanation of why the moral report was made. We will often explain the scientist's belief that a proton was present by appeal to the fact that one was. But, the argument goes, we will not explain the moral judge's belief that burning the cat is wrong by appeal to the wrongness of the act.

Harman elaborates on the problem with ethics by noting that "facts about protons can affect what you observe, since a proton passing through a cloud chamber can cause a vapor trail that reflects light to your eye in a way that, given

10. Harman, *The Nature of Morality*, pp. 6-7 (pp. 121-22 in this volume).

11. Of course, neither the scientist nor the moral judge need have a well-worked-out theory in order to make observations. The ability to form opinions (about protons, witches, or morals) as a direct result of perceptual experience is more a matter of effective training than of the conscious application of theory to experience.

your scientific training and psychological set, leads you to judge that what you see is a proton. But 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."¹² This emphasis on affecting (or failing to affect) an observer's perceptual apparatus suggests (mistakenly, I will argue) that the following *Causal Criterion* underlies the explanatory impotence attack on moral theory:

The only entities and properties we are justified in believing in are those that we are justified in believing have a causal impact on our perceptual apparatus.

Unless moral properties are causally efficacious and so figure as causes in the explanation of our making the observations that we do, moral theory will fail to meet this criterion's test. Even though the argument from explanatory impotence turns on a different (and more plausible) principle, the Causal Criterion deserves attention because of its intimate ties to the causal theories of knowledge and reference.

4. The Causal Criterion, knowledge, and reference

Any reasonable view of moral theory, and of the language(s) we use to formulate the theory, must (if moral theory is legitimate) be compatible with some account of how we come to *know* about moral properties and how the terms of the language come to *refer* to these properties. Assuming that the causal theories of knowledge and reference are substantially correct, the Causal Criterion is attractive simply because we could neither know about nor even refer to any class of properties that failed the Causal Criterion's test.¹³ Thus, by requiring that we believe in only those entities and properties with which we believe we can causally interact, the Causal Criterion encapsulates the demands of the causal theories of knowledge and reference.

According to the causal theory of knowledge, we can get evidence only about that to which we bear some appropriate causal connection.¹⁴ All our knowledge arises from the causal interaction of the objects of this knowledge with our bodies; anything outside all causal chains will be epistemically inaccessible.¹⁵ So, if

12. Harman, *The Nature of Morality*, p. 8 (p. 122 this volume).

13. Actually, the Causal Criterion will be attractive regardless of whether the causal theories are substantially correct, as long as we assume causal contact is a necessary condition for knowledge.

14. Goldman characterizes the appropriate connection in terms of there being a "reliable belief-forming operation." Alvin Goldman, "What is Justified Belief?" in *Justification and Knowledge*, ed. George Pappas (Dordrecht, 1979), pp. 1-23.

15. Mark Steiner defends mathematical entities from this objection in *Mathematical Knowledge* (Ithaca, N.Y., 1975) p. 10. See also Penelope Maddy, "Perception and Mathematical Intuition," *Philosophical Review* 89 (1980), 163-96; Paul Benacerraf, "Mathematical Truth," *Journal of Philosophy* 70 (1973), 661-79; Crispin Wright, *Frege's Conception of Numbers as Objects* (Aberdeen, 1983); and Philip Kitcher, *The Nature of Mathematical Knowledge* (Oxford, 1983).

moral properties are causally isolated—if they fail to meet the Causal Criterion—they will be unknowable. More important, if moral properties make absolutely no difference to what we experience, we can never even form reasonable beliefs about what they are like.

The causal theory of reference makes moral theory look all the more hopeless because it suggests that we cannot even successfully refer to, let alone know about, moral properties (if they fail the Causal Criterion). As the causal theory of reference would have it, words in our language refer because they stand at the end of a causal chain linking the speaker's use of the word to the thing to which the word refers. No appropriate chain can be established between speakers of a language (in this case a language containing moral terms) and causally isolated properties. Such properties will lie outside all causal chains, and so outside those causal chains which establish reference.¹⁶

Moral theory's trouble seems to be that the properties it ascribes to actions, people, and states of affairs, reflect no light, have no texture, give off no odor, have no taste, and make no sound. In fact, they do not causally affect our experience in any way. Were they absent, our experiences would be unchanged. Since we cannot interact with moral properties, there is no way for us to establish a causal chain between ourselves, our use of moral language, and moral properties. Consequently, our moral terms fail to refer.

So put, this criticism of moral theory is much too quick. Even the causal theory of reference allows success in establishing a referential tie between word and world by description as well as by ostension.¹⁷ It is true that ostension works in establishing reference only if the properties (or entities) referred to are causally present (since we can succeed in our ostensions only by locating something in space and time).¹⁸ However, we may still use descriptions to establish a referential link even to that from which we are causally isolated—as long as the appropriate terms of the description succeed in referring. If moral properties fail the Causal Criterion, we will not be able to refer to them by ostension, but we will nonetheless be able to refer to them as long as we can describe the moral properties in nonmoral terms. (Of course, if the description's terms were moral, they would

be no help in grounding the requisite referential link.)¹⁹ As a result, the causal theory of reference will serve to undermine ethics only if we cannot refer to moral properties by using nonmoral descriptions, and then only if moral properties are in fact causally isolated.²⁰ Any argument against moral facts using the causal theory of reference, then, must rely on some independent argument that shows that moral facts (if such there be) are both indescribable in nonmoral terms and causally isolated.

Plainly, our ability to refer to moral properties will be small consolation unless we can also secure evidence about the properties to which we refer. Successful reference may prove epistemically useless. So, even if we can succeed in referring to moral properties, the problems raised by the causal theory of knowledge remain.

Not surprisingly, these problems, too, are less straightforward than suggested so far. To tell against moral theory, the causal theory of knowledge (like the causal theory of reference) must be supplemented by an argument showing that moral properties are causally isolated (or that they do not exist).

Against theories concerned with abstract entities (like mathematics and Plato's Theory of Forms), the Causal Criterion, and the causal theories of knowledge and reference, apparently meet no resistance. The theories under attack grant right off that the entities in question are causally isolated (because outside space and time). That abstract entities fail the Causal Criterion appears to be a forgone conclusion.²¹

Against moral theories, in contrast, the charge of causal isolation meets with resistance. Unlike abstract entities, moral properties are traditionally thought of as firmly ensconced in the causal nexus: a bad character has notorious effects (at least when backed by power), and fair social institutions evidently affect the happiness of those in society. The ontology of moral theory will not be an unwitting accomplice in the causalist critique of ethics. Of course, moral theory's resistance does not establish that moral properties actually do satisfy the Causal Criterion (and so the causal theories of knowledge and reference); rather, the resistance

16. All this is compatible, of course, with there being a causal story of our use of moral language. Since we do live in a community of moral-language users, we are taught how to use moral words and we stand at the end of a causal (in this case, educational) chain that explains our use of moral terms. Despite there being such a causal story, if moral properties are causally isolated, our language will lack the grounding that would allow it to refer; the linguistic chain would lack an anchor.

17. See Saul Kripke, *Naming and Necessity* (Cambridge, Mass., 1980).

18. Incidentally, one may succeed in referring to an ordinary object, one located in space and time, even if the object does not actually have any causal impact on the referrer. Eyes closed, I may enter a room, point to my left and declare, "I'll bet that chair is brown." I will have referred to the chair (assuming one is there), and made a bet about its color, despite my neither bumping into it, seeing it, nor in any other way being causally affected by it. At most, successful ostension requires causal presence and not causal impact.

19. In this paper I shall leave unchallenged the (eminently challengable) assumption that there is some way to isolate moral from nonmoral language.

20. Note that we need not have naturalistic definitions in order to succeed in referring by description. Since the Causal Criterion rules out all causally inert properties, while the causal theory of reference allows reference to causally inert properties as long as they are describable, the strictures of the Causal Criterion actually go beyond those of the causal theory of reference.

21. Actually, even when applied to mathematical entities the game is not quite so easily won. For instance, Kurt Godel maintained that we have a mathematical intuition akin to visual perception that establishes a causal link, of sorts, between numbers and knowers (in "What is Cantor's Continuum Problem?" *American Mathematical Monthly* 54 (1947), 515–25). Penelope Maddy (in "Perception and Mathematical Intuition") has defended this possibility by appeal to recent theories of perception. In the process, she has argued that abstract mathematical entities (e.g., sets) will, contrary to initial appearances, satisfy the Causal Criterion.

imposes a barrier over which the causal critique of moral theory must climb. Some argument must be given for thinking moral properties fail to meet the Causal Criterion.

5. The Explanatory Criterion

Regardless of whether moral properties satisfy the Causal Criterion, there are good reasons for thinking the criterion itself too strong. To hold tight to the Causal Criterion (and the causal theories of knowledge and reference that support it) is to let go of some of our most impressive epistemological accomplishments; the claims of mathematics, as well as both the empirical generalizations and the laws of the physical sciences, all fail the criterion's test.

We never causally interact with numbers, for instance.²² So, if causal contact were really a prerequisite to knowledge (and reference), mathematical knowledge and discourse would be an impossibility. For similar reasons, empirical generalizations (like "all emeralds are green"), as well as natural laws (like the first law of thermodynamics), would fall victim to the Causal Criterion. Although these generalizations and laws may help *explain* why we experience what we do as we do, they *cause* none of our experiences. That all emeralds are green does not cause a particular emerald to be green, nor does it cause us to see emeralds as green.²³ These casualties of the Causal Criterion make it clear that we need to replace the Causal Criterion even if we wish to salvage its emphasis on the link between knowledge (or at least justified belief) and experience.²⁴

In forging a new criterion, we should concentrate on the reasons that might be given for thinking it reasonable to believe (as I assume it is) in the truth of many mathematical claims, empirical generalizations, and laws of physics. According to one standard line in the philosophy of science (one embraced by Harman, J. L. Mackie, Simon Blackburn, and many other critics of moral theory), the key to the legitimacy of these scientific and mathematical claims is the role they play in the explanations of our experiences. "An observation," Harman argues, "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."²⁵ Empirical generalizations and physical laws will likewise find

22. As Harman notes, "We do not and cannot perceive numbers . . . since we cannot be in causal contact with them." *The Nature of Morality*, p. 10 (p. 124 this volume).

23. As Harman argues in *Thought* (Princeton, N.J., 1973), p. 127. Adolf Grunbaum makes the same point in arguing that one scientific law may explain another, even though the first law does not *cause* the second. See "Science and Ideology," *Scientific Monthly* (July 1954), 13–19.

24. Harman recognizes the shortcomings of the Causal Criterion, and it is in his pointing them out that it becomes clear that he does not accept the criterion. See *Thought*, esp. pp. 126–32.

25. Harman, *The Nature of Morality*, p. 10 (p. 124 this volume), my emphasis. There is room, of course, to agree that if the truth of mathematical claims contribute to our best explanations these claims should be believed, while also holding that their truth does not so contribute. This is Hartry Field's position in *Science without Numbers* (Princeton, N.J., 1980).

their justification by appeal to their role in the best explanations of our experience; "scientific principles can be justified ultimately by their role in explaining observations."²⁶ The legitimacy of a theory seems to ride on its explanatory role and not on the causal impact of its ontology. From these points we can extract the *Explanatory Criterion*.

The only hypotheses we are justified in believing are those that figure in the best explanations we have of our making the observations that we do.

Significantly, the Explanatory Criterion retains, even reinforces, the empiricist's demand that epistemology be tied to experience; not only does justification turn on experiential testability, it now requires an *explanatory* link between the truth of our beliefs and our experiences as well. Accordingly, an acceptable theory must do more than have observational consequences; it must also contribute to our explanations of why we make the observations we do.

Two versions of the Explanatory Criterion should be distinguished: the first sets necessary and sufficient conditions for reasonable belief, the second sets only necessary conditions. In its stronger version, the criterion would say:

A hypothesis should be believed *if and only if* the hypothesis plays a role in the best explanation we have of our making the observations that we do.

In its weaker version the criterion would say instead:

A hypothesis should be believed *only if* the hypothesis plays a role in the best explanation we have of our making the observations that we do.

Or, in its contrapositive (and more intuitively attractive) form:

A hypothesis should not be believed if the hypothesis plays no role in the best explanation we have of our making the observations that we do.

Accepting the stronger version of the criterion involves endorsing what has come to be called 'inference to the best explanation'. In this guise, the criterion licenses inferring the truth of a hypothesis from its playing a role in our best explanations of our experiences. To be even remotely plausible, of course, some bottom limit must be set on the quality of the explanations that would be allowed to countenance inferences to the truth of the hypotheses invoked. Despite their being the best we have, our explanations can be so bad that we may be quite sure they are wrong. It would be a mistake to infer the truth of a hypothesis from its being part of our best—but obviously flawed—explanation. Even with a quality

26. Harman, *The Nature of Morality*, p. 9 (p. 123 this volume).

constraint, though, the strong version of the Explanatory Criterion is hopelessly liberal because we have such good grounds for thinking that the best explanations we can come up with, at any given time, are not right.²⁷ In light of these difficulties, I shall concentrate on the weaker version of the criterion; it raises all the same difficulties for moral theory without endorsing *carte-blanche* inferences to the best explanation.

The problem with moral theory is that moral principles and moral properties appear not to play a role in explaining our making the observations we do. All the explanatory work seems to be done by psychology, physiology, and physics.²⁸ A scientist's observing a proton in a cloud chamber is evidence for her theory because the theory explains the proton's presence and the scientist's observation better than competing theories can. The observation of a proton provides observational evidence for a theory because the truth of that observation is part of the best explanation we have of why the observation was made. A moral 'observation' does not appear to be, in the same sense, observational evidence for or against any moral theory, since (as Harman puts it) the truth or falsity of the "moral observation seems to be completely irrelevant to any reasonable explanation of why that observation was made."²⁹

Underlying the Explanatory Criterion is the conviction that confirmation mirrors explanation: theories are confirmed by what they explain. Added to this view of confirmation is the stipulation, motivated by an empiricist epistemology, that we should assume to exist only what we need to explain *our experiences*.

Put generally, some fact confirms whatever principles and hypotheses are part of the best explanation of the fact. So, the fact that some observation was made will confirm whatever is part of the best explanation of its having been made. But the making of the observation will not provide *observational* evidence for a theory unless the observation itself is accurate. And we will have grounds for thinking an observation accurate, on this view, only when its being accurate forms a part of our best explanation of the observation having been made. Thus, embedded within this overarching view of confirmation is a more specific account of observational confirmation. An observation will provide confirming observational evidence for a theory, according to this account, only to the extent that it is reasonable to explain the making of the observation by invoking the theory while also treating the observation as true.³⁰

27. The one reasonable application of this strong version of the Explanatory Criterion would tie its use to the explanations reached at the ideal limit of inquiry—an explanation we will almost surely never get. At this *Piercean* limit, there is sense to saying we can infer the truth of the hypotheses invoked by the (very) best explanation; for only if there is some such link with epistemology will truth be accessible. So used, though, the principle will never actually countenance any of our inferences.

28. According to Harman, "Moral hypotheses never help explain why we observe anything. So we have no evidence for our moral opinions." *The Nature of Morality*, p. 13; see also p. 8, (p. 123 this volume).

29. Harman, *The Nature of Morality*, p. 7 (p. 122 this volume).

30. Theories find observational confirmation only from accurate observations, and some particular theory will find observational confirmation from an accurate observation only if the theory also plays a

role in explaining the making of that observation. Nonetheless, the making of some observation *O* will confirm a theory *T* even if the observation is false as long as *T* explains why the false observation was made. Even supposing an observation inaccurate, then, the making of the observation will be confirming evidence (but not *observational* evidence) for whatever theories contribute to the best explanation of the making of that (false) observation. When the report is false, however, it will be the making of the observation, and not its content, that serves to confirm our explanatory theories; and it will be the accurate observation that the false observation was made (and not the false observation itself) that provides observational support for our explanatory theories.

For this reason, moral facts, and moral theory, will be vindicated (in the eyes of the Explanatory Criterion) only if they figure in our best explanations of at least some of the accurate observations we make. Unfortunately, as Harman argues, "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."³¹ Of course, moral facts would be acceptable, according to the Explanatory Criterion, as long as they were needed to explain the making of some observation or other (regardless of whether it is a moral observation). Just as mathematics is justified by its role in explaining physical (and not mathematical) observations, moral theory might similarly be justified by its role in explaining some nonmoral observations. But the problem with moral theory is that moral facts seem not to help explain the making of *any of our observations*.

Importantly, the problem is not that moral facts explain nothing at all (they may explain other moral facts); the problem is that regardless of whether they explain something, they do not hook up properly with our abilities to detect facts. Even if there are moral facts and even if some of these facts would help to explain others, none will be epistemically accessible unless some help to explain our making some of the observations we do. No matter how perfect the fit between the content of our moral judgments and a moral theory, no matter how stable and satisfying a reflective equilibrium can be established between them, the theory will not gain observational confirmation unless it enters into the best explanation of why some of our observations are made. We will be justified in accepting a moral theory on the basis of our observations only if we have reason to believe our observations are responsive to the moral facts. And we will have reason to believe this only if moral facts enter into the best explanations of why we make the observations we do. To be legitimized, then, moral facts must explain certain nonmoral facts; specifically, moral facts must explain our making observations.³²

In order to highlight the problem faced by moral theory, it is a good idea to go back to the (dis)analogy between the scientist's making the observation "there's a proton" and the moral judge's making the observation "there's a bad action."

31. Harman, *The Nature of Morality*, p. 6 (p. 121 this volume).

32. Implicit in the Explanatory Criterion, then, is the conviction that legitimate theories must be linked to an acceptable theory of observation. As Putnam argues, "It is an important and extremely useful constraint on our theory itself that our developing theory of the world taken as a whole should include an account of the very activity and process by which we are able to know that a theory is correct." *Reason, Truth and History* (Cambridge, 1981), p. 132.

Our best explanation of why the scientist made the observation she did will make reference to her psychology, her scientific theory, the fact that a vapor trail appeared in the cloud chamber, and *the fact that a proton left the trail*. Our best explanation of why the moral judge made the observation he did will likewise make reference to his psychology, his moral theory, the fact that a cat was set on fire, and even (when the explanation is more fully elaborated) to the fact that the cat and the kids were partially composed of protons. Yet our explanation will not make reference to the (purported) fact that burning the cat was wrong: "It seems to be completely irrelevant to our explanation whether [the judge's moral] judgment is true or false."³³ That burning the cat was wrong, if it was wrong, appears completely irrelevant to our explanation of the judge thinking it wrong.³⁴

6. Explanatory relevance and explanatory potency

The explanatory critique of moral theory seems to rest on the claim that moral facts are *irrelevant* to explanations of our observations. So, to flesh out the problem, we need a test for explanatory irrelevance. Nicholas Sturgeon proposes the following: "If a particular assumption is completely irrelevant to the explanation of a certain fact, then the fact would have obtained, and we could have explained it just as well, even if the assumption had been false."³⁵ With this in mind, Sturgeon argues that, for those who are not already moral skeptics, moral facts will prove to be explanatorily relevant.

Sturgeon's argument runs as follows. To decide whether the truth of some moral belief is explanatorily relevant to the making of an observation, we must consider a situation in which the belief is false, but which is otherwise as much like the actual situation as possible. Then we must determine whether the observation would still have been made under the new conditions. If so, if the observation would have been made in any case, then the truth of the moral belief is explanatorily irrelevant (the observation would have been made even if it had been false); otherwise its truth is relevant. If a supervenience account of moral properties is right (so that what makes a moral judgment true or false is some combination of physical facts), then for some true judgment to have been false, or some false

33. Harman, *The Nature of Morality*, p. 7 (p. 122 this volume).

34. Of course, that the burning of the cat was wrong might be part of the moral judge's (as opposed to our) best explanation of why he made the observation he did, just as for some people the best explanations they had of their observations made reference to phlogiston.

35. Sturgeon, "Moral Explanations," p. 245 of this volume. As Sturgeon recognizes, the test has its limits. It will not be a reliable indicator of explanatory relevance when dealing with two effects of the same cause; neither effect would have occurred without the other (because each would have occurred only if the cause of the other had), even though neither explains the other. And it will not be reliable when using 'that-would-have-had-to-be-because' counterfactuals; it may be that if Reagan had lost the presidential election that would have had to be because he failed to get enough votes, even though his being elected is not explanatorily relevant to his getting enough votes. For other limitations, see Warren Quinn, "Truth and Explanation in Ethics," *Ethics* 96 (1986), 524-44.

judgment true, the situation would have had to have been different in some physical respect.³⁶ Consider the hoodlums' cat-burning. According to Sturgeon, "If what they are actually doing is wrong, and if moral properties are, as many writers have held, supervenient on natural ones, then in order to imagine them not doing something wrong we are going to have to suppose their action different from the actual one in some of its natural features as well."³⁷ Whether we would still judge the action wrong given these changes is a contingent matter that turns on how closely tied our moral judgments are to the morally relevant physical features of the situation. In the case of a curmudgeon, who thinks badly of kids as a matter of principle (averring that "kids are always up to no good"), changing the moral (and so the nonmoral) features of the situation will probably not change his moral judgment. For him, the truth of his judgment is irrelevant to the explanation of his having made it. Fortunately, though, many people do not share this bias and are therefore more attuned to the evidence. Such people would have different opinions had the hoodlums found their entertainment in more acceptable ways (say, by petting rather than incinerating the cat). For those to whom the difference would make a difference, part of the explanation of their judgment would be that burning the cat is wrong.³⁸

Notice that the same contingency attaches to the scientist's sighting of a proton. Had the proton not been there, whether the scientist would have thought it was depends on how closely tied her scientific judgments are to the relevant features of the situation. If she is a poor researcher, she might well have reported the proton's presence had there really been only a passing reflection. Again fortunately, many scientists are well attuned to the difference between proton trails and passing reflections. At least these scientists would have made different reports concerning the presence of a proton had the proton been absent. For those to whom the difference would make a difference, part of the explanation of their judgment will be that a proton passed.³⁹

36. At this stage, the relevant feature of a supervenience account of moral properties is that if the moral properties of something are changed, then so must be some nonmoral properties; there is no holding the nonmoral properties fixed while altering the moral properties, as there can be no moral difference without a nonmoral difference.

37. Sturgeon, "Moral Explanations," p. 247 of this volume. Which nonmoral facts will have to be altered to change the moral facts is, obviously, open to dispute.

38. As Sturgeon emphasizes, "Hitler's moral depravity—the fact of his really having been morally depraved—forms part of a reasonable explanation of why we believe he was depraved" ("Moral Explanations," p. 234 of this volume.). Had he not been depraved we very likely would not have thought him depraved; for he would not have done all the despicable things he did, and it is his having done such things that leads us to our condemnation.

39. Our background theories will clearly play a central role in determining explanatory relevance. In the cat-burning case, we will rely on our moral theory in deciding what nonmoral features of the world would have been different had the hoodlums' activities been unobjectionable. Similarly, in the passing proton case, we will rely on our scientific theory in deciding what physical features of the world would have been different had a proton not passed. Such a reliance on background theories will certainly offend a thoroughgoing skeptic. But attacks on moral theory are interesting only if some of our views survive the skeptic's arguments, so I shall assume we may legitimately rely at least sometimes on our background theories.

Sturgeon holds that what separates his own position from Harman's is a differential willingness to rely on a background moral theory in evaluating the question: Would the moral observation have been made even if the observation had been false? Sturgeon assumes the observation to be true (relying as he does on his background moral theory), and believes that for it to be false, some nonmoral features of the situation must be assumed different (because the moral properties supervene upon natural properties). When these nonmoral features are changed, he points out, the moral judgments (along with the explanations of why they were made) will often change as well. Harman, though, does without the background moral theory, so he has no reason to think that if the observation were false, anything else about the situation (including the observer's beliefs) would have been different. Consequently, he holds that the observation would have been made regardless of its truth.

As a result, Sturgeon concludes that Harman's argument is not an independent defense of skepticism concerning moral facts, for its conclusion apparently rests on the assumption that our moral judgments are false (or, more accurately, on the assumption that moral theory cannot be relied on in estimating how the world would have been if it had been morally different).⁴⁰

Unfortunately for moral theory, Sturgeon's argument fails to meet the real challenge. The force of the explanatory attack on moral theory may be reinstated by shifting attention from explanatory *irrelevance* to explanatory *impotence*, where

a particular assumption is explanatorily impotent with respect to a certain fact if the fact would have obtained and we could have explained it just as well *even if the assumption had not been invoked in the explanation* (as opposed to: "even if the assumption had been false").

By charging explanatory impotence, rather than explanatory irrelevance, the explanatory challenge to moral theory survives the admission that we hold our moral theories dear. It also survives the supervenience account's provision of a necessary link between moral and nonmoral properties. For the question becomes: Do we honestly think appealing to moral facts in our explanations of moral judgments strengthens our explanations one bit? Behind this question is the worry that we have been profligate with our theory building (or, perhaps, that we've been unnecessarily and unwholesomely nostalgic about old, and now outdated, theories). The concern is that acknowledging moral facts adds nothing to our ability to explain our experiences. Everything we might reasonably want to explain can, it seems, be explained equally well without appeal to moral facts.⁴¹

40. Sturgeon, "Moral Explanations," pp. 249–51 of this volume. Independently, John McDowell has made essentially the same point concerning skeptical attacks on moral explanations in his "Values and Secondary Qualities," in *Morality and Objectivity*, ed. Ted Honderich (London, 1985), pp. 110–29, reprinted in this volume.

41. Which, of course, is not to say that we can explain everything that we might reasonably want to explain.

If these worries are well founded, if moral facts are explanatory 'fifth wheels', and if we accept the Explanatory Criterion for justified belief, then moral facts will become merely unjustified theoretical baggage weighing down our ontology without offering compensation. In the face of this threat, pointing out that we happen to rely on moral facts in explaining people's behavior is not sufficient to justify believing in (even supervenient) moral facts; actual reliance does not establish justified belief.

To see the force of the explanatory challenge, imagine that a belief in witches becomes popular among your friends. Imagine, too, that your friends teach each other, and you, how to give 'witch explanations'. You 'learn' that the reason some bound women float when tossed into ponds, and others do not, is that the floaters are witches, the others not; that the mysterious deaths of newborns should be attributed to the jealous intervention of witches; and so on. No doubt, with enough practice you could become skillful at generating your own witch explanations; so skillful, perhaps, that in your unreflective moments you would find yourself offering such explanations. In order to assuage your philosophical conscience, you might entertain a sort of supervenience account of being a witch. Then you might comfortably maintain that being a witch is explanatorily relevant (in Sturgeon's sense) to your observations. All this might come to pass, and still you would be justified in thinking there are not really any witches—as long as you could explain the floatings, the deaths, and whatever else, just as well without appealing to the existence of witches. Presumably, the availability of such alternative explanations is just the reason we should not now believe in witches.

Certainly, things could turn out otherwise; we might find that witch explanations are actually the best available. We might discover that postulating witches is the only reasonable way to account for all sorts of otherwise inexplicable phenomena. Should that happen, our conversion to a belief in witches would, of course, be quite justified.

The question is, to which witch scenario are our moral explanations more analogous? Are our appeals to moral properties just intellectually sloppy concessions to effective socialization, or do we really strengthen our explanatory abilities by supposing that there are moral properties? This is a substantial challenge, and one not adequately answered by the observation that we often rely on moral properties to explain behavior.

Two points about the explanatory challenge deserve emphasis. First, the challenge recognizes conditions under which a belief in moral facts, or witches, would be legitimate. Specifically, the Explanatory Criterion would take these beliefs to be justified if their truth figured in the best explanation of why we have the experiences we do. Second, the challenge will not be met simply by pointing out that witches, or moral facts, do figure in some of our best explanations of the world. For unless these explanations of the world can be properly linked to *our* experiences of the world, there will be no way for us to justify accepting some of the explanations rather than others. The truth of one or another will make no difference to our experience and will be epistemically inaccessible.

7. Supervenience and lenience

The problem with concentrating on explanatory relevance, rather than on explanatory potency, is that it makes a defense of moral properties (and belief in witches) too easy. It permits as justified the introduction of any properties whatsoever, so long as they are construed as supervenient upon admittedly explanatory properties.

The Explanatory Criterion, when interpreted as demanding explanatory potency (rather than mere relevance), promises a stricter standard, one which might separate those properties we are justified in believing instantiated from mere pretenders. Yet the criterion must be interpreted in a way that acknowledges, as justified, belief in two kinds of properties: those which can be reductively identified with explanatorily potent properties and those we have independent reason to think supervene upon, without being strictly reducible to, explanatorily potent properties. These properties demand special attention because, at least initially, they appear explanatorily expendable—despite our belief in them being justified.

For instance, though all the explanatory work of water may be better accomplished by H₂O; all the explanatory work of color, by the wavelengths of light; and all that of psychological states, by neurophysiological states of the brain, we are nonetheless justified in believing that the oceans are filled with water, that roses are red, and that people feel pain and have beliefs. Any criterion of justified belief that would rule these beliefs out as unjustified is simply too stringent. The difficulty (for those attacking moral theory) is to accommodate these legitimate beliefs without so weakening the Explanatory Criterion as to reintroduce excessive leniency.

Of course, some beliefs may plausibly find justification by relying on reductions: because water is H₂O, the justification of our belief in H₂O (by appeal to its explanatory potency) serves equally well as a justification of our belief in water. In cases where identification reductions are available, explanatory potency might well be transitive.⁴²

Where identification reductions are not available, however, things become trickier; and it is here that the Explanatory Criterion runs into problems. It seems straightforwardly true that roses are red, for example, but our best explanations of red-rose reports might well make reference to certain characteristics of roses, facts about light, and facts about the psychological and perceptual apparatus of perceivers, but not to the *redness* of the roses (and not to any particular feature of the roses that can be reductively identified with redness). Despite this, the availability of such explanations expands our understanding of colors; it does not show there

42. Yet there is at least some question as to whether the reductive hypothesis itself satisfies the Explanatory Criterion. What, after all, do we explain with the help of the reductive hypothesis that we could not explain just as well by assuming the reduced claims fail to refer? See Quinn, "Truth and Explanation in Ethics," for more on the tension between the Explanatory Criterion and reductive hypotheses.

are no colors. Similar points hold not just for those properties traditionally characterized as secondary qualities, but for *all* nonreducible properties.⁴³

Recognizing this, Harman attempts to make room for nonreducible properties. In discussing colors, he maintains that they satisfy the demand for explanatory potency because "we will sometimes refer to the actual colors of objects in explaining color perceptions if only for the sake of simplicity. . . . We will continue to believe that objects have colors because we will continue to refer to the actual colors of objects in the explanations that we will in practice give."⁴⁴ Thus, pragmatic tenacity is supposed by Harman to be enough to establish explanatory potency; now the criterion will allow, as explanatorily potent, those properties and entities to which we appeal in our best explanations, plus those that are precisely reducible to properties or entities appealed to in our best explanations, plus those that are pragmatically tenacious.

Relying on such a lenient interpretation of the Explanatory Criterion, Harman is able to treat moral facts as threatened only by resorting to what is patently false: he is forced to argue that moral facts are not "useful even in practice in our explanations of observations."⁴⁵ If nothing else, however, moral facts are useful, at least in practice, when explaining our observations. Many very useful, and frequently offered, explanations of events in the world (and so our observations of those events) make reference to moral facts. Mother Teresa's goodness won her a Nobel Prize; Solidarity is popular because of Poland's oppressive political institutions; millions died in Russia as a result of Stalin's inhumanity; people are starving unnecessarily because of the selfishness of others; unrest in Soweto is a response to the injustice of apartheid. Even if such explanations could eventually be replaced by others that appeal only to psychological, social, and physical factors, without mention of moral facts, the moral explanations would still be useful in just the way talk of colors remains useful even in light of theories of light. If mere pragmatic tenacity is enough to legitimize color properties, then it ought to be enough to legitimize moral properties.

If the Explanatory Criterion is to challenge the legitimacy of moral theory, it must require more than pragmatic tenacity for justification. Yet, almost certainly, any stronger requirement that remains plausible will countenance moral properties. For the Explanatory Criterion will be plausible only if it allows belief in those properties needed both to identify and to explain the natural regularities that are otherwise explicable only in a piecemeal fashion as singular events (and not as

43. There are some significant differences between moral properties and secondary qualities, not least of which is that we can learn to ascribe secondary qualities without having any idea as to what properties they supervene upon, whereas learning to ascribe moral properties requires an awareness of the properties upon which they supervene. This difference will stand in the way of treating moral properties as strictly analogous to secondary properties. But I think it won't underwrite any plausible version of the explanatory criterion that is still strong enough to rule out moral properties. See Quinn, "Truth and Explanation in Ethics."

44. Harman, *The Nature of Morality*, p. 23.

45. *Ibid.*

instances of regularities). Consider Hilary Putnam's example of the peg (square in cross section) that will pass through the square hole in a board, but not through the round hole.⁴⁶ To explain a single instance of the peg's going through one hole, we might offer a microstructural description of the peg and the board in terms of the distribution of atoms, and then appeal to particle mechanics. But even if we could eventually work out such an explanation, it would suffer from a serious drawback; it would only explain why the particular peg went through a particular hole at a particular time. The explanation will be of no help when we are faced with another board and peg, or even with the same board and peg a moment later (when the distribution of atoms has changed). The explanation will not extend to new cases. And the properties appealed to in giving an explanation at the level of ultimate constituents will be useless in trying to identify and explain the general fact that pegs of a certain size and shape (whether made of wood, or plastic, or steel) will fail to go through holes of a certain size and shape (whether the holes are in a piece of wood, or plastic, or steel). This general fact will be identifiable and explicable only if we appeal to certain macrostructural features of pegs and holes.

In the same way, certain regularities—for instance, honesty's engendering trust or justice's commanding allegiance, or kindness's encouraging friendship—are real regularities that are unidentifiable and inexplicable except by appeal to moral properties. Indeed, many moral virtues (such as honesty, justice, kindness) and vices (such as greed, lechery, sadism) figure in this way in our best explanations of many natural regularities. Moral explanations allow us to isolate what it is about a person or an action or an institution that leads to its having the effects it does. And these explanations rely on moral concepts that identify characteristics common to people, actions, and institutions that are uncapturable with finer-grained or differently structured categories.

Of course, even if moral properties do have a role in our best explanations of natural regularities, we might still wonder whether these properties are anything more, anything over and above, psychological properties and dispositions of individuals, and we might wonder to what extent these 'vices' and 'virtues' have any normative authority. For all that has been said so far, we might have no good reason to think the 'virtues' worthy of cultivation and the 'vices' worthy of condemnation. So even if moral properties ultimately satisfy the demands of the Explanatory Criterion (once we get a reasonable interpretation of its requirements), we will at most have established that certain people, actions, and institutions have those properties we label 'moral'. We will not yet have shown that there is any reason to care about the properties or that some of the properties are better than others.⁴⁷

46. "Philosophy and Our Mental Life," in *Mind, Language, and Reality* (Cambridge, 1975), pp. 291–303.

47. Just as reductions of the mental to the physical fail to capture intentionality, reductions of the moral to the mental fail to capture justifiability.

As long as we concentrate on which properties satisfy the Explanatory Criterion and which do not, the distinctive value of moral properties will remain elusive. The structure of a compelling defense of their value will emerge only after we turn our attention to the presuppositions of the Explanatory Criterion itself.

8. The evaluation of explanations

As the Explanatory Criterion would have it, which hypotheses are justified and which are not will depend crucially on our standards of explanation, since it is by figuring in the *best* available explanations that a hypothesis finds justification. No argument that depends on the Explanatory Criterion will get off the ground unless some explanations are better than others. This poses a dilemma for those who suppose that the Explanatory Criterion will support the wholesale rejection of evaluative facts. Either there is a fact of the matter about which explanations are best, or there is not. If there is, then there are at least some evaluative facts (as to which explanations are better than others); if not, then the criterion will never find an application and so will support no argument against moral theory.

If we say that astronomers, and not astrologers, make the appropriate inferences from what is seen of the constellations or that evolutionary theorists, and not creationists, have the best explanation of the origin of our species, we will be making value judgments. In trying to legitimize these judgments by appealing to our standards of explanation, a reliance on values becomes inescapable (even if the values appealed to are not themselves mentioned in our explanations).

The obvious response to this point is to embrace some account of explanatory quality in terms, say, of simplicity, generality, elegance, predictive power, and so on. One explanation is better than another, we could then maintain, in virtue of the way it combines these properties.⁴⁸

When offering the list of properties that are taken to be the measures of explanatory quality, however, it is important to avoid the mistake of thinking the list wipes values out of the picture. It is important to avoid thinking of the list as eliminating explanatory quality in favor of some evaluatively neutral properties. If one explanation is better than another in virtue of being simpler, more general, more elegant, and so on, then simplicity, generality, and elegance cannot themselves be evaluatively neutral. Were these properties evaluatively neutral, they could not account for one explanation being better than another.

If we are to use the Explanatory Criterion, we must hold that some explanations really are better than others, and not just that they have some evaluatively neutral properties that others do not. Any attempt to wash evaluative claims out as

48. For discussions of the (often conflicting) criteria for explanatory value, see Paul Thagard "The Best Explanation: Criteria for Theory Choice," *Journal of Philosophy* 75 (1978), 76–92; and William Lycan, "Epistemic Value," *Synthese* 64 (1985), pp. 137–64.

psychological or sociological reports, for instance, will fail—we will not be saying what we want, that one explanation is *better* than another, but only (for example) that we happen to like one explanation more or that our society approves of one more. What the Explanatory Criterion presupposes is that there are evaluative facts, at least concerning which explanations are better than others, regardless of whether these facts explain any of our observations.

Even assuming that the Explanatory Criterion presupposes the existence of some *evaluative* facts, the question remains whether we have any good reason for thinking there are moral facts as well. We might be convinced that some explanations really are better than others but still deny that some actions or characters or institutions are better than others. Significantly, though, once it has been granted that some explanations are better than others, many obstacles to a defense of moral values disappear. In fact, all general objections to the existence of value must be rejected as too strong. Moreover, whatever ontological niche and epistemological credentials we find for explanatory values will presumably serve equally well for moral values.⁴⁹

Without actually making the argument, I shall briefly sketch one of the ways one might defend the view that there are moral values. The aim of such an argument would be to show that some actions, characters, or institutions are better than others—just as some explanations are better than others.

This defense of moral values rests on recognizing and stressing the similarities between the evaluation of actions, and so on, and the evaluation of explanations. The crucial similarity is that in defending our evaluations (whether of actions, institutions, or explanations) we must inevitably rely on a theory that purports to *justify* our standards of evaluation as over against other sets of (moral or explanatory) standards. In both cases, we will be engaged in the process of justifying our judgments, not of explaining our experiences. The analogy to keep in mind here is not that between moral theory and scientific theory but that between moral theory and scientific epistemology.⁵⁰

Since we must regard certain evaluative claims (those concerning which explanations are better than others) as true, we will be justified in believing those parts of value theory that support our standards of explanatory value. Just as we take the explanatory role of certain hypotheses as grounds for believing the hypotheses, we must, I suggest, take the justificatory role of certain evaluative principles as grounds for believing the principles. If the principles are themselves not reasonably believed, they cannot support our particular evaluations of explanations; and

49. Of course this leaves open the possibility that more specific attacks may be leveled at moral values; the point is just that once epistemic values are allowed, no general arguments against the existence of values can work.

50. Here I part company with other moral realists (for example, Boyd, Sturgeon, and Railton) who seem to hold that moral theory should be seen as being of a piece with scientific theory. See Richard Boyd, "How to Be a Moral Realist" in this volume; Nicholas Sturgeon, "Moral Explanations" in this volume; and Peter Railton, "Moral Realism," *Philosophical Review* 95 (1986), 163–207.

if we can have no grounds for thinking one explanation better than another, the Explanatory Criterion will be toothless.

Thus, if evaluative facts are indispensable (because they are presupposed by the Explanatory Criterion), we can invoke what might be called *inference to the best justification* to argue for abstract value claims on the grounds that they justify (and at the same time, explain the truth of) our lower-level epistemological judgments. And these very same abstract evaluative principles might well imply lower-level, distinctly moral principles and particular moral judgments. If so, then in defending moral values, we might begin with evaluations of explanations, move up (in generality and abstraction) to principles justifying these evaluations, then move back down, along a different justificatory path, to recognizably moral evaluations of actions, characters, institutions, and so on. That is, to argue for a given moral judgment (for example, that it is better to be honest than duplicitous), we might show that the judgment is justified by some abstract evaluative principle that is itself justified by its relation to our standards of explanatory quality (which are indispensable to our application of the Explanatory Criterion). In this way, particular moral judgments and more general moral principles might find their legitimacy through their connection with the indispensable part of value theory that serves to justify our judgments of explanatory quality.

To take one (optimistic) example: Imagine that we justify believing in some property by appeal to its role in our best explanation of some observations, and we then justify our belief that some explanation is the best available by appeal to our standards of explanatory quality, and finally, we justify these standards (rather than some others) by appealing to their ultimate contribution to the maximization of expected utility. Imagine, also, that having justified our standards of explanatory value, we turn to the justification for cultivating some moral property (for example, honesty). The justification might plausibly appeal to its contribution to the cohesiveness of one's society, and we might in turn justify cultivating properties conducive to the cohesiveness of society by appeal to the benefits available only within society. Finally, we might justify these as benefits by appeal to their maximizing expected utility. Appeal to the maximization of expected utility would then serve both as the best justification for certain standards of explanatory value and as the best justification for cultivating particular moral properties. It would justify both our belief that some particular explanation is better than another and our belief that some moral properties (for example, honesty) are better than others (for example, duplicity). We could then deny the justifiability of moral judgments only by denying the justification of our evaluative judgments of explanations.

So, in pursuing justifications for our standards of evaluation, we might discover that the justificatory principles we embrace have as consequences not only evaluations of explanations but also recognizably moral evaluations of character or behavior or institutions. Justificatory principles might come most plausibly as a package deal carrying both explanatory and moral evaluations in tow.

No doubt this picture is overly optimistic. Most likely, the justificatory principles invoked in justifying particular standards of explanatory quality will not be so neatly tied with the justifications available for having or developing certain (recognizably moral) properties or with the justifications available for condemning other (recognizably immoral) properties. In following out the two lines of justification—that is, in justifying particular evaluations by appeal to principles; which we in turn justify by appeal to more general principles—we may never arrive at a single overarching justificatory principle. Indeed, it is highly unlikely that we will ever get such a principle either in epistemology or in moral theory.⁵¹ It is even less likely that we will ever find a single principle that serves for both.

Although inference to the best justification legitimizes both the lower-level standards of explanatory value (simplicity, generality, and so on) and—more important—the very process of justification, the substantive principles the process engenders will probably vary according to what is being justified. When we are justifying a belief that some property is instantiated, one set of justificatory principles will come into play; when we are justifying the having or cultivating of some property, a completely different set of justificatory principles may prove relevant. The two paths of justification might neither coincide nor converge.

Yet a failure of convergence would not undermine moral justifications. The legitimacy of moral theory does not require any special link between explanatory and moral justifications. What it does require is that moral properties figure both as properties we are justified in believing exemplifiable and as properties we are justified in cultivating.

In constructing explanatory and justificatory theories, we may discover any of four things: (1) that moral properties are neither possessable nor worth possessing, in which case (I assume) moral theory loses its point; (2) that moral properties (for example, honesty, kindness) are possessed by some but that there is no justification for thinking some better than others, in which case only an unexciting conclusion will have been established—like atomic weights, virtue and value would exist, and claims involving them would have a truth value, but they would be normatively inert; (3) that we have no reason to believe moral properties are exemplified, but we do have reason to cultivate them, in which case a unique version of the is/ought distinction will have been established—there are no instantiated moral properties, even though there ought to be; or finally (4) that moral properties are acutally possessed and (some) are worth possessing, in which case moral theory will have found its strongest defense. Which of these four positions is right can be settled only against the background of an accepted justificatory theory.

51. In "Coherence and Models for Moral Theorizing," *Pacific Philosophical Quarterly* 66 (1985), 170–90, I argue that we have good reason for rejecting any proposed unifying fundamental principle we might find.

Of course, whether we are justified in believing moral properties are both possessable and worth possessing is an open question. Yet it is a legitimate question, and it is a question that can be answered only by engaging in moral theorizing; that is, only by attempting the justifications and seeing where they lead.