Peter Railton

Explanation and Metaphysical Controversy

I

A seachange has occurred in the study of explanation.[1] As recently as a decade ago, students of explanation had a fairly standard way of proceeding. They had before them a dominant theory of explanation, C. G. Hempel's covering-law account. They would begin by producing counterexamples to it and could go on either to construct arguments and epicycles to escape these counterexamples or to propose abandonment of all or part of Hempel's account, perhaps also advancing some replacement of their own devising that better fit the examples given. This all had the air of a well-defined activity, and it had gone on for a long time. Part of what gave the activity its apparent definiteness was the existence of a large and carefully drawn target at which to aim—Hempel's account.[2] Part, too, was a tacit understanding of the rules of the game: one would adduce intuitive judgments of sample explanations, intuitions one expected one's readers to share, and then one would use these intuitions to test analyses. An acceptable analysis, like an acceptable grammar, should get the intuitions of native informants right.

These rules functioned at the same time as limits, for they defined a way of proceeding in the analysis of explanation that prevented philosophers from wandering off into other areas, such as metaphysics. Two philosophers of explanation could draw up to the same table, lay out for inspection their examples and their analyses, produce at appropriate times their favorite counterexamples to each other and their various strategies for handling them, try on these grounds to convince each other, and then depart, without having breathed a word about metaphysical disputes in the philosophy of science, such as the growing debate between realists and irrealists. Except in a few polemical places, theories of explanation were described by their formal features—"covering-law," "why-question," "speech-act," "statistical-relevance"—and did not come prefixed with such metaphysical codes as "empiricist," "pragmatist," or "realist." Yet at the table sat empiricists, pragmatists, and realists.

This activity began to lose its sense of purpose once dissatisfaction with the

Hempelian account became widespread and no new orthodoxy emerged to take its place as a focus of discussion. Many philosophers grew weary and turned to other questions, including questions about realism and irrealism. Interestingly, within *that* dispute the concept of explanation proved to be indispensable: one side often claimed that a realist interpretation of scientific theory was justified by inference to the best explanation; the other side often responded that the realist's posits could do no explanatory work, since they yielded no empirical predictions beyond those already afforded by the observational reduction of the theory.

For a while, it seemed as though explanation were able to play so central a role in this debate precisely because there had ceased to be any widely accepted idea of what explanation is. Each side could use the notion for its purposes because, in the absence of any agreed-upon account, nothing prevented them from doing so. Dissatisfaction with such a state of affairs may in some measure have contributed to the seachange in studies of explanation. For, after a low ebb, the flow of writing on explanation has once again quickened, but in the opposite direction. Bas van Fraassen's *The Scientific Image*[3] and Wesley Salmon's *Scientific Explanation and the Causal Structure of the World*,[4] to mention only two prominent examples, are entirely explicit about placing analysis of explanation in a metaphysical setting, the one irrealist, the other, realist. The accounts they give of explanation, while of course making use of familiar strategies of exampling and counterexampling, are defended on grounds that do not purport to be innocent of metaphysics, and that involve metaphysically driven reinterpretations of intuitions.[5]

What else should one expect? It is inconceivable that a notion such as explanation could fail to depend crucially upon one's most general picture of the world and its ways. A pair of examples may suffice.

First, it seems in retrospect that the empiricist view that the subject matter of science is in the first and last resort experiential, that no individual experience ever intrinsically points to another, and that therefore neither science nor the philosophy of science can admit such a notion as physical necessity, accounts in part for the lasting appeal of the Hempelian doctrine that explanations must be deductive inferences from covering laws.[6] At least since Aristotle, explanation has been thought to involve demonstrating the *necessity* of the phenomenon to be explained. Thus, Hempel wrote in his first major paper on explanation that "explanation . . . aims at showing that the event in question was not a 'matter of chance', but was to be expected in view of certain antecedent or simultaneous conditions."[7] But if individual experiences never on their own necessitate other experiences, how is explanation via demonstration from "antecedent or simultaneous conditions" to proceed? The answer is that a statement of a general character is needed to establish the linkage: the fact that a given piece of copper wire was heated cannot by itself necessitate the fact that this wire's electrical resistance increased, but it could do so in the presence of a generalization stating that the resistance of a copper wire always increases as a function of temperature. Necessity of the explanandum given "simultaneous conditions" is established, but the relation is among descriptions of these events, not the events themselves, and the necessity involved is logical, not physical—the Aristotelian thought has found expression without exceeding the bounds of sense as delimited by empiricism.

If one did not confine oneself to the empiricist's world of individual experiences, or if one otherwise made room for the idea of necessary connections among events "in the world," then to show the necessity of a particular phenomenon one would not have to resort to law-based deduction. Indeed, the order of explanation might be reversed: one could explain the generalization as holding in virtue of the necessary features of concrete physical systems, much as some now think that the relative frequencies displayed by physically indeterministic systems are to be explained as accumulated outcomes of individual chance events.

This mention of indeterminism brings me to the second example. According to the Hempelian account of probabilistic explanation, probabilistic explanations are analogous to more familiar nonprobabilistic explanations in two important ways: they involve empirical laws essentially and they show that the explanandum phenomenon was to have been expected. The degree of law-based expectation is however weakened to accommodate the statistical character of probabilistic laws—outright demonstration is not required for explanation, only inference with "practical certainty." Yet, according to Hempel, probabilistic explanations also show three crucial disanalogies with nonprobabilistic explanations: they are inductive rather than deductive in form, they are relativized to an epistemic context, and they must satisfy a principle of maximal specificity relative to that context.[8] This has some rather odd results. Many of the explanations of chance phenomena in contemporary physics and biology are of events with low probability given the initial conditions, and these would be ruled out by Hempel's criteria. Moreover, if our evidence currently supports a lawlike statistical generalization assigning high probability to an explanandum, then even if this evidence is limited, a valid explanation can be given; if our evidence improves, but in a way that lowers the probability attributed to the explanandum, then we must say that explanation now is precluded, yet all the while continuing to hold that the old explanation was strictly correct. Isn't the implication of the new evidence instead that the old explanation was not strictly correct? Hempel is adamant in the case of nonprobabilistic explanation that changes in evidence can do no more than make a potential explanation *seem* right or wrong, just as he would think it a confusion to say that a statement about, say, the price of a slave on a given date in colonial Havana fluctuates in truth value as material on deposit in tropical archives deteriorates.[9]

Why should probabilistic explanation be so different from nonprobabilistic explanation, in Hempel's eyes, and why different in just these ways? The answer may reflect a metaphysical assumption. Suppose that one were a determinist at heart, and thought that in principle all indeterminism could be removed from our theories by the discovery of hidden variables. What view might one have of the probabilistic explanations offered by contemporary science? One might think that they are part of an inductive process, steps along the road to genuine, deductive explanation. They could serve to summarize the current state of knowledge about possible causal factors, and to approximate deduction as best that permits. Of course, if the factors known at a given time failed to determine a high expectation value for an outcome, then science clearly would be missing some very crucial variable or variables, so that an inductive inference based upon these factors would not be very acceptable as a proxy for explanation. On the other hand, if the known factors conferred a high probability upon the explanandum, even when compared to all extant sample populations, then we might think that the induction comes closer to approximating explanation. There would be no need to be specific—as, indeed, Hempel rather uncharacteristically is not—about exactly where the line demarcating "high probability" is to be drawn, since we would be certain that we were not in possession of a true explanation until all hidden variables were revealed, at which time we should be able in principle to bring the probability of the explanandum arbitrarily close to one. Thus, if one were a closet determinist, or if one at least had not extricated oneself from determinist ways of thought, one naturally would gravitate toward a Hempelian model of probabilistic explanation: its peculiarities would be virtues.[10]

Suppose by contrast that one did not have such a picture, but thought instead that the world's physical and social processes might be irreducibly indeterministic, albeit in an orderly way, with stable, law-abiding probabilities that manifest themselves in highly predictable relative frequencies. Then why should it matter for explanation whether the probability of the phenomenon to be explained is low (e.g., radioactive decay by long-lived isotopes of uranium) or high (e.g., radioactive decay by unstable isotopes of actinium)? And why should probabilistic explanation be epistemically relativized if nonprobabilistic explanation is not? If we discover that our current theory omitted some variables relevant to determining the probabilistic state of a system, then our current theory would be to that extent wrong, now and always, about why the system behaves as it does.[11]

In these two cases there is an evident coincidence between the requirements laid down by a model of explanation, on the one hand, and substantive metaphysical assumptions, on the other. Coincidence is not explanation, and I do not purport to have given explanations. Rather, I want simply to enter the remark that these coincidences went largely unnoticed in the vast debate over the Hempelian models of explanation.[12]

Things have changed, and overtly metaphysical discussion has now become much in evidence in the theory of explanation. Welcome as this development is, it does very much change the rules for the philosophical study of explanation. To say that one's background picture of the world is involved in one's *conception* of explanation is to suggest that one's intuitions about particular kinds or instances of purported explanation may not constitute a body of neutral data for testing theories of explanation—exampling and counterexampling will never be quite what they used to be. Equally, it is to suggest that there may not be a unitary, substantial *concept* of explanation to analyze, or, perhaps more accurately, that the concept of explanation is rather thin, too slight, perhaps, to be asked to resolve deep philosophical disputes.[13]

II

If we come to believe that the analysis of explanation is not a metaphysically neutral activity, then we must appreciate how this affects the debate between realists and irrealists. I feel this particularly keenly as a realist interested in the philosophical study of explanation who finds himself uncomfortable with certain uses of "inference to the best explanation" to defend realism.

Typically, a realist about the external world will say that the existence of such a world affords the best explanation of the stability, coherence, and so on, of our sensations. A realist about science will say that the approximate truth of current theory—where this means something like its approximate correspondence, under a literal interpretation, to the world—is the best explanation of the predictive and manipulative successes of science. In such claims it appears that we are being given a special sort of confirmation of the existence of the realist's posits, confirmation stemming from their contribution to explanation. Now I do not for a moment doubt that near the heart of realism lies a concern about explanation, but I am inclined to think that if the issue is to be put in a way that does not beg the question against the irrealist, then what is at stake must be considerations that, while making some contribution to confirmation, make their primary contribution to the scientific rationale and epistemic warrant of realism in other ways. It will take me most of this essay to say what that might mean.[14]

Consider contemporary science, with its theoretical vocabulary of 'cells', 'genes', 'atoms', 'energy', and the like. This theory is the result of a protracted history of experimentation and innovation, and has been vastly successful in practice. We are inclined to say that our evidence for this theory is observational: the manifest experimental and practical successes. Yet for familiar reasons, we can at best give only a partial interpretation of such terms as 'cell' and 'energy' in the observational vocabulary.[15] We therefore face the question: What stance should we adopt toward these theoretical terms and the sentences containing them?

Two answers to this question are realism and irrealism, but these terms have become desperately nonspecific. Let us try to fix ideas for the purposes of this paper, without worrying about the comprehensiveness of our category scheme. The realist with whom I will be concerned interprets at least some theoretical terms literally, and holds that we sometimes have good scientific reasons for believing the sentences containing them to be literally true.[16] The irrealists I will have in mind are characterized by the part of realism they deny. The first sort of irrealist does not interpret theoretical terms literally and therefore holds that sentences containing them do not have literal truth-values. The second sort interprets theoretical terms literally but holds that we never have good scientific reasons for believing the sentences containing them to be literally true.

Irrealism of the first sort directs itself against semantic realism, usually on the grounds of an empiricist theory of meaning: only statements with implications for the actual or possible course of experience have genuine cognitive content, and they have it only insofar as they have such implications.[17] This claim finds its defense in the argument that for language to be learnable and to be a medium of communication, it must be conditioned to publicly observable states of affairs, not unobservable ones. Once certain metalinguistic vapors have evanesced, according to this view, the cognitive content of any statement or system of statements is seen to be the difference to the observable world that the truth of the statement would make. Where there is no such difference, there is no genuine statement, but only, at best, noncognitive sorts of linguistic function. Let us call this sort of irrealism *observationalism*, avoiding the term 'verificationism' and its association with certain rather definite views about criteria of observability.

Irrealism of the second sort directs itself against the realist's claim that we sometimes have good scientific reasons for believing the literal content of theoretical statements to be true. Here empiricism is wielded in the first instance as an epistemic rather than semantic doctrine: all our evidence is observational; theoretical statements, interpreted literally, involve claims not only about the course of experience, but also about things in principle unobservable; therefore the evidence we have for our theories is at most evidence only that they are correct in their claims about what is observable, i.e., that they are empirically adequate.[18] Let us call this sort of irrealism *agnosticism*.[19] I will mostly be concerned with the agnostic in what follows, although the argumentative strategy employed can readily be extended to the observationalist.

Both sorts of irrealist are concerned to show that all genuinely scientific functions of theory can be carried out without (what they view as) the extreme resort of realism. For example, an irrealist may appeal to Craig's theorem to demonstrate that we can in principle derive from scientific theory (in conjunction with its interpretive system) an axiomatizable reduced theory, couched entirely in the observational vocabulary, which effects all the same deductive connections among observables as the original theory. Call this the *observational theory*.[20]

Now the realist typically counters that there are scientific functions served by the unreduced theory that cannot be served as well by the observational theory alone. In particular, the unreduced theory affords *good explanations* of the observational theory's regularities. Thus, the observed ratios of combination in chemical compounding can be explained by the microtheory of atomic structure and bonding. If we did not accept the reality of atoms, electrons, etc., we would be left only with a large number of brute observational regularities. Therefore, the argument goes, by inference to the best explanation we can say that the observational evidence supports going beyond the observational theory and believing as well the unreduced theory.

The task of assessing this argument is not made easier by the fact that so little has been said explicitly about what the criteria of "better explanation" might be.[21] We will have to consider several possible notions of "better explanation" and ask of each whether it could do the job.

Sometimes "better explanation" seems to mean something like this: one explanation is better than another if it allows us to see the explanandum as more likely.[22] The obvious cases spring to mind. Suppose you are staying at a midtown hotel and the large office building across the street, which had been ablaze with light, suddenly goes altogether dark. Simultaneously, the noisy air-conditioning towers on its roof fall silent. What would explain this? Well, the ballasts in fluorescent lamps, though highly dependable, fail at random every so often. Similarly for the electric motors that drive air-conditioning compressors. And similarly for the large step-down transformers that bring power to modern office buildings. One explanation of the sudden changes observed in the office building is that the ballasts in all the fluorescent lamps independently failed at once, and at the same moment for unrelated reasons the motors in the air-conditioning towers broke down. This explanation makes the event observed a very remarkable coincidence. An alternative explanation is that the power supply to the entire building has been interrupted because the transformer in the basement shorted out. This explanation makes the event observed one which still may not be terribly common, but one which is not so grotesquely unlikely. Hence, we deem it the better explanation of the two considered, and take it to receive more support from the evidence than its competitor. In a similar way, some realists have said that the approximate truth of scientific theory, under a realist interpretation, is the only hypothesis that does not make the success of science a coincidence of cosmic scale.[23]

Stated as a principle of evidence, we might put this sort of argument as follows. Since noncoincidental explanations show their explananda to have been less unlikely, they therefore receive more support from the occurrence of the explananda than their competitors. This is akin to a maximum likelihood theory of confirmation.[24]

Now let us see how this argument might be used to compare the support for our original theory as opposed to its observational reduction. Suppose we have observed certain regularities in chemical combination. The observational theory will contain as a proper part a large conjunction of observational regularities regarding chemical combination, but the original theory will also contain an account of underlying atomic structures and mechanisms that implies these regularities, and makes their conjunction likely. So, if these regularities are observed, the argument goes, then the original theory should be better confirmed than the observational theory.

The irrealists, however, have a ready response. Since the observational theory implies this conjunction of observed regularities, it naturally assigns that conjunction a high likelihood. Admittedly, it does so by a rather quick logical deduction rather than by a complex derivation from underlying structures and mechanisms. But still, if the original theory implies the conjunction, then the observational theory will do so as well.[25] The mere conjunction of these observed regularities may have low *a priori* probability in the absence of an underlying theory, but our comparison is not between belief in the underlying theory and belief in no theory; instead, it is between belief in the underlying theory and belief only in the observational theory.[26]

In reply the realist may complain that the observational theory does not give us a good explanation of the conjunction, but now 'better explanation' must mean something more than "likelihood-enhancing hypothesis."

III

One important recent suggestion is that explanation proceeds by the reduction of independent phenomena; theories are explanatory in virtue of the unification they effect among diverse phenomena. This suggestion, which has taken various forms,[27] has seemed to some especially congenial to realism, for it gives a role to the postulation of underlying structures and mechanisms beyond the mere entailment of observations. Thus, when physical theory treats of such *prima facie* disparate phenomena as fixed compounding ratios, emission and absorption spectra, the conductivity of metals, and the periodic table of the elements, by providing a unifying physical model of the atom, it achieves a substantial reduction in the number of phenomena that must independently be taken as basic. At issue is not whether the theory supplies a higher likelihood for these diverse phenomena than does its observational reduction, but whether it renders these phenomena more comprehensible or better understood[28] in virtue of tracing them to a common structural and causal basis. This common basis is characterized theoretically, and if we were to eliminate scientific commitment to atoms, electrons, etc., we would lose the physical model by means of which unification is achieved. This affords a sense in which the unreduced theory provides a better explanation of observations than the observational theory, and if in science we infer to the best explanation of the data, then, it seems, we will infer to the unreduced theory, not merely to its observational reduction.

Irrealists are entitled to find this argument somewhat puzzling—at least, in the mouth of a realist. For the realist wishes to give a literal construal of physical theory. If inference to the best explanation is to be an account of confirmation and if confirmation is a matter of (something like) epistemic probability of truth relative to evidence, then this, plus literal construal, requires that the realist take the criteria of better explanation to be indicators of increasing epistemic probability of literal truth. But why should the fact that a theory renders phenomena more comprehensible to us constitute evidence of that theory's truth? Some contemporary philosophers have sought to build criteria of comprehensibility into truth, as for example in coherence theories of truth. It is open to realists to take such a view, and to make explanatory unification part of coherence and therefore part of truth. But at least one mark of the sort of realism with which we are here concerned has been resistance to the idea that considerations of comprehensibility have a hand in determining truth. Realists of this stripe have held that the world and its explanatory structure could be very unlike what we think them to be, or could comprehend them being. It would be a major revision of such realism to accept synthetic *a priori* constraints of comprehensibility upon the nature of reality.

The realist may instead insist that, in speaking of the reduction of independent phenomena as yielding comprehensibility, he is not employing a notion tied to the character of human thought. Core elements of quantum mechanics may be much harder for us to understand than any finite fragment of the complex array of independent observational regularities it might entail. Nonetheless, there is a structural sense in which quantum mechanics renders this complex array more comprehensible, by showing it to be the multifaceted manifestation of a limited number of elementary entities satisfying a small set of quite general laws.

It begins to look as if possessing "comprehension" in this structural sense is nothing more than possessing unifying explanatory accounts, so perhaps it is best to state this defense of realism directly in terms of a unification criterion of *explanation* without invoking any notion of comprehensibility. And once more, the irrealist is entitled to puzzlement. If unification provides a criterion of explanation, and if explanation is evidence of truth, then unification is evidence of truth. Yet how does the realist know *a priori* that the world we inhabit is a unified one? In the case of the office building blacking out, we find the unifying explanation afforded by a power failure quite plausible, but it might for all that be false. The simultaneous blackout might have been the coincidental result of multiple independent causal chains, and, if so, then the single-power-failure explanation certainly would not be "better," though it would be more unifying. Similarly, the realist cannot rule out *a priori* the possibility that such chemical phenomena as combination ratios, spectral lines, conductivity, and so on, are in fact due to a host of independent underlying entities and mechanisms. But then unification cannot be the *criterion* which separates explanations from nonexplanations, or better from worse. This conclusion appears strengthened by the difficulty of seeing how a realist could settle a dispute between competing conceptions of unification.

Nonetheless, it might be argued, unification may be *evidence* for explanation, which then is evidence for truth. After all, in the past, unified physical theories have been sought, and, perhaps surprisingly, found. These theories have subsequently been replaced by other, still more unified theories. And these more unified theories have survived the most rigorous and extensive testing. Do we not then have *a posteriori* evidence that nature is unified? Indeed, do we not have an *a posteriori* means of settling disputes about what sorts of unification count for scientific explanation?

This argument manages to beg the question the realist originally turned to explanation to answer. For it simply assumes that the observational success of these theories is strong evidence that their theoretical claims are literally true—that nature really is unified in the ways they have claimed.[29] But suppose that the history of science had been slightly different: instead of the theories actually adopted, other theories had been proposed and accepted, theories with the same observational content, but which achieved different degrees of unification, or achieved unification in some different sense. We would, under this supposition, now find ourselves saying of *these* theories that their observational success entitles us to take these quite different degrees or kinds of unification as confirmation of the truth of those explanations characterized by such degrees or kinds of unification. Experience, it seems, cannot by itself single out one degree or kind of unification as privileged to confirm literal truth when an indefinite number of theories—exhibiting an indefinite array of degrees or kinds of unification—are all compatible with the evidence.

This objection might receive a familiar response: theory testing is always a matter of deciding among actual rival hypotheses, not of comparing an indefinite number of possible hypotheses. Historically, the number of rivals has always been small, and, within this limited competition, those hypotheses that have emerged triumphant have exhibited certain definite degrees and kinds of unification. This is sufficient to justify us in appealing to such degrees and kinds of unification when assessing potential explanations.

But again, the realist is making use of a consideration with no obvious connection to literal truth. Have we the miraculous gift that the mere fact that we entertain certain hypotheses but not others (others, that is, that would also be compatible with the data but that would exhibit different degrees or kinds of unification) counts as a reliable sign of their literal truth?

The irrealists may at this point re-enter the argument. If the realist is saying that unification is a pragmatic desideratum in theories, and that it is impractical to consider all possible hypotheses, so that we are well advised to seek theories that are simple in ways similar to successful theories in the past, then they can heartily agree, and welcome the realist's return to his senses. Moreover, if the realist is saying that there is some sort of linkage between such unification and explanation, it is open to the irrealists to concur. Where the irrealists will of course balk is at any attempt by the realist to market such straightforwardly pragmatic considerations as if they were indicators of truth in some nonepistemic sense. Such advertising is surely false, and it begins to seem possible that the realist will find himself in the embarrassing position that his antagonists are better able than he to give an account of inference to the best explanation. For irrealists have no difficulty in treating unification as at once a merely practical virtue of a theory and a criterion of explanation—they can treat explanation itself as a pragmatic matter. The realist, by contrast, can establish a criterial connection between explanation and unification only at the expense of throwing into doubt the existence of an intrinsic linkage between explanation and truth.

IV

This is more than the realist can bear. How, he asks, can irrealists draw upon the unification a theory effects among observational phenomena when they have replaced belief in the theory with belief only in its observational reduction?

Our irrealists, however, are entitled to hang on to the original theory so long as they make no use of it inconsistent with the fact that they abjure commitment to its literal truth. They are as free to use it instrumentally—without compromising their irrealism—as they are to use any other instrument of science. They may, for example, employ it as a mechanism for generating predictions and retrodictions, or otherwise assisting in the gathering, sorting, and ordering of observational data. One of the useful features of the theory, from this standpoint, is that it provides an especially economical way of organizing experience. If, as Mach seems to have thought, explanation *is* economical organization of experience by mental constructs not assumed to have any literal reflection in reality, then even a Machian irrealist is able to appeal to the apparatus of the unreduced theory to generate unifying explanations.

In response, the realist might pull *Aspects of Scientific Explanation* down from the bookshelf and point out that Hempel requires that the explanans of a successful explanation be true. Or he might argue that the intuitive notion of explanation-by-unification involves commitment to the existence of the reducing entities or properties, or to the truth of the underlying principles. No genuinely explanatory unification is achieved, he could claim, when one says that a wide array of observational regularities crop up *as if* there were some unifying underlying structures and mechanisms, any more than genuinely extrasensory perception is achieved when one realizes that it is *as if* an attentive friend has read one's mind.

I find this appeal to intuition congenial, but I cannot expect the irrealist to share my feeling. Here we have come to one of those points, prefigured in section I, where issues about the nature of explanation depend upon, and thus are incapable of resolving, substantive metaphysical issues. Whatever one's view of the form of theoretical explanation, Hempelian or unificationist, if one includes among the material criteria of explanation a condition of literal truth of theoretical premises, one risks begging the question against irrealism. Of course, realist explanations are "better explanations" if theoretical explanations must be interpreted *ontically*;[30] but, by the same token, irrealist explanations are "better explanations" if an ontic interpretation of theory is incoherent or indefensible. Our irrealists may help themselves to any of a variety of theories of explanation – whether something as simple as the idea that explanations are no more than economical symbolizations, or something as intricate as recent work on the pragmatics of answers to why-questions, or something as familiar as Hempel's own deductive-nomological account (suitably reinterpreted on the matter of truth). On such accounts, it may be possible to view the derivations afforded by a partially interpreted unreduced theory, or by a theory held to be no more than a model useful for scientific purposes, as legitimate explanations.

V

Is the debate over explanation between realist and irrealists thus a standoff, with each side able to do no more than beg the question against the other? To push beyond this point, we must locate some considerations that both sides find to some degree compelling, but that nonetheless tell one way rather than the other. The irrealists might think that there are such considerations, and that they tell against realism. For if realist arguments involving inference to the best explanation are seen as question-begging, then the agnostic's argument – to the premises of which our irrealists happily assent and our realist seems, despite some grumbling, obliged to listen – once again threatens to strip realism of any epistemic respectability.

The agnostic argues as follows. First, all the (internally available) evidence we will ever have for our theories is observational evidence.[31] Second, it is possible in principle to generate a theory – the observational theory – that contains all and only the observational implications of our original theory.[32] Third, since the original theory implies the observational theory but not conversely, the original theory could not be more likely to be true than observational theory. Fourth, since the original theory and the observational theory assign the same likelihood to any collection of potential observations, it could not be the case that the original theory is better confirmed by actual observations than the observational theory. Fifth, since the observational theory is thus weaker than the unreduced theory, and since it, like our evidence, does not extend beyond the observable, we will always have more reason to believe the observational theory than the unreduced theory. Sixth and finally, since there will always be an indefinite number of unreduced theories compatible with any set of observations, however large, we will never have sufficient evidence to accept any one of these unreduced theories as true, or nearly so.

Although this is the agnostic's argument, it may be borrowed by the observationalist, who can, with its help, pose a dilemma for realists: to the extent that the observationalist's semantic argument (section II) succeeds in showing that statements about unobservables lack full cognitive content, then *a fortiori* we will never have good reason to accept them as literally true; to the extent that the argument fails, then the agnostic's argument can be brought in to show that nonetheless we still will never have good reason to accept such statements as literally true.[33] Other irrealists may make other arguments and draw other conclusions, but this argument is available to our irrealists. Against them, what has our realist to say?

Recently, realists have expended some ingenuity trying to construct an argument to the effect that it is wrong to suppose that the stronger, unreduced theory must always fare worse in confirmation.[34] Consider again the conjunction of observable regularities about chemical combination, spectral lines, conductivity, etc. contained in the observational theory. If a new observation is made about spectral lines, this may help confirm the conjuncts concerning spectral lines, but it will not be relevant to the conjuncts concerning chemical combination. Similarly, observations of chemical combination will not help confirm conjuncts concerning spectral lines. By contrast, in the presence of the unreduced theory, observed regularities of spectral behavior can help confirm underlying hypotheses about the atom, which in turn imply regularities in chemical combination. If these regularities are then observed to obtain, that adds confirmation to the underlying theory, which then can confer its enhanced confirmation upon the other observational regularities it implies, e.g., those concerning spectral behavior. So, given the same evidence, the presence of the unreduced theory appears to permit more confirmation than is possible for the observational theory alone: its very strength—in postulating unobservable phenomena that connect observable regularities—allows it to be better tested by the same data.

However, since the unreduced theory is equivalent to the conjunction of its observational component—i.e., the observational theory—and the unreduced remainder, we should find this conclusion difficult to accept. How could *A* & *B* be better confirmed by a given body of evidence than *A* alone?

Recall that our irrealists are allowed to employ the unreduced theory in any way that does not require that it be believed to be literally true. In particular, they are allowed to use it to generate inferences from one set of observations to another, inferences *as if* the unreduced theory were true.[35] Whatever confirmation the unreduced theory might pick up of course also accrues to the hypothesis that observation will be *as if* the unreduced theory were true. And the *as if* hypothesis will do as much by way of logically relating various sorts of evidence as does the unreduced theory itself.[36] Any evidence we have that observation will be *as if* the unreduced theory were true is evidence that, for example, if we observe certain spectral behavior, we should expect certain behavior in chemical combination. If such chemical behavior is observed, that will further support the *as if* hypothesis, and this hypothesis can then confer its additional support upon its deductive implications, such as the existence of certain regularities in spectral behavior. But now note that the *as if* hypothesis is weaker than the unreduced theory. Relative to any given body of positive evidence, it should be better confirmed than the unreduced theory, and so able to supply that much more additional support to the observational theory. The irrealists, then, have built a better bootstrap.

This same conclusion can be reached in a way that may be more intuitive, albeit less careful. The observational theory does not involve the claim that there *are no* underlying mechanisms; rather, it is silent on this question. We may thus think of the unreduced theory as *one* way the unobservable world might be that would make the observational theory come out true: there might be atoms, they might behave as the unreduced theory says, and so on.[37] Since the truth of the unreduced theory is one way that the observational theory might be true, then any evidence for the unreduced theory is automatically evidence for the observational theory. But notice that there are in addition many other ways in which the observational theory might be true, corresponding to all the possible ways unobservable reality might be and still yield the regularities of the observational theory. Of course, we may not find these other theories as plausible as familiar atomic theory, but it hardly seems possible to say that the evidence rules decisively against their truth, since in fact their truth is compatible with all our evidence. So however slight our confidence that any one of these competitors might be true, still, our fallibilism requires us to give their disjunction some nonzero probability. But then we must conclude that the observational theory is better supported than the unreduced theory, for it will inherit not only the support of the unreduced theory, but also the support of all its empirically equivalent competitors.[38]

Perhaps the realist should give up tugging at his bootstraps, and accept the fact that the unreduced theory could not be more likely to be true than the observational theory. Still, he might hope that the unreduced theory is nonetheless sufficiently well confirmed by the evidence that it has passed some threshold of acceptance. If, however, we take confirmation to be a measure of epistemic probability of literal truth relative to the evidence, this seems a forlorn hope. For there simply will be too many theories that fit all existing evidence to allow the unreduced theory to achieve high confirmation. And as soon as we introduce criteria such as simplicity, unification, entrenchment, etc. to limit the field of alternatives, we break the connection between degree of confirmation and probability of truth—unless we are prepared to make *a priori* assumptions about the simplicity and unity of the world, or about the aptness of our concepts to it.[39]

A final realist strategy fails for similar reasons. The realist may insist that if science has been spectacularly successful in establishing observational theories, then it is an epistemic desideratum that science give a noncoincidental explanation of this success, i.e., we have epistemic grounds for preferring a theory that yields such an explanation over one that does not.[40] This strategy differs from inference to the best explanation since it sets forth an internal explanatory desideratum, which could be interpreted according to realist or irrealist theories of explanation.

Suppose then that we compare the observational theory with the unreduced theory. Let us say that unreduced, current science gives a noncoincidental explanation of the success of scientific practices at yielding a theory of observables by showing how the experimental procedures and apparatus of current science are able to detect the unobservable phenomena causally responsible for the regularities of experience, and how the inferential patterns of current scientists enable them to translate these inputs into a fairly accurate map of the world around them. We have already seen how irrealists might defend the *general* claim that the observational theory is able to offer explanations, though depending upon the sort of analysis of explanation an irrealist provides, this general claim may not suffice to show that *particular* explanations of the unreduced theory will be preserved. So it is possible that the unreduced theory's explanation of the success of science will not be preserved, and thus possible that we will have at least this reason for preferring realism about the unreduced theory on epistemic grounds.

There are, however, irrealist analyses that would preserve all the explanations of the unreduced theory (suitably interpreted, of course), some of which have been mentioned above. And in any event the realist is able to see that there will exist many distinct yet observationally equivalent theories capable of giving noncoincidental explanations of the observational success of science. Thus both a theory that space is Euclidean and universal forces are present and a theory that space-time is curved could afford noncoincidental explanations of the success of current physics. So the desideratum seems unlikely to narrow the field of competing theories sufficiently to allow the unreduced theory to acquire confirmation at or near a reasonable threshold of acceptance.[41]

Our realist, then, appears to suffer by the comparison with our irrealists. He is in some difficulty explaining the nature or role of inference to the best explanation, and he is committed to a theory that it seems he must concede to be both less well confirmed than the observational theory and far from a threshold of confirmation-driven acceptance.

VI

However, the realist may wonder whether the arguments made thus far really are on the side of our irrealists, and whether, if not, the debate might be a closer call than it now appears to be.

Suppose that we have an effective specification in observational terms of the range of circumstances in which an *observation* takes place.[42] We then can work our way through the recursively generated observational theory culling out all statements concerning relations among observable states of affairs where the description of the states of affairs does not imply that an observation occurs. This leaves as a residue a theory, call it the *manifest theory*, which says all that the observational theory says about observed observables—past, present, or future—but which is altogether silent about unobserved observables. Whereas the observational theory says things (something like) 'All swans are white' as well as 'All swan observations are white swan observations,' the manifest theory says only the latter.[43]

The manifest theory bears the same relation to the observational theory that the observational theory bears to the original, unreduced theory. First, all the evidence we ever will have for our theories is observed evidence. Second, it is possible in principle to generate a theory—the manifest theory—that contains all and only the implications for actual observation of the observational theory.[44] Third, since the observational theory implies the manifest theory, but not conversely, it could not be more probable that the observational theory is true than that the manifest theory is true. Fourth, since the observational theory and the manifest theory assign the same likelihood to any collection of actual observations, it could not be the case that the observational theory is better confirmed than the manifest theory. Fifth, since the manifest theory is thus weaker than the observational theory, and since it, like our evidence, does not extend beyond the observed, we will always have more reason to believe the manifest theory than the observational theory. Sixth and finally, since there will always be an indefinite number of ways in which the unobserved portions of the observable world might be that nonetheless are compatible with any particular manifest portion, we can hardly expect any one of these observational theories to be sufficiently probable relative to the evidence to warrant belief in its literal truth.

The observational theory, despite its modesty in comparison to the unreduced theory, makes sweeping claims about undetected observable states of affairs, including states of affairs that would be so distant in time or space as not to be epistemically accessible to us. If we never have a reason for believing a theory that goes beyond whatever evidence we now have or ever will have, a theory facing myriad competitors that will never be ruled out by the evidence, then we never have a reason for believing the observational theory. That is, we never have reason for believing our theory to be empirically adequate, rather than merely manifestly adequate.

VII

Although they must recognize the force of these arguments in favor of the manifest theory, scientific irrealists are perhaps unlikely to deflate their commitments and embrace manifestationalism. Why not? For our irrealists, it is a fortuitous matter rather than a deep fact which points in the history and expanse of the universe are occupied by observers, depending as it does upon such trivial matters as who happens to be where and looking at what when. To confine scientific theory to claims about experiences and refrain from generalization to unobserved (yet "in principle observable") states of affairs would be to carve out one part of the world and ignore the rest, imposing a boundary of little inherent scientific interest despite its connection with greater confirmation. Our irrealists see it as part of the ambition of science to give a theory adequate not only to that which is experienced, but that which could be experienced.[45]

Perhaps an irrealist could defend belief in the observational theory on the ground that it is the most reasonable projection outward from the boundary of the manifest theory, since it merely involves lifting the restriction to observations. Such a response would only show that the irrealist had forgotten his own arguments against the realist who would defend the unreduced theory as the most reasonable projection of the observational theory, involving as it does no more than lifting the restriction to observables. The difficulty in both cases lies in saying what 'reasonable projection' might mean. It cannot mean "the projection that best fits the evidence," for there are indefinitely many projections from observed to unobserved, or observable to unobservable, compatible with all the evidence. And it cannot mean "the simplest projection," for we have no rationally favored account of simplicity—indeed, were simplicity in some chosen sense held to be *a priori* evidence for the literal truth of a theory, we would in effect be claiming to have *a priori* knowledge that the world is simple in that sense. This would hardly be consistent with the empiricist scruples of our irrealists or the naturalist scruples of our realist. Similarly, our irrealists would fall victims to their own arguments if they were to attempt an *a posteriori* validation of a criterion of simplicity, or if they were to appeal to inference to the best explanation as a way of undercutting the manifest theory. For example, all of the arguments made earlier to show that irrealists are able to preserve explanation (under an appropriate interpretation, to be sure) can now be made on behalf of the manifestationalist.

Where might an irrealist turn to justify his conception of the proper scope of scientific theory? Recall that we began by considering two irrealists, one observationalist, the other agnostic. Thus far, the manifestationalist has used the agnostic's epistemic argument to challenge irrealism. Perhaps irrealism might find support in the observationalist's semantic argument. Unfortunately, the manifestationalist can with equal success turn the observationalist's argument against *him*. Only actual experiences—and not "in principle observables"—have any real effect shaping speakers' acquisition and use of language. In the teaching of language one can of course *say* that the word 'pippin' is meant to pick out unobserved as well as observed apples, but the only tests of competence with 'pippin' involve actual, experienced circumstances. After a certain number of tests our pupil may say "now I can go on . . . ," but the dispositional competence this reflects is notoriously underdetermined by any actual history of instruction, and the only constraint upon future use will be his manifest experience of success or failure in communicating or meeting expectations. Therefore, according to this argument, only those statements whose conditions of assertion are exhausted by manifest experience can have determinate cognitive content.

Scientific irrealists thus face a problem of their own making. If they accept the agnostic's argument, then it seems they must conclude that the observational theory could not be epistemically warranted. If they accept the observationalist's argument, then it seems they must conclude that the observational theory exceeds the bounds of sense. So the irrealists' positions would appear to be self-destabilizing: the logic of their criticisms of realism should by rights carry them past observational reduction and at least as far as manifest reduction. Arguably, it should carry them further still: the logical terminus may be phenomenalism, perhaps even phenomenalism of the present moment.

Yet familiar difficulties in accommodating the content and practice of modern science have led those irrealists with whom we have been concerned, namely *scientific* irrealists, away from phenomenalism. When Vienna discovered that scientific laws are not finitely verifiable, it ultimately decided to give up the criterion of finite verifiability rather than abandon the idea that propositions expressing laws are cognitively meaningful. When Vienna discovered that statements about physical objects could not be translated without loss into the language of sense-contents, it ultimately decided to shift the basic vocabulary from the language of sense-contents to the physical-thing language. And when Vienna discovered that explicit definitions or sentence-by-sentence translations of the theoretical expressions of science were not possible within the language of observable properties of physical things, it ultimately decided to give up the requirement of full interpretability in favor of partial interpretability. At each of these points a trade-off was made against confirmation in an effort to preserve the scope of science—scope adequate to encompass whatever the truth might be about the lawful features of the observable properties of the physical world, the rendering of which was taken to be the aim of science.

These eminently sensible moves inched scientific irrealism off its argumentative foundations, and care was not taken to ensure restabilization. We have noticed two points of possible collapse. First, without a theory of meaning that goes beyond the constraints of manifest conditions, it is unclear how the observational theory's unrestricted quantification over observables could have cognitive content. Second, without a theory of epistemic warrant or rationality in belief-formation that goes beyond confirmation, it is unclear how we could ever be warranted or rational in believing that the observational theory is literally true.

Let us make our job more manageable by setting aside the first, or semantic, sort of instability for the remainder of this paper. What might be done by a scientific irrealist to contend with the second, the epistemic?

Here follows one proposal, not the only one and perhaps not the best, but one that has the advantage of incorporating a perspective on epistemology and belief-formation that places the issues in a different light from discussions based upon confirmation.

There are at least two ingredients in knowledge: truth and warrant. The equation of epistemic warrant with degree of confirmation (as understood here) has the unnerving implication that even if our current scientific theory is correct—or very nearly so—in its claim about observables, still, belief in these claims could not be knowledge. We might have true belief—but not warrant, for the epistemic probability of the highly informative observational theory relative to our evidence will always be low.

Why is warrant important for knowledge? In part because true opinion might be reached by arbitrary, unreliable means. But now suppose that the experimental and inferential procedures followed by scientists in coming to believe the observational theory are in fact such as to be highly reliable in the production of true belief about observables. On this supposition, scientists are excellent detectors of the observable world, and it is not accidental that they have come to believe a true theory of the observable. If that is the way things are, then one of the worries about warrant has been addressed even though the theory arrived at is not well confirmed.

But warrant involves other worries as well. We can imagine epistemic agents following reliable belief-forming practices "in the dark," without being able to see themselves as reliable, or to understand the nature of this reliability and its basis. So let us further suppose—a supposition of slight difficulty, as we noted—that the irrealist is able to make nonliteral use of the unreduced theory, or literal use of the observational theory, to give good explanations of our reliability at detecting observables. A second worry about warrant can then be addressed even by a theory that is not well confirmed.

Still, a third worry remains. We could be reliable, able to see and understand our reliability, and yet unable to demonstrate our reliability *with high epistemic probability on the basis of internally available evidence*. Confirmation is just the thing to address this worry, for it is a measure of epistemic probability relative to evidence. Yet we have seen how little would be left of scientific theory were we to insist that it be well confirmed. The irrealist may be well advised to consider whether this third worry is one worry too many.

If the irrealist were to accept a reliabilist account of epistemic warrant, then the manifestationalist's challenge might be blocked, for the irrealist would then be able to say that although the observational theory is below a threshold of confirmation-driven theory acceptance—as virtually any interesting general theory must be—it might nonetheless be the product of a practice that is above a threshold of reliability in belief-formation. It is perhaps unusual to combine scientific irrealism with epistemic externalism, but then the usual sort of scientific irrealism has just been seen to have some difficulty in saying how it might ever be possible that (the observational reductions of) scientific theories could be objects of knowledge. (Recall that we have set aside irrealisms that are revisionist about truth.)

One source of resistance to epistemic externalism is the sense that it can establish no more than the bare possibility of knowledge, and thus may be unable to provide a rationale for any particular set of belief-forming practices. Reasons for belief seem bound to be internal even if warrant for belief is not, and where is the externalist to look for scientific reasons for belief? He must, it would appear, look to the goals of science. Securing confirmation for what one believes is certainly a goal of science, but if one simply wanted highly confirmed beliefs, one could achieve this by the expedient of restricting one's beliefs to tautologies and observation reports. Irrealists have tended to say that among the goals of science is attainment of an empirically adequate theory, but that claim is bound to be controverted by the manifestationalist. If irrealists are to have a convincing defense of their conception of the aim of science, and of their corresponding view of legitimate scientific belief-forming practices, they must do so by reference to uncontroversial scientific goals.

It would seem uncontroversial that *scientists seek theories that are predictively, manipulatively, and explanatorily successful over as wide a range of experience as possible, and moreover seek genuine knowledge of the natural world, that is, beliefs that are (at least) both true and warranted.* We need not attempt to decide which among these ends are intrinsic, which instrumental, so long as none of the parties to our debate wishes to dispute them. The manifestationalist can accept them, for although he confines belief to the realm of experience, he is prepared to assign theories literal truth over that realm, and to provide a suitable interpretation of such notions as explanation and warrant. Nor need the realist dispute them, for although his first loves may be truth and explanation, he has no doubts about the centrality to scientific practice of the aims of successful prediction and control.

Let us say, then, that a belief-forming practice is scientifically rational to the extent that it advances these ends. However, while reasons may be internal, we may not always be able to assess them internally. We are, for example, unable to generate internally ratings of belief-forming practices in terms of their truth-conduciveness. Moreover, uncontroversial ends are subject to controversial interpretation: manifestationalists, irrealists, and realists will differ over such notions as warrant and explanation, so no agreed upon rating in these terms is possible. Let us begin, then, with this goal: *success at prediction and manipulation over a wide range of experience.* If one belief-forming practice could be shown to be superior to others in this regard, that would furnish it a rationale in terms of the ends of science. Because this is but one of the goals of science, only a *prima facie* rationale is possible, but because predictive and manipulative success may be the most important of those scientific goals amenable to uncontroversial assessment, such a rationale could carry considerable weight.

To see what is at stake here, and how this dispute might be carried out, we need to say what a belief-forming practice involves. It is a matter of which beliefs, or what degrees of belief, are formed in response to new information. For example, how much, if at all, does one allow one's observation that all *F*s thus far sampled are *G* to strengthen one's belief that all *F*s are *G*s? does one—or does one not—restrict this generalization to manifest or observable *F*s and *G*s? does one invest significant belief in all those hypotheses that are consistent with present observations, or only in some? how does one select certain properties or classes of properties for projection rather than others (e.g., green vs. grue)? does one come to accept beliefs as true or approximately true, and if so, when? and so on.

The belief-forming practices endorsed by the manifestationalist require that we remain agnostic among all hypotheses about observables consistent with observations, while irrealists think such suspension of belief about unobserved observables is excessive, and that the sort of evidence scientists now possess can justify investing a high degree of belief in one rather than another manifestly equivalent theory of the observable.

What sort of difference would it make to science to adopt irrealist rather than manifestationalist belief-forming practices? At first glance, no difference, for observables come within the grasp of science only during observation. Moreover, a manifestationalist is able make instrumental use of the observational theory. He can without compromising his manifestationalism believe that experience is *as if* the observational theory held, and thereby can take advantage of whatever inductive and deductive systematization that theory effects among the observed.

Nonetheless, if we think of scientists situated at a certain point in the development of theory, then which sorts of proposals they will make, or which experiments they will undertake, and therefore which conclusions they will in fact reach, will be affected by the character of their individual and, as it were, collective credence functions at that time, and by the dynamical character of these credence functions over time—that is, by their belief-forming practices. If a community of scientists is agnostic about what goes on in unobserved states of affairs, and resists accepting substantive theories about such states, then its behavior in the formulation and testing of hypotheses can be expected to differ from a community that believes and revises theories about unobserved observables on the strength of observations. For example, a community of manifestationalists would not infer from a series of experiments that they should believe a principle of conservation requiring unobserved magnitudes to have determinate values. Nor would it seem appropriate to them to invest a very substantial portion of research resources in an experimental program the design of which presupposes determinate values for unobserved magnitudes. They are, after all, rational decision-makers, and their willingness to commit resources will reflect their estimation of the expected values of outcomes, which in turn will reflect their credence functions. But their credence functions are agnostic over the whole range of unfalsified theories of the unobserved, and this would hardly yield a convincing reason for concentrating resources in any particular venture whose prospective worth is highly dependent upon how things are with the unobserved.

Of course, nothing would prevent manifestationalists from thinking that actual observations show regularities *as if* (for example) certain conservation principles extending to the unobserved held. But the question is whether, if they had no significant tendency to believe any principle of this kind, they would be inclined to consider or develop the idea that observation behaves *as if* any particular conservation principle were true, and so be less likely to discover those regularities of actual experience that one would expect were things *as if* it were true.

This concern will afford a scientific rationale if, historically, beliefs in theoretical principles involving unobserved magnitudes—and commitments to experimental procedures motivated by such beliefs—have played a large role in the development of predictively and manipulatively successful theories. It is impossible to know how scientific practice would have differed in its character or accomplishment had scientists allocated their degrees of belief in strict accord with degrees of confirmation instead of accepting stronger, less agnostic theories. Yet it does seem evident that belief in powerful theories about observables has been pervasively involved in the development and testing of successful scientific theories.

Scientists, like the rest of us, have a remarkable thirst for underconfirmed, overambitious theories, theories that strive to be comprehensive. Scarcely is a fraction of possible data points in hand than smooth curves are drawn connecting the points and racing outward in both directions. This is *not* an application of a well-confirmed second-order theory according to which predicates apply uniformly over time. For there is a much better confirmed second-order theory which tells us only that those predicates science in fact keeps track of apply uniformly over observed time. Any perfectly general second-order theory of uniformity would just be another underconfirmed, overambitious theory.

It is better to admit straight out that we are beings who are strongly inclined to hold opinions about how the world is, even when it is not observed, and who view agnosticism about what we take to be meaningful questions as a kind of cost. Scientists, especially, seem so inclined. We do not know exactly what role this inclination has played in the development of science, but we do know that this development has been highly successful. To abstain from belief about all states of affairs not observed would not only frustrate the ambition of obtaining a theory with wide scope, but would also have unknown effects upon the future progress of science.

If one is to have beliefs about what is literally true where no observation occurs, as our irrealists do, then one will have to sacrifice confirmation. Such a trade-off will seem especially worth making if one also believes that theory development might thereby be enhanced. And it does seem that the development of existing theories—theories which have manifestly done well—has been tightly bound up with the prevalence of these more ambitious beliefs. If so, then a rationale in terms of uncontroversial scientific goals is at hand for irrealist rather than manifestationalist belief-forming practices—why abandon something that is working in favor of something untried?

VIII

The manifestationalist may concede that a non-question-begging rationale of this kind for irrealist belief-forming practices is available, but insist that the rationale is of an entirely practical rather than epistemic character. Even if it does succeed in giving a scientific reason for believing the observational theory over and above the manifest theory, it does not do so by showing that this stronger belief is *epistemically warranted*, only that it is useful.

The irrealist may reply that the predictive and manipulative success of general theories about observables does provide epistemic warrant of a kind the manifestationalist must accept, since it affords some *confirmation* of observational theory. Moreover, it affords some confirmation that the belief-forming practices that led to the observational theory are reliable. But the manifestationalist will quickly point out that, given the range of manifestly equivalent competitors, the confirmation is weak.

Yet it would be question-begging against our (revisionist, externalist) irrealists to confine epistemic warrant to confirmation. Thus the manifestationalist cannot say in advance that the rationale provided for irrealist belief-forming practices is without epistemic force, for some of the very features of scientific belief-forming practices picked out by the rationale may contribute to its reliability, and thus be warrant-conferring. This gives the irrealist an answer to the manifestationalist's charge that, given the sort of evidence available to science, the observational theory *could not* be an object of knowledge. It could be, and, if it is not terribly wrong in its claims, it is.

Has the irrealist thus restabilized his position? He may have a way of resisting collapse into manifestationalism, but he has strengthened his position only by weakening the argument he used against realism. If he is to resist realist pressures, he must show that scientific irrealism is both an appropriate equilibrium point in the trade-off between confirmation and scope and the natural stopping place for belief-forming practices given the rationale afforded by the aims of science. It is not clear that he can do either.

As far as confirmation is concerned, the observational theory is of course more probable relative to the evidence than the unreduced theory, but neither is remotely well confirmed. What could make unobservables seem the right stopping place for extending the scope of scientific belief? After all, we could make substantial gains in confirmation by restricting scope still further.

This particular stopping place would seem appropriate if it corresponded to a point of special epistemic significance. Traditionally, irrealists have appealed to the notion of *epistemic access* to explain the significance of in-principle observability. We have, it is said, more direct access to the observable. But unobservable states of affairs in our neighborhood may be more directly accessible than in-principle observable states of affairs elsewhere. It can be easier for us to detect proximate photons, or changes in inertial force, or collisions of air molecules, than it is for us to detect observable states of affairs in distant galaxies, the remote past, or the uncertain future. An eye accustomed to darkness can register the arrival of a photon, an inner ear in equilibrium can sense inertial accelerations, and a child can hear molecular collisions against the walls of a conch shell. Such detection is of course indirect and inferential by the time it reaches the level of self-conscious perception, but it would be a very naïve picture of the operation of the human perceptual system to think that we see proximate middle-sized objects noninferentially. And, of course, all judgments about unobserved observable events will be inferential, often highly so. The observable thus does not uniformly enjoy an inherent advantage in directness of access over the unobservable. Similar remarks apply to *reliability* of detection.

Moreover, the irrealist's claim that only statements about what is in-principle observable are *testable* cannot establish a convincing epistemic distinction. Van Fraassen argues that an unreduced theory's "vulnerability to future experience consists *only* in that the claim of its empirical adequacy is thus vulnerable," but of course in this sense the vulnerability of a theory resides only in its claim to manifest, not empirical, adequacy. Thus van Fraassen's barb that realism is "but empty strutting and posturing" rather than "courage under fire" would apply to irrealism as well.[46] Happily, however, the barb pricks neither, since theories about observables and unobservables alike can face the barrage of experience—as corporate bodies if need be—whenever we perform an experiment for the outcome of which they have implications.

The distinction between observables and unobservables does not, it seems, mark off a divide of special epistemic interest with regard to directness of access, potential reliability of detection, or liability to test. Nor does this distinction appear to be of special scientific interest. Just as the irrealist pointed out that the realm of the observed is essentially fortuitous from the standpoint of large-scale scientific theorizing, so may the realist point out that it is a fortuitous matter rather than a deep fact that certain objects are of a size or nature to be observable to humans without technical assistance. No doubt within that part of evolutionary psychology that treats of the phylogeny of human sensory receptors there will be interesting things to be said about the nature of our thresholds of discrimination, but these thresholds are unlikely to define a category with far-reaching significance in natural science as a whole.

The irrealist was offered an externalist response to the question of how perfectly general statements about observables might be objects of knowledge. Yet without an argument that assigns statements about the observable a special status with regard to reliability, it will be difficult for the irrealist who would make use of this response to explain why he draws the line of epistemic tolerance to include unobserved observables but to exclude all other unobservables. If the observational theory, despite its low degree of confirmation, could nonetheless be warranted because it was arrived at by reliable means, then the unreduced theory might for the same reason be warranted as well: scientific belief-forming practices might be quite reliable with regard to viruses, cells, molecules, and so on. It will be a contingent matter whether warrant gets as far as the observational theory, and whether, if it does, it goes on to further parts of the unreduced theory. No quick, *a priori* argument, such as the agnostic's argument promised to be, will be able to secure the epistemic respectability of irrealism, or the epistemic indecency of realism.

Nonetheless, the irrealist might argue, belief that the unreduced theory is true involves substantially greater exposure to the possibility of error than belief only in the observational theory. This much the agnostic's argument usefully shows. Should we not, then, see whether we can do without it? After all, even though a rationale has been suggested for belief-forming practices strong enough to yield the observational theory, much stronger practices would be required to reach the unreduced theory. Since avoidance of error counts among the uncontroversial values of science, a rationale might be constructed for eschewing such strong belief-forming practices. The cost seems small for a large reduction in epistemic risk: we need only give up (fallible) belief that the unreduced theory is (approximately) literally true, and we may go on (fallibly) believing that it is (approximately) empirically adequate.

Yet the costs of altogether abandoning these strong belief-forming practices may in fact be quite high. In the first place, error has disvalue for science thanks in part to the value of truth. If scientific methods are reasonably reliable with regard to unobservables of various kinds, then we could lose more in truth than we gain in reduction of falsehood. Indeed, we might do better with regard to avoidance of error by giving up projection to certain remote or exotic observables and keeping projection to humdrum, local unobservables like the micro-organisms on this planet.

In the second place, even if jettisoning some particular part of our theory at a given time would reduce epistemic risk at that time, it might have a different effect over time, for—as the irrealist argued against the manifestationalist—scientific practices in the formulation and testing of hypotheses are highly dependent upon what scientists believe to be the case. Altering scientific credence functions from belief to agnosticism on questions about the nature and behavior of unobservables might dramatically alter theory testing and development, with unpredictable effects upon the reliability of scientific practices.

If one believed the unreduced theory's claims, literally interpreted, one would believe that certain entities really exist and certain mechanisms really are at work. This inevitably would influence one's behavior as a scientist. For example, if one thought we had reason to believe that there are such things as viruses, that they are made of protein, that they interact with antibodies in certain ways, and so on, one would have a reason to design and carry out certain experiments (but not others), to draw certain conclusions from the outcomes of these experiments (but not others), and to attempt theoretical innovations consistent with the existence and functioning of the mechanism in question (but not others).

If by contrast one allowed oneself to believe nothing stronger than that the regularities among observables are *as if* viruses existed, had certain properties, and so on, this would be only a weak reason for such a sharp focus in experimental design and theoretical innovation, for there are infinitely many ways the world might be and still be *as if* certain mechanisms were at work. Analogies are cheap, and there would be no inconsistency or even incoherence in deploying different analogies at different points in the testing or development of a theory, or in restricting the scope of an analogy that failed to fit certain aspects of the phenomenon under study. Which principles are likely to occur to an investigator, or to attract sufficient commitment to bring about much investigation or development, will depend upon what he believes to be true. To adapt the irrealist's own example, a conservation principle that requires unobservable entities to take on certain magnitudes would not come to be believed by an irrealist, and yet, were this principle not to be believed, a large-scale regularity in what *is* observable might fail to be noticed.

A realist would be able to extend the response offered to irrealists without begging questions because the features of theory testing and development that involve literalism about unobservables are part of a scientific tradition that has been successful in terms recognizable to the irrealist, that is, successful at extending the scope and accuracy of prediction and control. It would be difficult to show that these achievements could have been reached had the community of scientists not invested substantial belief in the existence of viruses, cells, molecules, atoms, and the like.[47]

It is perhaps a sense of the difficulty of establishing such a counterfactual that leads van Fraassen, an advocate of an agnostic form of irrealism, to recommend to scientists "total immersion" in the unreduced theory. He seeks, however, to distinguish total immersion from outright belief; according to him, the former relation to a theory involves " 'bracketing' its ontological implications"—although not when these concern the existence of observables, whether observed or not.[48] What is this attitude? "Total immersion" sounds like—and seems to be taken to be[49]—acting exactly as if the theory were literally true when it comes to doing science. Now doing science is not just going through motions; it is inextricable from one's credence function and belief-forming practices, for it involves designing experiments, making inferences, deciding which hypotheses are worth pursuing, deciding when to withhold judgment and when not. Of course, scientists who *act on the belief that observation behaves as if the unreduced theory were literally true* are able to bracket ontological commitments, but for that very reason their practice can be expected to differ from that of scientists who *act on the belief that the unreduced theory is literally true*. Yet the achievements of science may owe something—perhaps a good deal—to the fact that brackets have long been removed from such things as viruses, cells, molecules, and atoms.

Van Fraassen makes the intriguing suggestion that a scientist expressing his "epistemic commitment" is "stepping back" from his theory.[50] This retreat is reminiscent of a now unfashionable image of epistemology as prior to, and on a different footing from, on-going science. But fashions come and go. So one must ask: if such a retreat were to take place, and if a scientist were to conclude upon reflection that he had no business being "epistemically committed" to unobservables, what would his credence function and his belief-forming practices look like when he returns to science? He must make decisions about large, expensive, long-lived experimental programs whose selection from the realm of all possible experimental programs depends for its rationale upon quite strong views about the properties unobservables actually have. Yet he is "epistemically committed" to full-fledged agnosticism about the accuracy of these views. Mustn't he regard any such program as an irrational stab in the dark? And mustn't he regard the success of such a program, should it arrive, as no reason to believe that the views upon which it is based are accurate—beyond their empirical (really, manifest) adequacy? Would that sort of attitude be compatible with the total immersion van Fraassen recommends? And is it the sort of attitude that has characterized contemporary scientific beliefs about viruses, cells, molecules, and the like?[51] Perhaps the natural description of the attitude van Fraassen seeks to encourage is not 'immersion plus agnosticism' but 'realism plus fallibilism.' The latter is self-critical, the former, it would seem, self-undermining.

If literalism about various unobservables has been a central feature of theory testing and development in modern science, then there may be a rationale in terms of the goals of science for adopting belief-forming practices strong enough to support holding significant degrees of belief in detailed hypotheses about the nature and behavior of unobservables. And the rationale is the same the irrealist offered: why fix what is not broken?

As with the matter of epistemic warrant and reliability, the dispute between realist and irrealist over the rationality of belief-forming practices would become an empirical question: how much of a difference has it made to the success of science that realist interpretations of unobservables have been accepted so widely and pushed so hard? Here, too, there is no short, *a priori* argument (this time, about rationality in belief-formation) to settle the matter.

I have sketched a response to manifestationalism that is available to an irrealist who will countenance epistemic externalism, and have argued that this same response could in turn be adapted by realists to provide them a response to irrealism. My aim has not been to advocate this response to either realists or irrealists, but rather to indicate its existence and point out that it does allow us to see these two views as occupying points along a continuum. Moreover, it enables us to see the debate between them as not merely question-begging, as the debate over explanation has often been. For, as the two views have been understood here, both embrace the idea that even when the beliefs of scientists extend to that which is not actually given in experience, this may yet be rational (in light of the aims of science) and warranted. In particular, it may be rational and warranted when these ambitious beliefs (and the ambitious belief-forming mechanisms from which they result) play an important role in a successful tradition of theory testing and development. No questions are begged when we pose the dispute between realist and irrealist in terms of where along the path of forming beliefs about unobserved phenomena they think—or better, predict—either rationale or warrant will give out.

Both camps recognize that the scientific ambition is not merely to be able to answer lots of questions about the world, but to answer them well. So both realist and irrealist want good reasons for making any sort of trade-off against confirmability. Religious and natural teleologies answer many questions that contemporary science leaves hanging, and thus are attractive for their potential informativeness. The difficulty of such teleologies, in light of the present discussion, is that over the last few centuries theoretical traditions *un*constrained by literalism about teleological commitments have done vastly better in promoting the development of theories that refine and extend our success at prediction and control. Even though it is a goal of science to achieve a comprehensive system of beliefs, advancement in this regard is always viewed as having attendant costs, assessed in terms of increased risk of error and all that goes with it. Those costs can be seen as scientifically rational only so long as an appropriately large gain in theory development has accompanied them.

IX

> I do not consider leaps of faith or belief in things unseen, arrived at for whatever reason, necessarily irrational—only the pretense that we are rationally compelled (e.g., through arguments concerning explanatory value) to embrace more than strict empiricism prescribes.
>
> B. C. van Fraassen[52]

Let us end with a story. You are visiting a friend who lives in a city spread along the sea at the foot of a sierra. A chain of small, rocky islands extends out into the ocean, linked to each other and to the shore by a series of picturesque bridges.[53] Each island commands a fine vista of the city and its setting, although the view becomes more spectacular the farther out one walks. Your friend, however, does not venture beyond one of the middle islands.

"I always stop *here*," he says with cheery firmness.

"Why?" you ask, "Do you prefer a less sweeping view?"

"Oh, no," he replies, "it isn't that. It's just that I worry about support, and so don't think you should cross any more bridges than, according to reason, you strictly have to."

"Well, we didn't strictly *have to* come out this far," you note. "As far as support is concerned, we could have stopped at one of the inner islands, or stayed on the shore in the first place—that's better supported than any bridge."

"But if you don't come at least this far, you will never see the lovely, wide view of the city."

You are puzzled. "If you think a lovely, wide view is worth crossing a few bridges, why isn't a lovelier, wider view ever worth crossing a few more? After all," you add, "if you always stop here, you will never see the mountains behind the city."

The story is left for the reader to finish.

Notes

1. I would like to thank Richard Boyd, Risto Hilpinen, Alan Goldman, David Lewis, Richard Miller, and Tim Maudlin for helpful comments on an earlier draft. I am grateful to Lawrence Sklar and Nicholas White for numerous helpful conversations.

2. See especially, C. G. Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965).

3. Oxford: Clarendon, 1980.

4. Princeton: Princeton University Press, 1984.

5. Revisionism with regard to intuition is certainly not altogether new, but, characteristically, it previously did not find expression in overtly metaphysical language. Thus, Hempel wrote in response to complaints that his D-N model allowed unintuitive explanations in which the phenomena mentioned in the explanans do not cause or "bring about" the explanandum phenomenon: "while such considerations may well make our earlier examples of explanation, and all causal explanations, seem more natural or plausible, it is not clear what precise construal could be given to the notion of factors 'bringing about' a given event." *Aspects of Scientific Explanation*, p. 353. The point is put as a matter of analysis, but 'precise construal' in the context is in fact elliptical for "precise construal in terms of the categories of analysis permissible under empiricism—observables, observable patterns of observables, extensional logic, and so on."

6. There are, of course, other ways of accounting for this appeal. Cf. Hempel's remarks in *Aspects of Scientific Explanation* about fitting explananda "into a nomic nexus" (p. 488). Along a similar line, see also the remarks on the "nomothetic" character of scientific explanatory practice in P. Railton, "Probability, Explanation, and Information," *Synthese* 48 (1981): 233–56.

7. "The Function of General Laws in History," first printed in 1942, reprinted in *Aspects of Scientific Explanation*, p. 235.

8. See Hempel, *Aspects of Scientific Explanation*, pp. 376–410.

9. For Hempel's rejection of epistemic relativity in the case of nonprobabilistic explanation, see Hempel and P. Oppenheim, "Studies in the Logic of Explanation," reprinted in *Aspects of Scientific Explanation*, where it is said of a nonprobabilistic case similar to the one discussed in the text that "the ampler evidence now available makes it highly probable that the explanans is not true, and hence

that the account in question is not—and never was—a correct explanation" (pp. 248–49). See also *Aspects of Scientific Explanation*, pp. 388 and, especially, 488: "the concept of deductive-nomological explanation requires no such relativization."

10. For further discussion, see P. Railton, "Taking Physical Probability Seriously," in M. Salmon (ed.), *The Philosophy of Logical Mechanism* (Dordrecht: D. Reidel, forthcoming).

11. See P. Railton, "A Deductive-Nomological Model of Probabilistic Explanation," *Philosophy of Science* 48 (1978): 206-26.

12. An interesting exception with regard to the first example is to be found in a discussion of the Humean background of the covering-law model by Israel Scheffler, *The Anatomy of Inquiry: Philosophical Studies in the Theory of Science* (New York: Alfred A. Knopf, 1963), pp. 19-20 . Counterexamples to the Hempelian account based upon "singular causal statements" might have furnished raw material for substantive metaphysical discussion, but instead both Hempel and his critics tended to keep discussion of singular causal statements within the realms of linguistic analysis and epistemology.

13. For the distinction between a concept and a conception, as applied to justice, see John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971), pp. 5ff.

14. Roughly, what I have in mind by *confirmation* is a measure of the epistemic probability that a theory—perhaps taken as an interpreted whole—is true relative to a given body of evidence. The clearest example of a theory of confirmation in this sense is a theory of logical probability, which gives what is in effect a measure of the semantic information of a hypothesis and the semantic information of a given body of evidence and an explication of the logical relation between the two. The evidence will confirm the hypothesis to the extent that it agrees with the information content of the hypothesis.

15. I am assuming here and elsewhere that it is possible to give an interesting account of the theory/observation distinction, and that this account will take the form of distinguishing observable properties—or predicates—from theoretical ones. There are reasons for worrying about just what this might involve, reasons that do, I think, play an important role in the argument for realism. But I would like at this point to concentrate on a set of issues that can perhaps most readily be seen if, for the sake of argument, a theory/observation distinction of the sort just mentioned is allowed.

16. It is, of course, a vexed question how one should formulate the idea of a literal interpretation. It is not enough to say that under the interpretation the phrase " 'virus' refers to viruses" comes out true, since the phrase 'refers to viruses' must—so to speak—itself be understood literally, for example along the lines characteristic of (nonvacuous) correspondence theories.

17. Again, we should set aside difficulties of characterization. For the sake of argument, we will assume some definite sense can be attached to the idea of the observational implications of a statement or set of statements. See also note 20, below.

18. Cf. the "constructive empiricism" of van Fraassen in *The Scientific Image*.

19. The controversy between realists and irrealists often involves a controversy over the nature of truth, but I will assume that our realist and our irrealists share a (nonvacuous) correspondence theory of truth, although the observationalist thinks that sentences containing theoretical vocabulary may, owing to partial interpretation, fail to have truth-value. I will only consider in passing irrealists who propose nonstandard theories of truth.

20. Use of Craig's theorem presupposes that scientific theory can be put into axiomatized form and that its interpretive system can be effectively specified. Moreover, for the *observationalist* to make use of Craig's construction he must employ a purely syntactic notion of logical consequence. Once again, for the sake of argument, let us set aside some doubts that might be raised about the irrealists' assumptions. Here, too, we ignore the worry that, except relative to a *ceteris paribus* clause we do not know how to complete, theories may not have determinate observational consequences. Let us also ignore Hempel's influential argument that the observational theory will not preserve the *inductive* systematization affected by the unreduced theory. We will return to the question of inductive systematization, below. For Hempel's argument, see his essay "The Theoretician's Dilemma," reprinted in *Aspects of Scientific Explanation*, especially pp. 214–17.

Those who employ a model-theoretic conception of theories, e.g., van Fraassen, may prefer to speak not of the observational theory (constructed in a syntactic, Craigian way) but of the observational *substructures* of a theory. What I say of the observational theory will, however, be applicable *mutatis mutandis* to the unreduced theory's observational substructures. For example, I will make nothing of the unwieldy character of a Craigian reduction or of the question whether a Craigian observational theory might in fact contain statements describing or otherwise referring to unobservables, albeit using an observational vocabulary to do so.

21. An early paper of Harman's is perhaps typical. He writes, "There is, of course, a problem about how one is to judge that one hypothesis is sufficiently better than another. Presumably such a judgment will be based on considerations such as which hypothesis is simpler, which is more plausible, which explains more, which is less *ad hoc*, and so forth. I do not wish to deny that there is a problem about explaining the exact nature of these considerations; I will not, however, say anything more about this problem." From Gilbert Harman, "The Inference to the Best Explanation," *The Philosophical Review* 74 (1965): 88-95. For an interesting discussion of possible criteria, see Paul Thagard, "The Best Explanation: Criteria for Theory Choice," *The Journal of Philosophy* 75 (February 1978): 76-92.

22. This construal is suggested by many of the examples in the literature, as well as by Harman's remarks concerning statistical explanation in *Thought* (Princeton: Princeton University Press, 1973), pp. 137ff: "the greater the statistical probability an observed outcome has in a particular chance set-up, the better that set-up explains that outcome." In the example that follows, the various explanations might be thought of as involving chance mechanisms.

23. See J. J. C. Smart, *Between Science and Philosophy* (New York: Random House, 1968), pp. 150-51.

24. See, for example, Jerrold Aronson, *A Realist Philosophy of Science* (New York: St. Martin's Press, 1984). One might further complicate this account by the introduction of prior probabilities. Some of the difficulties the realist would create for his argument by introducing prior probabilities *in this connection* will be discussed below.

25. Things get more complicated when we consider relations weaker than outright logical implication. Let us, again, avoid both complication and attendant worries.

26. It should be emphasized that this latter belief does not involve commitment to the *non*existence of underlying mechanisms. The former belief, of course, involves the latter, since the underlying theory implies the observational theory.

27. See for example Michael Friedman, "Explanation and Scientific Understanding," *The Journal of Philosophy* 71 (1974): 250-61, and Philip Kitcher, "Explanation, Conjunction, and Unification," *The Journal of Philosophy* 73 (1976): 207-12.

28. Friedman characterizes his motivation in part as seeking to give a theory of explanation that appropriately links this notion to a conception of understanding. See his "Explanation and Scientific Understanding," p. 251.

29. It is important to see that this objection is not a form of blanket skepticism about induction. The argument allows one to infer from past observational success to future observational success. It questions whether one can infer from observational success—past, present, or future—to theoretical accuracy.

30. For discussion of an "ontic" conception of explanation and its contrast with, among others, an "epistemic" conception, see Salmon, *Scientific Explanation*, pp. 84-111, 121-24.

31. The (vague) qualification 'internally available' is necessary since a realist may opt for a causal theory of evidence, according to which unobservable entities can enter into our evidence in virtue of their physical effects upon us, even when the causal origin of these effects is not "internally available" to their recipients. But the bald claim that we have reason to believe that unobservable phenomena are in this causal sense part of our evidence would beg the question against irrealism. By contrast, it is not question-begging for the irrealist to *begin* from premises involving internally available evidence alone, with the aim of reaching as a conclusion the claim that we never have reason to believe that a theory of unobservables is literally true. For although the realist may hold that evidence *need not* be internally available, it appears that if he is to provide any guidance for actual epistemic agents, he will have to provide an account of scientific theory from an internal point of view. Such an account necessarily would start out from what is "available." We will see whether it must also end there. (The qualification about internal availability will often be taken to be understood below, and omitted.)

32. The realist's suspensions of disbelief remain in effect.

33. The observationalist's primary argument will be addressed below only in passing.

34. The argument that follows is in the spirit of a number of recent proposals, especially that of Clark Glymour, in *Theory and Evidence* (Princeton: Princeton University Press, 1980), pp. 162-63. Although the argument as given by Glymour is not in the form of an inference to the best explanation, he does suggest a link between his "bootstrap" strategy and explanatoriness: "Informally, the body of evidence may be explained in a uniform way in one theory but have to be explained in several different ways in the second theory. We prefer the first." (p. 353; a typographical error in the original has been corrected)

35. There is an awkwardness here. The observationalist may be unable to give full cognitive meaning to the unreduced theory's claims, and so to that extent must profess that there could be no fact corresponding to the truth or falsity of that theory. To this extent, one supposes, he professes not to know what it would be like for that theory to be true or false. He does, however, admit to understanding this much of what it would be like for the unreduced theory to be true: what would be the case in the realm of the observable. Hence he is able to attach a definite sense to the idea that the world is *as if* the unreduced theory were true—i.e., that the theory implies successful observational predictions—though not perhaps the sense attached to this idea by the realist or by an agnostic who claims to have no principled difficulty in understanding what it would be like were the unreduced theory literally true. Of course, both the realist and the agnostic may claim to be up against psychological limitations in attempting to understand what state of affairs corresponds to, say, a quantum-mechanical superposition.

36. These relations will be effected syntactically according to the observationalist; they may be effected semantically according to the agnostic.

37. The carelessness of this argument, at least, when employed by the observationalist, is conspicuous here. For this is precisely the sort of supposition that the observationalist might deny has full cognitive meaning. See note 35, above. Still, the observationalist might be able to claim that he has some notional understanding of what the unreduced theory says regarding viruses, gas molecules, etc. on the strength of something akin to semantic analogy. (See Lawrence Sklar, "Semantic Analogy," *Philosophical Studies* 38 (1980): 217-34.)

38. David Lewis has suggested to me a much less cumbersome way of putting this point for any irrealist who allows the unreduced theory to take on truth-value. Call the observational theory O and the unreduced theory T. T implies O, and so is equivalent to O & T. Now the observational theory is equivalent to O & (T v -T), or to (O & T) v (O & -T). But then we can see at once that, given evidence that does not rule decisively against -T, O must be better confirmed than T, since O picks up not only the confirmation of O & T, but also the confirmation of O & -T.

39. At one point, when discussing scientific grounds for preferring one of two empiricially equivalent theories over the other, Glymour claims that this preference "need not be founded on *a priori* conceptions about how the world is or probably is" (p. 353). His alternative explanation is that we prefer "better tested theories," in his sense. Yet he concedes in the concluding paragraph of *Theory and Evidence* that he has given no argument for expecting theories better tested in *his* sense to be more likely to be true (p. 377).

40. Richard Boyd follows a related approach in "Realism, Underdetermination, and a Causal Theory of Evidence," *Noûs* 7 (1973): 1-12, especially pp. 3ff.

41. It should be clear that the realist cannot claim that this desideratum is a *constraint* upon which theories might be true—for the observational success of science might after all be coincidental. Rather, it is a condition which, if satisfied by a theory, would contribute to the likelihood of that theory relative to evidence that includes observational success.

42. It may not be easy to do this, but it should be no harder than (indeed, it is likely to be a part of) giving an effective specification of the notion of potential observability.

43. Just as, for the sake of argument, we extended considerable charity to irrealists in formulating their account of the observational theory—see, e.g., notes 15 and 20—, so will we extend charity to manifestationalists in formulating *their* account of the manifest theory. Much more work is needed before it can be judged whether either theory can satisfactorily be formulated.

44. Just as the observational theory does not say *everything* about the course of experience that the unreduced theory says (since the unreduced theory says, for example, that certain observations indicate the presence of unobservable entities), so the manifest theory does not say *everything* about the course of experience the observation theory says (since the observational theory says, for exam-

ple, that certain observations are instances of generalizations extending to the unobserved). This partly shows that one must be more careful of one's wording than I have been.

45. Cf. van Fraassen, "*Science aims to give theories which are empirically adequate* . . . I must emphasize that this refers to *all* the phenomena; these are not exhausted by those actually observed, nor even by those observed at some time, whether past, present, or future." (*The Scientific Image*, p. 12) He goes on to give a rough characterization of 'observable' as "X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it." (*Ibid.*, p. 16) Despite his use of the indicative mood, this plainly is a modal rather than a manifest notion.

46. B. C. van Fraassen, "Empiricism in the Philosophy of Science," in P. M. Churchland and C. A. Hooker, *Images of Science: Essays on Realism and Empiricism* (Chicago: The University of Chicago Press, 1985), pp. 254-55.

47. Note that this extension of the irrealist's argument does not make the question-begging assumption that earlier scientific theories were approximately correct in their claims about unobservables, only that they were (in effect) interpreted in scientific practice as making claims about unobservables, and that this feature of practice played a significant role in subsequent theory testing and development.

48. *The Scientific Image*, pp. 81, 12.

49. *The Scientific Image*, p. 82.

50. *The Scientific Image*, p. 82.

51. It is no simple matter to characterize contemporary scientific beliefs about unobservables. It suffices for the realist's argument if there are some areas of current science where ontological brackets have been removed and literalism plays a crucial role in theory testing and development. Few biologists would, even in reflective moments, bracket the existence of cells, and few chemists would bracket molecules. Indeed, many in these disciplines would say that we have successfully observed both. But few physicists would say they are entirely sure what removing the brackets from quantum mechanics would involve. What is striking, however, and what the realist's argument turns on, is a pattern of experimentation and theory development extending across various disciplines, a pattern that would make sense under literalist assumptions, but appear irrational under the assumption of genuine agnosticism. Even in the face of quantum mysteries, it is commoner to hear physicists express doubts about *what* the psi-function describes than *whether* there is some hard-to-conceive but nonetheless real unobservable state of affairs represented by it, and it is commoner still to hear physical hypotheses and experimental designs defended on grounds that involve reference to mechanisms and entities, literally understood. Literalism tends to be pursued whenever there is a robust enough sense of what could be going on in terms deemed "physically realistic."

52. "Empiricism in the Philosophy of Science," p. 286.

53. I am grateful to David Lewis for help in improving this image.