Observation, Experiment, and Hypothesis in Modern Physical Science, ed. Peter Achinstein and Owen Hannaway (Cambridge MA: MIT Press, 1985)

3

Observations, Explanatory Power, and Simplicity: Toward a Non-Humean Account

Richard N. Boyd

Introduction

Two Puzzles about Confirmation

Truisms from empiricist philosophy of science often turn out to be false, but one such truism is certainly true: Scientific knowledge is experimental knowledge. It is characteristic of scientific research that observational evidence plays a decisive role in the resolution of the issue between contending hypotheses, and whatever sort of objectivity scientific inquiry has depends crucially on this feature of the scientific method. It may be disputed what the limits of experimental knowledge are, or how theory-dependent observations are, or how conventional or "constructive" scientific objectivity is, but it is not a matter for serious dispute that the remarkable and characteristic capacity scientific methodology has for the resolution of disputed issues and for the establishment of instrumental knowledge is strongly dependent upon the special role it assigns to observation. In some way, observations permit scientists to use the world as a kind of court to which issues can be submitted for resolution. However "biased" the court may be, the striking success of scientific methodology in identifying predictively reliable theories must be in significant measure a reflection of that court's role. Call a theory *instrumentally reliable* if, and to the extent that, it yields approximately accurate predictions about observable phenomena. Similarly, call methodological practices instrumentally reliable if, and to the extent that, they contribute to the discovery and acceptance of instrumentally reliable theories. It is unproblematical that the crucial role of observation in science contributes profoundly to the instrumental reliability of scientific methodology.

Once this special epistemological role of observations is recognized, it is natural to investigate other features of scientific methodology by comparing or contrasting the role they play with the special role played by observation. In this chapter I apply this strategy to two features of scientific methodology. The first of these is the systematic preference that scientific methodology dictates in favor of explanatory theories. The second goes by several names; what I have in mind is the methodological preference for theories having the property or properties that philosophers typically call simplicity or parsimony and scientists often call elegance (or, perhaps, beauty) instead. The standards for theory assessment (call them the nonexperimental standards) required by these features of scientific methodology are, at least apparently, so different from those set by the requirement that the predictions of theories must be sustained by observational tests that it is, initially at least, puzzling what they have to do with the rational scientific assessment of theories or with scientific objectivity.

Simplicity, Explanatory Power, and Projectability: Why the Puzzles Are Serious

When we think of scientific objectivity, two importantly different features of scientific practice seem to be at issue: *intersubjectivity* (the capacity of scientists to reach a stable consensus about the issues they investigate and to agree about revisions in that consensus in the light of new data or new theoretical developments) and *epistemic reliability* (the capacity of scientists to get it (approximately) right about the things they study). If we focus exclusively on the first component of scientific objectivity, then the role of the preference for explanatory theories and for simple theories may not seem especially puzzling. Suppose that, for whatever reason, scientists prefer simple and explanatory theories. Perhaps the preference for simplicity reflects a basic psychological law and the preference for explanatory theories reflects a feature of graduate training in science; the source of the preferences does not matter. Suppose as well that, as a result of common indoctrination in their professional training (a common "paradigm" in Kuhn's sense), scientists share basically the same standards of explanatory power and relative simplicity. Under these conditions, the methodological preference for explanatory and simple theories could as readily contribute to the production of a stable scientific consensus as could scientists' common recourse to the results of observation. Indeed, the contribution to the

establishment of consensus might be greater, since the consensus-making effects of appeals to observations sometimes depend upon considerable luck or ingenuity in the design of experiments or in the making of relevant observations in nature.

Similarly, even if we focus on the second component of scientific objectivity (the capacity of scientists to get it right in their views about the world), some features of the contribution of the nonexperimental standards of theory assessment to scientific objectivity may seem unpuzzling. Suppose that we follow Kuhn (as we should) in holding that judgments of explanatory power and simplicity are determined by standards embodied in the current research tradition or "paradigm" (Kuhn 1970). Suppose, further, that we follow Kuhn (as we should not; see Boyd 1979, 1982, 1983) in holding that the theoretical structure of the world that scientists study (its fundamental ontology, basic laws, and so on) is constituted or constructed by the adoption of the paradigm. In that case the contribution of nonexperimental standards to the epistemic reliability of scientific methodology with respect to theoretical knowledge will seem unproblematical. After all, it would be hardly surprising that paradigm-determined standards of the acceptability of theories should be a reliable guide to the truth about a paradigm-determined world.

When we turn to the question of the contribution of such standards to the epistemic reliability of scientific methods with respect to our general knowledge of observable phenomena—that is, their contribution to the instrumental reliability of those methods—the situation is quite different. In the first place, the instrumental reliability of scientific methodology cannot be plausibly explained solely on the basis of the supposed paradigmatic construction of reality postulated by Kuhn and others. The fact that anomalous experimental results (results that contradict the expectations dictated by the theoretical tradition or "paradigm" in theoretically intractable ways) occur repeatedly in the history of science and are important in initiating "scientific revolutions" (Kuhn 1970) is sufficient to show that the capacity of scientists to get it right in their predictions about observable phenomena cannot be explained by assuming that the observable world is "constituted by" or "constructed from" the paradigm that determines their methodology. The data from the history of science simply do not permit such an interpretation (Boyd 1983).

Moreover, nonexperimental criteria of theory acceptability are absolutely crucial to the methodology by which scientists achieve instrumental knowledge (Boyd 1973, 1979, 1982, 1983, forthcoming). Briefly, this is so for two reasons. In the first place, nonexperimental criteria determine which theories are taken to be "projectable" in Goodman's (1973) sense. Of the infinitely many generalizations about observables that are logically compatible with any body of observational evidence, only the (typically quite small) finite number of generalizations that correspond to theories that are simple, are explanatory, and otherwise satisfy nonexperimental criteria are candidates for even tentative confirmation by those observations. Thus, many possible and experimentally unrefuted generalizations about observables are simply ruled out by such criteria (Boyd 1972, 1973, 1979, 1980, 1982, 1983, forthcoming; van Fraassen 1980).

To make matters more puzzling, in the testing of hypotheses that have been identified in this way as projectable, scientific methodology requires that a theory be tested under circumstances that are identified by other projectable rival theories as circumstances in which its observational predictions are likely to prove false. From the extraordinarily large body of predictive consequences of a proposed theory we identify those few whose testing is adequate for its confirmation by pitting the proposed theory against its few rivals that satisfy the nonexperimental criteria. To a very good first approximation this is the fundamental methodological principle governing the assessment of experimental evidence in science (Boyd 1972, 1973, 1979, 1980, 1982, 1983, forthcoming). Both judgments of projectability and assessments of experimental evidence for claims about observables thus depend on nonexperimental criteria of the sort that I am discussing. They play a crucial epistemic role in scientific methodology, and thus, like the practice of subjecting theories to observational tests, they contribute to the epistemic reliability that characterizes scientific objectivity.

The same point may be put in another way. Van Fraassen (1980, p. 88) discusses the various nonexperimental theoretical "virtues" and concludes that they should be treated as pragmatic rather than epistemic constraints on theory acceptability: "In so far as they go beyond consistency, empirical adequacy, and empirical strength, they do not concern the relation between theory and the world, but rather the use and usefulness of the theory; they provide reasons to prefer the theory independently of questions of truth." What we have just seen is that this approach is not tenable. We cannot think of the nonexperimental virtues as additional purely pragmatic criteria of theory acceptability above and beyond the criterion of empirical adequacy, for they are

essential components in the methodology we have for assessing empirical adequacy. They may also be desirable "independently of questions of truth" (although I doubt it); however, what is striking about their methodological role is precisely that they are central to the ways we assess observational evidence for the truth of generalizations about observables.

We really do have an epistemological puzzle, then. On the one hand, it seems pretty clear that scientific objectivity depends crucially upon the practice of deciding scientific issues by referring those issues to adjudication by the world via experimental or observational testing of proposed theories. That this practice should contribute to both components of scientific objectivity seems unproblematical. On the other hand, it appears that judgments of the aesthetic or cognitive merits of theories play a role in establishing the epistemic reliability of scientific practice comparable to that played by the criterion of experimental confirmation—indeed, such considerations seem to be part of the very methodology by which adequate experimental confirmation is defined. We need to ask how nonexperimental criteria of this sort can play a role so similar to that played by observations in sound scientific practice.

Traditional Empiricist Approaches to the Puzzle

Traditional logical empiricist philosophy of science treats the two nonexperimental criteria I am discussing quite differently. In the case of simplicity and related criteria, I think it would be fair to characterize the approach of logical empiricists as varying, depending upon whether they were doing abstract epistemology of science or applied philosophy of science. In the former case, simplicity was almost always treated as a purely pragmatic theoretical virtue. Often the rationality of preferring simple theories (all other things being equal) was explained in terms of rational allocation of time: It was more rational to investigate first the computationally less complex theories rather than those whose testing would require longer and more difficult computations. Variations on this theme of simplicity as a factor in intellectual economy are characteristic of the pragmatic treatment of the issue within twentieth-century logical empiricism.

In the context of applied philosophy of science—the examination of epistemological and logical issues surrounding particular issues in the various sciences—the situation was quite different. In general, logical empiricists treated issues of scientific methodology more descriptively when they undertook to do applied philosophy. That is, they identified methodologically important features of scientific practice, which they characterized in relatively nonanalytical terms (such as "simplicity," "parsimony," or "coherence"). They then cited the standards set by such features in offering solutions to philosophical problems in particular sciences. What they tended not to do, in such contexts, was emphasize the "rational reconstruction" of methodological principles in the light of the verificationist accounts of scientific knowledge and scientific language that formed the basis of their more abstract philosophical investigations. There is little doubt that this departure from verificationist strictures in applied philosophy of science was a reflection of the fact that the anti-realist perspective dictated by verificationism cannot serve as the basis for an adequate account of the epistemology or the semantics of actual science (Boyd 1972, 1973, 1979, 1980, 1982, 1983, forthcoming). In any event, the general pattern of departure from strict verificationism in applied philosophy of science was clearly manifested in many applications of the methodological principle of preference for simpler theories. In dealing with actual disputes in science, logical empiricist philosophers of science typically took the preference for simpler theories as a basic principle in the epistemology of science and cheerfully cited it as relevant to the determination of answers to questions that were plainly substantive rather than pragmatic. If this practice admits a coherent philosophical rationalization within the empiricist tradition, its rationalization probably lies in positions like that of Carnap (1950), according to which many substantive questions are held to be intelligible only when they are understood as arising within a theoretical perspective that is itself purely conventional and is chosen on essentially pragmatic grounds. Positions of this sort are anticipations of "constructivist" positions in the philosophy of science, such as those of Hanson (1958) and Kuhn (1970), and they are probably best thought of as intermediate between verificationist anti-realism and the anti-realism of these latter positions (Boyd 1983). In any event, no matter how their philosophical practice might be rationalized, logical empiricists routinely treated the methodological preference for simple theories as though it were on a par with more obviously epistemic norms of the scientific method when they were dealing with philosophical issues arising out of actual scientific theories or scientific practices.

V In the case of explanatory power, standard logical-empiricist accounts have all been variations on a single basic account, the deductive-nomological (D-N) theory of explanation, which has been employed both

in the abstract analysis of scientific methodology and in applications to particular scientific issues. The key idea is that what it is for a theory to explain an event is that it is possible to carry out an *ex post facto* prediction of the event from the theory together with suitable specifications of conditions antecedent to the event in question. The explanatory power of a theory consists in its capacity to serve as the basis for such "retrodictions." As logical empiricists knew, the adoption of this sort of analysis of explanatory power affords what appears to be a neat (indeed, elegant and even simple) solution to the puzzle of the relationship between scientific objectivity and the methodological principle of preference for explanatory theories. A successful explanation by a theory of some fact has just the same logical form as the confirmation by that fact of an experimental prediction of the theory. An explanation amounts to a demonstration that some event that has occurred previously can be retrospectively interpreted as an experimental test of the theory on which the explanation is based—a test which the theory passes. Thus, it is hardly surprising that the observational testing of theories and the practice of preferring explanatory theories should play similar roles in establishing scientific objectivity; they are the same practice, except for the largely irrelevant retrospective character of the latter. The methodological preference for explanatory theories is just a special case of the more general preference for theories that have survived experimental testing.

Three features of the D-N account make it especially attractive and V plausible:

- It has the consequence—plausible in the light of the integrative nature of scientific understanding—that the explanatory power of a theory depends upon the theoretical setting in which it is applied. That it has this consequence is a reflection of the acknowledged role of previously established "auxiliary hypotheses" in the derivation of testable (or applicable) observational consequences from a given scientific theory.
- It is appropriate to the conception of causation prevailing in the philosophical tradition in which it arises. This is so because the D-N account is simply a verificationist "Humean" gloss on the "unreconstructed" preanalytic conception that to explain an event is to say how it was caused.
- It portrays the methodological preference for explanatory theories as a special case of a general epistemic principle, of which the principle

dictating a preference for theories whose observational predictions have been confirmed is also a special case.

To these we should add a feature that almost all logical empiricists intended as a feature not only of the D-N account of explanation but also of their accounts of all other features of scientific methodology:

• Philosophical accounts of scientific methodology should all honor the distinction between the "context of discovery" and the "context of justification." In particular, they should invoke principles of deductive logic and statistical reasoning, but not principles of inductive logic of the sort that might be thought to provide rational principles for the invention or discovery of scientific theories. Accounts of the nature of theory confirmation should be entirely independent of contingent empirical claims about how theories are invented.

Despite its attractiveness, the D-N account of explanation proved vulnerable to a number of *prima facie* objections. These fall into three rough categories. In the first place, there seem to be clear-cut cases of scientific theories that explain events even though they do not yield deterministic predictions of their occurrence. Second, there are retrodictions from laws that fit the D-N account of explanations but do not seem to be genuinely explanatory. Finally, even where the laws in question appear to be deterministic, there are clear-cut cases of explanations in which it seems doubtful that the explanation is founded on information sufficient to allow the deduction of a retrodiction of the explained event.

It will not be my aim here to examine in detail any of these objections to the D-N account, or the rebuttals to them, since I hope to raise difficulties for the D-N account of quite a different sort. Suffice it to say that the first of the objections has typically been met by requiring only that there be a statistical prediction of the explained event deduced from the laws in question. (For criticisms of this approach and a defense of a related alternative see Salmon 1971.) The second objection has typically been met by holding that the apparently deficient D-N "explanations" are indeed explanations that their apparent deficiency reflects merely their failure to meet purely pragmatic standards of, for example, practical or current theoretical interest. Against the third sort of objection, the typical reply has been (depending on the case at issue) either to identify suitable "tacit" premises to make the deductive prediction of the event possible or to assimilate the case to the statistical version of the D-N account. I think it is a fair summary of the literature in the empiricist tradition to say that the first and the third of the objections we are considering have been seen as the more pressing and that the treatment of the second objection in terms of pragmatic considerations has typically been taken to have been largely successful. For the sake of argument, I will assume throughout the chapter that an adequate empiricist solution to the first of these problems exists. I will speak of "predictions" or of "retrodictions" deducible from scientific theories on the understanding that these terms cover the relevant sort of statistical prediction or retrodiction in cases where deterministic predictions or retrodictions are not possible.

The Aims of the Chapter

The D-N account of explanation and the "Humean" account of causation from which it derives are, in their numerous variants, the most durable legacy of the tradition of logical positivism within professional philosophy. (No doubt extreme noncognitivism in ethics is even more durable if we consider the thinking of those who are not professional philosophers.) What I intend to show here is, first, that these legacies of positivism are inadequate even as first approximations to the epistemic task of explaining how considerations of explanatory power are able to play a methodological role analogous to that played by observational testing in science. An adequate explanation of this phenomenon requires that we adopt an account of explanation appropriate to a scientific-realist conception of scientific theories and scientific knowledge (see Smart 1963; Putnam 1975b; Boyd 1972, 1973, 1979, 1980, 1982, 1983, forthcoming).

I shall argue that nevertheless logical empiricists were right in proposing an account of explanation having the first three features mentioned above: that it portray the explanatory power of a theory as depending upon the theoretical setting in which it is applied, that it be consonant with an appropriate account of causation, and that it treat explanatory power as epistemically relevant in the same way that success in making observational predictions is. I will offer a realist account of a wide class of scientific explanations that meet these criteria and that avoid the difficulties plaguing the D-N account and its variants.

I will indicate how the realist account of explanation can be extended to a closely analogous treatment of the other nonexperimental criterion of theory acceptability we are considering: simplicity. Indeed, I will argue that, in an extended but well-motivated sense of the term, both simplicity and explanatory power are "experimental" criteria of theory acceptability. They reflect indirect theory-mediated evidential considerations that can be accounted for only from the perspective of scientific realism.

Finally I shall argue that these realist treatments of the nonexperimental criteria show that the fourth feature of logical empiricist accounts of scientific methodology—the sharp distinction between context of invention and context of confirmation—cannot be sustained. An adequate account of the epistemic role of observations or of the nonexperimental criteria of theory acceptability requires that we countenance inductive inferences at the theoretical as well as the observational level. The epistemic reliability of such inferences depends both upon logically contingent facts about the particular theoretical tradition that human invention has produced and upon logically contingent psychological and social facts about the capacity of scientists to employ that tradition in the invention of future theories. No account of the epistemology of science that is independent of contingent claims about the social and psychological foundations of scientific practice can be adequate to the task of explaining how the epistemic evaluation of scientific theories works. The epistemology of science must be "naturalized" in a way that requires that the sharp distinction between theory invention and theory confirmation be rejected.

The Humean Conception of Explanation

The "Humean" Conception of Causation in Recent Empiricism

According to Hume's philosophical definition, a cause is "an object precedent and contiguous to another, and where all objects resembling the former are placed in like relations of precedency and contiguity to those objects that resemble the latter." Hume's reasons for adopting this definition are as close to twentieth-century verificationism as one can get in early empiricism. His account has the property (characteristic of later verificationist analyses of scientific notions) that, according to the analysis it provides, the cognitive content of a causal statement is a simple generalization of the cognitive content of the observation statements that are seen as providing evidence for it. No inference from

observed regularities to natural necessities or causal powers is required for the confirmation of causal statements.

The version of Hume's account that prevails in twentieth-century empiricist philosophy is significantly different. Roughly, this account holds that an event e_1 causes an event e_2 just in case there are natural laws L and statements C describing conditions antecedent to e_2 such that from L and C, together with a statement reporting the occurrence of e_1, a statement describing the subsequent occurrence of e_2 can be deduced. This account, with variations intended to rule out "trick" cases and to accommodate statistical laws, has proved to be the most durable of the doctrines of logical positivism. The contemporary empiricist account is of course fundamentally verificationist in its content and its justification. "Metaphysical" commitment to such insensibilia as causal powers, underlying mechanisms, hidden essences, and natural necessity is eliminated in favor of the "rational reconstruction" of causal notions in terms of deductive subsumption under natural laws. As in the case of Hume's original formulation, the effect is to make the cognitive content of causal statements closely related to the cognitive content of the observation statements that support them. On an empiricist conception, the nonstipulative cognitive content of natural laws is exhausted by the observational predictions deducible from them, since scientific knowledge cannot extend to "unobservables." Moreover, confirmation of a body of laws consists solely in the experimental confirmation of just those predictions. Thus, the cognitive content of a body of laws consists in a predicted pattern in observations, and evidence for the laws consists in observations that instantiate the pattern in question. Just as in the case of Hume's analysis, events are causally related if they instantiate an appropriate pattern in observable phenomena and evidence for a causal claim consists of confirmation of instances of that pattern. What is different is the way the two "Humean" analyses of cause characterize the relevant patterns in observable phenomena.

The difference in formulation between Hume's account and the account that logical positivists adopted in his name reflects two important features of recent empiricist philosophy of science. In the first place, the contemporary formulations reflect the emphasis recent empiricists have placed on employing the results of modern logical theory in the "rational reconstruction" of scientific concepts. Where Hume's "natural" definition clarifies his philosophical definition by reference to the natural disposition of the mind to form associations of ideas, the contemporary definition refers instead to the logical integration of propositions into deductive systems. More important for our purposes is a special case of this sort of reconstruction: the syntactic conception of "lawlikeness." Hume's account of causation is incomplete without some answer to the question of what respects of resemblances are relevant in applying the definition he offers. It is rather plain that Hume's answer is provided by the "natural" definition of causation: The respects of resemblance that "count" are just those to which the mind naturally attends in forming general beliefs about property correlations. Logical positivists quite rightly rejected this particular form of philosophical naturalism. In its place they substituted an appeal to the notion of a natural law. Respects of resemblance "count" just in case they are the respects of similarity indicated as relevant by natural laws. Now, for any two nonsimultaneous events there will be some true general statement about events from which one can deduce a prediction of the occurrence of the subsequent event if one is given as an additional premise a statement reporting the occurrence of the antecedent event. If by a natural law one were to understand simply a true general statement about events, the contemporary "Humean" definition of causation would have the absurd consequence that causation amounts simply to temporal priority. The positivists' solution was to distinguish "lawlike" from non-"lawlike" generalizations and to understand the natural laws to be just the true lawlike generations. It was understood that lawlikeness should be a syntactic property of sentences—in particular, that it should be an *a priori* question which sentences were lawlike, although of course it would be an empirical question which of these were true (and thus laws). The problem of characterizing those generalizations that are lawlike is just the same problem as characterizing those generalizations that are "projectable" (Goodman 1973) or those kinds, relations, and categories that are "natural" (Quine 1969). In each case the question is which patterns in empirical data should be thought of *prima facie* as instantiations of causal regularities.

Within recent empiricist philosophy (see, e.g., Goodman 1973; Quine 1969) there have been proposed variations on the traditional positivist conception of lawlikeness according to which judgments of lawlikeness or projectability are not *a priori*. Successful inductive generalizations governed by particular judgments of projectability may be taken to provide empirical evidence in favor of the projectability judgments themselves, whereas unsuccessful inductive inferences may tend to disconfirm the projectability judgments upon which they depend. It

will be important for us to establish just what variations on the traditional positivist conception of lawlikeness are compatible with the contemporary Humean conception of causation.

The Humean definition of causation, whether in its eighteenth-century or in its twentieth-century version, is essentially an eliminative definition. It is not an analysis of what we (as scientist or as laypersons) ordinarily take ourselves to mean when we talk about causal relations. No doubt we would ordinarily paraphrase causal statements in such terms as "makes happen," "brings about," or "necessitates," or in terms that refer to underlying mechanisms or processes. The Humean conception rejects definitions of causation in such terms not because they inadequately capture our preanalytic conceptions but rather because our preanalytic conceptions are held to be epistemologically defective. Neither natural necessitation nor most of the underlying mechanisms or processes to which we would ordinarily refer in paraphrasing causal statements are observable. Therefore, on the empiricist conception, knowledge of such phenomena—if there are any—is impossible. Our preanalytic conceptions of causation, if taken literally, would render knowledge of causal relations likewise impossible. The Humean definition of causation offers a remedy for this difficulty by "rationally reconstructing" our causal concepts in noncausal terms. Reference to suspect unobservable entities, powers, or necessitations is reduced to reference to patterns in observable data. This is the whole point of, and the sole justification for, the Humean definition. The appropriateness of various conceptions of lawlikeness must be assessed in the light of this essentially verificationist justification for the Humean definition. An analysis of lawlikeness—whatever its independent merits might be—is inappropriate for the formulation of a Humean definition of causation unless it is itself compatible with the verificationist project of reducing causal talk and other talk about insensibilia to talk about regularities in the behavior of observables.

This constraint is important because, just as our preanalytic inclination would be to paraphrase causal statements in terms of natural necessitation or underlying mechanisms, our preanalytic conception of the distinction between natural laws and accidental generalizations is probably equally infected with unreduced causal notions. We might, for example, propose to define as lawlike those generalizations that attribute the observable regularities they predict to the operation of a fixed set of underlying mechanisms, or perhaps to consider lawlike those generalizations that attribute the predicted observable regularities to underlying mechanisms that are relevantly similar to those already postulated in well-confirmed generalizations. Some such definition of lawlikeness might well be the correct one (indeed, I think that the latter proposal is very nearly right), but no such definition would be appropriate for the formulation of the contemporary version of the Humean definition of causation. If lawlikeness is already a causal notion, then the Humean definition fails to accomplish the desired eliminative reduction of causal notions to noncausal observational notions and is thus without any philosophical justification. It must be emphasized that any analysis of lawlikeness that referred to unobservable "theoretical entities" and "theoretical properties" or to unobservable underlying mechanisms or processes would be just as inappropriate for a formulation of the Humean definition as one that talked explicitly about "natural necessity." Such "secret powers" or hidden "inner constitutions" of matter have always been the paradigm cases of the sort of alleged causal phenomena reference to which the Humean definition of causation is designed to eliminate. To appeal to unobservable constituents of matter and their unobservable theoretical properties (such as mass, charge, and spin) is precisely to engage in a twentieth-century version of Locke's appeal to insensible corpuscles and their various "powers." Unreduced reference to, say, the charge of electrons *just is* reference to an unobservable causal power of one of the unobservable participants in the unobservable mechanisms underlying causal relations among observables. Reference to phenomena of this sort is precisely what the Humean definition must eliminate.

Similar considerations dictate a closely related additional constraint on definitions of lawlikeness suitable for formulations of the Humean definition. Suppose that a definition of lawlikeness were proposed that involved no unreduced reference to causal notions or to theoretical entities. Such a definition of lawlikeness might still prove inappropriate for the Humean definition of causation if in order to determine whether or not a statement fell under it one would have to rely on inferences from premises that themselves involve irreducible reference to causal notions or to theoretical entities. After all, the whole point of the Humean definition is to render causal statements confirmable even on the assumption that knowledge of unobservable phenomena is impossible. If judgments of lawlikeness can be made only on the basis of premises thus supposed to be unknowable, then the Humean project fails. As we shall see, this proves to be the case.

Explanation and the Humean Definition of Causation

For a wide class of cases, an explanation of an event is provided by a statement saying how the event was caused. On the Humean definition of causation, saying how an event was caused amounts to deductive subsumption of the event under natural laws together with specifications of antecedent conditions—in other words, deductive retrodiction of the event from initial condition statements and laws. The preanalytic conception of a wide class of explanations reduces to the deductive-nomological conception upon Humean rational reconstruction. This fact provides the only good reason there has ever been to accept the D-N account of explanation; to a good first approximation, the D-N account *just is* the Humean definition of causation.

As the recent empiricist conceptions of causation and lawlikeness depart significantly from our preanalytic conceptions, so the D-N account of explanation departs from our unreconstructed conception of explanation. Without doubt our preanalytic understanding of the central cases of scientific and everyday explanation would, if spelled out, invoke unreduced notions of causation and of causal processes and mechanisms. If unreconstructed causal talk were philosophically unobjectionable (as I shall eventually argue, it should be), there would be no reason whatsoever to adopt the alternative D-N account. Indeed, the considerable difficulty defenders of the D-N account and its variants have had in accommodating paradigm cases of explanations (and of nonexplanations) to the definitions of explanation they have offered indicates just how far from compelling (or even plausible) the D-N account would be were it not for the verificationist objections to unreduced causal notions.

The Humean roots of the D-N account are evident in the literature, albeit in a somewhat unexpected way. A survey of the classical early papers defending and elaborating the D-N account and its variants (e.g., Hempel and Oppenheim 1948; Hempel 1965; Feigl 1945; Popper 1959) indicates that in the typical case Hume is never mentioned but it is taken for granted that the D-N account is appropriate for straightforward causal explanations. In Hempel and Oppenheim 1948 and in Hempel 1965 the Humean analysis (not so described) is very briefly appealed to in the case of causal explanations. Hempel and Oppenheim adduce the requirement that the explanans must have empirical content in support of the D-N account, and Feigl insists that it is possible to "retain the valuable anti-metaphysical point of view" in rival conceptions of explanation while adopting the D-N definition instead. None of the early authors, however, spend much time elaborating these plainly verificationist and Humean justifications for the D-N account. Instead, insofar as the account is defended in detail, they defend its extension to less clear-cut cases (teleological, motivational, or statistical explanations, for example). They take it for granted that, perhaps with the help of a few verificationist "reminders," the reader will agree that the D-N account of explanation is appropriate for ordinary causal explanations and will find controversial only its extension to other sorts of explanation. The unself-conscious Humeanism in these early papers is striking, but the fact that it is unself-conscious merely makes it clearer that the philosophical justification for the D-N account lies in the fact that it represents the Humean rational reconstruction of the notion of causal explanation. At least for the central case of causal explanation, no other philosophical justification is to be found.

This situation persists in the more recent literature on the D-N account. I think it is fair to say that, insofar as more recent philosophers have defended the D-N account or its variants, their strategy has been to offer rebuttals to a variety of putative counterexamples to the account. These rebuttals have often been extraordinarily ingenious. It is nevertheless true that if we had no Humean and verificationist reasons for accepting the D-N account in the first place these rebuttals to counterexamples would not by themselves constitute a good reason to accept the account. The situation remains that the D-N account of explanation is a preanalytically implausible analysis whose philosophical justification lies in the presumed need to rationally reconstruct causal notions in noncausal terms. Without this verificationist and anti-"metaphysical" premise, the D-N account would be philosophically indefensible.

Humean Explanation and Evidence

I suggested above that the D-N account of explanation has three philosophical virtues: that it portrays the explanatory power of a theory as depending upon the more general theoretical setting in which it is applied, that it rests on a theory of causation appropriate to the philosophical tradition in which it arises, and that it portrays the methodological preference for explanatory theories as a special case of the same methodological principle that dictates a preference for theories whose predictions have been observationally confirmed. We have just seen that the second of these claims is true, that at least for the central

case of causal explanations the D-N account *is* the Humean definition of explanation. The first is true just because the D-N account was understood in light of the principle of "unity of science" according to which a variety of different well-confirmed theories may legitimately be employed conjointly in making observational predictions. This principle is exactly the principle that entails the employment of "auxiliary hypotheses" in deducing the observational predictions that are to be tested in order to confirm or disconfirm a proposed theory. It is worth remarking that the "unity of science" principle is ineliminable if the D-N account of explanation is to be even remotely plausible. Even in the most typical and straightforward cases of causal explanation it is usually true that the event explained will not be retrodictable from the primary explanatory theory unless additional well-confirmed theories are also employed as premises. This point is as unchallenged in the empiricist literature as the corresponding point about the necessity of "auxiliary hypotheses" in the testing of theories.

Let us now turn to the third of these features. When an event is explained, the theories that are said to be explanatory on the D-N account are those that are employed in the retrodiction of the event. Thus, every successful explanation of an observable event has just the same logical form as a successful observational test of the relevant explanatory theories. This happy result is no surprise. It is characteristic of Humean conceptions of causation that the occurrence of a cause followed by its effect should be an instance of, and thus evidence for, the law or regularity whose existence is asserted by the appropriate "rationally reconstructed" causal statement. It would thus appear that the D-N account of explanation solves the epistemological puzzle about the evidential relevance of explanatory power as a nonexperimental criterion of theory acceptance. Really, according to the D-N account, the methodological preference for explanatory theories is not a nonexperimental criterion. Instead it is the special case of the criterion that dictates preference for experimentally tested theories—the case that applies to experimental (or observational) evidence whose epistemic relevance is recognized only after the relevant observation has been made. For one of the nonexperimental criteria, at least, the puzzle appears to be resolved. The elegance of this proposed solution is surely one of the most attractive features of the D-N account.

Nevertheless, it is extremely important to recognize that—even by empiricist standards—the D-N solution to the puzzle of the evidential role of explanatory power is incomplete. Recall that explanatory power is only one of a number of apparently nonexperimental criteria of theory acceptability. Even if it should turn out that the explanatory power of a theory is just a matter of its experimental confirmation by belatedly recognized observational tests of its own predictions, the fact remains that there are some genuinely nonexperimental criteria that are central to scientific methodology. This must be the case, since, as Goodman demonstrated, judgments regarding the confirmation of theories require prior (though perhaps tentative and revisable) judgments of their projectability. The genuine nonexperimental criteria are just those that legitimately play a role in projectability judgments. It follows, of course, that no account of the epistemology of science that does not say something about the epistemic (as opposed to the purely pragmatic) role of the genuine nonexperimental criteria can be complete. What is striking is that the D-N account presupposes an appropriate solution to this problem because the notion of projectability or lawlikeness is appealed to in the very formulation of the D-N account. There are two possibilities: Either judgments of lawlikeness are simply judgments of explanatory potential (a plausible enough view in light of actual scientific practice) or there are additional or different components of such judgments. In the first case, the D-N account of explanation cannot be a complete account of the epistemic role of judgments of explanatory power, since it presupposes a nonexperimental role for such judgments. In the second case, the D-N account succeeds in the project of providing a Humean anti-metaphysical analysis of the epistemic role of such judgments only if a similarly Humean reconstruction is possible for the genuinely nonexperimental criteria. In either case, the view that the D-N account of explanation succeeds in offering an account of the epistemic role of judgments of explanatory power presupposes the possibility of providing a similarly Humean account of whatever nonexperimental criteria of theory acceptability there are.

Thus, providing Humean accounts of the evidential relevance of nonexperimental criteria and of explanatory power are not two independent tasks of empiricist philosophy of science; success in the former is a prerequisite for success in the latter. Once this fact is recognized, one can see that there is a significant *prima facie* difficulty in the traditional empiricist program. The traditional empiricist treatment of nonexperimental criteria other than explanatory power has been to treat such criteria either as purely pragmatic and thus epistemically irrelevant (as is typical when such criteria are described as "simplicity" or "parsimony") or as purely syntactic and thus conventional (as in

the case of typical treatments of lawlikeness or projectability). As we have seen, there is no reason to believe that the epistemic contribution of nonexperimental criteria to the instrumental reliability of scientific methodology can be accounted for on solely pragmatic or conventionalistic grounds. The question is thus raised whether an adequate treatment of nonexperimental criteria is possible within the empiricist tradition, as the D-N account of explanation requires. In the next section I shall argue that the answer is no.

Toward a Non-Humean Alternative

Why a Non-Humean Alternative Is Needed

There is an extraordinarily rich and interesting literature in which various versions of the D-N account of explanation are criticized and defended with respect to their applicability to a wide range of kinds of explanation. What is characteristic of this literature is that philosophers have debated the applicability of the D-N account with respect to particular examples of scientific, historical, or psychological explanations that might be thought to resist subsumption under the D-N conception. Many of the criticisms of the D-N account represented in this literature are extremely important, as are many of the replies in its defense. Nevertheless, what I propose to do here is not to review this important literature but instead to argue directly against the D-N account on the grounds that the Humean definition of causation—which is its only philosophical basis—can now be seen to be wholly inadequate. I propose to adopt this strategy for two reasons. In the first place, it seems to me that the fact that for causal explanations the D-N account is an utterly straightforward application of the Humean definition of causation means that, unless a critique of the Humean definition is developed, the effectiveness of any criticisms of the D-N account for such cases will necessarily be reduced in the light of the support the D-N account receives from so well established a philosophical doctrine as the Humean definition. Moreover, it seems to me that recent criticisms of empiricist philosophy of science (including, of course, criticisms of the D-N account) have permitted us to develop enough anti-empiricist insights in the philosophy of science that a useful direct criticism of the Humean roots of the D-N account is now possible. What I propose to do in the remainder of this section is offer two sorts of criticisms of the Humean definition of causation (one more technical and the other more epistemological), propose and defend alternative conceptions of causation and of explanation that are in the tradition of "scientific realism" rather than in the tradition of logical empiricism, and indicate briefly how the proposed alternative conception of explanation would apply to the problem of explaining the epistemic role of explanatory power in scientific methodology. Since my aim is to indicate how recent critiques of logical empiricism and its variants can be extended to a treatment of the issue of explanation, I will rely heavily on recent work (including some of my own) that is critical of empiricism. I will usually sketch the main philosophical arguments involved, but I will not attempt to defend in detail the anti-empiricist positions upon which my critique of the Humean conceptions of causation and explanation depends. The present work is intended as a contribution to a developing realist critique of empiricism, not as an entirely self-contained refutation of the empiricist conception of explanation.

A Technical Criticism

Suppose that *L* is a set of strictly deterministic natural laws that hold in some possible world, C is a specification of initial conditions in that world at some fixed time, and *e* is an event subsequent to that time such that for systems governed by *L* an outcome just like *e* is necessary whenever initial conditions satisfying the specification obtain. It is part of any reasonable conception of causation—certainly it is part of any typical empiricist conception—that the conditions satisfying C (or, perhaps, some proper subset of them) constitute the total cause of *e*. Thus, on the Humean definition of causation, it should be true that the occurrence of *e* will be deductively retrodictable from *L* together with C. But this need not be the case. If the determining function defined by *L* is not general recursive in finitely many additional variables (representing physical constants), then it will certainly not be the case that such deductive retrodictions are always possible, since if they were they would provide a general recursive computation procedure for the determining function. (I am here assuming that *L* is itself recursively specifiable, and I am ignoring a more complicated possible case involving infinitely many physical constants for which a similar result can be obtained; see Boyd 1972 for the details.) Possible laws with this embarrassing property exists, and there is no general reason to suppose that the fact that the determining functions they define are not effectively computable for all possible initial conditions precludes their experi-

mental confirmation (again, see Boyd 1972). Thus, the Humean conception of causation is not universally applicable, even in those cases in which discoverable deterministic laws are governing. Exactly the same difficulties can arise in cases in which nondeterministic statistical laws govern the world in question. Thus, neither the contemporary Humean conception of causation nor any of the natural modifications of it that are appropriate for statistical laws can possibly represent a scientifically appropriate analysis of the concept of causation.

It might be thought that this difficulty could be remedied by taking the Humean analysis to rest upon the empirical assertion that the actual laws of nature define general recursive determining functions (or their statistical analogs). Of course this response would be entirely out of character with traditional empiricist philosophical methodology, which sought to provide rational reconstructions of scientific concepts that were justifiable *a priori*, but there is a long, if poorly developed, tradition of philosophical naturalism within empiricism, so we should certainly consider whether this particular appeal to the empirical might salvage the contemporary empiricist analysis of causation. It is very doubtful that it can. The reason is twofold. In the first place, even if it were demonstrated that the currently accepted laws of nature define determining functions that are general recursive (which may well be true), the well-established truism that all the laws we currently accept are likely to be only approximately true would prevent our immediately concluding that the true laws of nature have this property. It might seem that this difficulty could be overcome simply by inferring from the recursiveness of the determining functions defined by the currently accepted laws that the true natural laws probably also define recursive functions. This piece of inductive inference is, however, quite dubious methodologically. It is extremely doubtful that the hypothesis that the determining function defined by natural laws is recursive is itself a projectable hypothesis; it is doubtful, in other words, that one can reasonably infer this conclusion from the premise that various approximately correct "laws" define recursive determining functions. This is so because recursive determining functions are rather rare (if we treat physical constants as variables and thus count the number of basic forms of such laws, then there are only countably many of them) and each of them is approximated arbitrarily well by a continuum of extremely well-behaved (say infinitely differentiable) nonrecursive functions. Moreover, there is no reason to believe that the computability of determining functions has any physical significance. It is thus extremely risky to take the approximate truth of laws with recursive determining functions as even *prima facie* evidence that the true laws of nature specify such functions.

A somewhat more promising response might be to propose a natural revision in the formulation of the Humean definition itself. Suppose that, instead of requiring that the occurrence of the effect be deductively predictable from the laws together with relevant specifications of initial conditions, we require that statements completely specifying the effect hold in all of the intended models of the laws together with the specifications of initial conditions. At least for a great many cases (including all those discussed in Boyd 1972), a suitable notion of "intended interpretation" is available, and the Humean definition so modified will therefore identify causal relations in the appropriate way. The difficulties with this proposal are philosophical rather than mathematical. In order for the proposed definition of causation to be Humean, it would have to be the case that—on the conception of cause it advances—causal relations could be discerned in nature without recourse to knowledge of unobservable "theoretical entities" and their causal powers. There are extremely good reasons to believe that this is not so. The question of the confirmation of laws not all of whose observational consequences are deducible from them raises, in an especially clear way, a general problem in the experimental confirmation of theories. In general, when we accept the observational confirmation of finitely many of the infinitely many observational predictions of a theory as constituting sufficient evidence for its tentative confirmation, we are tacitly relying on some solution to what might be called the general problem of "sampling" in experimental design. By this I mean the problem of deciding, for any particular proposed theory, which reasonably small finite subsets from among the infinite set of observational predictions it makes are "representative samples" in the sense that observational confirmation of all their members would constitute good evidence for the approximate truth of the rest of the theory's observational consequences. The significance of this problem is especially easy to see when we consider the special case of theories not all of whose intended observational consequences are computationally available. In order to confirm such a theory, we would have to assure ourselves that from among the computationally available predictions of the theory a suitable representative sample can be formed.

I have argued (Boyd 1972, section 7) that for some theories with some computationally unavailable consequences it would be possible

to reliably identify such representative samples by employing available theoretical knowledge of (typically unobservable) underlying mechanisms to determine under what various sorts of conditions the theory would be likely to fail, and by finding computationally available predictions of the theory regarding conditions of these various sorts. Such a strategy would permit confirmation of such theories even though the determining functions they define are not recursive, but it would not do so within the constraints required by a Humean conception of scientific knowledge. Prior theoretical knowledge of underlying unobservable causal mechanisms would be essential for the confirmation of such theories. Thus, the revised definition of causation we are considering would fail to be Humean, in that it would not portray causal knowledge as independent of knowledge of unobservable causal factors.

I have also argued (Boyd 1973, 1979, 1982, 1983, and especially Boyd forthcoming) that—for theories in general and not just for those with nonrecursive determining functions—no alternative to this procedure for solving the problem of sampling exists. I conclude, therefore, that the second response to the primarily technical objection to the Humean definition of causation also fails. The technical criticism is apparently successful.

It is, nevertheless, a good methodological practice in philosophy to be cautious in accepting primarily technical criticisms of broadly significant philosophical theses. Such theses often admit of unanticipated reformulations that are sufficient to avoid particular technical criticisms. The more important criticisms are those suggesting that the thesis in question rests upon a fundamental philosophical mistake. Such a criticism of the Humean definition is suggested by the epistemological rebuttal just offered to the possible revision we were considering. If the methods of actual scientific practice for resolving questions about sampling in experimental design rely upon prior (approximate) theoretical knowledge of unobservable causal factors, then, in particular, knowledge of such factors is actual and therefore possible. Thus, the empiricist conception that experimental knowledge cannot extend to unobservable causal powers and mechanisms must be mistaken and the philosophical justification of the Humean definition of causation rests upon a false epistemological premise. There is indeed considerable evidence that almost all the significant features of the methodology of recent science rest ultimately upon knowledge of unobservable causal powers and mechanisms (see Putnam 1975a, 1975b; Boyd 1973, 1979, 1982, 1983, forthcoming), and thus that the empiricist reservations about experimental knowledge of unobservable causal powers and mechanisms are profoundly mistaken.]In the next section, I will explore in greater detail the consequences for the Humean conception of causation and explanation of the failure of the empiricist conception of experimental knowledge.

The Epistemological Inadequacy of the Humean Definition

The Humean definition of causation and the associated D-N account of explanation require acceptance of the "unity of science" principle and presuppose a Humean (that is, nonrealistic) but nevertheless epistemic account of the nonexperimental criteria of theory acceptance that determine judgments of lawlikeness and projectability. In fact, no satisfactory epistemological account of the "unity of science" principle is compatible with the empiricist's denial that we can have knowledge of unobservable causal powers and mechanisms, and no Humean account of lawlikeness and projectability can be epistemologically adequate.

Consider first the "unity of science" principle. Neither the Humean definition of causation nor the D-N account of explanation is even remotely plausible unless it is understood that the laws under which the caused or explained event is to be subsumed can be drawn from several different scientific disciplines or subdisciplines and conjointly applied in predicting an event. It must be possible for two laws that have been quite independently confirmed by specialists working in different areas to both be premises in the sort of deductive prediction to which the Humean definition and the D-N account refer. Moreover, it must of course be epistemically legitimate that independently confirmed laws be conjointly applied in this way to make observational predictions. After all, the point of the Humean conceptions we are considering is to reduce knowledge of causal relations and of explanations for events to knowledge of predictable regularities in the behavior of observable phenomena. Only if predictions obtained in accordance with the "unity of science" principle are epistemically justified would beliefs about the behavior of observables established by the sorts of deductive prediction we are considering constitute knowledge.

There is a further Humean requirement that applications of the "unity of science" principle must meet if the Humean conceptions of causation and explanation are to be justified. We have already seen that the

Humean conceptions are philosophically untenable if judgments of lawlikeness or projectability involve knowledge of unreduced causal factors. In a similar way, the Humean conceptions would be philosophically untenable if the applications of the "unity of science" principle upon which their plausibility depends themselves presupposed knowledge of unobservable causal factors. In fact this proves to be the case. The argument (see Putnam 1975b; Boyd 1982, 1983, forthcoming) can be summarized as follows: The principle that independently confirmed theories can legitimately be conjointly applied in making predictions about observables must presuppose some sort of judgments of univocality for the nonobservational (or "theoretical") terms occurring in the theories in question. Without the requirement that all the theoretical terms occurring in the conjointly applied theories occur univocally in their conjunction, the "unity of science" principle would dictate the absurd conclusion that one should expect approximately true observational predictions from the conjunction of well-confirmed physical theories of force together with well-confirmed theories of the role of force in international affairs even if one does not disambiguate the various occurrences of the lexical item "force" in the conjunction. Moreover, the principle for assessing univocality cannot be that theoretical terms are nonunivocal whenever they occur in different theories; such a principle would result in a very significant underestimation of the scope of the "unity of science" principle in actual scientific practice and therefore would be inappropriate for the defense of the Humean conceptions we are considering.

Once it is recognized that theoretical terms from quite different theories are sometimes to be counted as occurring univocally, it becomes clear that the "unity of science" principle makes a striking epistemological claim. Suppose that T_1 and T_2 are two theories from quite different scientific disciplines in whose conjunction no theoretical terms occur ambiguously. Suppose further that in the experimental confirmation of these theories neither was ever employed as an "auxiliary hypothesis" in the testing of the other. There will thus have been no *direct* experimental test of the conjunction of the two theories, except insofar as the predictive reliability of each of them taken independently has been tested by prior experiments. Nevertheless, the "unity of science" principle maintains, we are justified in expecting the conjunction of the two theories to be instrumentally reliable even in the absence of direct experimental tests, provided only that the univocality constraint on their constituent theoretical terms is satisfied. The univocality constraint is thus supposed to do real epistemic work in making possible what may be thought of as the *indirect* confirmation of the instrumental reliability of the conjunction of T_1 and T_2.

A good way of seeing what is going on is to consider what an empiricist might plausibly take the confirmation of a theory to amount to. Since no knowledge of theoretical entities is supposed to be possible, it would be initially natural for the empiricist to hold that when a theory is confirmed all that is confirmed is the approximate instrumental reliability of the theory itself. Recognition of the crucial role of auxiliary hypotheses in the testing of theories suggests replacing this instrumentalist conception with a broader one according to which the experimental confirmation of a theory amounts to the confirmation of the conjoint reliability of the theory together with the other theories that have been employed as auxiliary hypotheses in testing it. The "unity of science" principle requires a much broader conception. Experimental confirmation of a theory is supposed to constitute evidence for its instrumental reliability even when it is applied conjointly with other well-confirmed theories not even discovered at the time the evidence for the first theory was assessed! Something over and above the instrumental reliability of the conjunction of the theory with actually employed auxiliary hypotheses—something over and above even the instrumental reliability of the theory taken conjointly with currently established theories—is supposed to be confirmed when the theory is properly tested. That "something," the knowledge over and above the instrumental knowledge that has been directly confirmed, is represented in the theoretical structure of the theory, and the rule for extracting it is to make deductive predictions from the theoretical sentences in the theory in question together with the theoretical sentences that represent the similar "excess knowledge" in other well-confirmed theories. There is no plausible explanation of the instrumental reliability of this sort of instrumental knowledge-extraction procedure other than that provided by a realist conception of theory confirmation according to which confirmation of theories involves confirmation of the approximate truth of their theoretical claims as well as their observational ones. On such a conception, judgments of univocality for theoretical terms are judgments of co-referentiality, and what the "unity of science" principle licenses is deductive inferences from the partly theoretical knowledge embodied in independently tested theories to conclusions about the behavior of observables. Univocality judgments are crucial in establishing that the nonobservational subject matter of the two theories is

really the same when it appears to be. (As a matter of fact, the situation is even more complicated: The unity of science also involves inductive inferences from theoretical knowledge. This additional consideration strengthens the case for a realist construal of both theory confirmation and univocality judgments; see Boyd forthcoming.)

It follows not only that knowledge of unobservable causal factors is possible but also that it is presupposed by the "unity of science" principle. The principle is tenable only on the assumption that knowledge of theoretical entities is possible, and it presupposes that the univocality judgments for theoretical terms that scientists actually make are reliable judgments about reference relations between theoretical terms and theoretical entities. This, in turn, requires that scientists have reliable knowledge of causal relations between unobservable causal factors and their own use of language. The "unity of science" principle thus presupposes just the sort of knowledge that the Humean conceptions are designed to "rationally reconstruct" away, and the Humean conceptions are thus philosophically indefensible.

Similar arguments show that judgments of lawlikeness and projectability are likewise infected with essentially non-Humean commitments to knowledge of unobservable causal factors.[7] I have argued for this and related claims elsewhere (Boyd 1973, 1979, 1982, 1983, forthcoming); the basic argument can be summarized as follows.

We have seen that the solution to the problem of sampling in experimental design in mature sciences presupposes prior knowledge of unobservable "theoretical entities" or causal factors. In fact, the solution to this problem is intimately related to the solution to the problem of projectability. Roughly, theories are projectable just in case there is some *prima facie* reason to believe that they might be (approximately) true and thus some reason to treat them as live candidates for confirmation by observational evidence. The methodological rule for the solution to the sampling problem is this: Test a proposed theory under circumstances representative of those identified by other projectable theories about the same issues as those under which its predictions are most likely to be wrong. The theory-dependent judgments that go into solving the problem of sampling are just special cases of judgments of projectability.

In fact, as Kuhn (1970) has correctly maintained, this pattern of the dependence of scientific methodology on the ontological picture presented by the received theoretical tradition infects all the important principles of scientific methodology. For example, another important question in experimental design is what factors must be controlled for in setting up the experimental conditions. There are infinitely many factors about which it is logically possible that they could interfere with the intended functioning of experimental apparatus. We identify the relatively few factors that must be controlled for by applying our existing theories of underlying (and typically unobservable) causal mechanisms to identify those sorts of interference that it is reasonable to believe might operate in the relevant experimental conditions. Here too the methodological principle we employ is very intimately connected with judgments of projectability. To a very good first approximation, the rule we employ is that factors should be controlled for that are suggested by those logically possible theories that are themselves projectable. Again our judgments of projectability turn out to be theory-dependent judgments relying on the accounts of unobservable causal factors represented by our best confirmed theories.

For each of the theory-dependent principles of scientific methodology we can ask what explains its contribution to the instrumental reliability of scientific practice. In each case, the only plausible explanation is given by the realist conception that in making such judgments we employ the approximate knowledge of observable and unobservable causal factors reflected in existing theories in order to establish methods for improving our knowledge of both observable and "theoretical" entities. Theoretical understanding of unobservable causal factors enjoys a dialectical relationship with the development and improvement of methods for improving theoretical understanding itself. In particular, judgments of projectability require knowledge of unobservable causal factors. Thus, the appeal to projectability in the Humean definition of causation deprives that definition of its Humean content and hence of its only philosophical justification.

It will be useful to consider one important rebuttal to the position I have just taken. In "Natural Kinds" (1969), Quine sometimes describes the natural kinds in mature sciences as issuing from theory "full-blown." When Quine writes in this way, his account of natural kinds (and thus of projectability) seems very close to the realist and anti-Humean conception just discussed. In other places he seems to prefer to treat the identification of projectable theories or predicates as involving "second-order induction about induction." He says: "We establish the projectability of some predicate, to our satisfaction, by trying to project it. . . . In induction, nothing succeeds like success" (p. 129). This formulation suggests that projectability judgments might be thought of as *a posteriori*

judgments involving consideration only of observable phenomena. After all, the instrumental reliability thus far displayed by some particular inductive strategy (with its particular judgments of projectability) is an observable phenomenon, and Quine's suggestion appears to be that we can (at least tentatively) identify projectable theories or predicates by looking at which ones have figured in successful inductive inferences in the past. No consideration of unobservable phenomena appears to be involved.

It is important to see what the philosophical consequences would be if this conception of projectability judgments could be sustained. We have already seen that projectability judgments play an essential epistemic role in establishing the instrumental reliability of scientific methodology, and that therefore it is not adequate to treat projectability judgments as purely conventional or to offer a purely pragmatic account of their rationale. As Quine and Goodman both recognized, projectability judgments must have some sort of empirical basis in order for their epistemic role to be explicable. The proposal that Quine appears to be making would, if it were successful, provide an adequate account of that epistemic role without invoking knowledge of unobservable causal factors. The Humean conceptions of causation and explanation would therefore succeed in offering a reductive analysis of causal notions. Moreover, a nonrealist account of the epistemic role of projectability judgments would undermine the arguments rehearsed earlier in this section to the effect that experimental knowledge of unobservable causal factors is possible. The Humean definition of causation and the D-N account of explanation would indeed be philosophically justified, and the project of the present essay would be misconceived.

I have discussed the second-order induction about induction interpretation of projectability at length elsewhere (Boyd 1972, section 2.3; forthcoming, part III; and especially 1983, section 8). Roughly, the flaw in the proposal lies not in the claim that projectability judgments can be thought of as a species of second-order induction about induction but rather in the conception (which may not have been Quine's) that such inductions are independent of knowledge of unobservable causal factors. The problem is that such inductions—like all inductions—depend upon projectability judgments, and the projectability judgments upon which they depend involve just the appeals to knowledge of unobservable causal factors that (the version we are considering of) second-order induction about induction is supposed to eliminate.

The simplest among several ways to see this is to realize that the inductive inferences about induction we are considering are supposed to take as premises claims that previous inductive inferences guided by certain standards of projectability have been largely successful and to conclude that future inductive inferences guided by the same standards will also tend to be successful. The premises of such inferences admit of two interpretations. On one reading, the premises state that, generally speaking, the theories that have been accepted in accordance with the relevant conception of projectability are instrumentally reliable; they make approximately true predictions about the behavior of observable phenomena. On this reading, knowledge of the premises requires knowledge about various theories that their observational predictions—future as well as past, untested as well as tested—are approximately true. On the second and weaker reading, all the premises state is that it has been largely true that those observational predictions that follow from theories accepted in accordance with the relevant conception of projectability and have been subjected to experimental test have proved approximately true. Only on the weaker reading do the premises report observed facts.

Pretty plainly, the actual structure of the inference from observable data to the conclusion that the relevant conception of projectability is epistemically reliable must be thought of as proceeding in two steps. From the information provided by the premises on the weaker reading (from the predictive successes in actually tested cases of the theories in question) we first infer the instrumental reliability of a suitable number of the relevant theories. These conclusions about instrumental reliability, together with the presumably uncontroversial premise that the standards of projectability in question dictated the acceptance of those theories, then confirm the premises we are considering on their stronger reading. It is from those premises on their stronger reading that we reasonably conclude (in the second step) that the relevant conception of projectability is epistemically reliable.

There is nothing whatsoever wrong with this pattern of inductive inference about inductive procedures. Indeed, reliance on just such inferences is essential not only in science but also in the practice of epistemologists of science. It is true, however, that the first step in this inference consists simply in a number of cases of inferring the instrumental reliability of a theory on the basis of experimental evidence. It is just this sort of inference that as we have seen, depends upon projectability judgments grounded in knowledge of unobservable causal factors. Second-order induction about induction thus presupposes such knowledge and cannot form the basis for a Humean reconstruction of projectability judgments.

The Humean definition aims to rationally reconstruct causal notions in noncausal terms. The philosophical justification for this project rests upon the epistemological premise that experimental knowledge of unobservable causal factors is impossible. The epistemological premise is false, and the rational reconstruction is in any event unsuccessful. There is no reason to accept the Humean definition. The D-N account of explanation is—for those cases in which it is most plausible—simply an application of the Humean definition, and thus it is also without philosophical justification. There is no reason to reject the preanalytic conception that, for a wide and central class of cases, to explain an event or a recurring phenomenon is to say something about how it is caused. Nor is there any reason to think that the empiricist analyses of causation and explanation rest on, or provide, an even approximately accurate conception of the nature of causal knowledge in science or in any other area of inquiry.

The Semantics of "Cause" and "Explain"

It is a puzzling fact that many philosophers who reject the empiricist conclusion that knowledge of unobservables is impossible and who are sympathetic to scientific realism rather than to verificationism or instrumentalism nevertheless employ the Humean definition of causation or the D-N account of explanation, even in cases in which the phenomena caused or explained are unobservable. The principal explanation of this phenomenon, I believe, is that the Humean definition and the D-N account were so widely accepted during the time when empiricism dominated the philosophy of science that they now have the status of established philosophical maxims whose initial justification has been forgotten. The fact that the rejection of an empiricist conception of experimental knowledge in favor of a realist conception leaves these positions without any philosophical justification may have gone largely unnoticed. I am inclined to think, however, that there is an additional explanation for the durability of these two pieces of philosophical analysis in the face of widely accepted criticisms of their empiricist foundations. I think that philosophers may believe that we need to have some analysis of the meaning of terms such as "cause" and "explain" and that the definitions that arise from the Humean tradition may serve as good first approximations to such meaning analyses. Of course, that such definitions fail to be reductive and thus fail to meet empiricist standards would not, by itself, show that they are inappropriate as nonreductive philosophical analyses.

It nevertheless seems plain that the Humean definitions are strikingly inadequate. In the case of the Humean definition of cause, what seems to be the primary causal notion gets defined in terms of the highly derivative causal-epistemological notion of projectability. Instead, it would seem that the revealing definition would go more nearly in the opposite direction, projectability being defined in terms of knowledge of causal powers and mechanisms. The D-N account of explanation of course inherits these difficulties; moreover, there are notorious difficulties in assimilating clear-cut cases of explanation to the D-N model. It would appear that neither of these definitions is a very promising beginning for a philosophical analysis of the relevant concept. It might be thought that they have the advantage of reminding us that our concept of causation is related to something like a conception of determinism. (It had better not be exactly like one if our conception is to correspond to something real; that is why there are statistical versions of both definitions.) The D-N account of explanation might also be thought of as setting a standard for complete explanations appropriate to the conception that something like determinism is involved in causation.

Against these claims one may reply, first, that the technical criticism of the Humean definitions presented above show that in any event they embody the wrong analysis of determinism (see Boyd 1972). Moreover, in any event we want the question of the relationship between causation and determinism to be spelled out by research in the various sciences and social sciences rather than by so abstract a definitional specification. In a similar way, we should want the relevant methodological standards of completeness of explanations to be determined (in a theory-dependent way) by the aims and the findings of the various special sciences. (Indeed, it is difficult to imagine what scientific activity would, even with a suitable idealization, require us to seek explanations complete in the D-N sense even in a deterministic world.) Finally, as we shall see in the next section, in order to account for the evidential import of explanatory power we need not assimilate explanation to the sort of retrodiction provided by "complete" D-N explanations.

But, someone might ask, if the Humean definitions of causation and explanation are rejected in the name of scientific realism then what does the realist propose as an alternative account of the semantics of these and other causal notions? The answer dictated by the realist considerations offered here has several components. First, of course, it is not to be expected that any significant causal notions are adequately definable in noncausal terms. That is just what the critique of the Humean definition establishes. Second, it is quite doubtful that there are philosophically interesting analytic definitions of scientifically important causal notions, even in terms of each other. As the change in our conception of the relation between causation and determinism induced by the acceptance of quantum mechanics indicates, there is no reason to believe that proposed philosophical analyses or definitions of causal notions will be immune in principle from amendment in the light of new theoretical discoveries.

It is nevertheless clear that informative philosophical analyses of many causal or partly causal notions are possible. I take it that such analyses are in some sense empirical because they depend upon empirical facts about causal phenomena and about our practices regarding them and because they are revisable in the light of new discoveries in these areas. Nevertheless, they appear to lie squarely within the province of philosophy. Analyses of such causal notions as explanation, projectability, reference, and knowledge I take to be in this category. About the less derivative causal notions, such as "(total) cause," "causal power," "interaction," "mechanism," and "possibility," it seems less likely that informative analyses of the sort that philosophers typically seek are available; it might be that "cause" and "causal power" are somehow interdefinable, but it is doubtful that whatever definition might be available would prove very informative to someone who wanted to know, e.g., what causal powers are. Informative definitions or analyses in these latter cases, I suggest, are not primarily a matter of conceptual analysis, even on the understanding suggested above according to which conceptual analysis is a kind of empirical enterprise. Instead, the informative analyses or definitions of more basic causal notions are to be established by theoretical inquiry in the various sciences and social sciences. What causation is and what causal interaction amounts to are theoretical questions about natural phenomena (to reject the Humean project is just to admit that causal relations, powers, and interactions really are features of nature), so it is hardly surprising that answers to them should depend more upon the empirically confirmed theoretical findings of the various sciences than should answers to more abstract (and more typically philosophical) questions about the nature of knowledge, reference, or explanation.

The distinction between the two sorts of questions is one of degree. "Conceptual analysis," when done well, has an ineliminable empirical component, and the more foundational questions in the various sciences are typically philosophical questions as well, often requiring the special analytical techniques of philosophy for their resolution. But even though the distinction is one of degree, the fact that definitional questions about fundamental causal notions fall on the side nearer to the various empirical sciences dictates an important conclusion: that such notions as "(total) cause," "causal power," and "interaction" are like the notions of various natural kinds in that they possess no analytic definitions, no "nominal essences." They are defined instead by natural definitions or "real essences" whose features are dictated by logically contingent facts about the way the world is. From a realist perspective, this is hardly surprising. Natural kinds and categories lack stipulative *a priori* definitions precisely because, in order to play a reliable role in explanation and induction, natural kinds and categories must be defined in ways that reflect the particular causal structure the world happens to possess (Putnam 1975a, 1975b; Boyd 1979, 1982, 1983, forthcoming). For exactly the same reason, of course, the definitions of our causal notions must also reflect *a posteriori* facts about the nature of causation.

It follows that the reference of terms referring to fundamental causal notions is not fixed by analytic definitions; there are none. Instead, such terms are like natural-kind terms, theoretical terms in the particular sciences, and other terms with "natural" rather than analytic definitions, in having their reference fixed by epistemically relevant causal relations between occasions of their use and instantiations of the causal phenomena to which they refer. (For discussions of the epistemic character of reference see Boyd 1979, 1982.) It follows that in order to account for the semantics of causal terms we need no such analyses of their meaning as the Humean definitions provide. To hold that some largely *a priori* conceptual analysis must provide definitions for such terms is to fall victim to an outmoded empiricist conception of the semantics of scientific and everyday language.

Explanation and Evidence

At least for many central cases (and for the cases the D-N account is designed to fit), an explanation of an event is an account of how it was caused. In all but the most atypical cases the account will be partial: Not all the causally determining factors will be indicated, nor will the relevant mechanisms be fully specified. The D-N account is typically extended to cover the cases of explanations for laws or regularities in nature. On the D-N conception, to explain a law or a regularity is to deduce the law or a statement of the regularity from other laws, together with statements of appropriate boundary conditions. It is not entirely clear that this standard extension of the D-N account is really appropriate to the Humean task of reducing causal notions to noncausal ones. One should argue that, inasmuch as the possible knowledge reflected in the explained theory is supposed to be exhausted by its observational predictions, all the consistent Humean should require by way of an explanation is the deduction of those observational consequences from the explaining theory. In any event, the realist conception of explanation also generalizes (even more naturally) to cases of the explanation of laws or regularities: To explain a law or a regularity is to give an account (presumably partial) of the causal factors, mechanisms, processes, and the like that bring about the regularity or the phenomena described in the law.

It is a consequence of the Humean account that all explanations of particular events have a certain level of generality "built in" in virtue of reference to the relevant laws; of course, this is just what the Humean conception of causation requires. The realist conception of causation and explanation does not rule out the possibility of singular causal relations that are not instances of more general patterns; it leaves such issues to the findings of the various special sciences. Nevertheless, it does appear that, given what we know about causal relations and about the sorts of causal explanations that are actually discovered, the Humean conception is in this respect right or very nearly right. Scientific explanations of individual events do, almost always, extend to cover similar cases, actual and counterfactual. In consequence, our conception will be appropriate for the central cases of causal explanation if we think of explanations as being provided by small theories describing the causal factors that determine, or the causal mechanisms or processes that underlie, some class of phenomena. I will use the term *explanation* to refer to such theories. Explanations will of course differ considerably in the extent to which they are complete in their identification of causative factors, in the specificity with which they describe underlying mechanisms or processes, in the level of numerical precision with which they characterize the relations between such factors, and in other respects. Part of the task of a theory of causal explanations is to say how the epistemological significance of an explanation is influenced by factors such as these.

Suppose that a theory *E* is an explanation for some phenomenon *p*. It would be natural to understand the terms "explains" and "explanatory power" so that it is just *E* that is therefore said to explain *p* and so that it is just the explanatory power of *E* that is thereby demonstrated. Neither scientific usage nor scientific practice conforms to this picture. *E* might well be said to explain *p*, but scientific practice dictates our taking the explanatory success of *E* as grounds for saying of other more general theories that they explain *p*. Indeed, under the circumstances envisioned, we would not ordinarily speak of the explanatory power of *E* being manifested at all; instead, *E*'s being an explanation of *p* would ordinarily be taken to indicate the explanatory power of those other, more general theories and to provide evidential support for them.

Consider for example the explanation of the "law" of fixed combining ratios, according to which in a certain class of reactions chemical elements combine in fixed ratios by weight. An explanation is provided by a theory that says that this phenomenon is produced by the underlying tendency for atoms to combine in fixed ratios by number, together with the claim that atoms of an element all have the same weight. (I ignore here the issue of isotopes.) What chemists and historians of chemistry correctly say is that this explanation indicated that the atomic theory of matter could explain the phenomenon in question; it demonstrated the explanatory power of the atomic theory and thus provided evidence for it. Similarly, consider the occurrence of subcutaneous degenerate hind limbs in the larger constrictors. An explanation for this phenomenon is provided by a theory according to which these limbs are the vestigial remnants of the ordinary hind limbs of reptiles ancestral to the snakes, which were gradually lost through the process of natural selection. Insofar as this explanation is accepted, it indicates the explanatory power of the Darwinian conception of the origin of species and provides evidence for it.

In these cases we can see a pattern that is utterly typical. An explanation for a particular phenomenon will typically draw upon the resources of some more general theory. It will appropriate the theoretical resources of the broader theory (entities, mechanisms, processes, causal powers, physical magnitudes, and so on), and it will employ, and often elaborate upon, these resources in describing how the phenomenon in question is caused. The dependence of minor theories and explanations

upon the theoretical resources of larger theories has been amply documented by Kuhn (1970), who describes the dependence of research in "normal science" upon the ontological picture dictated by the most general theories in the theoretical tradition or "paradigm." (Note that in thus agreeing with Kuhn one need not accept the constructivist conception of scientific knowledge he so ably defends; see Boyd 1979, 1983.) Adequate explanations of particular phenomena are taken to be indicative of the explanatory power of the more general theories whose resources they exploit and to provide additional evidence for those theories. The evidential relevance of explanations does not depend upon its being possible to retrodict or deductively subsume the explained phenomena from the explanation itself or from the explanation together with the relevant general theory(ies) (together with auxiliary hypotheses). The examples of explanations just mentioned illustrate the last point. In the case of the explanation of degenerate limbs in the large constrictors (as in the case of almost all similar evolutionary explanations) we lack altogether the resources for retrodiction, but the fact that the Darwinian theory provides the resources for explanations in such cases properly counts as evidence for it nevertheless.

The case of the "law" of fixed combining ratios is more complicated, because the more precise formulations of the "law" were developed simultaneously with its explanation. Nevertheless, it seems clear that the capacity to predict previously unnoticed instances of the "law" or to deduce an adequate formulation of it emerged quite slowly as the explanation became more detailed, and probably not until Mendelev's work on the periodic table was anything remotely approximating the sort of explanation anticipated by the D-N model available. Despite this fact, it is also clear that the cogency of earlier versions of the atomic explanation, from Dalton's early-nineteenth-century work on, were rightly taken to indicate the explanatory power of the atomic theory of matter and to constitute some evidence in its favor.

There are indeed cases in which the explained phenomenon (or rather a statement describing it) is better thought of as a premise than as a conclusion in the testing of the theory that constitutes its explanation and thereby indicates the (evidentially relevant) explanatory power of the theory upon which it itself is based. The following sort of situation is commonplace (though perhaps not typical): Some general theory *T*, which is projectable (and thus already supported by some theory-mediated empirical evidence), postulates mechanisms of a certain sort as causally relevant in a broad class of related phenomena; however, *T* is not sufficiently well developed to permit prediction of such phenomena. *T* itself does not specify in detail the mechanisms underlying particular cases of the phenomena in question, but it does specify theoretically important general descriptions under which (if *T* is right) such mechanisms will fall. For some phenomenon *p* of the relevant sort, an explanation *E* is proposed. *E* says that *p* is produced by certain more precisely specified mechanisms of the general sort prescribed by *T*. Despite its greater precision, even *E* is inadequate (given available auxiliary hypotheses) to reliably predict occurrences of *p*. Nevertheless, it is predictable from *E* that if an instance of *p* occurs then various experimentally distinguishable symptoms of the operation of the mechanisms specified by *E* will be present. Experimental confirmation of *E* consists in producing or finding occurrences of *p* and testing for the relevant symptoms. The experimental test of *E* consists, not in finding occurrences of *p* where and when *E* predicts their occurrence (since it makes no such predictions), but in finding the relevant symptoms when and where they are predicted to occur, given *E* and the occurrence of *p* (together, presumably, with other auxiliary hypotheses) as premises! The success of such predictions tends to confirm *E* and, less directly, *T*.

Cases of this sort are routine where *T* is a general chemical theory about complex and predictively intractable reaction mechanisms, *E* is a proposed application of *T* to the case of a particular sort of reaction, and the symptom in question indicates the presence of a reaction byproduct that can, on the basis of well-established chemical theories, be taken to be distinctive of the particular reaction mechanisms postulated by *E*. What all the sorts of cases we have examined suggest is that when an explanation *E* of a phenomenon *p* provides evidence for a more general theory *T* by indicating that *T* has explanatory power, what is crucial is that *E* be testable largely independent of *T* and that the approximate truth of *E* constitute good reason for believing the approximate truth of *T*. It appears not to matter very much whether the occurrence of *p* is itself otherwise significantly confirmatory of *E* or *T*, much less whether it is predictable from *E* or *T*.

What we need to know is how this sort of confirmatory evidence for theories upon whose resources successful explanations are based is related to the sort of confirmatory evidence provided by the experimental confirmation of observational predictions made from the theories themselves. Understanding this will be easier if we understand better the confirmatory relationship between theories and those of their observational predictions whose experimental confirmation supports them.

In mature sciences all theory confirmation is theory-mediated. As we have seen, theories are not confirmable at all unless they are projectable, and projectability judgments are theory-dependent judgments of plausibility. The confirmation of an observational prediction of a projectable theory does not count significantly toward its confirmation unless theory-determined considerations indicate that it is a relevant test (that is, roughly, unless it tests the theory against a projectable rival). Particular experiments do not count as well designed (and thus are not potentially confirmatory or disconfirmatory) unless there are appropriate controls for the possible experimental artifacts that are indicated as relevant by previously established background theories. No piece of experimental evidence counts for or against a theory except in the light of theoretical considerations dictated by previously established theories. (For discussions of these and the following points see Boyd 1972, 1973, 1979, 1982, 1983, forthcoming.)

The theoretical considerations that thus bear on theory confirmation are themselves evidential considerations. The fact that a theory is plausible in the light of well-confirmed theories is evidence that it is approximately true. This is so because the evidence for the well-confirmed theories that form the basis for the plausibility judgment is evidence for their approximate truth (as scientific realists insist) rather than just for their empirical adequacy (as empiricists typically maintain) and because the inferential principles by which conclusions about the plausibility of proposed theories are drawn from the previously established theories are themselves determined by previously acquired theoretical knowledge. The dialectical development of theoretical knowledge and of methodological principles extends to the principles by which plausible inferences are made from wholly or partly theoretical premises to theoretical conclusions. Judgments of theoretical plausibility reflect inductive inferences at the theoretical level (that is, inferences from previously acquired theoretical knowledge to inductively justified theoretical conclusions). These inferences proceed according to theory-determined assessments of projectability, just as inferences from observational data to theoretical conclusions do. (See especially Boyd 1983, Boyd forthcoming.)

The evidence for a theory provided by its being plausible in the light of previously established background theories is every bit as much empirical evidence as is the evidence provided by experimental tests of the theory's observational predictions; the empirical basis for this evidence consists of the various observations involved in the confirmation of the relevant background theories. Call empirical evidence for a theory *direct* if it is provided by experimental tests of observational predictions drawn from the theory itself, and *indirect* if it is obtained by inductive inferences at the theoretical level from other theories that have themselves been confirmed by experimental tests. The important consequence of a realist conception of scientific epistemology is that the distinction between direct and indirect empirical evidence is of no fundamental significance. The observations that provide direct experimental evidence for a theory provide significant evidence at all only because of indirect evidential considerations in support of the theory itself (viz. the bases of the judgment that it is projectable) and in support of various other theories (those of its logically possible rivals also judged projectable, the theoretically plausible accounts of possible experimental artifacts, and so on). Thus, direct experimental evidence is only superficially direct. Moreover, indirect empirical evidence can be very strong evidence indeed. The fact that a theory provides theoretically plausible accounts of a very large number of phenomena in a way in which none of its plausible rivals do can, under appropriate circumstances, constitute genuinely confirmatory evidence for it even though almost no evidence for it is provided by direct tests of its observational predictions. This was the situation of the Darwinian theory of the origin of species until very recently, and it is the current situation of many astronomical theories. Observations can often provide striking confirmation of a theory not by confirming a prediction of the theory but by ruling out its theoretically plausible alternatives. It may be true, nevertheless, that in many important cases direct empirical evidence for a theory plays an especially important confirmatory role, and there may even be some general methodological reason why this should be so. But the distinction between direct and indirect evidence cannot be an epistemologically fundamental one. They are two closely related and interpenetrating cases of the same epistemological phenomenon.

It will now be clear, I believe, how the evidence for a theory that arises from a demonstration of its explanatory power is to be understood. Let T be a theory, E an explanation that draws upon the resources of T, and p the phenomenon that E explains. Evidence (direct or indirect) for E will demonstrate the explanatory power of T just in case (given the available background theoretical knowledge and the inductive standards it determines) the way in which E draws on the theoretical resources of T is such that E's being approximately true provides inductive reason to believe that T is also approximately true. It will not matter in any fundamental way whether or not the evidence for E includes successful prediction of p. In any event, the evidence thus provided for T will be a perfectly ordinary case of indirect empirical evidence of the sort we have just been examining. Assessments of explanatory power are just one species of assessment of indirect theory-mediated empirical evidence. There is nothing going on when we prefer explanatory theories over and above what goes on in all cases in which we prefer theories that are supported by (necessarily partly indirect) empirical evidence.

Let us return to the D-N account of explanation. Its three most attractive features were that it rested upon an appropriate account of causation, that it indicated that the explanatory power of a theory depends upon its integration into a larger body of well-confirmed theories, and that it portrays the preference for explanatory theories as a special case of the preference for theories supported by observational evidence. In all these respects the D-N account is right. What we have seen is that the weakness of the D-N account lies not in the unworkability of the above three features but in the mistaken Humean conceptions of causation and evidence upon which the D-N account rests. When we adopt a realist conception of causal relations and causal powers as real features of the world, an account of the integration of theories that countenances inductive integration of theoretical knowledge as well as conjoint deductive prediction, and an account of empirical evidence that recognizes the crucial methodological role of such inductive integration, we are able to preserve the best features of the D-N account while avoiding the insuperable difficulties to which empiricist accounts of scientific methodology invariably fall victim.

Other Epistemological Issues

Simplicity and Parsimony

Traditional logical-empiricist accounts assimilate the methodological preference for explanatory theories to a preference for empirically tested theories but typically treat the other nonexperimental standards for the acceptability of a theory as purely conventional or pragmatic. At least, that is the typical "official" empiricist position. In applied philosophy of science, logical empiricists often treated considerations of "simplicity" and "parsimony" as though they had evidential weight. In this latter case, I believe, empiricists were basically right.

One of the striking things about the methodological judgments philosophers of science assimilate to the categories of "simplicity" and "parsimony" is the extent to which they are more complex than those descriptions would suggest. Scientists do not, as a general rule, prefer the simplest from among the empirically unrefuted theories about some natural phenomenon. They quite often—and without any misgivings—reject theories as too simple (or perhaps as too simpleminded) even when they fit the data that have already been examined. There are whole disciplines in which "single-factor" theories are held up to methodological derision, and there are even more disciplines in which this would be true were single-factor theories seriously proposed. Similarly, the principle of parsimony, or Occam's razor, seems to be applied quite unevenly. In many fields, at particular moments in their histories, scientists quite cheerfully postulate new entities in order to account for new empirical discoveries rather than making other theoretical accommodations equally compatible with the data in question. What plainly happens in these cases is that theoretical reasons legitimate the unsimple or unparsimonious theoretical choices. Thus, judgments of simplicity and parsimony are—like judgments of explanatory power—theory-dependent.

We know, moreover, that if (as seems plausible) judgments identified as simplicity or parsimony judgments are important factors in judgments of projectability then such judgments cannot be merely conventional or pragmatic; they must play an epistemic role in scientific practice. What I suggest is that judgments of "simplicity" and of "parsimony" are simply special cases of judgments of theoretical plausibility. When a proposed theory assimilates new data into our existing theoretical framework via a modification that is (according to the evidential standards dictated by that framework) warranted by those data, we see the modification as a simple one (in the sense that it does not introduce epistemologically needless modifications into theories we already take to be well confirmed) and we somewhat misleadingly describe the theory itself as simple. Similarly, we reject proposed theories that accommodate new data by postulating theoretically implausible new entities, and we misleadingly characterize our preference as being for parsimonious theories in general.

If this suggestion is basically right (I invite the reader to consider various actual cases), then the methodological preferences we typically misdescribe in terms of simplicity and parsimony are simply special cases of the methodological preference for theories that are supported by inductive inferences at the theoretical level from the approximate theoretical knowledge we already have. But that principle, as we have seen, amounts to a preference for theories supported by indirect experimental evidence. In the case of the principles of "simplicity" and "parsimony" just as in the case of the principle that we should prefer explanatory theories, all that is really going on is a recognition of the role of indirect evidence in science. The nonexperimental criteria of theory acceptability, which initially appear puzzling, turn out to be nonexperimental only in the sense that they do not reflect the assessment of direct experimental evidence. The logical positivists were right in their applied philosophy of science when they took these principles to be evidentially relevant, but their anti-realist Humean conception of scientific knowledge prevented them from seeing why they were right.

Contexts of Discovery and Confirmation

I have suggested above that three characteristic features of the D-N conception of explanation that help to explain its philosophical plausibility actually represent important insights of the empiricist tradition in the philosophy of science—insights that can be extended to the cases of other nonexperimental criteria as well, but insights that an empiricist as opposed to a realist conception of scientific knowledge cannot successfully assimilate. The development of a consistent realist conception of the epistemic role of nonexperimental (or, better, indirectly experimental) criteria of theory acceptability permits us to examine the cogency of another distinctive feature of empiricist philosophy of science: the traditional logical-empiricist claim that the epistemology of science need concern itself with the logic of confirmation but not with the principles of reasoning by which scientific theories are invented or discovered. On the logical-empiricist conception, the latter issue belongs to psychology and to the social study of science but not to the philosophy of science.

Part of what empiricists meant when they held that issues about the context of discovery were irrelevant to the philosophy of science was that philosophers of science need not develop a formal inductive logic to account for the discovery of theories. No doubt they were right in this respect; there is no reason to believe that what is ordinarily meant by an inductive logic would provide an even remotely adequate account of theory discovery in science. They also meant that philosophers of science need not concern themselves with all the details of psychological theories about how theories are discovered. Here too they were no doubt right. What is striking, however, is that some quite important empirical issues about theory discovery are irremediably central to an adequate epistemology of science.

It is a central part of the business of the philosophy of science to answer the fundamental epistemological question of why the methods of science are epistemically reliable. We have just had the occasion to examine in some detail two important features of those methods. First, the problem of sampling in experimental design is solved by the requirement that a proposed theory be tested under experimental circumstances that pit it against those alternatives to it that are theoretically plausible (and thus evidentially supported by inductive inference) given the body of previously established theories. Second, it can sometimes count as overwhelmingly confirmatory evidence for a theory that it is the basis for theoretically plausible explanations of a wide variety of phenomena that none of its otherwise plausible rivals can explain equally well. In each of these sorts of cases, the epistemic reliability of the relevant methodological practice depends on its being true in the (not too) long run that, when a proposed theory is in fact seriously mistaken, among its theoretically plausible rivals there will be theories that are relevantly closer to the truth and that can serve to identify the errors in the first theory or to challenge its exclusive claim to explanatory power with respect to the relevant class of natural phenomena.

It is, of course, impossible to assess the theoretical plausibility of theoretical proposals unless someone thinks them up. Failures of theoretical imagination can thus render the methodological practices we are discussing epistemologically unreliable in particular cases. The scientists who test a proposed theory against all the available theoretically plausible alternative theories will be employing an epistemically reliable testing strategy only on the assumption that the imaginative capacity of the scientific community is sufficient, so that theories near the truth in relevant respects will appear among those theories. Similarly, the scientists who accept a theory because it displays an apparent explanatory capacity utterly unmatched by any of the available plausible rivals will be reasoning reliably only if the imaginative capacity of the scientific community is up to the task of inventing a rich enough class of theoretically plausible rivals. Rival theories that would be theoretically plausible if we were only able to invent them and to understand them well enough to assess their theoretical plausibility play no methodological role unless we actually possess and display the relevant imaginative and cognitive capacities.

It is true, even on an empiricist conception of the matter, that successful science depends upon facts about our intellectual and imaginative capacities. Even if theory confirmation did not depend upon those capacities, we would not succeed in science unless we were able to think up suitably accurate theories to test. What we have just seen is that the same dependence of success upon our imaginative and cognitive capacities infects our ability to reliably confirm or disconfirm the theories we have already invented. The epistemic reliability of our scientific practices depends not only upon our possession of a suitably approximately true body of background theories but also upon our having quite contingent psychological capacities for exploiting these theoretical resources. This fact has three quite different implications for the philosophy of science.

First, it seems plausible that something somewhat like an "inductive logic" of theory invention may be epistemically important in science. It is probably true that theory invention (and creativity in general) involves finding new combinations of previously understood ideas and concepts. It is also true that inductive inferences at the theoretical level favor theoretical proposals that are relevantly similar (where the relevant respects of similarity are themselves theory-determined) to proposals that have already been established. It would be quite surprising if the respects of similarity to previous theories involved in theory invention were not fairly closely related to the respects of similarity determined to be epistemically relevant by the previous theories themselves. Indeed, if there were no relevant relations between the two it would be hard to see how our methodological practices thus far would have been epistemically reliable. The "logic of confirmation" must be somehow related to psychologically real inductive procedures for theory invention if scientific practice is to be epistemically reliable at all. The question "Just what is the relationship?" is simultaneously a question in empirical psychology and a question in the epistemology of science.

Second, a recognition of the role of theoretical imagination in epistemically reliable scientific practice opens up important possibilities in applied philosophy of science. Consider the question of the role of social prejudice in the practice of scientists. It has been traditional, on discovering that some figure in the history of science reached conclusions on scientific matters that we can, in retrospect, see as having been determined by inaccurate racial or sexual stereotypes, to conclude that the figure in question must have failed to employ the scientific method conistently. No doubt this is right in many cases, but reflection upon the crucial epistemic role of theoretical imagination in the evidential assessment of theories suggests an alternative hypothesis that may prove more accurate in many actual historical cases. The scientist may well have adhered scrupulously to the dictates of sound scientific methodology; all the available theoretically plausible alternatives to the now objectionable conclusions may have been taken quite seriously in assessing the evidence for them. The epistemic unreliability of the scientist's procedures may have stemmed, not from a failure to be methodologically scrupulous, but rather from socially determined failures of imagination on the part of the scientific community as a whole. In cases where this explanation is the right one, there may be no culpable methodological failure at all. Avoidance of socially prejudiced conclusions in such cases will depend either on political and social changes affecting the imaginative capacity of researchers or (perhaps) on extraordinary leaps of imagination, which are not part of normal scientific practice. In important ways, then, good scientific methodology is not prejudice-proof even when practiced with the greatest possible care, which is not to say that good methodological practice does not in the very long run help to overcome social prejudice. It is an instructive exercise to see how well or badly this model fits the various cases of social prejudice in biology described by Gould (1981).

Third, the importance of scientists' imaginative capacity for the epistemic reliability of scientific methodology illustrates in a striking way what is perhaps the most surprising feature of the realist conception of scientific knowledge. The epistemic reliability of scientific methods is logically contingent. It depends upon the historically contingent emergence of relevantly approximately true theoretical traditions (Boyd 1982, 1983, forthcoming) and also upon logically contingent features of our individual and collective capacities for theoretical imagination. Thus, principles of scientific methodology are not defensible *a priori* but have empirical presuppositions. The philosophy of science is an empirical discipline, not an *a priori* one. Indeed, this is probably true of philosophical inquiry generally. Here again is a conclusion with which Hume might have agreed, although it is true for distinctly non-Humean reasons.

References

Boyd, R. 1972. "Determinism, Laws and Predictability in Principle." *Philosophy of Science* 39: 431–450.

Boyd, R. 1973. "Realism, Underdetermination and a Causal Theory of Evidence." *Nous* 7: 1–12.

Boyd, R. 1979. "Metaphor and Theory Change." In *Metaphor and Thought*, ed. A. Ortony. Cambridge University Press.

Boyd, R. 1980. "Materialism Without Reductionism: What Physicalism Does Not Entail." In *Readings in Philosophy of Psychology*, vol. 1, ed. N. Block. Cambridge, Mass.: Harvard University Press.

Boyd, R. 1982. "Scientific Realism and Naturalistic Epistemology." In *PSA 80*, vol. 2. East Lansing, Mich.: Philosophy of Science Association.

Boyd, R. 1983. "On the Current Status of the Issue of Scientific Realism." *Erkenntnis* 19: 45–90.

Boyd, R. Forthcoming. "Lex Orendi est Lex Credendi." In *Images of Science: Scientific Realism Versus Constructive Empiricism*, ed. Churchland and Hooker. University of Chicago Press.

Carnap, R. 1950. "Empiricism, Semantics and Ontology." *Revue Internationale de Philosophie* 3: 20–40.

Feigl, H. 1945. "Operationism and Scientific Method." *Psychological Review* 52: 250–259.

Goodman, N. 1973. *Fact, Fiction and Forecast*, third edition. New York: Bobbs-Merrill.

Gould, S. J. 1981. *The Mismeasure of Man*. New York: Norton.

Hanson, N. R. 1958. *Patterns of Discovery*. Cambridge University Press.

Hempel, C. G. 1965. "Aspects of Scientific Explanation." In Hempel, *Aspects of Scientific Explanation*. New York: Free Press.

Hempel, C. G., and P. Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15: 98–115.

Kripke, S. 1972. "Naming and Necessity." In *The Semantics of Natural Language*, ed. G. Harman and D. Davidson. Dordrecht: Reidel.

Kuhn, T. 1970. *The Structure of Scientific Revolutions*, second edition. University of Chicago Press.

Popper, K. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.

Putnam, H. 1975a. "The Meaning of 'Meaning.'" In Putnam, *Mind, Language and Reality*. Cambridge University Press.

Putnam, H. 1975b. "Explanation and Reference." In Putnam, *Mind, Language and Reality*. Cambridge University Press.

Quine, W. V. O. 1969. "Natural Kinds." In Quine, *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Salmon, W. 1971. *Statistical Explanation and Statistical Relevance*. University of Pittsburgh Press.

Smart, J. J. C. 1963. *Philosophy and Scientific Realism*. London: Routledge and Kegan Paul.

van Fraassen, B. 1980. *The Scientific Image*. Oxford: Clarendon.