Sample selection from:

Steven M. Cahn, ed., *Philosophy for the 21st Century - A Comprehensive Reader* (Oxford: Oxford UP 2003).

PART 3

Philosophy of Science

Part B: pages 265-86

The Truth Doesn't Explain Much

NANCY CARTWRIGHT

Nancy Cartwright is professor of philosophy and chair of the Centre for Philosophy of Natural and Social Science at the London School of Economics and Political Science. She is a major contributor to contemporary philosophy of science and philosophy of social science. Her books include Nature's Capacities and Their Measurement *and* The Dappled World: A Study of the Boundaries of Science.

INTRODUCTION

Scientific theories must tell us both what is true in nature, and how we are to explain it. I shall argue that these are entirely different functions and should be kept distinct. Usually the two are conflated. The second is commonly seen as a by-product of the first. Scientific theories are thought to explain by dint of the descriptions they give of reality. Once the job of describing is done, science can shut down. That is all there is to do. To describe nature—to tell its laws, the values of its fundamental constants, its mass distributions—is ipso facto to lay down how we are to explain it.

This is a mistake, I shall argue; a mistake that is fostered by the covering-law model of explanation. The covering-law model supposes that all we need to know are the laws of nature—and a little logic, perhaps a little probability theory—and then we know which factors can explain which others. For example, in the simplest deductive-nomological version, the covering-law model says that one factor explains another just in case the occurrence of the second can be deduced from the occurrence of the first given the laws of nature. . . .

A good deal of criticism has been aimed at Hempel's original covering-law models. Much of the criticism objects that these models let in too much. On Hempel's account it seems we can explain Henry's failure to get pregnant by his taking birth control pills, and we can explain the storm by the falling barometer. My objection is quite the opposite. Covering-law models let in too little. With a covering-law model we can explain hardly anything, even the things of which we are most proud—like the role of DNA in the inheritance of genetic characteristics, or the formation of rainbows when sunlight is refracted through raindrops. We cannot explain these phenomena with a covering-law model, I shall argue, because we do not have laws that cover them. Covering laws are scarce.

Many phenomena which have perfectly good scientific explanations are not covered by any laws. No true laws, that is. They are at best covered by ceteris paribus generalizations—generalizations that hold only under special conditions, usually ideal conditions. The literal translation is "other things being equal"; but it would be more apt to read "ceteris paribus" as "other things being *right.*"

Sometimes we act as if this does not matter. We have in the back of our minds an "understudy" picture of ceteris paribus laws: ceteris paribus laws are real laws; they can stand in when the laws we would like to see are not available and they can perform all the same functions, only not quite so well. But this will not do. Ceteris paribus generalizations, read literally without the "ceteris paribus" modifier, are

Reprinted from *How the Laws of Physics Lie* (Oxford: Oxford University Press, 1983).

false. They are not only false, but held by us to be false; and there is no ground in the covering-law picture for false laws to explain anything. On the other hand, with the modifier the ceteris paribus generalizations may be true, but they cover only those few cases where the conditions are right. For most cases, either we have a law that purports to cover, but cannot explain because it is acknowledged to be false, or we have a law that does not cover. Either way, it is bad for the covering-law picture.

1. CETERIS PARIBUS LAWS

When I first started talking about the scarcity of covering laws, I tried to summarize my view by saying "There are no exceptionless generalizations." Then a friend asked, "How about 'All men are mortal'?" She was right. I had been focusing too much on the equations of physics. A more plausible claim would have been that there are no exceptionless quantitative laws in physics. Indeed not only are there no exceptionless laws, but in fact our best candidates are known to fail. This is something like the Popperian thesis that *every theory is born refuted.* Every theory we have proposed in physics, even at the time when it was most firmly entrenched, was known to be deficient in specific and detailed ways. I think this is also true for every precise quantitative law within a physics theory.

But this is not the point I had wanted to make. Some laws are treated, at least for the time being, as if they were exceptionless, whereas others are not, even though they remain "on the books." Snell's law (about the angle of incidence and the angle of refraction for a ray of light) is a good example of this latter kind. In the optics text I use for reference (Miles V. Klein, *Optics*),[1] it first appears on page 21, and without qualification:

> *Snell's Law:* At an interface between dielectric media, there is (also) *a refracted ray* in the second medium, lying in the plane of incidence, making an angle θ_t with the normal, and obeying Snell's law:
>
> $$\sin\theta/\sin\theta_t = n_2/n_1$$
>
> where v_1 and v_2 are the velocities of propagation in the two media, and $n_1 = (c/v_1)$, $n_2 = (c/v_2)$ are the indices of refraction.

It is only some 500 pages later, when the law is derived from the "full electromagnetic theory of light," that we learn that Snell's law as stated on page 21 is true only for media whose optical properties are *isotropic.* (In anisotropic media, "there will generally be *two* transmitted waves.") So what is deemed true is not really Snell's law as stated on page 21, but rather a refinement of Snell's law:

> *Refined Snell's Law: For any two media which are optically isotropic,* at an interface between dielectrics there is a refracted ray in the second medium, lying in the plane of incidence, making an angle θ_t with the normal, such that:
>
> $$\sin\theta/\sin\theta_t = n_2/n_1.$$

The Snell's law of page 21 in Klein's book is an example of a ceteris paribus law, a law that holds only in special circumstances—in this case when the media are both isotropic. Klein's statement on page 21 is clearly not to be taken literally. Charitably, we are inclined to put the modifier 'ceteris paribus' in front to hedge it. But what does this ceteris paribus modifier do? With an eye to statistical versions of the covering law model . . . we may suppose that the unrefined Snell's law is not intended to be a universal law, as literally stated, but rather some kind of statistical law. The obvious candidate is a crude statistical law: *for the most part,* at an interface between dielectric media there is *a* refracted ray . . . But this will not do. For *most* media are optically anisotropic, and in an anisotropic medium there are *two* rays. I think there are no more satisfactory alternatives. If ceteris paribus laws are to be true laws, there are no statistical laws with which they can generally be identified.

2. WHEN LAWS ARE SCARCE

Why do we keep Snell's law on the books when we both know it to be false and have a more accurate refinement available? There are obvious pedagogic reasons. But are there serious scientific ones? I think there are, and these reasons have to do with the task of explaining. Specifying which factors are explanatorily relevant to which others is a job done by science over and above the job of laying out the laws of nature. Once the laws of nature are known, we still have to decide what kinds of factors can be cited in explanation.

One thing that ceteris paribus laws do is to express our explanatory commitments. They tell what kinds of explanations are permitted. We know from the refined Snell's law that in any isotropic medium, the angle of refraction can be explained by the angle of incidence, according to the equation $\sin\theta/\sin\theta_t = n_2/n_1$. To leave the unrefined Snell's law on the books is to signal that the same kind of explanation can be given even for some anisotropic media. The pattern of explanation derived from the ideal situation is employed even where the conditions are less than ideal; and we assume that we can understand what happens in *nearly* isotropic media by rehearsing how light rays behave in pure isotropic cases.

This assumption is a delicate one . . . For the moment I intend only to point out that it *is* an assumption, and an assumption which (prior to the "full electromagnetic theory") goes well beyond our knowledge of the facts of nature. We *know* that in isotropic media, the angle of refraction is due to the angle of incidence under the equation $\sin\theta/\sin\theta_t = n_2/n_1$. We *decide* to explain the angles for the two refracted rays in anisotropic media in the same manner. We may have good reasons for the decision; in this case if the media are nearly isotropic, the two rays will be very close together, and close to the angle predicted by Snell's law; or we believe in continuity of physical processes. But still this decision is not forced by our knowledge of the laws of nature.

Obviously this decision could not be taken if we also had on the books a second refinement of Snell's law, implying that in any anisotropic media the angles are quite different from those given by Snell's law. But laws are scarce, and often we have no law at all about what happens in conditions that are less than ideal.

Covering-law theorists will tell a different story about the use of ceteris paribus laws in explanation. From their point of view, ceteris paribus explanations are elliptical for genuine covering law explanations from true laws which we do not yet know. When we use a ceteris paribus "law" which we know to be false, the covering-law theorist supposes us to be making a bet about what form the true law takes. For example, to retain Snell's unqualified law would be to bet that the (at the time unknown) law for anisotropic media will entail values "close enough" to those derived from the original Snell law.

I have two difficulties with this story. The first arises from an extreme metaphysical possibility, in which I in fact believe. Covering-law theorists tend to think that nature is well-regulated; in the extreme, that there is a law to cover every case. I do not. I imagine that natural objects are much like people in societies. Their behaviour is constrained by some specific laws and by a handful of general principles, but it is not determined in detail, even statistically. What happens on most occasions is dictated by no law at all. This is not a metaphysical picture that I urge. My claim is that this picture is as plausible as the alternative. God may have written just a few laws and grown tired. We do not know whether we are in a tidy universe or an untidy one. Whichever universe we are in, the ordinary commonplace activity of giving explanations ought to make sense.

The second difficulty for the ellipsis version of the covering-law account is more pedestrian. Elliptical explanations are not explanations: they are at best assurances that explanations are to be had. The law that is supposed to appear in the complete, correct D-N explanation is not a law we have in our theory, not a law that we can state, let alone test. There may be covering-law explanations in these cases. But those explanations are not our explanations; and those unknown laws cannot be our grounds for saying of a nearly isotropic medium, "$\sin\theta_t \approx k(n_2/n_1)$ *because* $\sin\theta = k$."

What then are our grounds? I assert only what they are not: they are not the laws of nature. The laws of nature that we know at any time are not enough to tell us what kinds of explanations can be given at that time. That requires a decision; and it is just this decision that covering-law theorists make when they wager about the existence of unknown laws. We may believe in these unknown laws, but we do so on no ordinary grounds: they have not been tested, nor are they derived from a higher level theory. Our grounds for believing in them are only as good as our reasons for adopting the corresponding explanatory strategy, and no better.

3. WHEN LAWS CONFLICT

I have been maintaining that there are not enough covering laws to go around. Why? The view depends on the picture of science that I mentioned earlier. Science is broken into various distinct domains: hydrodynamics, genetics, laser theory, . . . We have many detailed and sophisticated theories about what happens within the various domains. But we have little theory about what happens in the intersection of domains. . . .

For example, (ceteris paribus) adding salt to water decreases the cooking time of potatoes; taking the water to higher altitudes increases it. Refining, if we speak more carefully we might say instead, 'Adding salt to water while keeping the altitude constant decreases the cooking time; whereas increasing the altitude while keeping the saline content fixed increases it' . . .

But neither of these tells what happens when we both add salt to the water and move to higher altitudes.

Here we think that probably there is a precise answer about what would happen, even though it is not part of our common folk wisdom. But this is not always the case. I discuss this in detail in the next essay. Most real life cases involve some combination of causes; and general laws that describe what happens in these complex cases are not always available. Although both quantum theory and relativity are highly developed, detailed, and sophisticated, there is no satisfactory theory of relativistic quantum mechanics. A more detailed example from transport theory is given in the next essay. The general lesson is this: where theories intersect, laws are usually hard to come by.

4. WHEN EXPLANATIONS CAN BE GIVEN ANYWAY

So far, I have only argued half the case. I have argued that covering laws are scarce, and that ceteris paribus laws are no true laws. It remains to argue that, nevertheless, ceteris paribus laws have a fundamental explanatory role. But this is easy, for most of our explanations are explanations from ceteris paribus laws.

Let me illustrate with a humdrum example. Last year I planted camellias in my garden. I know that camellias like rich soil, so I planted them in composted manure. On the other hand, the manure was still warm, and I also know that camellia roots cannot take high temperatures. So I did not know what to expect. But when many of my camellias died, despite otherwise perfect care, I knew what went wrong. The camellias died because they were planted in hot soil.

This is surely the right explanation to give. Of course, I cannot be absolutely certain that this explanation is the correct one. Some other factor may have been responsible, nitrogen deficiency or some genetic defect in the plants, a factor that I did not notice, or may not even have known to be relevant. But this uncertainty is not peculiar to cases of explanation. It is just the uncertainty that besets all of our judgements about matters of fact. We must allow for oversight; still, since I made a reasonable effort to eliminate other menaces to my camellias, we may have some confidence that this is the right explanation.

So we have an explanation for the death of my camellias. But it is not an explanation from any true covering law. There is no law that says that camellias just like mine, planted in soil which is both hot and rich, die. To the contrary, they do not all die. Some thrive; and probably those that do, do so *because* of the richness of the soil they are planted in. We may insist that there must be some differentiating factor which brings the case under a covering law: in soil which is rich and hot, camellias of one kind die; those of another thrive. I will not deny that there may be such a covering law. I merely repeat that our ability to give this humdrum explanation precedes our knowledge of that law. On the Day of Judgment, when all laws are known, these may suffice to explain all phenomena. But in the meantime we do give explanations; and it is the job of science to tell us what kinds of explanations are admissible.

In fact I want to urge a stronger thesis. If, as is possible, the world is not a tidy deterministic system, this job of telling how we are to explain will be a job which is still left when the descriptive task of science is complete. Imagine for example (what I suppose actually to be the case) that the facts about camellias are irreducibly statistical. Then it is possible to know all the general nomological facts about camellias which there are to know—for example, that 62 percent of all camellias in just the circumstances of my camellias die, and 38 per cent survive. But one would not thereby know how to explain what happened in my garden. You would still have to look to the *Sunset Garden Book* to learn that the *heat* of the soil explains the perishing, and the *richness* explains the plants that thrive.

5. CONCLUSION

Most scientific explanations use ceteris paribus laws. These laws, read literally as descriptive statements, are false, not only false but deemed false even in the context of use. This is no surprise: we want laws that unify; but what happens may well be varied and diverse. We are lucky that we can organize phenomena at all. There is no reason to think that the principles that best organize will be true, nor that the principles that are true will organize much.

NOTE

1. Miles V. Klein. *Optics* (New York: John Wiley and Sons. 1970). p. 21. italics added. θ is the angle of incidence.

The New Riddle of Induction

NELSON GOODMAN

Nelson Goodman (1906–1998) made important contributions to aesthetics, epistemology, metaphysics, logic, and the philosophy of science. His books include *The Structure of Appearance, Languages of Art,* and *Ways of Worldmaking.*

Confirmation of a hypothesis by an instance depends rather heavily upon features of the hypothesis other than its syntactical form. That a given piece of copper conducts electricity increases the credibility of statements asserting that other pieces of copper conduct electricity, and thus confirms the hypothesis that all copper conducts electricity. But the fact that a given man now in this room is a third son does not increase the credibility of statements asserting that other men now in this room are third sons, and so does not confirm the hypothesis that all men now in this room are third sons. Yet in both cases our hypothesis is a generalization of the evidence statement. The difference is that in the former case the hypothesis is a *lawlike* statement; while in the latter case, the hypothesis is a merely contingent or accidental generality. Only a statement that is *lawlike*—regardless of its truth or falsity or its scientific importance—is capable of receiving confirmation from an instance of it; accidental statements are not. Plainly, then, we must look for a way of distinguishing lawlike from accidental statements.

So long as what seems to be needed is merely a way of excluding a few odd and unwanted cases that are inadvertently admitted by our definition of confirmation, the problem may not seem very hard or very pressing. We fully expect that minor defects will be found in our definition and that the necessary refinements will have to be worked out patiently one after another. But some further examples will show that our present difficulty is of a much graver kind.

Suppose that all emeralds examined before a certain time *t* are green. At time *t*, then, our observations support the hypothesis that all emeralds are green; and this is in accord with our definition of confirmation. Our evidence statements assert that emerald *a* is green, that emerald *b* is green, and so on; and each confirms the general hypothesis that all emeralds are green. So far, so good.

Now let me introduce another predicate less familiar than "green." It is the predicate "grue" and it applies to all things examined before *t* just in case they are green but to other things just in case they are blue. Then at time *t* we have, for each evidence statement asserting that a given emerald is green, a parallel evidence statement asserting that that emerald is grue. And the statements that emerald *a* is grue, that emerald *b* is grue, and so on, will each confirm the general hypothesis that all emeralds are grue. Thus according to our definition, the prediction that all emeralds subsequently examined will be green and the prediction that all will be grue are alike confirmed by evidence statements describing the same observations. But if an emerald subsequently examined is grue, it is blue and hence not green. Thus although we are well aware which of the two incompatible predictions is genuinely confirmed, they are equally well confirmed according to our present definition. Moreover, it is clear that if we simply choose an appropriate predicate, then on the basis of these same observations we shall have equal confirmation, by our definition, for any prediction whatever about other emeralds—or indeed about anything else.[1] As in our earlier example, only the predictions subsumed under lawlike hypothesis are genuinely confirmed; but we have no criterion as yet for determining lawlikeness. And now we see that without some such criterion, our definition not merely includes a few unwanted cases, but is so completely ineffectual that it virtually excludes nothing. We are left once again with the intolerable result that anything confirms anything. This difficulty cannot be set aside as an annoying detail to be taken care of in due course. It has to be met before our definition will work at all.

Nevertheless, the difficulty is often slighted because on the surface there seem to be easy ways of dealing with it. Sometimes, for example, the problem is thought to be . . . [that we are], making tacit and illegitimate use of information outside the stated evidence: the information, for example, that different samples of one material are usually alike in conductivity, and the information that different men in a lecture audience are usually not alike in the number of their older brothers. But while it is true that such information is being smuggled in, this does not by itself settle the matter as it settles the matter of the ravens. There the point was that when the smuggled information is forthrightly declared, its effect upon the confirmation of the hypothesis in question is immediately and properly registered by the definition we are using. On the other hand, if to our initial evidence we add statements concerning the conductivity of pieces of other materials or concerning the number of older brothers of members of other lecture audiences, this will not in the least affect the confirmation, according to our definition, of the hypothesis concerning copper or of that concerning this lecture audience. Since our definition is insensitive to the bearing upon hypotheses of evidence so related to them, even when the evidence is fully declared, the difficulty about accidental hypotheses cannot be explained away on the ground that such evidence is being surreptitiously taken into account.

A more promising suggestion is to explain the matter in terms of the effect of this other evidence not directly upon the hypothesis in question but *in*directly through other hypotheses that *are* confirmed, according to our definition, by such evidence. Our information about other materials does by our definition confirm such hypotheses as that all pieces of iron conduct electricity, that no pieces of rubber do, and so on; and these hypotheses, the explanation runs, impart to the hypothesis that all pieces of copper conduct electricity (and also to the hypothesis that none do) the character of lawlikeness—that is, amenability to confirmation by direct positive instances when found. On the other hand, our information about other lecture audiences *dis*confirms many hypotheses to the effect that all the men in one audience are third sons, or that none are; and this strips any character of lawlikeness from the hypothesis that all (or the hypothesis that none) of the men in *this* audience are third sons. But clearly if this course is to be followed, the circumstances under which hypotheses are thus related to one another will have to be precisely articulated.

The problem, then, is to define the relevant way in which such hypotheses must be alike. Evidence for the hypothesis that all iron conducts electricity enhances the lawlikeness of the hypothesis that all zirconium conducts electricity, but does not similarly affect the hypothesis that all the objects on my desk conduct electricity. Wherein lies the difference? The first two hypotheses fall under the broader hypothesis—call it "*H*"—that every class of things of the same material is uniform in conductivity; the first and third fall only under some such hypothesis as—call it "*K*"—that every class of things that are either all of the same material or all on a desk is uniform in conductivity. Clearly the important difference here is that evidence for a statement affirming that one of the classes covered by *H* has the property in question increases the credibility of any statement affirming that another such class has this property; while nothing of the sort holds true with respect to *K*. But this is only to say that *H* is lawlike and *K* is not. We are faced anew with the very problem we are trying to solve: the problem of distinguishing between lawlike and accidental hypotheses.

The most popular way of attacking the problem takes its cue from the fact that accidental hypotheses seem typically to involve some spatial or temporal restriction, or reference to some particular individual. They seem to concern the people in some particular room, or the objects on some particular person's desk; while lawlike hypotheses characteristically concern all ravens or all pieces of copper whatsoever. Complete generality is thus very often supposed to be a sufficient condition of lawlikeness; but to define this complete generality is by no means easy. Merely to require that the hypothesis contain no term naming, describing, or indicating a particular thing or location will obviously not be enough. The troublesome hypothesis that all emeralds are grue contains no such term; and where such a term does occur, as in hypotheses about men in *this room*, it can be suppressed in favor of some predicate (short or long, new or old) that contains no such term but applies only to exactly the same things. One might think, then, of excluding not only hypotheses that actually contain terms for specific individuals but also all hypotheses that are equivalent to others that do contain such terms. But, as we have just seen, to exclude only hypotheses of which *all* equivalents contain such terms is to exclude nothing. On the other hand, to exclude all hypotheses that have *some* equivalent containing such a term is to exclude everything; for even the hypotheses

All grass is green

has as an equivalent

All grass in London or elsewhere is green.

The next step, therefore, has been to consider ruling out predicates of certain kinds. A syntactically universal hypothesis is lawlike, the proposal runs, if its predicates are "purely qualitative" or "nonpositional". This will obviously accomplish nothing if a purely qualitative predicate is then conceived either as one that is equivalent to some expression free of terms for specific individuals, or as one that is equivalent to no expression that contains such a term; for this only raises again the difficulties just pointed out. The claim appears to be rather that at least in the case of a simple enough predicate we can readily determine by direct inspection of its meaning whether or not it is purely qualitative. But even aside from obscurities in the notion of 'the meaning' of a predicate, this claim seems to me wrong. I simply do not know how to tell whether a predicate is qualitative or positional, except perhaps by completely begging the question at issue and asking whether the predicate is "well-behaved"—that is, whether simple syntactically universal hypotheses applying it are lawlike.

This statement will not go unprotested. "Consider," it will be argued, "the predicates 'blue' and 'green' and the predicate 'grue' introduced earlier, and also the predicate 'bleen' that applies to emeralds examined before time *t* just in case they are blue and to other emeralds just in case they are green. Surely it is clear," the argument runs, "that the first two are purely qualitative and the second two are not; for the meaning of each of the latter two plainly involves reference to a specific temporal position." To this I reply that indeed I do recognize the first two as well-behaved predicates admissible in lawlike hypotheses, and the second two as ill-behaved predicates. But the argument that the former but not the latter are purely qualitative seems to me quite unsound. True enough, if we start with "blue" and "green", then "grue" and "bleen" will be explained in terms of "blue" and "green" and a temporal term. But equally truly, if we start with "grue" and "bleen", then "blue" and "green" will be explained in terms of "grue" and "bleen" and a temporal term; "green", for example, applies to emeralds examined before time *t* just in case they are grue, and to other emeralds just in case they are bleen. Thus qualitativeness is an entirely relative matter and does not by itself establish any dichotomy of predicates. This relativity seems to be completely overlooked by those who contend that the qualitative character of a predicate is a criterion for its good behavior.

Of course, one may ask why we need worry about such unfamiliar predicates as "grue" or about accidental hypotheses in general, since we are unlikely to use them in making predictions. If our definition works for such hypotheses as are normally employed, isn't that all we need? In a sense, yes; but only in the sense that we need no definition, no theory of induction, and no philosophy of knowledge at all. We get along well enough without them in daily life and in scientific research. But if we seek a theory at all, we cannot excuse gross anomalies resulting from a proposed theory by pleading that we can avoid them in practice. The odd cases we have been considering are clinically pure cases that, though seldom encountered in practice, nevertheless display to best advantage the symptoms of a widespread and destructive malady.

We have so far neither any answer nor any promising clue to an answer to the question what distinguishes lawlike or confirmable hypotheses from accidental or non-confirmable ones; and what may at first have seemed a minor technical difficulty has taken on the stature of a major obstacle to the development of a satisfactory theory of confirmation. It is this problem that I call the new riddle of induction.

THE PERVASIVE PROBLEM OF PROJECTION

At the beginning of this lecture, I expressed the opinion that the problem of induction is still unsolved, but that the difficulties that face us today are not the old ones; and I have tried to outline the changes that have taken place. The problem of justifying induction has been displaced by the problem of defining confirmation, and our work upon this has left us with the residual problem of distinguishing between confirmable and non-confirmable hypotheses. One might say roughly that the first question was "Why does a positive instance of a hypothesis give any grounds for predicting further instances?"; that the newer question was "What is a positive instance of a hypothesis?"; and that the crucial remaining question is "What hypotheses are confirmed by their positive instances?"

The vast amount of effort expended on the problem of induction in modern times has thus altered our afflictions but hardly relieved them. The original difficulty about induction arose from the recognition that anything may follow upon anything. Then, in attempting to define confirmation in terms of the converse of the consequence relation, we found ourselves with the distressingly similar difficulty that our definition would make any statement confirm any other. And now, after modifying our definition drastically, we still get the old devastating result that any statement will confirm any statement. Until we find a way of exercising some control over the hypotheses to be admitted, our definition makes no distinction whatsoever between valid and invalid inductive inferences.

The real inadequacy of Hume's account lay not in his descriptive approach but in the imprecision of his description. Regularities in experience, according to him, give rise to habits of expectation; and thus it is predictions conforming to past regularities that are normal or valid. But Hume overlooks the fact that some regularities do and some do not establish such habits; that predictions based on some regularities are valid while predictions based on other regularities are not. Every word you have heard me say has occurred prior to the final sentence of this lecture; but that does not, I hope, create any expectation that every word you will hear me say will be prior to that sentence. Again, consider our case of emeralds. All those examined before time t are green; and this leads us to expect, and confirms the prediction, that the next one will be green. But also, all those examined are grue; and this does not lead us to expect, and does not confirm the prediction, that the next one will be grue. Regularity in greenness confirms the prediction of further cases; regularity in grueness does not. To say that valid predictions are those based on past regularities, without being able to say *which* regularities, is thus quite pointless. Regularities are where you find them, and you can find them anywhere. As we have seen, Hume's failure to recognize and deal with this problem has been shared even by his most recent successors.

As a result, what we have in current confirmation theory is a definition that is adequate for certain cases that so far can be described only as those for which it is adequate. The theory works where it works. A hypothesis is confirmed by statements related to it in the prescribed way provided it is so confirmed. This is a good deal like having a theory that tells us that the area of a plane figure is one-half the base times the altitude, without telling us for what figures this holds. We must somehow find a way of distinguishing lawlike hypotheses, to which our definition of confirmation applies, from accidental hypotheses, to which it does not.

Today I have been speaking solely of the problem of induction, but what has been said applies equally to the more general problem of projection. As pointed out earlier, the problem of prediction from past to future cases is but a narrower version of the problem of projecting from any set of cases to others. We saw that a whole cluster of troublesome problems concerning dispositions and possibility can be reduced to this problem of projection. That is why the new riddle of induction, which is more broadly the problem of distinguishing between projectible and non-projectible hypotheses, is as important as it is exasperating.

NOTE

1. For instance, we shall have equal confirmation, by our present definition, for the prediction that roses subsequently examined will be blue. Let "emerose" apply just to emeralds examined before time t, and to roses examined later. Then all emeroses so far examined are grue, and this confirms the hypothesis that all emeroses are grue and hence the prediction that roses subsequently examined will be blue. The problem raised by such antecedents has been little noticed, but is no easier to meet than that raised by similarly perverse consequents.

The Structure of Scientific Revolutions

THOMAS S. KUHN

Thomas S. Kuhn (1922–1996) was an historian and philosopher of science. His book *The Structure of Scientific Revolutions* is a seminal work on the nature and importance of scientific development and scientific revolutions that revolutionized the field of philosophy of science.

What are scientific revolutions and what is their function in scientific development? . . . [S]cientific revolutions are here taken to be those noncumulative developmental episodes in which an older paradigm is replaced in whole or in part by an incompatible new one. There is more to be said, however, and an essential part of it can be introduced by asking one further question. Why should a change of paradigm be called a revolution? In the face of the vast and essential differences between political and scientific development, what parallelism can justify the metaphor that finds revolutions in both?

One aspect of the parallelism must already be apparent. Political revolutions are inaugurated by a growing sense, often restricted to a segment of the political community, that existing institutions have ceased adequately to meet the problems posed by an environment that they have in part created. In much the same way, scientific revolutions are inaugurated by a growing sense, again often restricted to a narrow subdivision of the scientific community, that an existing paradigm has ceased to function adequately in the exploration of an aspect of nature to which that paradigm itself had previously led the way. In both political and scientific development the sense of malfunction that can lead to crisis is prerequisite to revolution. Furthermore, though it admittedly strains the metaphor, that parallelism holds not only for the major paradigm changes, like those attributable to Copernicus and Lavoisier, but also for the far smaller ones associated with the assimilation of a new sort of phenomenon, like oxygen or X-rays. Scientific revolutions . . . need seem revolutionary only to those whose paradigms are affected by them. To outsiders they may, like the Balkan revolutions of the early twentieth century, seem normal parts of the developmental process. Astronomers, for example, could accept X-rays as a mere addition to knowledge, for their paradigms were unaffected by the existence of the new radiation. But for men like Kelvin, Crookes, and Roentgen, whose research dealt with radiation theory or with cathode ray tubes, the emergence of X-rays necessarily violated one paradigm as it created another. That is why these rays could be discovered only through something's first going wrong with normal research.

This genetic aspect of the parallel between political and scientific development should no longer be open to doubt. The parallel has, however, a second and more profound aspect upon which the significance of the first depends. Political revolutions aim to change political institutions in ways that those institutions themselves prohibit. Their success therefore necessitates the partial relinquishment of one set of institutions in favor of another, and in the interim, society is not fully governed by institutions at all. Initially it is

Reprinted from *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962), by permission of the publisher.

crisis alone that attenuates the role of political institutions as we have already seen it attenuate the role of paradigms. In increasing numbers individuals become increasingly estranged from political life and behave more and more eccentrically within it. Then, as the crisis deepens, many of these individuals commit themselves to some concrete proposal for the reconstruction of society in a new institutional framework. At that point the society is divided into competing camps or parties, one seeking to defend the old institutional constellation, the others seeking to institute some new one. And, once that polarization has occurred, *political recourse fails*. Because they differ about the institutional matrix within which political change is to be achieved and evaluated, because they acknowledge no supra-institutional framework for the adjudication of revolutionary difference, the parties to a revolutionary conflict must finally resort to the techniques of mass persuasion, often including force. Though revolutions have had a vital role in the evolution of political institutions, that role depends upon their being partially extrapolitical or extrainstitutional events.

The remainder of this essay aims to demonstrate that the historical study of paradigm change reveals very similar characteristics in the evolution of the sciences. Like the choice between competing political institutions, that between competing paradigms proves to be a choice between incompatible modes of community life. Because it has that character, the choice is not and cannot be determined merely by the evaluative procedures characteristic of normal science, for these depend in part upon a particular paradigm, and that paradigm is at issue. When paradigms enter, as they must, into a debate about paradigm choice, their role is necessarily circular. Each group uses its own paradigm to argue in that paradigm's defense.

The resulting circularity does not, of course, make the arguments wrong or even ineffectual. The man who premises a paradigm when arguing in its defense can nonetheless provide a clear exhibit of what scientific practice will be like for those who adopt the new view of nature. That exhibit can be immensely persuasive, often compellingly so. Yet, whatever its force, the status of the circular argument is only that of persuasion. It cannot be made logically or even probabilistically compelling for those who refuse to step into the circle. The premises and values shared by the two parties to a debate over paradigms are not sufficiently extensive for that. As in political revolutions, so in paradigm choice—there is no standard higher than the assent of the relevant community. To discover how scientific revolutions are effected, we shall therefore have to examine not only the impact of nature and of logic, but also the techniques of persuasive argumentation effective within the quite special groups that constitute the community of scientists.

To discover why this issue of paradigm choice can never be unequivocally settled by logic and experiment alone, we must shortly examine the nature of the differences that separate the proponents of a traditional paradigm from their revolutionary successors. That examination is the principal object of this section and the next. We have, however, already noted numerous examples of such differences, and no one will doubt that history can supply many others. What is more likely to be doubted than their existence—and what must therefore be considered first—is that such examples provide essential information about the nature of science. Granting that paradigm rejection has been a historic fact, does it illuminate more than human credulity and confusion? Are there intrinsic reasons why the assimilation of either a new sort of phenomenon or a new scientific theory must demand the rejection of an older paradigm?

First notice that if there are such reasons, they do not derive from the logical structure of scientific knowledge. In principle, a new phenomenon might emerge without reflecting destructively upon any part of past scientific practice. Though discovering life on the moon would today be destructive of existing paradigms (these tell us things about the moon that seem incompatible with life's existence there), discovering life in some less well-known part of the galaxy would not. By the same token, a new theory does not have to conflict with any of its predecessors. It might deal exclusively with phenomena not previously known, as the quantum theory deals (but, significantly, not exclusively) with subatomic phenomena unknown before the twentieth century. Or again, the new theory might be simply a higher level theory than those known before, one that linked together a whole group of lower level theories without substantially changing any. Today, the theory of energy conservation provides just such links between dynamics, chemistry, electricity, optics, thermal theory, and so on. Still other compatible relationships between old and new theories can be conceived. Any and all of them might be exemplified by the historical process through which science has developed. If they were, scientific development would be genuinely cumulative. New sorts of phenomena would simply disclose order in an aspect of nature where none had been seen before. In the evolution of science new knowledge would replace ignorance rather than replace knowledge of another and incompatible sort.

Of course, science (or some other enterprise, perhaps less effective) might have developed in that fully cumulative manner. Many people have believed that it did so, and most still seem to suppose that cumulation is at least the ideal that historical development would display if only it had not so often been distorted by human idiosyncrasy. There are important reasons for that belief. . . . Nevertheless, despite the immense plausibility of that ideal image, there is increasing reason to wonder whether it can possibly be an image of *science*. After the pre-paradigm period the assimilation of all new theories and of almost all new sorts of phenomena has in fact demanded the destruction of a prior paradigm and a consequent conflict between competing schools of scientific thought. Cumulative acquisition of unanticipated novelties proves to be an almost non-existent exception to the rule of scientific development. The man who takes historic fact seriously must suspect that science does not tend toward the ideal that our image of its cumulativeness has suggested. Perhaps it is another sort of enterprise.

If, however, resistant facts can carry us that far, then a second look at the ground we have already covered may suggest that cumulative acquisition of novelty is not only rare in fact but improbable in principle. Normal research, which *is* cumulative, owes its success to the ability of scientists regularly to select problems that can be solved with conceptual and instrumental techniques close to those already in existence. (That is why an excessive concern with useful problems, regardless of their relation to existing knowledge and technique, can so easily inhibit scientific development.) The man who is striving to solve a problem defined by existing knowledge and technique is not, however, just looking around. He knows what he wants to achieve, and he designs his instruments and directs his thoughts accordingly. Unanticipated novelty, the new discovery, can emerge only to the extent that his anticipations about nature and his instruments prove wrong. Often the importance of the resulting discovery will itself be proportional to the extent and stubbornness of the anomaly that foreshadowed it. Obviously, then, there must be a conflict between the paradigm that discloses anomaly and the one that later renders the anomaly law-like. The examples of discovery through paradigm destruction . . . did not confront us with mere historical accident. There is no other effective way in which discoveries might be generated.

The same argument applies even more clearly to the invention of new theories. There are, in principle, only three types of phenomena about which a new theory might be developed. The first consists of phenomena already well explained by existing paradigms, and these seldom provide either motive or point of departure for theory construction. When they do . . . the theories that result are seldom accepted, because nature provides no ground for discrimination. A second class of phenomena consists of those whose nature is indicated by existing paradigms but whose details can be understood only through further theory articulation. These are the phenomena to which scientists direct their research much of the time, but that research aims at the articulation of existing paradigms rather than at the invention of new ones. Only when these attempts at articulation fail do scientists encounter the third type of phenomena, the recognized anomalies whose characteristic feature is their stubborn refusal to be assimilated to existing paradigms. This type alone gives rise to new theories. Paradigms provide all phenomena except anomalies with a theory-determined place in the scientist's field of vision.

But if new theories are called forth to resolve anomalies in the relation of an existing theory to nature, then the successful new theory must some-where permit predictions that are different from those derived from its predecessor. That difference could not occur if the two were logically compatible. In the process of being assimilated, the second must displace the first. Even a theory like energy conservation, which today seems a logical superstructure that relates to nature only through independently established theories, did not develop historically without paradigm destruction. Instead, it emerged from a crisis in which an essential ingredient was the incompatibility between Newtonian dynamics and some recently formulated consequences of the caloric theory of heat. Only after the caloric theory had been rejected could energy conservation become part of science.[1] And only after it had been part of science for some time could it come to seem a theory of a logically higher type, one not in conflict with its predecessors. It is hard to see how new theories could arise without these destructive changes in beliefs about nature. Though logical inclusiveness remains a permissible view of the relation between successive scientific theories, it is a historical implausibility.

A century ago it would, I think, have been possible to let the case for the necessity of revolutions rest at this point. But today, unfortunately, that cannot be done because the view of the subject developed above cannot be maintained if the most prevalent contemporary interpretation of the nature and function of scientific theory is accepted. That interpretation, closely associated with early logical positivism and not categorically rejected by its successors, would restrict the range and meaning of an accepted theory so that it could not possibly conflict with any later theory that made predictions about some of the same natural phenomena. The best-known and the strongest case for this restricted conception of a scientific theory emerges in discussions of the relation between contemporary Einsteinian dynamics and the older dynamical equations that descend from Newton's *Principia*. From the viewpoint of this essay these two theories are fundamentally incompatible in the sense illustrated by the relation of Copernican to Ptolemaic astronomy: Einstein's theory can be accepted only with the recognition that Newton's was wrong. Today this remains a minority view.[2] We must therefore examine the most prevalent objections to it.

The gist of these objections can be developed as follows. Relativistic dynamics cannot have shown Newtonian dynamics to be wrong, for Newtonian dynamics is still used with great success by most engineers and, in selected applications, by many physicists. Furthermore, the propriety of this use of the older theory can be proved from the very theory that has, in other applications, replaced it. Einstein's theory can be used to show that predictions from Newton's equations will be as good as our measuring instruments in all applications that satisfy a small number of restrictive conditions. For example, if Newtonian theory is to provide a good approximate solution, the relative velocities of the bodies considered must be small compared with the velocity of light. Subject to this condition and a few others, Newtonian theory seems to be derivable from Einsteinian, of which it is therefore a special case.

But, the objection continues, no theory can possibly conflict with one of its special cases. If Einsteinian science seems to make Newtonian dynamics wrong, that is only because some Newtonians were so incautious as to claim that Newtonian theory yielded entirely precise results or that it was valid at very high relative velocities. Since they could not have had any evidence for such claims, they betrayed the standards of science when they made them. In so far as Newtonian theory was ever a truly scientific theory supported by valid evidence, it still is. Only extravagant claims for the theory—claims that were never properly parts of science—can have been shown by Einstein to be wrong. Purged of these merely human extravagances, Newtonian theory has never been challenged and cannot be.

Some variant of this argument is quite sufficient to make any theory ever used by a significant group of competent scientists immune to attack. The much-maligned phlogiston theory, for example, gave order to a large number of physical and chemical phenomena. It explained why bodies burned—they were rich in phlogiston—and why metals had so many more properties in common than did their ores. The metals were all compounded from different elementary earths combined with phlogiston, and the latter, common to all metals, produced common properties. In addition, the phlogiston theory accounted for a number of reactions in which acids were formed by the combustion of substances like carbon and sulphur. Also, it explained the decrease of volume when combustion occurs in a confined volume of air—the phlogiston released by combustion "spoils" the elasticity of the air that absorbed it, just as fire "spoils" the elasticity of a steel spring.[3] If these were the only phenomena that the phlogiston theorists had claimed for their theory, that theory could never have been challenged. A similar argument will suffice for any theory that has ever been successfully applied to any range of phenomena at all.

But to save theories in this way, their range of application must be restricted to those phenomena and to that precision of observation with which the experimental evidence in hand already deals.[4] Carried just a step further (and the step can scarcely be avoided once the first is taken), such a limitation prohibits the scientist from claiming to speak "scientifically" about any phenomenon not already observed. Even in its present form the restriction forbids the scientist to rely upon a theory in his own research whenever that research enters an area or seeks a degree of precision for which past practice with the theory offers no precedent. These prohibitions are logically unexceptionable. But the result of accepting them would be the end of the research through which science may develop further.

By now that point too is virtually a tautology. Without commitment to a paradigm there could be no normal science. Furthermore, that commitment must extend to areas and to degrees of precision for which there is no full precedent. If it did not, the paradigm could provide no puzzles that had not already been solved. Besides, it is not only normal science that depends upon commitment to a paradigm. If existing theory binds the scientist only with respect to existing applications, then there can be no surprises, anomalies, or crises. But these are just the signposts that point the way to extraordinary science. If positivistic restrictions on the range of a theory's legitimate applicability are taken literally, the mechanism that tells the scientific community what problems may lead to fundamental change must cease to function. And when that occurs, the community will inevitably return to something much like its pre-paradigm state, a condition in which all members practice science but in which their gross product scarcely resembles science at all. Is it really any wonder that the price of significant scientific advance is a commitment that runs the risk of being wrong?

More important, there is a revealing logical lacuna in the positivist's argument, one that will reintroduce us immediately to the nature of revolutionary change. Can Newtonian dynamics really be *derived* from relativistic dynamics? What would such a derivation look like? Imagine a set of statements, $E_1, E_2, \ldots, E_n$, which together embody the laws of relativity theory. These statements contain variables and parameters representing spatial position, time, rest mass, etc. From them, together with the apparatus of logic and mathematics, is deducible a whole set of further statements including some that can be checked by observation. To prove the adequacy of Newtonian dynamics as a special case, we must add to the E_i's additional statements, like $(v/c)^2 << 1$, restricting the range of the parameters and variables. This enlarged set of statements is then manipulated to yield a new set, $N_1, N_2, \ldots, N_m$, which is identical in form with Newton's laws of motion, the law of gravity, and so on. Apparently Newtonian dynamics has been derived from Einsteinian, subject to a few limiting conditions.

Yet the derivation is spurious, at least to this point. Though the N_i's are a special case of the laws of relativistic mechanics, they are not Newton's Laws. Or at least they are not unless those laws are reinterpreted in a way that would have been impossible until after Einstein's work. The variables and parameters that in the Einsteinian E_i's represented spatial position, time, mass, etc., still occur in the N_i's; and they there still represent Einsteinian space, time, and mass. But the physical referents of these Einsteinian concepts are by no means identical with those of the Newtonian concepts that bear the same name. (Newtonian mass is conserved; Einsteinian is convertible with energy. Only at low relative velocities may the two be measured in the same way, and even then they must not be conceived to be the same.) Unless we change the definitions of the variables in the N_i's, the statements we have derived are not Newtonian. If we do change them, we cannot properly be said to have *derived* Newton's Laws, at least not in any sense of "derive" now generally recognized. Our argument has, of course, explained why Newton's Laws ever seemed to work. In doing so it has justified, say, an automobile driver in acting as though he lived in a Newtonian universe. An argument of the same type is used to justify teaching earth-centered astronomy to surveyors. But the argument has still not done what it purported to do. It has not, that is, shown Newton's Laws to be a limiting case of Einstein's. For in the passage to the limit it is not only the forms of the laws that have changed. Simultaneously we have had to alter the fundamental structural elements of which the universe to which they apply is composed.

This need to change the meaning of established and familiar concepts is central to the revolutionary impact of Einstein's theory. Though subtler than the changes from geocentrism to heliocentrism, from phlogiston to oxygen, or from corpuscles to waves, the resulting conceptual transformation is no less decisively destructive of a previously established paradigm. We may even come to see it as a prototype for revolutionary reorientations in the sciences. Just because it did not involve the introduction of additional objects or concepts, the transition from Newtonian to Einsteinian mechanics illustrates with particular clarity the scientific revolution as a displacement of the conceptual network through which scientists view the world.

NOTES

1. Silvanus P. Thompson, *Life of William Thomson Baron Kelvin of Largs* (London, 1910), I, 266–81.
2. See, for example, the remarks by P. P. Wiener in *Philosophy of Science*, XXV (1958), 298.
3. James B. Conant, *Overthrow of the Phlogiston Theory* (Cambridge, 1950), pp. 13–16; and J. R. Partington, *A Short History of Chemistry* (2d ed.; London, 1951), pp. 85–88. The fullest and most sympathetic account of the phlogiston theory's achievements is by H. Metzger, *Newton, Stahl, Boerhaave et la doctrine chimique* (Paris, 1930), Part II.
4. Compare the conclusions reached through a very different sort of analysis by R. B. Braithwaite, *Scientific Explanation* (Cambridge, 1953), pp. 50–87, esp. p. 76.

Realism and the Theory-Dependence of Experimental Design

RICHARD N. BOYD

Richard N. Boyd is professor of philosophy at Cornell University. He is a leading exponent of scientific realism, and his articles include "On the Current Status of Scientific Realism," "Scientific Realism and Naturalistic Epistemology," and "Determinism, Laws and Predictability in Principle."

I. REALISM AND THE THEORY-DEPENDENCE OF EXPERIMENTAL DESIGN

In several papers . . . I have argued that scientific realism provides the only scientifically reasonable explanation for the reliability of certain important features of scientific methodology which are crucial in experimental design and in the assessment of experimental evidence. Roughly speaking, these are the features of scientific methodology relevant to the assessment of the "degree of confirmation" of a pro-

Reprinted from *Images of Science*, eds. Paul M. Churchland and Clifford A. Hooker (Chicago: University of Chicago Press, 1985), by permission of the publisher.

posed theory, given a body of observational evidence (if we choose to employ the standard empiricist terminology). In the present section, I want to expand upon this claim and to offer arguments for it in somewhat greater detail.

To begin with, it is important to understand what sort of reliability of methodology is to be explained. If scientific realism is true, then the methodological practices of science provide a reliable guide to approximate truth about theoretical matters and, no doubt, only scientific realism could provide a satisfactory explanation for this fact. But it would be question-begging to suggest that this provides any good reason to accept scientific realism; after all, only realists believe that the methodology of science is reliable in this sense, anyway. What I propose to do is to take advantage of the fact that antirealists in the philosophy of science are typically selective in their skepticism and to define the reliability of the methods of science in such a way that no questions are begged against the position of the typical antirealist. Call a theory instrumentally reliable if it makes approximately true predictions about observable phenomena. Call a methodology instrumentally reliable if it is a reliable guide to the acceptance of theories which are themselves instrumentally reliable. For the antirealist against whom my arguments are directed, it is uncontroversial that the actual methods of science are instrumentally reliable in this sense, although it may of course be a matter for philosophical dispute just which features of actual scientific practice explain this reliability. The arguments I am discussing here are directed against only the selectively skeptical antirealist; I have nothing to say to "the Skeptic."

Let us suppose that some scientific theory *T* has been proposed and that a body *E* of experimental results has been obtained which is consonant with the predictions of *T*. Imagine that *T* is an ordinary medium-sized theory of the sort which scientists routinely confirm or disconfirm in the course of what Kuhn calls "normal science." (Scientific realism must have something to say about the acceptance of large-scale paradigm-fixing theories as well, but it is a matter of controversy whether there is a reliable methodology for such cases, and so I want here to examine the more commonplace instances of theory testing. . . .) Questions regarding the extent to which *T* is confirmed by the evidence *E* may be fruitfully divided into three categories:

1. *The question of "projectability."* One of the things which Goodman has taught us is that something important about inductive inference can be learned by examining the *unrefuted* inductive generalizations which no one ought to accept. Goodman formulates the issue in terms of the projectability of predicates in simple inductive generalizations, but it is clear that the issue he raises is more general. We can think of any sufficiently general theory as representing the proposal to consider as projectable certain possible patterns in observable data, *viz.*, those patterns which the theory predicts. In general, the methodological acceptability of such a proposal will not depend soley upon the projectability of the individual predicates contained in the theory in question considered in isolation, but also on the structure of the theory itself. . . . *T* will receive significant evidential support from *E* only if *T* represents a projectable pattern in possible observational data.

2. *The question of experimental controls and experimental artifacts.* Suppose that *T* represents a projectable pattern in possible observational data, and suppose further that the data in *E* represent apparently confirming evidence for *T*, whatever this latter constraint might come to. It will still be methodologically inappropriate to accept *T* as well confirmed unless there is reason to believe that the experiments involved in the production of these data were well designed. There must have been experimental controls for the influence of factors irrelevant to the assessment of *T*; in particular, the data which appear to confirm *T* must not be artifacts of the design of the experiments in question rather than genuine tests of the empirical adequacy of *T*. The analogous constraint applies, of course, to the case in which *T* is apparently disconfirmed by *E*.

3. *The question of "sampling."* Suppose that *T* represents a projectable pattern and that the experiments whose results are reflected in *E* are individually well designed. If we now ask how well, or to what extent, *T* is confirmed by *E*, we face head on the methodological analogue of the pure epistemologist's problem of induction. *T* will typically have infi-

nitely many different observational consequences, and the problem of assessing the extent to which *E* confirms *T* comes down to the question of which (typically relatively small) finite subsets of those consequences are such that their confirmation bestows significant confirmation on all the rest. (For the realist, of course, the problem is broader; one needs to know which such subsets bestow significant confirmation on the theory taken literally as a description of [partly] unobservable reality. Here, as in the case of the definition of reliability of methodology, I frame the issue in a way which does not beg the question against the antirealist.)

We may, I think, frame this question in a revealing way. The question is whether the consequences of *T* which have been tested are—in an epistemically appropriate sense—*a representative sample* of all the observational consequences of *T*. We cannot have checked out all of the (epistemically) possible ways in which *T* could "go wrong" with respect to observational prediction; there are, after all, infinitely many such ways. What we want to know is whether the experimental studies in question involve a representative sample of those ways, so that, if *T* hasn't gone wrong where we've tested it, then we can be justified in believing that it isn't going to go (very far) wrong at all. (Actually, this description is somewhat idealized; in the actual history of science, well-established theories have often turned out to be very wrong indeed in some of their empirical predictions. What is important is that we—rightly—expect experimental confirmation of a theory to warrant our belief that it will prove instrumentally reliable in a wide range of applications whose limits we cannot set in advance. The problem of identifying a relevantly representative sample of the observational predictions of a proposed theory is hardly rendered easier by this complication.)

The Theory-Dependence of the Answers to These Questions

The ways in which scientists answer these fundamental questions regarding the assessment of experimental evidence are quite profoundly dependent upon their prior theoretical commitments. That this is generally true of scientific methodology is now uncontroversial; Kuhn, Quine, both H. Putnams, Goodman, Glymour, and van Fraassen have all emphasized this point without, of course, all drawing realist conclusions. It will be important for our purposes to examine in some detail the ways in which theoretical considerations are involved in answering the three questions about experimental evidence which we have just identified.

1. *Projectability.* Kuhn correctly insists that in mature sciences the basic form of solutions to particular research problems is tightly circumscribed by the theoretical and research tradition (the "paradigm"), and van Fraassen agrees that the acceptance of particular theories involves the scientist "in a certain sort of research programme." The proposed theory *T* whose degree of confirmation by *E* is to be estimated will not be a serious candidate for confirmation at all unless it arises as a proposed solution to some problem: the extension of an existing theory to some new area of application, perhaps, or the explanation of some particular phenomena or observations. The theoretical tradition very sharply constrains such proposals; a proposed solution is unacceptable unless it is *theoretically* plausible in the light of existing theories, unless it is one of the solutions suggested by the existing "paradigm." Only those patterns in observable data are considered projectable which correspond to theoretically plausible theoretical proposals. Two facts about the theory-dependence of such projectability judgments are important for our concerns.

In the first place, such judgments sharply limit the generalizations *about observables* which we take to be confirmable. Suppose, as is typically the case, that *T* is put forward to account for some particular (finite) set of observational data. Of course, there will be infinitely many possible theories which would accommodate those data. Even if we take two such theories to be equivalent if they are empirically equivalent (or, better, if their respective integrations into the existing theoretical tradition would be empirically equivalent), there will remain infinitely many equivalence classes, each representing one possible observational generalization from the initial data. The effect of our theory-dependent judgments of projectability is to restrict our attention to a quite small finite number of these possible generalizations. Only the generalizations in this small set are potentially confirmable by observations, given the prevailing standards for the assessment of scientific evidence.

Secondly, the projectability judgments in question are genuinely *theory*-dependent. The judgments of theoretical plausibility which these projectability judgments reflect depend upon the *theoretical* structure both of the proposed solutions and of the received theoretical tradition. Proposed problem solutions are plausible, for instance, when the unobservable mechanisms they postulate are relevantly similar to the mechanisms postulated in the received theoretical tradition, where the relevant respects of similarity are likewise dependent on the theoretical structures postulated in the tradition. If the received body of theories were replaced by some quite different but empirically equivalent body of theories, then judgments of theoretical plausibility would pick out quite different problem solutions as acceptable and thus typically identify quite different patterns as projectable. As Kuhn insists, the ontology of the received "paradigm" is crucial in determining the range of acceptable problem solutions (and thus the range of projectable patterns in data).

2. *Experimental artifacts.* Suppose that T is theoretically plausible and thus represents a projectable pattern in observable data, and suppose that the experimental results in E appear to support (or refute) T. If these results are really to be evidentially relevant, then there must be reason to think that the results favorable (or unfavorable) to T were not the result of features of the experimental situation which are irrelevant to the assessment of T. Of course, it is impossible to control for all epistemically possible experimental artifacts (of which there is an infinite number). Instead, we rely upon established theory to indicate the conditions under which the presence of experimental artifacts is to be suspected and the sorts of experimental controls which will permit us to avoid or discount for their effects. This is, I think, uncontroversial. It is also uncontroversial—although it is not much stressed in the literature—that our theory-dependent judgments in this area cut down the number of epistemically possible artifactual effects we actually control for from infinitely many to rather few.

What may be more controversial is whether or not these judgments are theory-dependent in the broader sense that they depend on the *theoretical* structure of the relevant background theories rather than just on their observational consequences. It might seem that they do not. Consider, for example, the commonplace that one must, in experiments involving electrical phenomena, control for the 60Hz hum induced by the alternating current in ordinary electrical wiring. Of course, the background theories which draw our attention to this sort of possible artifact have a complex theoretical structure, postulating electrons and electrical and magnetic fields and so forth. But, in order to know that we must shield various pieces of apparatus, all we need to know is that unless we do there will appear a certain sort of signal in our recording equipment superimposed on whatever signal comes from the preparation we are studying. If this sort of situation always obtains in cases of controlling for experimental artifacts, then it would appear that the methodological judgments which govern such controls do not depend on the theoretical structure of the relevant background theories.

Even if this were the case, there would, of course, be a significant way in which the identification of necessary experimental controls depends on the *theoretical* structure of the theories in the relevant theoretical tradition: the judgments of "projectability" which governed the acceptance of the generalizations about observables reflected in the currently accepted theories would have themselves depended on the *theoretical* structure of the earlier stages in the theoretical tradition. More importantly, it is by no means the case that the identification of relevant possible experimental artifacts depends solely on the observational consequences of the relevant background theories. This is so for two related reasons. In the first place, sound methodology often requires that we control for possible experimental artifacts whose effects are not by any means *predicted* by the received body of the theories but whose interference with the intended function of the experimental apparatus is *suggested* by those theories. Whatever may be the ultimate "rational reconstruction" of our practice, it is true that, in the typical case, the way in which the possible artifactual effects are suggested is that there are (typically unobservable) mechanisms postulated by the received theories about which it is *theoretically* plausible that either these mechanisms or mechanisms similar to them will produce the artifactual effects in question. Thus, we identify relevant possible experimental artifacts by something like "inductive" inference from theoretical premises, and the sorts of possible artifacts which we thereby identify depend dramatically on the theoretical structure of the theories which are the premises of these inferences.

We may see the same sort of theoretical-structure-dependent inferences in another methodologically important strategy for the identification of relevant possible experimental artifacts. Good methodology often requires that we control in one experimental situation E for some possible artifact A because we have already encountered similar artifacts A' in similar experimental situations E'. In the typical case, the possible artifact will be described in partly theoretical language, and the relevant respects of similarity (between E and E' and between A and A') will be determined by theoretical considerations—by considerations about the structure and effects of the unobservable mechanisms which the received theories postulate as operating in the relevant natural systems. In this case, too, whatever the ultimate reconstruction might be, the theoretical structure of the accepted theories, and not just their observational consequences, plays a crucial role in the identification of the relevant possible artifacts.

Two important points of similarity thus emerge about the way in which sound scientific methodology controls for the possibility of experimental artifacts and the way in which the problem of projectability is solved. In the first place, while there are infinitely many epistemically possible experimental artifacts which might affect any given experiment, scientific attention is paid to only a small finite number. In this regard, the identification of relevant possible experimental artifacts resembles the assessment of projectability: from an infinity of epistemic possibilities, the scientific method identifies a small finite number as methodologically relevant. The identification of relevant possible experimental artifacts resembles the solution to the problem of projectability in another crucial way: in each case the relevant methodology depends on the theoretical structure of the currently accepted scientific theories; were those theories replaced by others which are empirically equivalent but theoretically divergent, quite different methodological practices would be identified as appropriate. In both cases, scientists behave as though their methodology were determined by inductive inferences from the theoretical principles embodied in the received theoretical tradition.

3. *Sampling*. The pattern discernible in our examination of the ways in which scientific methodology handles the issues of projectability and experimental artifacts is even more striking in the case of the solution to the problem of "sampling." It is a fair statement of the most basic methodological principle governing the assessment of experimental evidence that a proposed theory T should be tested under conditions representative of those in which it is most reasonable to think that the theory will fail, if it's going to fail at all. The identification of these conditions rests upon *theoretical* criticism of that theory. The proposed theory T will, typically, postulate various mechanisms, entities, processes, etc., as factors in the phenomena to which it applies. Theoretical criticism involves the identification of alternative conceptions of the mechanisms, processes, etc., involved which are theoretically plausible—that is, which are suggested by the sorts of mechanisms, entities, etc., which are postulated by the received body of theories. These theoretically plausible alternatives to T will suggest circumstances in which the observational predictions of T might be expected to be wrong. It is under (representative instances of) these circumstances that T must be tested if it is to be well confirmed. This is the central methodological principle of experimental design.

Plainly, the methodological solution to the problem of sampling is theory-dependent. Moreover, whatever the "rational reconstruction" of this methodology might be, scientists do not in practice distinguish sharply between unobservable mechanisms, processes, entities, etc., and observable ones in identifying ways in which a proposed theory might reasonably be expected to fail. Indeed, inferences which look for all the world like inductive inferences from accepted premises about unobservables to conclusions about unobservables play an absolutely crucial role in the sort of theoretical criticism we are discussing. Thus, in the present case, as in the case of the methodological solutions to the problems of projectability and of experimental artifacts, the ways in which scientific methodology is theory-dependent are such that, if the existing body of theories were replaced by an empirically equivalent but theoretically divergent body of theories, our methodological judgments *regarding the "degree of confirmation" of generalizations about observables* would be profoundly different.

Projectability and Induction About Unobservables

The ways in which the features of scientific methodology just discussed depend on the theoretical structure of the received body of background theories may be seen more clearly if we consider a standard way in which philosophers in the tradition of logical positivism have treated the feature of theory testing which I have presented under the heading "Sampling." It has been widely recognized that at any given time in the history of science, and for any given problem or issue, there are typically only a very few theories "in the field" and contending for acceptance. The practice of testing a proposed theory against its most plausible rivals might, in this context, be seen as simply an application of the same pragmatic principle which dictates that, if there are only a very few brands of band saw available, one should evaluate each before making a purchase. This sort of description gives the appearance of reducing the methodological principle we have been discussing to a *merely* pragmatic level, denying it any special epistemic relevance.

Of course, such an interpretation *would not* deprive the practices we have been discussing either of their theory-dependence or of their epistemic importance. So long as the relevant rival theories are identified in the theory-dependent way described in the section on projectability, then the practice would be as theory-dependent as one could wish. Moreover, if testing proposed theories against instrumental rivals in the way suggested represented *the* methological solution to the problem of "sampling," then it has epistemic importance however much it may also have a purely pragmatic justification. What is most interesting in this context, however, is that the actual methodological practice of scientists departs in a revealing way from that suggested by the pragmatic account. In order for the pragmatic picture to have any plausibility, we must think of a proposed theory as competing against other possible predictive instruments roughly as powerful as itself. The rival "theories" against which it must be tested must be theories in the sense of fairly well developed systems with some significant predictive power. One does, after all, test band saws against other band saws.

While it is true that sound scientific methodology does require that a proposed theory should be tested against similarly well articulated rivals which are approximately equally theoretically plausible (roughly, that's what being a rival *theory* means, beyond having been invented in the first place), what is striking about the methodological practices which constitute the solution to the problem of sampling is that they may also require that a proposed theory be tested against a mere hunch, which has no deductive predictive consequences whatsoever. Suppose that a proposed theory T postulates a particular sort of unobservable mechanism as operating in the systems to which T applies, and imagine that T is sufficiently well worked out that (using well-established auxiliary hypotheses) it is possible to obtain experimentally testable predictions from T. Suppose, also, that theoretical criticism of T identifies alternative possible mechanisms, plausible in the light of received theories. Under these circumstances, it becomes necessary to try to pit T's conception of the matter against the alternative in some sort of experimental situation *even if* the alternative account is not nearly so thoroughly worked out as T and even if, for that reason, it yields (together with relevant auxiliary hypotheses) *no* definite predictions about observables at all. Under these circumstances, what sound methodology dictates is the identification of experimental circumstances under which the sorts of observations which it is *theoretically plausible* to expect given the alternative conception are different from those which one would expect given T's account of the relevant mechanisms.

By way of example, suppose that T provides an account of the reaction mechanisms for some biochemical process and that T is worked out in sufficient detail that it has (together with well-confirmed auxiliary hypotheses) significant deductive observational consequences. Suppose that a rival conception of the relevant reaction mechanisms is suggested by theoretically plausible considerations but that this rival conception is insufficiently well developed to have specific testable deductive consequences. It might nevertheless be possible to test *T* against the rival conception. Suppose that, in the case of better-studied systems, those systems to which mechanisms like those proposed by *T* and ascribed by the received theories are much more sensitive to some particular class of chemical agent than those to which mechanisms like those proposed in the alternative are ascribed (note here that the relevant respects of likeness will be determined by the content of theoretical descriptions of the systems in question, and the theoretical content of the relevant background theories, *and* that the class of chemical agents in question may similarly be theoretically defined). Under such circumstances, sound methodology will dictate subjecting the biochemical systems to which *T* applies to chemical agents in the relevant class; data indicating considerable sensitivity to such agents will be especially important for the confirmation of *T* precisely because they will constitute a test of *T* against the theoretically plausible rival conception of the relevant reaction mechanisms. Note that in the present case the rival conception need not have any *deductive* observational consequences regarding the experimental situations in question. Instead, reasoning by analogy *at the theoretical level* makes it *theoretically plausible* to expect low sensitivity if the rival conception is true. *T* is tested against a theoretically plausible hunch about how it might go wrong.

Indeed, the role of considerations of theoretical plausibility in theory testing can go even deeper; a proposed theory may be pitted against a theoretically plausible rival in a particular experimental setting even though neither the rival *nor* the proposed theory have (when taken together with appropriate well-confirmed auxiliary hypotheses) any deductive observational predictions about the results of the experiments in question! In the example we have been considering, the appropriateness of the experimental test in question does not depend on the theory *T*'s having any observational deductive predictions about the results of the experiment. All that is required is that it be *theoretically* plausible that a test of the sensitivity of the relevant biochemical systems to the specific chemical agent will provide an indication of which of the two accounts of reaction mechanisms (if either) is right. We may test *T* by pitting a hunch about the outcome of experimentation which is theoretically plausible given *T* (and the body of received theories) against an experimental hunch which is theoretically plausible on the assumption of the rival conception of reaction mechanisms. Even though we have assumed that *T* makes a significant number of deductive observational predictions, we need not assume that it makes any *deductive* predictions about the outcome of this crucial experimental test! In sciences which deal with complex systems, instances of theory testing which fit the model just presented are by no means uncommon. Indeed, it may be a good idea to ask whether in describing the instrumental application of theories (rather than their confirmation)—when defining empirical adequacy, for example—the idealization that it is the *deductive* observational consequences of a theory (together with auxiliary hypotheses) rather than its *inductive* consequences that are relevant may not be fundamentally misleading; but that is a topic for another paper.

In any event, what we may learn from these examples is that, in practice, inductive inferences in science extend to inferences with theoretical premises and theoretical conclusions. Just as there are theory-dependent judgments about which possible patterns in observables are projectable, so there are judgments about which patterns in the properties or behavior of "theoretical entities" are projectable. Just as there are theory-dependent judgments of the "degree of confirmation" of instrumental claims by empirical data, so there are theory-dependent judgments of the plausibility of various theoretical claims in the light of other considerations both empirical and theoretical. Indeed, whatever the correct philosophical analysis of this matter, scientific methodology does not dictate any significant distinction between inductive inferences about observables and what certainly look like inductive inferences about unobservables. Finally, and most strikingly, the very methodological principles which govern scientific induction about observables are, in practice, parasitic upon "inductive" inferences about unobservables.

An Argument for Scientific Realism

It will be evident how one may argue for scientific realism on the basis of the theory-dependence of experimental methodology. Consider the question, why are the methodological practices of science instrumentally reliable? Both scientific realists and (almost all) empiricists agree that these practices are instrumentally reliable, but they differ sharply in their capacity to explain this reliability. So theory-dependent are the most basic principles for the assessment of experimental evidence that it must be concluded that these are principles for applying the knowledge which is reflected in currently accepted theories as a guide to the proper methods for the evidential assessment of new theoretical proposals; any other conclusion makes the instrumental success of the scientific method a miracle.

According to the empiricist, the knowledge reflected in the existing body of accepted theories at any time in the history of science is entirely instrumental knowledge: the most we know on the basis of experimental evidence is that the existing body of theories is empirically adequate. Thus, the replacement of existing theories by an empirically equivalent set of theories would leave the knowledge they embody unchanged. Thus, the empiricist can explain the epistemic adequacy of only those theory-dependent features of scientific methodology whose dictates are preserved under the substitution, for the actual body of accepted theories, of any other empirically equivalent one. But, as we have just seen, *none* of the central methodological principles which govern the evaluation of scientific evidence have this property! The consistent empiricist cannot explain the instrumental reliability of the methodology which scientists actually employ.

The scientific realist, on the other hand, has no difficulty in providing the required explanation. According to the realist, existing theories provide approximate knowledge not only of relations between observables, but also of the unobservable structures which underlie observable phenomena. In applying theory-dependent evidential standards, scientists use existing theoretical (and observational) knowledge as a guide to the articulation and experimental assessment of new theories. The judgments of projectability, identification of experimental artifacts, and theoretical criticisms of proposed theories which look ever so much like inductive inferences *are* inductive inferences from acquired theoretical knowledge to new theoretical conclusions. When a theoretical proposal is theoretically plausible in the light of the existing theoretical tradition, what that means is that it is supported by an inductive inference at the theoretical level from previously acquired theoretical knowledge.

Judgments of "projectability" are thus just what they look like "preanalytically": they represent the identification of theoretical proposals for which there are good inductive reasons to believe that they are (approximately) true and thus for which there is good reason to believe that they will eventually be articulated into empirically adequate theories. The role of experimentation is to choose between the various theoretical proposals which pass this preliminary test for probable (approximate) truth.

Similarly, the judgments of theoretical plausibility by which possible experimental artifacts are identified turn out to be inductive inferences from theoretical knowledge which result in reliable assessments of the evidential likelihood that various unobservable factors will influence the outcome of experiments. Finally, the methodological solution to the problem of sampling really does consist in identifying—by reliable inductive inference from theoretical knowledge—the most plausible rivals to a proposed theory and the experimental conditions under which they can be effectively pitted against it. The reliability of scientific methodology in guiding induction about observables turns out to be largely parasitic upon the reliability of the methodology in applying existing theoretical knowledge to guide the establishment of new theoretical knowledge. . . . Only this explanation, the realist maintains, can account both for the reliability of the scientific method and for the fact that seemingly inductive reasoning about theoretical matters is so central to it.