'$\sim Fb \vee \sim Ga \vee Ha$'. Kim shows generally that his additional requirement blocks the proofs offered by Eberle, Kaplan, and Montague for the five theorems that "trivialize" the defintions (7.8) and (7.9). However, it would be desirable to ascertain more clearly to what extent the additional requirement is justifiable, not on the *ad hoc* ground that it blocks those proofs, but in terms of the rationale of scientific explanation.

Kaplan approaches the problem by formulating three very plausible requirements of adequacy for any analysis of the deductive type of explanation here to be explicated. He then shows that the analysis proposed in Part III does not satisfy those requirements jointly, and that the difficulties exhibited in the five trivializing theorems are linked to this shortcoming. Finally, he revises the definitions offered in Part III so that they meet the requirements of adequacy and avoid the difficulties we have been discussing. For the details of this illuminating contribution, the reader will have to consult Kaplan's article.

Notes

1. C. Hempel, "The Function of General Laws in History," in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press, 1965, pp. 35–48.
2. Hempel, "Aspects of Scientific Explanation," op. cit., pp. 331–496.
3. E. Nagel, *The Structure of Science*. New York, 1961, p. 58.
4. Op. cit., p. 59.
5. R. Eberle, D. Kaplan, and R. Montague, "Hempel and Oppenheim on Explanation," *Philosophy of Science* 28 (1961), pp. 418–28.
6. D. Kaplan, "Explanation Revisited," *Philosophy of Science* 28 (1961), pp. 429–36.
7. J. Kim, "Discussion: On the Logical Conditions of Deductive Explanation," *Philosophy of Science* 30 (1963), pp. 286–91.

3

Explanations, Predictions, and Laws

Michael Scriven

3. Preliminary Issues

3.1. Explanations: Introduction

I am going to take a series of suggested analytical claims about the logic of explanation and gradually develop a general idea of what is lacking in them, or too restrictive about them. I shall at each stage try to formulate criteria which will survive the difficulties while retaining the virtues of the current candidate. Eventually I shall try to draw the surviving criteria together into an outline of a new account of both explanation and understanding; but this will not be possible until I have encompassed the whole field of topics envisioned above. The questions with which I begin will seem quite unimportant; but they are in fact more significant than they appear because of the cumulative error in the standard answers to them.

3.2. Explanations as Answers to "Why" Questions

"To explain the phenomena in the world of our experience, to answer the question 'Why?' rather than only the question 'What?' . . ." With these words, Hempel and Oppenheim begin their monograph on scientific explanation.[1] Braithwaite says, "Any proper answer to a 'Why?' question may be said to be an explanation of a sort. So the different kinds of explanation can best be appreciated by considering the different sorts of answers that are appropriate to the same or to different 'Why?' questions."[2]

This happens to be a *non sequitur*, but it is the conclusion that I particularly wish to consider. The "answer-to-a-why-question" criterion could have been proposed in the absence of a serious attempt to think of counterexamples. "How can a neutrino be detected, when it has zero mass and zero charge?" is a perfectly good request for a perfectly standard scientific explanation. "What is it about cepheid variables that makes them so useful for the determination of interstellar distances?" Likewise, the same can be said of suitable Which, Whither, and When questions.

Not all, perhaps not even most, of the answers to such questions involve explanations, whereas it is perhaps true that most answers to Why questions are explanations. (But not all, as for example an answer that rebuts a presupposition of the question "Why do you persist in lying?" "I have *never* lied about this affair.") But explanations are also given when no questions are asked at all, as in the course of a lecture, or in correcting or supporting an assertion. The identifying feature of an explanation cannot, therefore, be the grammatical form of the question which (sometimes) produces it. Indeed, it is fairly clear that one does not teach a foreigner or a child the word "explanation" simply by reference to Why questions; so the authors quoted presumably had some prior (or at least alternative) notion of explanation in mind which enabled them to identify answers to Why questions as explanations. Should we not look for the meaning of that notion?

It is sometimes replied that our common notion of explanation is excessively vague, and it is therefore quite unrewarding to seek its exact meaning; far better to concentrate on some substantial concept which clearly does occur. This is a very good reply and represents a sensible approach, if only it can be shown to be true. This requires showing (a) that the ordinary notion *is* excessively vague, and (b) that the "substantial" alternative occurs often enough to justify any general conclusions about explanations which are inferred from studying it. I shall be arguing that neither of these seemingly innocuous premises can be established, and in consequence the analysis suggested by the reconstructionish authors is fundamentally unsatisfactory.

Explanation is undoubtedly a notion whose analysis must be sought in the practical foundation of language; but it is too much to hope one can identify explanations by such a simple linguistic device as the one suggested. Nor will it do to suppose that explanations are such that they are answers to *potential* Why questions; for then they are also potentially answers to What-about questions, How-possibly questions, etc. Thus, to take an example quoted by Hempel and Oppenheim, the question "Why does a stick half-immersed in water appear bent?" can readily be rephrased as "What makes a stick (in such circumstances) seem to be bent?" Indeed, such a question as "How can the sun possibly continue to produce so much energy with a negligible loss of mass?" is only with some difficulty rephrased as a Why question.[3] In sum, the grammatical indicators of explanations are complicated and none of them are necessary; some more illuminating and reliable criteria must be sought.

3.3. *Explanations as "More Than" Descriptions*

Another common remark in the literature is that explanations are more than descriptions. This is put by Hempel and Oppenheim in the following words: ". . . especially, scientific research in its various branches strives to go beyond a mere description of its subject matter by providing an explanation of the phenomena it investigates."[4] But if one goes on to examine their own examples of explanations one finds what seem to be simply complex descriptions.[5] Thus they offer an explanation of the fact that when "a mercury thermometer is rapidly immersed in hot water, there occurs a temporary drop of the mercury column, which is then followed by a swift rise." And the explanation consists of the following account: "The increase in temperature affects at first only the glass tube of the thermometer; it expands and thus provides a larger space for the mercury inside, whose surface therefore drops. As soon as by heat conduction the rise in temperature reaches the mercury, however, the latter expands, and as its coefficient of expansion is considerably larger than that of glass, a rise of the mercury level results."[6]

This is surely intended to be a narrative description of exactly what happens. The one feature which might suggest a difference from a "mere description" is the occurrence of such words as "thus," "however," and "results." These are *reminiscent* of an argument or demonstration, and I think partially explain the analysis proposed by Hempel and Oppenheim, and others. But they are not part of an argument or demonstration here, simply of an explanation; and they or their equivalents occur in some of the simplest descriptions. "The curtains knocked over the vase" is a description which includes a causal claim and it could equally well be put, style aside, as "The curtains brushed against the vase, *thus* knocking it over" (or ". . . resulting in it being knocked over"). The fact that it is an explanatory account is therefore not in any way a ground for saying it is not a descriptive account (cf. "historical narrative"). Indeed, if it was not descriptive of what happens, it could hardly be explanatory. The question we have to answer is how and when certain descriptions count as explanations. Explaining how fusion processes enable the sun to maintain its heat output consists exactly in describing these processes and their products. Explaining therefore sometimes consists simply in giving the *right* description. What counts as the right description? Tentatively we can consider the vague hypothesis that the right description is the one which fills in a particular gap in the understanding of the person or people to whom the explanation is directed. That there is a gap in understanding, or a misunderstanding, seems plausible since whatever an explanation *actually* does, in order to be called an explanation at all it must be *capable* of making clear something not previously clear, that is, of increasing or producing understanding of something. The difference between explaining and "merely" informing, like the difference between explaining and describing, does not, I shall argue, consist in explaining being something "more than" or even something intrinsically different from informing or describing, but in its being the appropriate piece of informing or describing, the appropriateness being a matter of its relation to a *particular context*. Thus, what would in one context be "a mere description" can in another be "a full explanation." The distinguishing features will be found, not in the verbal form of the question or answer, but in the known or inferred state of understanding and the proposed explanation's relation to it. To these, of course, the form of the question and answer are often important clues, though not the only clues. But this is only a rough indication of the *direction* of the solution to be proposed in this paper, and it may be that the notion of understanding will present us with substantial difficulties, quite apart from the problem of identifying the criteria for "closing the gap" in understanding (or rectifying the misunderstanding). However, let me remind the reader that understanding is *not* a subjectively appraised state any more than knowing is; both are objectively testable and are, in fact, tested in examinations. We may first benefit from examining the relation between explanation and another important scientific activity.

3.4. *Explanations as "Essentially Similar" to Predictions*

The next suggestion to be considered is a much more penetrating one, and although it cannot be regarded as satisfactory, the reasons for dissatisfaction are more involved. Quoting from Hempel and Oppenheim once more: ". . . the same formal

analysis . . . applies to scientific prediction as well as to explanation. The difference between the two is of a pragmatic character . . . It may be said, therefore, that an explanation is not fully adequate unless . . . if taken account of in time, [it] could have served as a basis for predicting the phenomenon under consideration."[7]

(3.41) The full treatment of this view will require some points that will only be made later in the paper; but we can begin with several rather weighty objections. First, there certainly seem to be occasions when we can predict some phenomenon with the greatest success, but cannot provide any explanation of it. For example, we may discover that whenever cows lie down in the open fields by day, it always rains within a few hours. We are in an excellent position for prediction, but we could scarcely offer the earlier event as an explanation of the latter. It appears that explanation requires something "more than" prediction; and my suggestion would be that, whereas an understanding of a phenomenon often enables us to forecast it, the ability to forecast it does not constitute an understanding of a phenomenon.

(3.42) Indeed, the forecast is simply a description of an event (or condition, etc.) given prior to its occurrence and identified as referring to a future time; whereas an explanation will have to do more than merely describe those *features of the thing to be explained that identify it*. (In this sense, it is more than a (particular) description.[8]) At the very least some other features of it must be mentioned, and often some reference is made to previous or (other) concurrent events and/or laws. Since none of this is required of a prediction, it seems rather extraordinary to suppose that the *contents* of a prediction are logically identical to those of an explanation. And our first point showed that the *grounds* for the two are often quite different, in that one can be inferred from a mere correlation and the other not.

Such cases also demonstrate that explaining something is by no means the same as showing it was to be expected, since the latter task can be accomplished without any explanation being given.[9] For our purpose, however, the crucial point is that, however achieved, a prediction is what it is, simply because it is produced in advance of the event it predicts; it is *intrinsically* nothing but a bare description of that event. Whereas an explanation of the event *must* be more than the *identifying* description of it, else to request an explanation of X (where "X" is a description, not a name) is to give an explanation of X. Of course, there is usually a difference of tense, but we could agree to this as a pragmatic difference. However, it is the *least* and not the *only* difference between explanations and predictions.

(3.43) There also seem to be cases where explanation is not in terms of temporally ordered and causally related events, and we are consequently never able to make predictions. These cases are common enough outside the physical sciences, for example, in explaining the rules of succession in an Egyptian dynasty, or the symbolism of a tribal dance. Within science there are of course all the cases of explaining a theory or mechanism or proof; these are normally dismissed by supporters of the Hempel and Oppenheim position, on the grounds that they are clearly a different kind of explanation, the explanation of *meaning*, not at all related to *scientific* explanation. While there is no doubt about the difference in procedure between explaining a theory and explaining some phenomenon in terms of the theory, it is not enough to appeal to intuition for support of the claim that they are not "fundamentally," that is, for all logical purposes, the same except for subject matter, much as definition in mathematics might be said to be fundamentally the same as

definition in the empirical sciences except for subject matter. In fact, it seems clear enough that one important element is held in common between the two "kinds" of explanation, viz. the provision of understanding. But is there not a great deal of difference between the kinds of understanding provided in the two cases?

Now, *not* understanding a theory may be due to not understanding what its assumptions are, to not understanding the meaning of some of its terms, or to not understanding how the derivations said to be possible from it are to be made. One might suppose this to be quite unlike not understanding why a stick half-immersed in water appears bent.

But instead of asking how we go about explaining a natural phenomenon, let us ask how we come to ask for an explanation, that is, what it is that we think *needs* explanation. It may seem that science is committed to the view that *everything* needs to be explained. Now it is clear that *everything* cannot be explained *every* time we give an explanation of some particular thing (or set of things) which is all we ever do in a given context. So we can rephrase our question as, What is it that needs explanation in a given context? It seems clear that it is those things which are not properly understood (by whomever the explanation is addressed to). Now, lack of understanding of a natural phenomenon may be due to the absence of certain information about the situation, to the presence of false beliefs about it, or to an inability to see the connections between what is understood and what is not understood. These are much the same kinds of difficulty as occur in not understanding a theory, although the information will be in one case about a verbal construction out of our knowledge and in the other about, for example, a mechanical construction out of our raw materials. However important the differences of subject may be, it is not obvious that the notion of understanding or explanation involved is in any important way different; and it is quite obvious that no predictions are possible on the basis of an explanation of the meaning of a theory (except, irrelevantly, those which the theory makes possible, if any—and it may be a theory whose advantages lie solely in its unifying powers). Certainly one should feel uneasy about any general claims of common logical structure for explanation and prediction which have to be defended by rejecting clear cases of explanation as "essentially different," without detailed examination. I shall argue that the differences are much less important than the similarities: in effect we are in both cases providing a series of comprehensible statements that have some of a wide range of logical relations to other statements. Lest this seem to be a proof of similarity by simply weakening the definition of 'similar,' I also try to show that the narrower definition is independently unsatisfactory.

(3.44) Again, we often talk of *explaining laws*: indeed half of Hempel and Oppenheim's examples are of this kind. Now, when we offer an explanation of Newton's Law of Cooling (that a body cools at a rate proportional to the difference between its temperature and that of its surroundings), we do so—according to Hempel and Oppenheim—by deriving this law from more general laws.[10] What predictions could be made which would have "the same formal structure" as this kind of explanation? The "pragmatic" difference between the two as they see it is essentially that explanation occurs after the phenomenon, and prediction before. But in the case of laws, which are presumably believed to hold at all times, what does it mean to talk of predicting the phenomenon? It is surely the case that the truth of Newton's

law is *simultaneous* with that of the more general laws from which it is derivable. We cannot speak of being able to predict the inclusion of the class of A's in the class of C's if we already know that A's are B's and B's are C's.

It may seem that this argument can only be countered by saying that a law is a generalization about a number of events and that to "predict a law" is to predict the outcome of experiments done to determine the pattern of these events. It is true that this is different from predicting an eclipse, where the actual event to which the prediction refers is in the future, not merely the discovery of its nature. However, (i) it is certainly true that *some*—but not all—events governed by most laws lie entirely in the future, and its truth depends on these and these are predictable in the usual sense. (ii) Certainly, too, we want to say that inferences about the past, which *generate* predictions about what archaeologists or geologists will discover, have exactly the same logical structure as inferences about the future, a fact well brought out by the practice of calling them postdictions or retrodictions. So, if *explaining a law* consists in explaining the overall pattern of events, past, present, and future, *predicting a law*, as it seems we might interpret it, could be regarded as compounded out of such predictions and postdictions. This interpretation represents at the very least an *extension of meaning*, since we cannot in the usual sense call inferences about the activities of the earth's crust in Jurassic times (which will be covered by geological laws) predictions. Although we could quite properly apply this term to inferences about what will be found by geologists upon searching in certain areas, this is *not* what the law is about (for else we must say that the law asserts something different every time something new is discovered by the geologists, there being that much less for them still to discover). Indeed, we land in a well-known swamp if we make this move; for the same argument makes all historical statements into statements about contemporary evidence and all statements about distant places into statements about local evidence, etc. It is the argument which confuses the *reference of* a statement with the *evidence for* a statement. So explanations of laws only have a correlate among predictions if we *extend* the meaning of the notion of prediction to include postdiction.

Now this extension may not seem very significant until one reminds oneself that the whole significance of the term "prediction" resides in the temporal relation of its utterance to the event it mentions. "Prediction" is a term defining a category of sentences, in the same way as "command," "argument," "description," etc.; it defines descriptive sentences in the future tense made when the tense is appropriate. The sentence uttered or written in making any given prediction can be repeated after the event has occurred as a perfectly good historical description, provided only that the tense of the verbs (or the corresponding construction) is changed, that is, *apart from tense*, predictions are not identifiable. So this extension of meaning amounts to an *elimination* of the meaning with which one began. We may agree that one procedure of inferring past events is essentially the same as one procedure of inferring future events, but we cannot possibly conclude that the *results* in the first case are essentially the same as predictions. This is like saying that analytic statements are essentially the same as synthetic statements since both can be inferred syllogistically. If the *only way* of inferring to such different kinds of statement was syllogistic, one *might* be more inclined to call them logically identical. One would still not be very impressed, since their logical character is written on their face, and appealing to their common ancestry cannot prove that all siblings are twins; it cannot

eliminate the obvious differences. But it is clear that there is nothing unique about the type of inference suggested here; predictions and postdictions can be obtained from arguments of virtually any logical form and also without any argument at all, as in the case of the expert but inarticulate diagnostician or the precognitive. I conclude that the explanation of laws has no proper counterpart among predictions, since there is no general concept of predicting laws; for (i) if what can be predicted is said to be the *discovery* of a law, this fails because the counterpart to explaining an *event* is predicting it, not its *discovery* (which would require laws about discovering laws); (ii) in the only other possible interpretation, a large number of the conclusions inferred are simply not predictions at all; (iii) even ignoring the first two points, nothing is more obvious than the difference in logical structure between the "prediction" "All A's are (or even, will be found to be) C's" and anything that might conceivably be said to be an explanation of it.

(3.45) I think little can be salvaged from the impact of this set of four points (3.41 through 3.44) against the 3.4 thesis, but I wish to indicate another series of difficulties which will help us to develop a constructive alternative position. The first involves a rather lengthy example, but the same example is of some assistance in dealing with the notion of cause as well as those of explanation and prediction. Suppose we are in a position to explain the collapse of a bridge as due to the fatigue of the metal in one of the thrust members. This is not an unusual kind of situation, and it is, of course, one where no prediction was *in fact* made. According to Hempel and Oppenheim, if this is a satisfactory explanation, then, if taken account of in time, it could have formed the basis for a prediction. (We can abandon the idea—presumably but incorrectly taken to be equivalent by Hempel and Oppenheim—that it would actually have the same logical structure as the prediction.) Let us examine in a little detail how this could be so.

We begin our search for the explanation with an eyewitness account that locates a particular girder as the first to go. We already know that there is a substantial deterioration in the elastic properties of carbon steel as it ages and is subjected to repeated compressions; we also know that the amount of this deterioration is not predictable with great reliability since it depends on the conditions in the original welding, casting, and annealing processes, the size and frequency of subsequent temperature changes to which the formed metal is subjected, the special stress to which it may have been subjected, for example, by lightning discharges which put heavy currents through it, and, of course, irregularities in and perhaps violations of the design load. The only way to deal with these sources of error is to "overdesign," that is, to make an allowance for the unpredictables and provide a safety factor on top of that. But the cost of materials and the pressure of competitive tenders puts limits on the size of such safety margins, and every now and then, as in the spectacular case of the Launceston Bridge, where the wind set up resonant vibrations, a failure occurs. In the present case, where internal rather than external circumstances are the significant factor in the failure, we obtain samples of the metal from the girder in question and discover that its elastic properties have substantially deteriorated. But as we do not have any exact data about the load at the time of failure, we cannot immediately prove that such a load would definitely have produced failure.

Now we go over the rest of the bridge carefully, searching for other possible causes of failure and find none. The bridge is of standard design, sited on good

bedrock, and well built. We do have good reason to suppose that the load-causing failure was no greater than the bridge had withstood on many previous occasions, though greater than the static load (assume standard traffic and moderate wind); so we are forced to look for the cause in the structural changes. In the light of all this information, we can have great confidence in our explanation of the failure as due to fatigue in the particular beam; but we simply do not have the data required for a prediction that the failure would take place on a certain date.

It is perfectly true that *if* we *also* had exact data about the load when the failure occurred, could obtain some exact and reliable elastic coefficients from the fatigued sample, were in no doubt about our theory, and found on calculation with the revised elastic coefficients that the load exceeded the residual strength, we could be *even more confident* of our explanation. But I have described a much more realistic situation in which we can still have a very high level of rational confidence in our explanation, a level which places it beyond reasonable doubt.

Now, in both cases—with and without the exact details—we *can* make some kind of a *conditional* prediction—that the bridge will fail *if* the load goes over a certain point, for obviously we can give some load which exceeds any known bridge's capacity. It must be noticed first that such a prediction has no practical interest at all except insofar as we can predict the occurrence of such loads. It is a conditional prediction not a categorical prediction, and if the only kind of prediction which is associated with explanations is conditional prediction, especially if they are of this "upper limit" kind, this is of very little interest indeed for scientists or engineers who cannot predict when the conditions are met, or who know they are very rarely met.

These considerations make us realize that the crucial element in the "duality thesis" about explanations and predictions is the existence of a specific correlative prediction for each explanation. Naturally, we can make a number of conditional predictions as soon as, or indeed *before*, we have any data about the material and form of the bridge; but these are independent of the particular circumstances of the failure. The "duality" claim presumably implies that to every different good explanation there corresponds a different prediction relating to precisely those circumstances to which the explanation applies. But it is easy enough to see that we can attain all reasonable certainty about an explanation with less evidence than is required to justify even a conditional prediction with the same *specific* reference.

In the bridge example, we have so far been much too profligate of our investigator's time. In fact, he knows very well that the only causes of failure other than a load in excess of anything for which the structure was originally designed are metal fatigue or external damage by, for example, corrosion, abrasion, or explosion. It is easy to check for the symptoms of external damage, it is relatively easy to judge that the load was not beyond the design limits. Consequently, he can almost certainly identify the cause of failure immediately as fatigue. In suitable circumstances, that is, with suitable evidence for the above statements, it is only a formality to go through with testing samples of the structural steel. Suppose, however, as a final check, we do a rough computation of Young's modulus for material in the beam and find it has substantially decreased and by much more than for the other beams; but we take no exact measurements of it, and none at all of the other elastic coefficients (which normally vary in the same direction as Young's modulus). Our hypothesized explanation has been very strongly confirmed. It is now beyond reasonable doubt. Now

what conditional prediction can we make? It seems there is only the extremely weak one that if sufficient substantial fatiguing takes place, and a somewhat higher load than normal is imposed, failure will take place.[11] Not only does such a "prediction" correspond to an indefinitely large number of explanations and hence fail to meet the uniqueness condition previously mentioned, but it is couched in such vague terms as to be almost wholly uninformative. Yet I wish to maintain that such a prediction is all that can be said to be correlated with some very well-established explanations.

Let us examine the most natural counterargument. This would consist in saying that no such explanation could ever be regarded as certain in view of its lack of precise support. Imagine an attorney for the steel company attacking this explanation in court. "How can you be sure that the metal fatigue was *enough* to produce a failure? You made no calculation and no measurements from which you could in any way infer that a bridge built of the same steel, in the same condition you found it, would fail even under *twice* the load impressed on it that windy night. Hence, I submit that no evidence for blaming the steel has been produced, and hence no evidence that the steel was at fault."

The weakness of this argument as far as *our* considerations are concerned is twofold. First, there is not the least difference between direct and indirect evidence for establishing a conclusion beyond reasonable doubt; indirect evidence is often more reliable and the distinction between the two is largely arbitrary. If only A, B, or C can cause X, and A and B are ruled out, it is unnecessary to show C is present; in this case, however, it was *also* shown that C was present, and the only debate is over whether *sufficient* C was present. The reply to such doubts is simply "What else, then?" A redoubtable prima-facie case has been made, and if not rebutted, it must be accepted.

Second, there are some grounds for doubting the significance of any "direct" test, which do not apply to the indirect evidence. Suppose we take all the exact measurements we can and make all the calculations we can, and they indicate that a bridge made of the metal tested would not collapse under any stress that seems likely to have occurred in the circumstances at the time of failure. Here we have two *conflicting* indications. Far from it being the case that the "direct" test affords the crucial test, we find it substantially less reliable than the other evidence. First the sample of metal tested is not known to be identical with that which failed: we take a sample adjacent to the *fracture*, but it is a very difficult matter to determine where the fracture begins, that is, where the *failure* occurs. It is quite certain that different spots on the same girder—and along the same fracture—are under very different stresses and hence at very different stages of fatigue. Second, the steps involved in going from the data on materials to conclusions about bridge strength involve a vast number of assumptions of various kinds, few of them more than approximations whose errors may in sum be fatal to the argument. For example, it is *possible* that exceptional conditions did prevail in local areas around the bridge structure, producing strains such as would not normally be associated with a moderate storm—a typical example is provided by the random development of wind resonance, which can build up a considerable, though not precisely known, extra force from a mild breeze that happens to be blowing in the right direction at the right velocity. Hence our "direct" calculations by no means settle the matter; and the recent examples of wing failure in the Electra airliner show that fatigue can be identified as the cause of failure even when exact theory is wholly inapplicable. The

moral of this example is that explanations can be supported by assertions about qualitative *necessary* conditions whereas even a conditional prediction requires quantitative *sufficient* conditions. (This point would of course be completely lost if one proceeds on the common assumption that causes are simply sufficient conditions.)

We have thus discovered that the "direct test" of the indirectly supported hypothesis is by no means immune to rejection. But the general issues about confirmation are not important here; it is the existence of *some* cases where we can have every confidence in an explanation and yet be in no position to make a prediction, even an applicable conditional prediction. This counterexample to the "duality" view is the analogue of the counterexamples already mentioned where we are in an excellent position to make a prediction but cannot produce an explanation. A simple and somewhat rough way of putting the point of the last example would be to say that a prediction has to say *when* something will happen, or *what* will (sometime) happen, a causal explanation only *what made it* happen. The first requires either attaching a time or range of times (unconditional prediction), or a value of some other variable (conditional prediction), to the description of an event, whereas the second often requires only giving a cause, that is, picking out (not estimating the size of) a variable, or another event.

Naturally there are some cases where more than these minimum requirements are available. Sometimes the nature of the problem is such that when the explanation is certain the prediction *is* possible: the Farnborough research into the fuselage fatigue of the De Havilland Comet airliners is a case in point. This was possible only because they had excellent data on the circumstances of the failure (from service records plus recovered instruments). The first type of case we have described is of central importance in the social sciences, because most of our knowledge of human behavior can be expressed only in necessary condition propositions or judgment propositions. Hence it enables us to explain but not predict with equal accuracy. We can confidently explain the migration of the Okies to California in terms of the drought in Oklahoma, though we could not have predicted it with any reliability. For we know (i) that there must have been a reason for migrating and (ii) that drought produces economic conditions which can provide such a reason, and (iii) that nothing else with such effects was present. But we do not know *how* much of a drought is required to produce a migration and hence could not have predicted this with any confidence. Hempel mistakenly regards this as grounds for doubting the explanation.[12] We must insist on making a distinction between a dubious explanation and one for which further confirmation—in the technical sense—is still possible: every empirical claim has the latter property.

(3.46) To summarize, in part. The idea that a causal explanation can only be justified by direct test of the conditions from which a prediction could be made is a root notion in the Hempel and Oppenheim treatment of explanation, and they try to give it a precise formulation. It is said that an explanation must have the form of a deduction from (a) causal laws (L_v) connecting certain antecedent conditions (C_v) to certain consequent conditions (E_v), plus (b) assertions that the conditions C_v obtained in the case under consideration, where we are trying to explain X, which is the sum of the conditions E_v. In the bridge example, we would have to show (by appeal to connections involving L_v) that material with the properties of the sample taken (C_1, C_2, ..., C_{n-1}) under the ambient conditions of the failure (C_n, C_{n+1}, ..., C_{m-1}) would lead to the behavior described, that is, X, the collapse of the bridge

(the bridge's design and state prior to failure being described in terms of C_m, C_{m+1}, ..., C_p). I have been arguing that an indirect approach may be just as effective, that is, one showing fatigue (C_1) to be a necessary condition for X under the circumstances (C_2, ..., C_p). This would involve appealing to a proposition of the form "If X occurs, then either A_1, B_1, or C_1 caused it" and showing that C_2, ..., C_p rule out A_1, and B_1. The main trouble with such laws for the thesis of Hempel and Oppenheim is that they do not permit any predictions of X, since the occurrence of X is required for their application. Nor can such laws be reformulated for predictive use, for they are quite different from "If A_1 or B_1 or C_1 occur, they will produce X," not just because one states necessary and the other sufficient conditions, but because the first does not and the second does require quantitative formulation if it is to be true—for it is obviously false that *any* degree of fatigue produces failure. These laws incidentally demonstrate that the duality thesis about explanation and prediction was actually a separate, fifth condition and not a consequence of the four conditions R_1–R_4.

(3.47) In concluding this discussion of the prediction criterion for explanations, I think it is worth mentioning some points which are neither wholly independent of those discussed above nor, it seems to me, quite so strong. First, it is a consequence of Hempel and Oppenheim's analysis that whatever we explain must be a true statement, since they explicitly require all statements in an explanation to be true. Now it is certainly not the case that all the predictions we make must be true: we often err in predicting the behavior of the stock market, the weather, and the ponies. This point is thought by Scheffler to show that explanations and predictions are different in this respect; but I take it to be mainly a difficulty with Hempel and Oppenheim's analysis of explanation. For one can talk of explaining things that do not happen, just as one can talk of the consequences of things that do not happen. "If you hadn't got here on time, I know who would have been responsible," the irate parent says to the almost-wayward daughter; "If the fourth stage had failed to (ever fails to) fire, you may be sure it would have been (will be) because of a valve failure in the fuel-line," the missile technician may say. This use is *derivative*, that is, it can be explained by reference to the commonest use; and in the commonest use I think we can agree that to say something is an explanation of X is to presuppose (in Strawson's sense) that X occurred. But this is not to say that in *all proper uses*, this can be inferred.

Apart from the case just cited, where it is known that X did not occur, known even by the giver of the explanation, and apart from explanations of events in fiction, there are other cases where this condition does not hold. In the modified phlogiston theory of about 1785, the explanation of the limited phlogistication of air when calcination occurs in a closed vessel was in terms of the finite capacity of a finite volume of air for absorbing phlogiston. The very phenomenon here explained does not exist, although even Cavendish thought it did. The explanation given is within the theory, of something described in *theoretical* terms. This is not to be confused with the case where we quite commonly put single quotation marks around the term "explanation," meaning that the term is not properly applicable, as when Conant says "an 'explanation' of metallurgy was at hand: Metallic ore (an oxide) + Phlogiston from charcoal → Metal."[13] For here we are referring to an incorrect explanation of something we know *does* occur, viz. smelting, and we know this not to be the correct explanation: compare the previous cases discussed where we know

the phenomenon *does not* occur. I am therefore unwilling to agree that all proper uses of the term "explanation" presuppose that the phenomenon explained occurs; though I would agree that in the primary use this is so.[14] I shall say something about the necessity for the truth of the body of the explanation itself in the next section.

In the primary use of explanation, then, we know something when we are called on for an explanation that we do not know when called on for a prediction, viz. that the event referred to has occurred. This is sometimes a priceless item of information, since it may demonstrate the existence or absence of a hitherto unknown strength of a certain power. Thus, to take a simpler example than the bridge case, a man in charge of an open-hearth furnace may be suspiciously watching a roil on the surface of the liquid steel, wondering if it is a sign of a "boil" (an occasionally serious destructive reaction) on the furnace lining down below or just due to some normal oxidizing of the additives in the mixture. Suddenly, a catastrophe: the whole charge drops through the furnace lining into the basement. It is now absolutely clear that there was a boil which had eaten through the lining: apart from sabotage (easily disproved by examination) there's no other possibility. But no prediction is possible to the event, using the data then available. This renders almost empty Hempel and Oppenheim's (and even Scheffler's) conclusion that explanations provide a basis for predictions. For "Had we known what was going to happen, we could have predicted it" is a vacuous claim. One might mutter something about "If the furnace was in exactly the same state *again* we could predict it would dump," but I have already pointed out that this is a virtually empty remark since we usually can't identify "exactly the same state"; it is simply a dubious determinist slogan, not even a genuine conditional prediction. Since it is technically entirely impossible to rebuild the furnace to the point where it is identical down to the temperature distribution in the mixture (a crucial factor) and the shape of the irregularities in the floor (also crucial), even if we knew these specifications, it will be pure chance if the conditions ever recur and when they do they won't be identifiable. Thus our grounds for thinking the determinist's slogan to be true—if we do—are entirely indirect, and the explanation certainly does not rest on subsumption under the slogan since we cannot even tell when the latter applies, whereas we can be sure the explanation is correct.

The problem of direct versus indirect confirmation which arises here is of great importance throughout structural logic. To say that "same cause, same effect" is a determinist's slogan is not to say it has *no* empirical content. It has, and it is actually false, as far as present evidence goes, though only to a small extent for macroscopic observations. (It is also not equivalent to the idea of determinism as universal law-governed behavior.) What it lacks is single-case applicability and hence direct confirmability when complex systems are involved—for it is often impossible to specify what counts as the "same conditions." It may still be felt on general grounds that unless we do know what counts as the "same conditions" in a given case, we cannot be sure of the proposed causal explanation in that case. The opposite thesis will be defended in a later section of this paper, and to prepare for it I shall need to make several further distinctions and points. At this stage, however, let me summarize by saying that any prediction specifically associated with an explanation is (i) often conditional, and (ii) either so general as to be almost empty or so specific as to refer to no other case, and (iii) often not assertible until it is known the event occurred, that is, not a true prediction.

3.5. Explanations as Sets of True Statements

It is not possible to claim that explanations can only be offered for events that actually occur or have occurred. They can be given for events in the future (Scheffler), for events in fiction, for events known not to occur, and for events wrongly believed to occur—and also for some laws, states, and relationships which are timeless (see above). Assuming Hempel and Oppenheim's analysis to be in other respects correct, it follows that in such cases some of the propositions comprising the explanation itself cannot be true, contrary to one of their explicit conditions.[15] The reason they give for this condition is a very plausible one, however, and it is of interest to see if a more general account can be given which will contain allowance for their point. They say: ". . . it might seem more appropriate to stipulate that the (explanation) has to be highly confirmed by all the relevant evidence rather than that it should be true. This stipulation, however, leads to awkward consequences. Suppose that a certain phenomenon was explained at an earlier stage of science by means of an (explanation) which was well supported by the evidence then at hand, but which had been highly disconfirmed by more recent empirical findings. In such a case, we would have to say that originally the explanatory account was a correct explanation, but that it ceased to be one later, when unfavorable evidence was discovered. This does not appear to accord with sound common usage, which directs us to say that . . . the account in question was not—and had never been—a correct explanation.[16]

It is roughly on these grounds that Conant puts the term "explanation" in quotes when he is referring to the phlogiston theory's account of calcination. For much the same reason we refer to an astrologer's remarks as an "explanation" of Henry Ford's successful business career.

But notice we can talk perfectly well about "two competing explanations" of some phenomenon in contemporary physics without feeling it improper to refer to both as explanations although only one can be true. And there certainly seem to be cases where we want to say, for example, that the Babylonian explanation of the origin of the universe was basically naturalistic, without using inverted commas. The best treatment of these cases, it seems to me, is to regard them as secondary uses which have become fairly standard, the notion of a secondary use being defined in terms of the fact that understanding it depends logically on understanding the primary use. But these are definitely proper uses and the term "explanation" is hence perhaps less a 'success word' or 'achievement word' than, for example, "knowledge" and "perception." We canot say of two contradictory claims that both are *known*, since this implies both are true. And this suggests a solution to our present problem.

The proper way of avoiding Hempel and Oppenheim's powerful argument is, I think, very simple; the secondary uses of "explanation" are legitimate but there are no such secondary uses of "correct explanation," the term which they substitute halfway through the argument. Remove the qualifying adjective "correct" and you will see that the argument is no longer persuasive. For consistency, this term must be and can be added to the occurrences of "explanation" in the premises. Overwhelming counterevidence does not necessarily lead us to abandon or even to put quotes around "explanation," but, as the argument rightly says, it does lead us to abandon the application of the term "correct explanation" (or "the explanation" which is often used equivalently). Hence we should regard Hempel and Oppenheim's anal-

ysis as an analysis of "correct explanation" rather than of "explanation," or "an explanation," and this is surely what they were most interested in. "Explanations," or "an explanation," or "his explanation," or "a possible explanation," do not *always* have to be true (or of the appropriate type, or adequate); they only need high confirmation, at some stage.

Doesn't the notion of confirmation come into the analysis of "correct explanation" at all? It is not part of the *analysis*, which only involves truth; but it is our only *means of access* to the truth. We have not got the correct explanation unless it contains only true assertions, but if we want to know which explanation is most likely to meet that condition, we must select the one with the highest degree of confirmation. Good evidence does not guarantee true conclusions but it is the best indicator, so we need no excuse for appealing to degree of confirmation. Moreover, we have no need to adopt the skeptic's position that all possibility of knowing when we have a correct explanation is by now beyond reasonable doubt, and to restrict "knowing" to cases of absolute logical necessity is to mistake the empty glitter of definitional truth for the fallible flame of knowledge. The notion of reasonable doubt is highly dependent on context, but highly unambiguous in a given context, and it sets the threshold level which distinguishes knowledge from likelihood. Anything that is to be called the "correct explanation" of something that is known to have happened must contain only statements from the domain of knowledge.

Now among the things we know are some statements about the probability of certain events under certain circumstances, for example, about the probability of throwing a six with a die that passes various specifiable tests. Could we not use such propositions as part of an explanation? Hempel and Oppenheim—in the papers cited—countenance the possibility that what they call statistical explanations may be of great importance but they neither undertake to discuss them nor, more significantly, restrict most of their conclusions about explanation in general in the way that would be appropriate if we do take seriously the claims of statistical explanations (which I include in the broader class of explanations based on probability statements). In *Minnesota Studies*, vol. 3, Hempel sets out an account of statistical explanation on which I comment later. Such explanations cannot be subsumed under Hempel and Oppenheim's original analysis as it stands, because no *deduction* of a nonprobability statement from them is possible, and it is hence impossible for them to explain any actual occurrence, since actual occurrences have to be described by non-probability statements. In particular we could make a (probable) prediction from such "laws," but could not—using the same premises—be said to explain the event predicted, if it does come about.

(3.51) It is of some importance to notice that Hempel and Oppenheim's analysis of explanation absolutely presupposes a descriptive language. For them there can be nothing to explain if there is no language, since the thing-to-be-explained is dealt with *only* via the "explanandum" which is its description in the relevant language. One suspects such a restriction immediately because there are clearly cases where we can explain without language, for example, when we explain to the mechanic in a Yugoslav garage what has gone wrong with the car. Now this is hardly a scientific explanation, but it seems reasonable to suppose that the scientific explanation represents a refinement on, rather than a totally different kind of entity from, the ordinary explanation. In our terms, it is the *understanding* which is the essential part of an explanation and the *language* which is a useful accessory for the process of

communicating the understanding. By completely eliminating consideration of the step from the phenomenon to the description of the phenomenon, Hempel and Oppenheim make it much easier to convince us that deducibility is a criterion of explanation. In fact, within the language there is only one other relation possible, viz. inducibility. We shall argue that good inductive inferribility is the only required relation involved in explanations, deduction being a dispensable and overrestrictive requirement which may of course sometimes be met. But a source of both error and understanding is left out of account in such a debate; for unexplained things are sometimes such that we do not describe them in asking for an explanation and such that they are explained *merely* by being described in the correct way regardless of deduction from laws.[17] (And on the other hand, sometimes a *true* description and deduction is not enough.)

3.6. *Explanations as Involving Descriptions of What Is to Be Explained*

Once we have realized the extraordinary difficulty there is in supposing that explanations and predictions have a common structure, it is natural to ask what the structure of an explanation really is. The "structure" of a prediction, we noticed, is simply that of a declarative statement using an appropriate future tense and any kind of descriptive language. It *may* indeed be of the form "*C* will bring about *X*" but is more usually of the form "*X* will occur" or "*X*, at time *t*." The structure of an argument, to take a further example, is such as to involve several statements which are put forward as bearing upon each other or upon some other statements in the relation of premises to conclusions. Now what is the structure of an explanation? A bridge's failure may have as its explanation the fatigue of the metal in a particular member or the overload due to a bomb blast. These appear to be a state and an event which could be held to be the cause of the event to be explained. A different account will have to be given of the explanation of laws, but for the moment we can profitably concentrate on the explanation of events, to which Hempel and Oppenheim devote a good deal of space, and which has some claim to be epistemologically prior to the explanation of laws.

Now there is a further apparent ambiguity about "explanation": it can either refer to the linguistic structure which describes certain states or events or to the states (or events) themselves. This kind of ambiguity even occurs in connection with such terms as "consequence," "concept," "cause," "inference," and "argument"; it is common throughout logic and best illustrated, perhaps, by the very term "fact." We shall usually be referring to the linguistic entity when we use the term "explanation," but clearly *neither* this entity nor its referents include whatever it is that is to be explained. In the simple but standard examples just given, the explanation, in this sense, is an assertion about a state or event that is entirely different from (assertions about) the state or event to be explained. But Hempel and Oppenheim say, "We divide an explanation into two major constituents . . . the sentence describing the phenomenon to be explained . . . [and] the class of those sentences which are adduced to account for the phenomenon."[18] The former is plainly not a constituent of the explanation at all (except where it is all of the explanation—see 3.5). Only if we find its consequences very confusing or inconvenient should we abandon such a clear distinction as this.

The first difficulty that strikes us about this version of the Hempel and Oppen-

heim account, then, is that it asserts all explanations of a phenomenon X consist in a deductive argumentlike structure, with a statement about X as the conclusion, whereas our simple examples above are merely statements about something or other that is held to be the cause of whatever is to be explained. And are there not occasions, on which one is going over—demonstrating—an explanation, when one does finish off by giving as the last step the description of what is to be explained? It is certainly not a common practice, scientifically or ordinarily, and even when it occurs, it only shows that part of a proof that something is an explanation of X may involve a description of X. The *explanation* of the photoelectric effect does not involve the description of the effect—this is presupposed by the explanation. The point may be minor, but it puts us on our guard, for we cannot be sure whether it may not have unfortunate consequences, analogous to those involved in saying that predictions have the same logical structure as explanations. In fact, we have already seen one error that results from this incautious amalgamation of (i) phenomena, (ii) their description, and (iii) their explanation, in 3.5. We could state part of it by saying that a sixth requirement is actually implicit in their account, viz. the requirement of accuracy and relevance of the description of X, which for them is part of the explanation of X.

In fact, the most serious error of all those I believe to be involved in Hempel and Oppenheim's analysis also springs from the very same innocuous-seeming oversimplification: the requirement of deducibility itself, plausible only if we forget that our concern is fundamentally with a phenomenon, not a statement. It may seem unjust to suggest that Hempel and Oppenheim amalgamate the phenomenon and its description (though certainly they do amalgamate the description and the explanation) when they make clear that the "conclusion" of the explanation is "a sentence describing the phenomenon (not that phenomenon itself)."[19] The justice of my complaint rests, not on their failure *ever* to make this distinction, but on their failure to be consistent in dealing with its consequences. For it is a consequence of this distinction that a nondeductive step is involved between the statements in an explanation and the phenomenon explained. And we may then ask why they should suppose deducibility to be the only logical relation in a good explanation. They never address themselves to this question directly, chiefly, I think, because they do not *realize* the consequences of the distinction they do once make. Attention to it would surely have led them to notice (i) cases of explanatory description (see 3.5), (ii) cases where the completeness or (iii) the uniqueness of the description are crucial in assessing the explanation (see 6.2). Only if we assume that getting as far as the description is getting to the phenomenon, that is, doing what an explanation is supposed to do, could we overlook such interesting cases. (I think the fact that "description" can be taken to *mean* "accurate description" also led them to overlook the independent importance of this requirement.)

3.7. The Last Two Conditions and a Summary of Difficulties

It is stated that the explanation "must contain general laws, and these must actually be required for the derivation. . . ." And finally, it is said that the derivation must be deductive, ". . . for otherwise the (explanation) would not constitute adequate grounds for (the proposition describing the phenomenon)."[20] We now have a general idea of Hempel and Oppenheim's model of explanation, which I have elsewhere

christened, for obvious reasons, 'the deductive model.'[21] I wish to maintain against it the following criticisms in particular, and some others incidentally;

1. It fails to make the crucial logical distinctions between explanations, grounds for explanations, predictions, things to be explained, and the description of these things.
2. It is too restrictive in that it excludes their own examples and almost every ordinary scientific one.
3. It is too inclusive and admits entirely nonexplanatory schema.
4. It requires an account of cause, law, and probability which are basically unsound.
5. It leaves out of account three notions that are in fact essential for an account of scientific explanation: context, judgment, and understanding.

These objections are not wholly independent, and I have already dealt with some of them.

4. Fundamental Issues

4.1. The Distinction Between Explanations and the Grounds for Explanations

It is certainly not the case that our grounds for thinking a plain descriptive statement to be true are part of the statement itself; no one thinks that a more complete analysis of "Gandhi died at an assassin's hand in 1953" would include "I read about Gandhi's death in a somewhat unreliable newspaper" or "I was there at the time and saw it happen, the only time I've been there, and it was my last sabbatical leave so I couldn't be mistaken about the date," etc. Why, then, should one suppose that our grounds for (believing ourselves justified in putting forward)[22] a particular explanation of a bridge collapsing, for example, the results of our tests on samples of the metal, our knowledge about the behavior of metals, eyewitness accounts, are part of the explanation? They might indeed be produced as part of a *justification* of (the claim that what has been produced is) the explanation. But surely an explanation does not have to contain its own justification any more than a statement about Gandhi's death has to contain the evidence on which it is based. Yet, the deductive model of explanation requires that an explanation include what are often nothing but the grounds for the explanation.

Not only linguistic impropriety but absolute impossibility is involved in the attempt to market the joint package as the "whole explanation" or "complete explanation." The linguistic impropriety is twofold: first, perfectly proper explanations would be rejected for the quite unjust reason that they did not contain the grounds on which they were asserted; second, the indefinite number of possible grounds for an explanation makes absurd the idea of a single correct explanation since there is, in terms of the model, nothing more or less correct about any one of the wide range of possible sets of deductively adequate true grounds. And clearly these are circumstances in which we do identify a particular account as "The correct explanation." The impossibility derives from the second impropriety. There is no sense in which one could ever provide a complete justification of an explanation, out of context;

for a justification is a defense against some specific doubt or complaint, and there is an indefinite number of possible doubts.

The deductive model apparently provides an answer to the latter objection in an interesting way. It prescribes that the only kind of justification required is deduction from general laws and specific antecedent conditions. Once this is given, a complete explanation has been given; until this has been done, only (at best) an "explanation sketch" has been given.

When we say that a perfectly good explanation of one event, for example, a bridge collapsing, may be no more than an assertion about another event, for example, a bomb exploding, might it not plausibly be said that this can only *be* an explanation if some laws are assumed to be true, which *connect* the two events? After all, the one is an explanation of the other, not because it came before it, but because it *caused* it. In which case, a full statement of the explanation would make explicit these essential, presupposed laws.

The major weakness in this argument is the last sentence; we can put the difficulty again by saying that, if completeness requires not merely the existence but the quoting of all necessary grounds, there are no complete explanations at all. For just as the statement about the bomb couldn't be an explanation of the bridge collapsing unless there was some connection between the two events, it couldn't be an explanation unless it was true. So, *if* we must include a statement of the relevant laws to justify our belief in the connection, that is, in the soundness of the explanation, then we must include a statement of the relevant data to justify our belief in the claim that a bomb burst, on which the soundness of the explanation also depends.[23]

Certainly in putting forward one event as an explanation of another in the usual cause-seeking contexts, we are committed to the view that the first event caused the second, and we are also committed to the view that the first took place. Of course, we may be wrong about either view and then we are wrong in thinking we have given the explanation. But it is a mistake to suppose this error can be eliminated by quoting further evidence (whether laws or data); it is merely that the error may be then located in a more precise way—as due to a mistaken belief in such and such a datum or law. The function of deduction is only to shift the grounds for doubt, though doubts sometimes get tired and give up after a certain amount of this treatment.

Perhaps the most important reason that Hempel and Oppenheim have for insisting on the inclusion of laws in the explanation is what I take to be their belief (at the time of writing the paper in question) that only if one had such laws in mind could one have any rational grounds for putting forward one's explanation. This is simply false as can be seen immediately by considering an example of a simple physical explanation of which we can be quite certain. If you reach for a cigarette and in doing so knock over an ink bottle which then spills onto the floor, you are in an excellent position to explain to your wife how that stain appeared on the carpet, that is, why the carpet is stained (if you cannot clean it off fast enough). You knocked the ink bottle over. This is the explanation of the state of affairs in question, and there is no nonsense about it being in doubt because you cannot quote the laws that are involved, Newton's and all the others; in fact, it appears one cannot here quote any unambiguous true general statements, such as would meet the requirements of the deductive model.

The fact you cannot quote them does not show they are not somehow *involved*, but the catch lies in the term "involved." Some kind of connection must hold, and if we say this *means* that laws are involved, then of course the point is won. The suggestion is debatable, but even if true, it does not follow that we will be able to state a law that guarantees the connection. The explanation requires that there be a connection, but not any particular one—just one of a wide range of alternatives. Certainly it would not be the explanation if the world was governed by *anti*gravity. But then it would not be the explanation if you *had not* knocked over the ink bottle—and you have just as good reasons for believing that you did knock it over as you have for believing that knocking it over led to (caused) the stain. Having reasons for causal claims thus does not always mean being able to quote laws. We shall return to this example later. For the moment, it is useful mainly to indicate that (i) there is a reply to the claim that one cannot have good reasons for a causal ascription unless one can quote intersubjectively verifiable general statements and (ii) there is an important similarity between the way in which the production of an appropriate law supports the claim that one event explains another, and the way in which the production of further data (plus laws) to confirm the claim that the prior event occurred supports the same claim. They are defenses against two entirely different kinds of error or doubt, indeed, but they are also both support for the same kind of claim, viz. the claim that one event (state, etc.) explains another.

This is perhaps obscured by the fact that when we make an assertion our claim is in full view, so to speak, whereas when we put forward an assertion *as* an explanation, its further role is entirely derived from the context, for example, that it is produced in answer to a request for an explanation, and so its further obligations seem to require explicit statement. This is a superficial view. All that we actually identify in the linguistic entity of a 'declarative statement' is the subject, predicate, tense, etc. We have no reason at all, apart from the context of its utterance, for supposing it to be *asserted*, rather than proposed for consideration, pronounced for a grammatical exercise, mouthed by an actor, produced as an absurdity, etc.[24] *That* it is asserted to be true we infer from the context just as we infer that it is proffered as the explanation of something else; and for both these tasks it may need support. We may concede that assertion is the *primary* role of indicative sentences without weakening this point.

It is in fact the case that considerations of context, seen to be necessary even at the level of identifying assertions and explanations themselves, not only open up another dimension of error for an explanation, that of pragmatic inappropriateness, but simultaneously offer a possible way of *identifying* the explanation of something, where this notion is applicable.

A particular context—such as a discussion between organic chemists working on the same problem—may make one of many deductively acceptable explanations of a biochemical phenomenon entirely inappropriate, and make another of exactly the right type. (Of course, I also wish to reject the criteria of the deductive model; but even if one accepted it, the consideration of context turns out to be *also* necessary. So its importance is not only apparent in dealing with alternative analyses.)

We may generalize our observations in the following terms. An explanation is sometimes said to be incorrect or incomplete or improper. I suggest we pin down these somewhat general terms along with their slightly more specific siblings as follows. If an explanation explicitly contains false propositions, we can call it *incorrect*

or *inaccurate*. If it fails to explain what it is supposed to explain because it cannot be "brought to bear" on it, for example, because no causal connection exists between the phenomenon as so far specified and its alleged effect, we can call it *incomplete* or *inadequate*. If it is satisfactory in the previous respects but is clearly not the explanation required in the given context, either because of its difficulty or its field of reference, we can call it *irrelevant, improper,* or *inappropriate*.

Corresponding to these possible failings there are types of defense which may be relevant. Against the charge of inaccuracy, we produce what I shall call *truth-justifying grounds*. Against the charge of inadequacy, we produce *role-justifying grounds,* and against the complaint of inappropriateness, we invoke *type-justifying grounds*. To put forward an explanation is to commit oneself on truth, role, and type, though it is certainly not to have explicitly considered grounds of these kinds in advance, any more than to speak English in England implies language-type consideration for a lifelong but polylingual resident Englishman.

The mere production of, for example, truth-justifying grounds does not guarantee their acceptance, of course. They may be questioned, and they may be defended further by appeal to further evidence; we defend our claim that a bomb damaged a bridge by producing witnesses or even photographs taken at the time; and we may defend the accuracy of the latter by producing the unretouched negatives and so on. The second line of defense involves *second-level grounds,* and they may be of the same three kinds. That they can be of these kinds is partly fortuitous (since they are not explanations of anything) and due to the fact that the relation of being-evidence-for is in certain ways logically similar to being-an-explanation-of. In each case, truth, role, and type may be in doubt; in fact, this coincidence of logical character is extremely important. We notice, however, that there is no similarity of any importance between these two and being-a-prediction of; where truth is not relevant in the same way, role is wholly determined by time of utterance and syntax, and only the type can be—in some sense—challenged.

4.2. *Completeness in Explanations*

The possibility of indefinitely challenging the successive grounds of an explanation has suggested to some people—*not* Hempel and Oppenheim—that a complete explanation cannot be given within science. Such people are adopting another use of "complete"—even less satisfactory than Hempel and Oppenheim's—according to which the idea of a complete explanation becomes not only foreign to science but in fact either wholly empty, essentially teleological, or capable of completion by appeal to a self-caused cause. Interesting though this move is in certain respects, it essentially requires saying that we can better understand something in the world by ultimately ascribing its existence and nature to the activities of a mysterious entity whose existence and nature cannot be explained in the same way, than by relating it to its proximate causes or arguing that the world has existed indefinitely. I shall only add that we are supposed to be studying scientific explanations, and if none of them are complete in this sense, we may as well drop this sense while making a note of the point—which is equivalent to the point that the causal relation is irreflexive and hence rather unexciting—for there is an important and standard use of "complete" which does apply to some suggested scientific explanation, and not to others, and is well worth analyzing.

Now, if some scientific explanations are complete—and think how a question in a physics exam may ask for a complete explanation of, for example, the effects noticed by Hertz in his experiments to determine whether electromagnetic waves existed—it cannot be because there is a last step in the process of challenging grounds, for there is no stage at which a request for further proof could not make sense. But *in any given context* such requests eventually become absurd, because in any scientific context certain kinds of data are taken as beyond question, and there is no meaning to the notions of explanation and justification which is not, directly or indirectly, dependent on a context. This situation is of a very familiar kind of logic. It makes perfectly good sense to ask for the spatial location of any physical object; and perfectly proper and complete answers will involve a reference to the location of some other physical objects. Naturally we can, and often do, go on to ask the further question concerning where these other objects are ("Where's Carleton College?" "In Northfield." "But where's Northfield?"). And no question in a series of this kind is meaningless, unless one includes the question "Where is the universe, that is, everything?" as of this kind. If one does include this question (which is the analogue of "Where did the universe come *from?*"), then the impossibility of answering it only shows something about the notion of position, and nothing about the incompleteness of our knowledge. If one excludes this question, the absence of a last stage in such a series does not show our inability to give anyone complete directions to the public library but ony that the notion of completeness of such descriptions involves context criteria.

Any request for directions logically presupposes that *some* directions can be understood; if no directions can be understood, then the proper request is for an account of the notions of position and directions. A *complete* answer has been given when the particular object has been comprehensibly related to the directions that are understood. Similarly, then, the request for an explanation presupposes that *something* is understood, and a complete answer is one that relates the object of inquiry to the realm of understanding in some comprehensible and appropriate way. What this way is varies from subject matter to subject matter just as what makes something better than something else varies from the field of automobiles to solutions of chess problems; but the *logical* function of explanation, as of evaluation, is the same in each field. And what counts as complete will vary from context to context within a field; but the logical category of complete explanation can still be characterized in the perfectly general way just given, that is, the logical function of "complete," as applied to "explanations," can be described. Hence the notion of the proper context for giving or requesting an explanation, which presupposes the existence of a certain level of knowledge and understanding on the part of the audience or inquirer, *automatically* entails the possibility of a complete explanation being given. And it indicates exactly what can be meant by the phrase "*the* (complete) explanation." For levels of understanding and interest define areas of lack of understanding and interest, and the required explanation is the one which relates to these areas and not to those other areas related to the subject of the explanation but perfectly well understood or of no interest (these would be explanations which could be correct and adequate but inappropriate). It is worth mentioning that the same analogy with spatial location (or evaluation) provides a resolution of the "problem of induction," as a limit case of a request for a "complete justification."

It is also clear that calling an explanation into question is not the same as—

though it includes—rejecting it as not itself explained. Type justifying involves more than showing relevance of subject matter, that is, topical and ontological relevance; it involves showing the appropriateness of the intellectual and logical level of the content; a proposed explanation may be inappropriate because it involves the wrong kind of true statements from the right field, for example, trivial generalizations of the kind of event to be explained, such that they fulfill the deductive model's requirements but succeed only in generalizing the puzzlement. One cannot explain why this bridge failed in this storm by appealing to a law that all bridges of this design in such sites fail in storms of this strength (there having been only two such cases, but there being independent evidence for the law, not quoted). This might have the desirable effect of making the maintenance boss feel responsible, but it surely does not explain *why* this bridge (or any of the other bridges of the same design) fails in such storms. It may be because of excessive transverse wind pressure, because of the waves affecting the foundations or lower members, because of resonance, etc.

So mere deduction from true general statements is again seen to be less than a sufficient condition for explanation; but what interests us here is that our grounds for *rejecting* such an explanation are not suspicions about its *truth* or its *adequacy*, which are the usual grounds for doubting an explanation, but only its failure to *explain*. Certainly it fails to explain if incorrect or inadequate, but then one feels it fails in a genuine attempt, that the slip is then between the cup and the lip; whereas irrelevance of type is a slip between the hand and the cup—the question of it being a *sound* explanation never even arises. One may react to this situation by declaring with Hempel and Oppenheim that the only *logical* criteria for an explanation are correctness and adequacy, the matter of type being psychological; or, as I think preferable, by saying that the concept of explanation is logically dependent on the concept of understanding, just as the concept of discovery is logically dependent on the concept of knowledge-at-a-particular time. One cannot discover what one already knows, nor what one never knows; nor can one explain what everyone or no one understands. These are tautologies of logical analysis (I hope) and hardly grounds for saying that we are confusing logic with psychology.

Having distinguished the types of difficulty an explanation may encounter, one can more easily see there is no reason for insisting that it is complete only if it is armed against them in advance, since (i) to display in advance one's armor against *all* possible objections is impossible and (ii) the value of such a requirement is adequately retained by requiring that scientific explanations be such that scientifically sound defenses of the several kinds indicated *be available* for them though not necessarily *embodied in* them. Since there is no special reason for thinking that true first-level role-justifying assumptions are any more necessary for the explanation than any others, it seems quite arbitrary to require that they should be included in a complete explanation; and it is quite independently an error to suppose they must take the form of laws.

Notes

1. "The Logic of Explanation" in *Readings in the Philosophy of Science*, H. Feigl and M. Brodbeck, eds. (New York: Appleton-Century-Crofts, 1953), p. 319. This is an abridgment of "Studies in the Logic of Explanation," *Philosophy of Science*, 15, 1948, pp. 135–75. All page references hereafter are to the later version.

2. *Scientific Explanation* (Cambridge: Cambridge University Press, 1953), p. 319.

3. The discussion of How possibly questions has been initiated and sustained by William Dray. See his *Laws and Explanation in History* (London: Oxford University Press, 1957), especially pp. 164ff.

4. Op. cit., p. 319.

5. For an alternative and acceptable interpretation of their remarks, see 3.4 below.

6. Op. cit., p. 320.

7. Op. cit., pp. 322–23.

8. And this salvages the theme which forms the title of 3.3—but at the expense of making that of 3.4 untenable. I am thus uncertain which interpretation to accept.

9. See also "Explanation, Prediction, and Abstraction" by Israel Scheffler, *British Journal for the Philosophy of Science*, 7: 293–309 (1957), where several of the points in this section are discussed.

10. The reader may be worried by the fact that his law is known to be only an approximation. This is true of almost all "laws," but we do give explanations of them, for example, of Kepler's laws, Snell's law. Now, for any such law, deducing it from any premises would simply show the premises to be inaccurate. Hence, explanation cannot require deduction from true premises. We must substitute a weaker requirement; there are several possibilities which appear to retain much of the Hempel analysis (but see 3.5 below).

11. Hempel, replying to this point in [*Minnesota Studies in the Philosophy of Science*, vol. 3], suggests another, more specific candidate. I comment on this later.

12. "The Function of General Laws in History," *Journal of Philosophy*, 39, (1942), 35–48.

13. *Harvard Case Histories in Experimental Science*, vol. I (Cambridge, Mass.: Harvard University Press, 1958), pp. 70, 110.

14. There are of course a number of terms besides "explanation" (for example, description (see 3.6), observation, insight) that are used in such a way that the description "incorrect (explanation)" can be synonymous with "not an (explanation) at all." The points just made do not, however, depend on this ambiguity.

15. Op. cit., pp. 321–22.

16. Op. cit., p. 322.

17. A common case is that when someone, greatly puzzled, asks What on earth is this? or What's going on here? and is told, for example, that it is an initiation ceremonial on which he has stumbled. Analogous cases in particle physics, engineering, and astronomy are obvious. The point of these examples is that understanding is roughly the perception of relationships and hence may be conveyed by any process which locates the puzzling phenomenon in a system of relations. When we supply a law, we supply part of the system; but a description may enable us to supply a whole framework which we already understand, but of whose *relevance* we had been unaware. We deduce nothing; our understanding comes because we *see* the phenomenon for what it is, and are *in a position* to make other inferences from this realization.

18. Op. cit., p. 321.

19. Op. cit., p. 321.

20. Op. cit., p. 321.

21. "Certain Weaknesses in the Deductive Model of Explanation," paper read at the Midwestern Division of the American Philosophical Association, May 1955.

22. I shall abbreviate some more precise formulations by omitting the words in parentheses where I think they are not essential.

23. Their model requires the truth of the asserted explanation, but it doesn't require the

inclusion of evidence for this. Instead of similarly requiring a causal connection, it actually requires the inclusion of one *special kind* of evidence for this. If it treated both requirements equitably, the model would be either trivial (causal explanations must be true and causally relevant) or deviously arbitrary (. . . must include *deductively* adequate grounds for the truth of any assertions and for the causal connection).

24. See Max Black's "Definition, Presupposition, and Assertion," in his *Problems of Analysis* (London: Routledge and Kegan Paul, 1954).

4

Statistical Explanation and Causality

Wesley C. Salmon

1. Statistical Explanation

The Nature of Statistical Explanation

Let me now, at long last, offer a general characterization of explanations of particular events. As I have suggested earlier, we may think of an explanation as an answer to a question of the form, "Why does this x which is a member of A have the property B?" The answer to such a question consists of a partition of the reference class A into a number of subclasses, all of which are homogeneous with respect to B, along with the probabilities of B within each of these subclasses. In addition, we must say which of the members of the partition contains our particular x. More formally, an explanation of the fact that x, a member of A, is a member of B would go as follows:

$$\begin{aligned} P(A.C_1,B) &= p_1 \\ P(A.C_2,B) &= p_2 \\ &\vdots \\ P(A.C_n,B) &= p_n \end{aligned}$$

where

$$A.C_1, A.C_2, \ldots, A.C_n \text{ is a homogeneous partition of } A \text{ with respect to } B,$$
$$p_i = p_j \text{ only if } i = j, \text{ and}$$
$$x \in A.C_k.$$

With Hempel, I regard an explanation as a linguistic entity, namely, a set of statements, but unlike him, I do not regard it as an argument. On my view, an explanation is a set of probability statements, qualified by certain provisos, plus a statement specifying the compartment to which the explanandum event belongs.

The question of whether explanations should be regarded as arguments is, I be-