Carl G. Hempel

ASPECTS of SCIENTIFIC EXPLANATION

And Other Essays in the Philosophy of Science

THE FREE PRESS, NEW YORK

COLLIER-MACMILLAN LIMITED, LONDON

inductive, nonconjunctiveness presents itself as an inevitable aspect of it, and thus as one of the fundamental characteristics that set I-S explanation apart from its deductive counterparts.

4. THE CONCEPTS OF COVERING-LAW EXPLANATION AS EXPLICATORY MODELS

4.1 General Character and Intent of the Models. We have by now distinguished three basic types of scientific explanative: deductive-nomological, inductive-statistical, and deductive-statistical. The first of these is often referred to as the covering-law model or the deductive model of explanation, but since the other two types also involve reference to covering laws, and since one of them is deductive as well, we will call the first more specifically the *deductive-nomological model;* analogously, we will speak of the others as the *inductive-statistical* and the *deductive statistical models of explanation.*

As is made clear by our earlier discussions, these models are not meant to describe how working scientists actually formulate their explanatory accounts. Their purpose is rather to indicate in reasonably precise terms the logical structure and the rationale of various ways in which empirical science answers explanation-seeking why-questions. The construction of our models therefore involves some measure of abstraction and of logical schematization.

In these respects, our concepts of explanation resemble the concept, or concepts, of mathematical proof (within a given mathematical theory) as construed in metamathematics. Let us note the principal points of resemblance.

In either case, the models seek to explicate the use and function of certain "explicandum" terms—'proof' and its cognates in one case, 'explanation' and its cognates in the other. However, the models are selective; they are not meant to illuminate all the different customary uses of the terms in question, but only certain special ones. Thus, metamathematical proof theory is concerned only with the notion of proof in mathematics. To put the theory forward is not to deny that there are other contexts in which we speak of proofs and proving, nor is it to assert that the metamathematical concepts are relevant to those contexts.

Similarly, to put forward the covering-law models of scientific explanation is not to deny that there are other contexts in which we speak of explanation, nor is it to assert that the corresponding uses of the word 'explain' conform to one or another of our models. Obviously, those models are not intended to reflect the various senses of 'explain' that are involved when we speak of explaining the rules of a contest, explaining the meaning of a cuneiform inscription or of a complex legal clause or of a passage in a symbolist poem, explaining how to bake Sacher torte or how to repair a radio. Explicating the concept of scientific explanation is not the same thing as writing an entry on the word 'explain' for the *Oxford English Dictionary*. Hence to deplore, as one critic does, the "hopelessness" of the deductive-nomological model on the ground that it does not fit the case of explaining or understanding the rules of Hanoverian succession[1] is simply to miss the intent of the model. And it is the height of irrelevance to point out that the deductive-nomological model presupposes that explanations are formulated in a "descriptive language," whereas "there are clearly cases where we can explain without language, e.g., when we explain to the mechanic in a Yugoslav garage what has gone wrong with the car."[2] This is like objecting to a metamathematical definition of proof on the ground that it does not fit the use of the word 'proof' in 'the proof of the pudding is in the eating', nor in '86 proof Scotch'. Wordless gesticulation intended to indicate to a Yugoslav mechanic what is wrong with the car indeed does not qualify as scientific explanation according to any of our models; but that is as it should be, for a construal of scientific explanation that did admit this case would thereby show itself to be seriously inadequate.

In support of the idea that all these different uses of the word 'explain' should be encompassed by an adequate analysis of explanation, Scriven has argued that they all have the same "logical function," about which he remarks: "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. . . ; but the *logical function* of explanation . . . is the same in each field."[3] But while the opening remark of this passage may well apply to many different kinds of explanation, neither it nor the rest of Scriven's remarks on the subject concern what could properly be called a *logical* aspect of explanation. Indeed, such expressions as 'realm of understanding' and 'comprehensible' do not belong to the vocabulary of logic, for they refer to the psychological or pragmatic aspects of explanation. We will consider these aspects in the next section and will see that when construed as observations about the pragmatics rather than the logic of explanation, characterizations such as Scriven's are quite relevant.

But the different ways of explaining contemplated by Scriven certainly cannot be said to have the same logical function. For, first, even the linguistic

1. Scriven (1959), p. 452.

2. Scriven (1962), p. 192. That such objections are irrelevant has been stressed also by Brodbeck (1962), p. 240. Some perceptive and stimulating comments on this issue and on other aspects of "the quarrel about historical explanation" will be found in Weingartner's article (1961).

3. Scriven (1962), p. 202, Italics the author's.

means which serve to indicate the subject matter of different kinds of explanation are of different logical character. For example, when an explanation is to indicate the "meaning" of a literary passage, a symbol, a work of art, and the like, the explanandum will be specified by means of a *noun-phrase* ('the ampersand sign', 'the first sentence of Genesis', 'the swastika'); whereas explanations of the kind we have been considering are concerned with facts, occurrences, events, uniformities—any one of which is properly characterized by means of a *sentence* (which appears as the explanandum-sentence in our schemata). Secondly, the problem of specifying meanings and that of stating the "causes" of an occurrence or perhaps the reasons for which an action was done surely are of different logical character; and the adequacy of the solutions proposed in each case must be judged by quite different criteria. The differences between the tasks to be accomplished by these and other kinds of explanation lie, in fact, precisely in differences between the logical structure of the corresponding kinds of explanation.

From the selectiveness of explicatory models of proof and of explanation let us now turn to another common feature. Metamathematical proof theory is not intended to give a descriptive account of how mathematicians formulate their proofs. Indeed the formulations that mathematicians actually offer will usually depart to some extent from that called for by rigorous and, as it were, "ideal" metamathematical standards. Yet those standards may be said to exhibit the logical structure and the rationale of mathematical demonstration and to provide criteria for the critical appraisal of particular proofs that might be proposed.

A proposed proof may then be found to depart from a given theoretical standard only in inessential ways; for example, by omitting as obvious certain intermediate steps in the argument; or by failing to mention certain premises, which are taken to be understood, and which can be specified explicitly if the need should arise. In such cases, we might say that the proof is *elliptically formulated*. On the other hand, the shortcomings may be crucial, as in the various proofs of the postulate of the parallels on the basis of the other postulates of Euclidean geometry.

In addition to providing standards for critical appraisal, the construction of rigorous concepts of mathematical proof has permitted the development of a powerful theory which has yielded far-reaching and often quite unexpected results concerning provability, decidability, and definability in mathematical systems of specified kinds.

Analytic models of scientific explanation, I think, can serve similar purposes, if on a much more modest scale. As for the possibility of general systematic developments, we might mention, for example, the results established by

Ramsey and by Craig[4] concerning the role and the possible dispensability, in the context of scientific explanation, of principles ostensibly referring to unobservable "theoretical" entities. These results, and whatever insight they convey into the logic of scientific procedure, could be achieved only by reference to a precisely formulated, and to some extent schematic, conception of scientific explanation.

4.2 Varieties of Explanatory Incompleteness

4.2.1 *Elliptic Formulation.* Like a proposed mathematical proof, a proposed explanation may be *elliptically formulated.* When we explain, for example, that a lump of butter melted because it was put into a hot frying pan, or that a small rainbow appeared in the spray of a lawn sprinkler because sunlight was reflected and refracted in the water droplets, we may be said to be offering elliptic versions of D-N explanations. Accounts of this kind forego mention of certain laws or particular facts that are tacitly taken for granted, and whose explicit inclusion in the explanans would yield a complete D-N argument. An elliptically formulated explanation may be said to be *incomplete,* but in a rather harmless sense.

4.2.2 *Partial Explanation.* Often, however, explanatory accounts exhibit a more serious kind of incompleteness. Here, the statements actually included in the explanans, even when supplemented by those which may reasonably be assumed to have been tacitly taken for granted in the given context, account for the specified explanandum only partially, in a sense I will try to indicate by an illustration.

In his *Psychopathology of Everyday Life,* Freud offers this description and explanation of a slip of the pen:

> On a sheet of paper containing principally short daily notes of business interest, I found, to my surprise, the incorrect date "Thursday, October 20th," bracketed under the correct date of the month of September. It was not difficult to explain this anticipation as the expression of a wish. A few days before, I had returned fresh from my vacation and felt ready for any amount of professional work, but as yet, there were few patients. On my arrival, I had found a letter from a patient announcing her arrival on the twentieth of October. As I wrote the same date in September, I may certainly have thought, "X ought to be here already; what a pity about that whole month!" and with this thought, I pushed the current date a month ahead.[5]

Clearly, this formulation of the intended explanation is elliptical in the

4. See Ramsey (1931), pp. 212-15, 231; and Craig (1956). *Cf.* also the discussion of these results in Hempel (1958), section 9.

5. Freud (1951), p. 64.

sense considered a moment ago; for it does not mention any laws or theoretical principles in virtue of which the subconscious wish, and the other particular circumstances referred to, could be held to explain the slip in question. However, the theoretical ideas that Freud proposes for the interpretation of such lapses strongly suggest that his explanation is governed by a general hypothesis to the effect that when a person has a strong, though perhaps subconscious, wish, then if he commits a slip of pen, tongue, or memory, the slip will take a form in which it expresses, and perhaps symbolically fulfills, that wish.

Even this vague statement is no doubt more definite than what Freud would have been willing to assert; and perhaps, despite Freud's deterministic leanings, it would be more appropriate to conceive of the key hypothesis as being of statistical form, and to regard the proposed explanation as probabilistic. But for the sake of the argument, let us take the hypothesis as stated and incorporate it into the explanans, together with particular statements to the effect that Freud did have the subconscious wish he mentions, and that in fact he was going to commit a slip of the pen. Even then, the resulting explanans enables us to infer only that the slip would take *some form or other* that would express, and perhaps symbolically fulfill, Freud's subconscious wish; but the explanans does not imply that the slip would take the specific form of writing "Thursday, October 20," on the calendar, next to the corresponding date for September.

But inasmuch as the class, say *F*, of slips taking this latter form is a proper subclass of the class, say *W*, of those slips of the pen which in some way express and perhaps symbolically fulfill the specified wish, we might say that the explanandum as described by Freud—i.e., that he made a slip falling into the class *F*—is explained at least in part by this account, which places the slip into the wider class *W*. Arguments of this kind might be called *partial explanations*. Many of the explanatory accounts offered in the literature of psychoanalysis[6] and of historiography are at most partial explanations in this sense: the explanans does not account for the explanandum-phenomenon in the specificity with which it is characterized by the explanandum-sentence, and thus, the explanatory force of the argument is less than what it claims or appears to be.

I think it is important and illuminating to distinguish such partial explanations, however widely they may be offered and accepted, and however fruitful and suggestive they may prove, from what might be called *deductively complete explanations*, i.e., those in which the explanandum as stated is logically implied by the explanans; for the latter do, whereas the former do not, account for the explanandum phenomenon in the specificity with which the explanandum

6. This holds true, I think, for the many, often highly suggestive, explanatory analyses included in Freud's *Psychopathology of Everyday Life*.

sentence describes it.[7] An explanation that conforms to the D-N model is, therefore, automatically complete in this sense; and a partial explanation as we have characterized it always falls short of being a D-N explanation.

In a statistical explanation, the explanans does not logically imply the explanandum. Are we then to qualify all such explanations as incomplete? Dray raises this question when he asks whether "an event can be *completely* explained (although perhaps in a different sense) without subsuming it under a universal law licensing its deduction, and consequently without showing that it had to happen."[8] The answer that statistical explanations are deductively incomplete would be an uninteresting truism. As is suggested by Dray's clause "although perhaps in a different sense", we are, rather, faced with the question whether the notion of explanatory completeness, which so far has been defined only in reference to proposed D-N explanations, might reasonably be broadened so as to become applicable also within the domain of probabilistic explanation. It seems inadvisable to construct an extended concept of explanatory completeness in such a way as to qualify all statistical explanations as incomplete. For this qualification carries with it connotations of a deficiency, and surely, we cannot regard statistical explanations simply as unsuccessful D-N explanations: they constitute an important type of explanation in their own right. To be sure, the early explanatory uses of statistical laws and theories, for example in nineteenth century physics, were often propounded in the belief that the micro-phenomena involved in the physical processes under study were all subject to strictly universal laws, and that the use of statistical hypotheses and theories was made necessary only by limitations in our ability individually to measure all those micro-phenomena, and then to perform the vast and complex computations that would be required to account for a given physical phenomenon in full microscopic detail. But this idea has gradually been abandoned: in certain

7. A partial explanation may evidently be more or less weak, depending on how much more extensive is the class within which the explanans places the given case (W in our illustration) as compared with the class to which the explanandum-sentence assigns it (F in our case). Furthermore, while some partial explanations are no doubt illuminating and suggest further research that might lead to a fuller explanatory account, there are other arguments that completely lack such merit even though they bear a formal resemblance to our illustration, and might for that reason be qualified as partial explanations. Suppose, for example, that b is F and also G, and that we have a D-N explanation of b being F. Then (save for certain trivial exceptions) the explanans of the latter will automatically afford a basis for a partial explanation of b being G; for it implies that b is F and hence that b is F or G: and the class characterized by 'F or G' contains G as a proper subclass. But I am not concerned here to explore the conditions under which partial explanations may prove fruitful; I simply wish to call attention to the fact that many explanatory accounts offered in the literature of empirical science have the formal characteristics of partial explanations, and that, as a consequence, they overstate the extent to which they explain a given phenomenon.

8. Dray (1963), p. 119.

areas of physics, such as quantum theory, laws of statistical form have come to be accepted as basic laws of nature. And whatever the future of scientific theorizing may hold in store, this development clearly reflects the realization that logically, statistical explanation is quite independent of the assumption of strictly universal laws and thus constitutes a mode of explanation *sui generis*. All this strongly suggests that under a reasonable extension of the idea of explanatory completeness, any explanation conforming to our statistical model should qualify as formally complete, for it assigns to the explanandum event described by the explanandum statement (or, more properly, to the explanandum statement itself) the logical probability called for by the logical relation between the explanans and explanandum statements. In this respect, such a statistical explanation is analogous to one which conforms to the D-N model, and which thus correctly claims that the explanandum is implied by the explanans (and hence has the logical probability 1 relative to the latter). In the light of this analogy, a proposed statistical explanation should be qualified as partial if the explanans confers the specified probability, not upon the explanandum sentence actually stated, but upon a weaker one related to it in the manner illustrated by our example from Freud. The idea may be illustrated very schematically by reference to that same example. Suppose that the general law we tentatively formulated as the presumptive basis of Freud's explanation were construed instead as a statistical law to the effect that in the presence of a strong though perhaps subconscious wish, the statistical probability is high that if a slip of the pen is committed it will take a form which expresses and perhaps symbolically satisfies that wish. Then Freud's account—now construed as claiming that the explanatory information adduced confers a high logical probability upon the explanandum statement—would count as a *partial statistical explanation;* for the explanans confers a high probability, not upon the statement that the particular slip fell within the class F defined earlier, but upon the weaker statement that the slip belonged to the class W.

4.2.3 *Explanatory Incompleteness vs. Overdetermination.* The considerations just presented are relevant also to the problem illustrated by the following example:[9] Suppose that rod r, made of copper ($C\,r$), is simultaneously subjected to heating ($H\,r$) and to longitudinal stress ($S\,r$), and that, in the process, the rod lengthens ($L\,r$). Then it is possible to formulate two different arguments, each of which

9. I am much indebted to my collegue at Princeton, Professor Arthur Mendel, of the Department of Music, who put to me some searching questions which made me aware of the problem here considered. In his paper (1962) Mendel takes as his point of departure a concrete problem in the history of music and by reference to it develops some illuminating general ideas concerning, among other things, the significance of the covering-law models for the explanatory objectives of the historian.

constitutes, by the standards we have suggested, a D-N explanation of why the rod lengthened. One of them will be based on the law that copper rods lengthen when heated; the other, on the law that copper rods lengthen when stressed. Schematically:

$$\frac{\begin{array}{c}(x)\,[(Cx \cdot Hx) \supset Lx] \\ Cr \cdot Hr\end{array}}{Lr} \qquad\qquad \frac{\begin{array}{c}(x)\,[(Cx \cdot Sx) \supset Lx] \\ Cr \cdot Sr\end{array}}{Lr}$$

It might be objected that—even granting the truth of all the premises—both accounts are unacceptable since they are "incomplete": each neglects one of the two factors that contributed to the lengthening. In appraising the force of this objection it is again important to be clear about just what is to be explained. If, as in our example, this is simply the fact that *Lr*, i.e., that *r* lengthened, or that there was *some* increase in the length of *r*, then, I think, either of the two arguments conclusively does *that*, and the charge of incompleteness is groundless. But if we wish to account for the fact that the length of the rod increased by so and so much, then clearly neither of the two arguments will do; for we would have to take into account both the temperature increase and the stress, and we would need quantitative laws governing their joint effect on the length of a copper rod. Such common locutions as 'explaining the increase in the length of a metal rod' have to be handled with care: they are ambiguous in that they refer to at least the two quite different tasks here distinguished.

Adopting a term that is often used in psychoanalytic theorizing, we might say that an event is *overdetermined* if two or more alternative explanations with nonequivalent explanans-sets are available for it. Thus, the occurrence of some lengthening in the copper rod *r* constitutes a case of *explanatory overdetermination* in virtue of the availability of the alternative explanations mentioned above. In this example, the alternative explanations invoke different laws (and consequently some different statements concerning particular facts). In another, perhaps less interesting, kind of situation which under our definition would likewise qualify as explanatory overdetermination, the alternative explanations rest on the same laws, but adduce different particular circumstances[10]. For example, the state of a deterministic physical system at time *t* can be explained, with the help of the relevant laws, by specifying the state of the system at any earlier time; potentially this permits infinitely many alternative explanations no two of which have logcially equivalent explanans-sets.

A problem that bears a certain resemblance to the one just considered has

10. On this point, *cf.* the remarks in Braithwaite (1953), p. 320.

been raised by Scriven, who illustrates it by the following example: In order to explain how a certain bridge came to be destroyed in wartime, "we could appeal to the law 'whenever an atom bomb is released accurately above a bridge and explodes with full force, the bridge is destroyed', plus the appropriate antecedent conditions." But it may also "be the case that whenever 1000 kilograms of dynamite are detonated on the main span of such a bridge it will be destroyed, and that the underground movement has applied just this treatment to this bridge with the attendant destruction occurring between the release and the arrival of the atomic bomb." Scriven holds that this invalidates the bomb explanation, "which cannot account for other features of the event, in this case the time of the destruction." He concludes that in order to rule out such explanations we must impose the requirement of total evidence, even on D-N explanations, in a more specific form which requires "that an explanation be acceptable for a phenomenon only so long as no facts are known about the circumstances surrounding the occurrence of the phenomenon which the explanation cannot accommodate."[11]

But surely the bomb explanation in Scriven's example is unacceptable because its explanans requires the assumption that when the pressure wave of the bomb reached the place in question, there was a bridge there that could be destroyed—an assumption that is false, since at that time the span had already been wrecked by dynamite. Hence, the contemplated bomb explanation is false in the sense specified in section 2, and no additional requirement is needed to disqualify it or other accounts of this kind.

Besides being unnecessary, the specific requirement Scriven suggests in order to rule out the bomb explanation and its likes is, I think, vastly too strong to be tenable. For neither in scientific research nor in our practical pursuits do we require of an acceptable explanation that it accomodate everything we know—or believe we know—about the facts surrounding the explanandum phenomenon. In the case of the bridge, for example, these facts may include a great deal of information about the shape, size, and location of the fragments after the destruction; perhaps the identities of the dynamiters; their objectives; and many other things. Surely we do not require that all of these details must be accounted for by any acceptable explanation of "how the bridge came to be destroyed."

Finally, the condition proposed by Scriven has nothing whatever to do with the requirement of total evidence; in particular, it is not a "more specific" version of it. And Scriven's contention that some such condition must be imposed even on explanations of deductive form because they do not automati-

11. Scriven (1962), pp. 229-30. See also a brief remark, which seems to have the same intent, in Scriven (1963a), pp. 348-49.

cally satisfy the requirement of total evidence[12] overlooks the straightforward proof to the contrary.[13]

4.2.4 *Explanatory Incompleteness and "Concrete Events"*. A scientific explanation, we noted earlier, may be regarded as a potential answer to a question of the form 'why is it the case that *p*?', where the place of '*p*' is occupied by an empirical sentence detailing the facts to be explained. Accordingly, both the deductive-nomological and the statistical models of explanation characterize the explanandum-phenomenon by means of a *sentence*, the explanandum-sentence. Take, for example, the explanation of individual facts such as that the length of a given copper rod *r* increased during the time interval from 9.00 to 9.01 A.M., or that a particular drawing *d* from a given urn produced a white ball: here the explanandum phenomena are fully described by the sentences 'the length of copper rod *r* increased between 9.00 and 9.01 A.M.' and 'drawing *d* produced a white ball'. And only when understood in this sense, as fully describable by means of sentences, can particular facts or events be amenable to scientific explanation.

But the notion of an individual or particular event is often construed in quite a different manner. An event in this second sense is specified, not by means of a sentence describing it, but by means of a noun phrase such as an individual name or a definite description, as, for example, 'the first solar eclipse of the twentieth century', 'the eruption of Mt. Vesuvius in A.D. 79', 'the assassination of Leon Trotsky', 'the stock market crash of 1929. ' For want of a better terminology, individual events thus understood will be referred to as *concrete events*,[14] and facts and events in the first sense here considered will be called sententially characterizable, or briefly, *sentential facts and events.*

The familiar question of whether individual events permit of a complete explanation is no doubt inspired to a large extent by the conception of an individual event as a concrete event. But what could be meant, in this case, by a complete explanation? Presumably, one that accounts for every aspect of the

12. Scriven (1962), p. 230.

13. In a deductively valid argument, the premises constitute conclusive grounds for asserting the conclusion; and whatever part of the total evidence is not included in the premises is irrelevant to the conclusion in the strict sense that if it were added to the premises, the resulting set of sentences would still constitute conclusive grounds for the conclusion. Or, in the terminology of inductive logic: the logical probability which the premises of a D-N argument confer upon the conclusion is 1, and it remains 1 if part or all of the total evidence is added to the premises.

14. I do not wish to suggest that the notion of concrete event here adumbrated is entirely clear; in particular, I do not know how to formulate a necessary and sufficient condition of identity for concrete events. Gibson's perceptive observations on "What is Explained," in (1960), pp. 188-190, are highly relevant to the issues we are about to examine here.

given event. If that is the idea, then indeed no concrete event can be completely explained. For a concrete event has infinitely many different aspects and thus cannot even be completely described, let alone completely explained. For example, a complete description of the eruption of Mt. Vesuvius in A.D. 79 would have to specify the exact time of its occurrence; the path of the lava stream as well as its physical and chemical characteristics—including temperatures, pressures, densities, at every point—and their changes in the course of time; the most minute details of the destruction wreaked upon Pompeii and Herculaneum; full information about all persons and animals involved in the catastrophe, including the fact that the remains of such and such victims, found at such and such places, are on display at a museum in Naples; and so on *ad infinitum*. It must also mention—for this surely constitutes yet another aspect of that concrete event—all the literature about the subject. Indeed, there seems to be no clear and satisfactory way at all of separating off some class of facts that do not constitute aspects of the concrete event here referred to. Clearly, then, it is quite pointless to ask for a complete explanation of an individual event thus understood.

In sum, a request for an explanation can be significantly made only concerning what we have called sentential facts and events; only with respect to them can we raise a question of the form 'why is it the case that *p*?'. As for concrete events, let us note that what we have called their aspects or characteristics are all of them describable by means of sentences; each of these aspects, then, is a sentential fact or event (e.g., that the eruption of Mt. Vesuvius in A.D. 79 lasted for so many hours; that it killed more than 1000 persons in Pompeii, and so on); with respect to such particular aspects of a concrete event, therefore, the question of an explanation can significantly be raised. And clearly, when we speak of explaining a particular event, such as the abdication of Edward VIII, we normally think only of certain aspects of the event as being under scrutiny; what aspects are thus meant to be singled out for explanatory attention will depend on the context of the inquiry.[15]

Though the issues here touched upon are perhaps discussed most frequently with special reference to historical events in their "individuality and uniqueness," the problems inherent in the notion of a concrete event are by no means limited to the historian's domain. An event such as the solar eclipse of July 20, 1963, also possesses an infinity of physical, chemical, biological, sociological,

15. As Max Weber remarks, with special reference to historical explanation: "When it is said that history seeks to understand the concrete *reality* of an 'event' in its individuality causally, what is obviously not meant by this . . . is that it is to . . . explain causally the concrete *reality* of an event in the totality of its individual qualities. To do the latter would be not only actually impossible, it would also be a task which is meaningless in principle." (Weber (1949), p. 169. Italics the author's.)

and yet other aspects and thus resists complete description and *a fortiori*, complete explanation. But certain aspects of the eclipse—such as the duration of its totality, and the fact that it was visible in Alaska and subsequently in Maine—may well be capable of explanation.

It would be incorrect, however, to summarize this point by saying that the object of an explanation is always a *kind* of event rather than an individual event. For a kind of event would have to be characterized by means of a predicate-expression, such as 'total solar eclipse' or 'volcanic eruption'; and since this sort of expression is not a sentence, it makes no sense to ask for an explanation of a kind of event. What might in fact be explained is rather the *occurrence of a particular instance of a given kind of event*, such as the occurrence of a total solar eclipse on July 20, 1963. And what is thus explained is definitely an individual event; indeed, it is one that is unique and unrepeatable in view of the temporal location assigned to it. But it is an individual *sentential* event, of course: it can be described by means of the statement that on July 20, 1963, there occurred a total solar eclipse. I agree therefore with Mandelbaum's rejection of Hayek's view that explanation and prediction never refer to an individual event but always to phenomena of a certain kind: "One would think that the prediction of a specific solar eclipse, or the explanation of that eclipse, would count as referring to a particular event even if it does not refer to all aspects of the event, such as the temperature of the sun, or the effect of the eclipse on the temperature of the earth, and the like."[16]

However, given this notion of explaining a particular occurrence of a solar eclipse or of a rainbow, etc., one can speak *derivatively* of a theoretical explanation of solar eclipses or rainbows in general: such an explanation is then one that accounts for any instance of an eclipse or a rainbow. Thus, the notion of explaining particular instances of a given kind of occurrence is the primary one.

4.2.5 *Explanatory Closure: Explanation Sketch.* Perhaps yet another conception of completeness might seem pertinent to the idea of explanation; we shall call it explanatory closure. An explanatory account would be complete in this sense if for every fact or law it invoked, it contained in turn an explanation. In an account with explanatory closure, nothing would be left unexplained. But completeness in this sense obviously calls for an infinite regress in explanation and is therefore unachievable; to seek such completeness is to misunderstand the nature of explanation.

At any stage in the development of empirical science, certain (presumptive) facts will be unexplainable; in particular, those expressed by the most fundamental laws or theoretical principles accepted at the time, those for which no

16. Mandelbaum (1961), p. 233.

explanation by means of a "deeper" theory is at hand. But while unexplained, these ultimate principles need not be unsupported, for, as hypotheses in empirical science, they will have to be susceptible to test, and it may well be that suitable tests have in fact provided strongly supporting evidence for them.

We have by now considered several ways in which a proposed explanation may deviate from the standards incorporated into our analytic models. In some cases, what is intended as an explanatory account will diverge even more strongly from those standards. A proposed explanation, for example, which is not explicit and specific enough to be reasonably qualified as an elliptically formulated explanation or as a partial one, can often be viewed as an *explanation sketch*, i.e., as presenting the general outlines of what might well be developed, by gradual elaboration and supplementation, into a more closely reasoned explanatory argument, based on hypotheses which are stated more fully and which permit of a critical appraisal by reference to empirical evidence.

The decision whether a proposed explanatory account is to be qualified as an elliptically formulated deductive-nomological or statistical explanation, as a partial explanation, as an explanation sketch, or perhaps as none of these is a matter of judicious interpretation. It calls for an appraisal of the intent of the given account and of the background assumptions that may have been left unstated because they are taken to be understood in the given context. Unequivocal criteria of adjudication cannot be formulated for this purpose any more than for deciding whether a given informally stated argument which does not meet reasonably strict standards of deductive validity is to count as nevertheless valid but enthymematically formulated, or as fallacious, or as a sound inductive argument, or perhaps, for lack of clarity, as none of these.

Among the various respects here considered in which a proposed explanation or demonstration may fall short of the logical standards incorporated into some nonpragmatic model of explanation or proof, there are several which can be characterized only by reference to the knowledge, interests, intentions, and so forth of the persons who propose the arguments in question or of those to whom they are addressed; hence, the corresponding concepts are essentially pragmatic. This is true, for example, of the notions of enthymeme, of elliptically formulated explanation, and of explanation sketch.

4.3 Concluding Remark on The Covering-Law Models. We have found, then, that the explanatory accounts actually formulated in science and in everyday contexts vary greatly in the explicitness, completeness, and precision with which they specify the explanans and the explanandum; accordingly, they diverge more or less markedly from the idealized and schematized covering-law models. But, granting this, I think that all adequate scientific explanations and their everyday counterparts claim or presuppose at least implicitly the deductive or inductive subsumability of whatever is to be explained under general laws or theoretical principles.[17] In the explanation of an individual occurrence, those general nomic principles are required to connect the explanandum event with other particulars, and it is by such nomic connection that the latter acquire the status of explanatory factors. In the explanation of general empirical regularities, the nomic principles invoked express more comprehensive uniformities of which those to be explained are strict or approximate specializations. And the covering-law models represent, as far as I can see, the basic logical structure of the principal modes of such explanatory subsumption.

The construal here broadly summarized is not, of course, susceptible to strict "proof"; its soundness has to be judged by the light it can shed on the rationale and force of explanatory accounts offered in different branches of empirical science. Some of the ways in which this construal of explanation may prove illuminating have already been suggested in the course of developing the covering-law models and characterizing their intended function; other such ways should come into view as we proceed, and particularly when we turn, in later sections, to an analysis of certain peculiar explanatory procedures that seem to be at variance with the covering-law construal of explanation.

5. PRAGMATIC ASPECTS OF EXPLANATION

5.1 INTRODUCTORY REMARKS. Very broadly speaking, to explain something to a person is to make it plain and intelligible to him, to make him understand it. Thus construed, the word 'explanation' and its cognates are *pragmatic* terms: their use requires reference to the persons involved in the process of explaining. In a pragmatic context we might say, for example, that a given account A explains fact X to person P_1. We will then have to bear in mind that the same account may well not constitute an explanation of X for another person P_2, who might not even regard X as requiring an explanation, or who might find the

17. This idea needs to be sharply distinguished from another one, which I am not proposing, namely, that any empirical phenomenon can be explained by deductive or inductive subsumption under covering laws. The idea here suggested is that the logic of all scientific explanations is basically of the covering-law variety, but not that all empirical phenomena are scientifically explainable, and even less, of course, that they are all governed by a system of deterministic laws. The question whether all empirical phenomena can be scientifically explained is not nearly as intelligible as it might seem at first glance, and it calls for a great deal of analytic clarification. I am inclined to think that it cannot be given any clear meaning at all; but at any rate, and quite broadly speaking, an opinion as to what laws hold in nature and what phenomena can be explained surely cannot be formed on analytic grounds alone but must be based on the results of empirical research.

account A unintelligible or unilluminating, or irrelevant to what puzzles him about X. Explanation in this pragmatic sense is thus a relative notion: something can be significantly said to constitute an explanation in this sense only for this or that individual.

Quite similarly, the word 'proof' and its cognates can be used in a pragmatic sense which requires reference to the producers and the recipients of the arguments in question. For example, an argument Y that proves a simple geometrical theorem to the complete satisfaction of a tyro may be entirely unacceptable, and thus not a proof at all, for a mathematician; and conversely, what for the mathematician is a sound and illuminating proof may be unintelligible or pointless to the beginner. Generally, whether a given argument Y proves (or explains) a certain item X to a person P will depend not only on X and Y, but quite importantly also on P's beliefs at the time as well as on his intelligence, his critical standards, his personal idiosyncrasies, and so forth.

The pragmatic aspects of proof form an interesting and important subject for empirical investigation. Piaget, for example, has devoted a great deal of effort to the psychological study of the standards of proof in children of different ages. But for the purposes of mathematics and logic as objective disciplines, we clearly need a concept of proof which is not subjective in the sense of being relative to, and variable with, individuals; a concept in terms of which it makes sense to say that a given argument Y is a proof of a given sentence X (in a theory) without making any mention of persons who might take cognizance of Y. Concepts of proof which have this character can be defined once the mathematical discipline in reference to which the concept is to be used has been suitably formalized.

The case of scientific explanation is similar. For scientific research seeks to account for empirical phenomena by means of laws and theories which are objective in the sense that their empirical implications and their evidential support are independent of what particular individuals happen to test or to apply them; and the explanations, as well as the predictions, based upon such laws and theories are meant to be objective in an analogous sense. This ideal intent suggests the problem of constructing a nonpragmatic concept of scientific explanation—a concept which is abstracted, as it were, from the pragmatic one, and which does not require relativization with respect to questioning individuals any more than does the concept of mathematical proof. It is this nonpragmatic conception of explanation which the covering-law models are meant to explicate.

To propound those models is therefore neither to deny the pragmatic "dimension" of explanation nor to belittle its importance; nor, of course, is it to claim that people will find an explanatory account illuminating or satisfactory only as far as it conforms to one of the covering-law models. To explain a given phenomenon to a person, it will often suffice to call to his attention some particular fact of which he has not properly taken cognizance. This is presumably true of the man mentioned in a newspaper story some years ago who was puzzled to find that it got cold in his house whenever he happened to watch a television program in winter. All he had to be told by way of explanation was that the television set was directly under the thermostat,and thus warmed the latter and shut the heating off. Thus the quest for an explanation is often a quest for the "cause" of the puzzling occurrence, in the loose sense here illustrated. The questioner who accepts a particular causal account as satisfactory will sometimes have background information of a nomological kind—e.g., about the way a thermostat works—which might justify the causal attribution. In other cases, he may lack such information and might still be satisfied by the explanation: the pragmatic conditions for the acceptability of a proposed explanation do not coincide with the logic-systematic ones that the covering-law models are meant to explicate. When the relevant laws are more or less clearly understood and taken for granted by the questioner, it would of course be incorrect to say that his question had the pragmatic function of eliciting covering laws; but it is neither incorrect nor superfluous to make reference to such laws if the logic of the account, and especially the explanatory force of the particular facts mentioned in it, is to be made explicit.

In other contexts—for example, frequently in scientific research—the pragmatic concern prompting the quest for an explanation may be the desire to discover laws or theoretical principles covering a given class of phenomena. And in yet other cases, the questioner may be aware of the requisite particular data and laws but may need to be shown how the explanandum can be derived from this information.[1]

But to call attention to the important pragmatic facets of explanation and to indicate the diverse procedures that may be appropriate in different cases to dispel the perplexity reflected in someone's quest for an explanation is not to show that a nonpragmatic model of scientific explanation must be hopelessly inadequate, just as analogous arguments concerning the notion of proof cannot show that nonpragmatic models of proof must be sterile and unilluminating. As is well known, the contrary is the case.

It is therefore beside the point to complain that the covering-law models

1. In an interesting discussion of what are, to a large extent, pragmatic aspects of explanation, Scriven uses the term 'derivation-explanation' for an explanation that consists simply in demonstrating this derivability, and he gives an illustration from the history of science, which shows that the derivation may well present considerable mathematical difficulties and may thus be hard to discover. (Scriven 1959, pp. 461-62).

do not closely match the form in which working scientists actually present their explanations. Those formulations are generally chosen with a particular kind of audience—and thus with particular pragmatic requirements—in mind. This is true also of the way in which mathematicians present their proofs; but the metamathematical theory of proof quite properly abstracts from these pragmatic considerations.[2]

5.2 EXPLAINING HOW-POSSIBLY. An important pragmatic aspect of explanation is reflected in Dray's distinction of "explaining why-necessarily" an event occurred and "explaining how-possibly" an event could have occurred.[3] A D-N explanation might be regarded as adequate for the former purpose; to accomplish the latter is quite a different task, as we will now see.

If a friend tells me that at a party he attended last New Year's Eve his teaspoon promptly melted when he put it into a cup of hot punch, I might ask: how could this possibly have happened—metal does not melt at so low a temperature. Similarly, the news that the *Andrea Doria* had sunk as a result of a collision gave rise to the question how this could possibly have happened, considering that the ship was equipped with the most advanced safety devices and was operated by experienced seamen.

As these examples illustrate, we will normally ask how *X* could possibly have occurred only if, as Dray puts it, "what we know seems to rule out the possibility of the occurrence which is to be explained,"[4] i.e., if some of the beliefs we hold concerning relevant matters of fact seem to us to make it impossible or at least highly improbable that *X* should have occurred; herein lies the pragmatic aspect of the question. To give a satisfactory 'how-possibly' explanation, it will be necessary, therefore, to ascertain the empirical assumptions underlying the question and then to show either that some of these are false or else that the questioner was mistaken in thinking that those assumptions warranted his belief that *X* could not have occurred. In the case of the teaspoon, it might suffice to point out that some metals, such as Wood's alloy, do melt at the temperature of hot punch; and a full covering-law explanation might be achieved by establishing that the teaspoon in question had indeed been one of those made from Wood's alloy for the use of practical jokers.[5]

2. *Cf.* also the comments on this point in section 1 of Bartley's paper (1962), in which Popper's presentation of the deductive model is defended against this charge. For some observations in a similar vein, see Pitt (1959), pp. 585-86.

3. *Cf.* Dray (1957), pp. 158 ff.

4. Dray (1957), p. 161.

5. In a review of Dray's book, Passmore (1958) goes so far as to say that "to answer a 'how possibly' question, unless with a mere guess, is to sketch in a 'why-necessarily' explanation." While this observation seems basically sound, it should, I think, be liberalized so as to call for the sketching either of a 'why-necessarily' or else of a 'why-probably' explanation.

If, as in the case of the *Andrea Doria*, the question 'How could *X* possibly have occurred?' springs from assumptions that seem to make the occurrence of *X* highly improbable but not logically to preclude it, then an appropriate answer may consist in pointing out that the questioner is mistaken in some of his factual assumptions or in the belief that his assumptions make the occurrence of *X* very improbable: these two possibilities are analogous to those considered in the previous illustration. But in addition, we have here a third possibility, suggested also by our earlier discussion of the logic of statistical explanation: all of the questioner's relevant assumptions might be true, and his belief that they make the occurrence of *X* very improbable may be correct. In that event, the perplexity expressed by the questioner's 'how could it possibly have happened?' may be resolvable by broadening the questioner's total evidence, i.e., by calling to his attention certain further facts whose addition to those previously taken into account will render the occurrence of *X* less improbable.

Similar observations apply to questions of the form 'why is it not the case that *p*?', which might well be rephrased as 'how-possibly' questions: 'How could it possibly be the case that not-*p*?'. Questions such as 'Why doesn't the Leaning Tower of Pisa topple over?' or 'Why don't the antipodes fall off the earth?', 'If reflection in a plane mirror interchanges right and left, why not also top and bottom?' will normally be raised only if the questioner entertains certain assumptions concerning relevant empirical matters which seem to him to make it certain or, at any rate, highly probable that the specified phenomenon should occur. A pragmatically adequate answer again will have to clear up the empirical or logical misapprehensions underlying this belief.

And, of course, explanation-seeking questions of the standard type 'Why is it the case that *p*?' are often, though by no means invariably, prompted by the belief that *p* would not be the case—a belief which, again, may seem to the questioner to be more or less strongly supported by certain other empirical assumptions which he accepts as being true. And in this event, the questioner may not feel satisfied if he is simply offered, say, a covering-law explanation of

Someone who asks how *X* could possibly have happened will not, as a rule, be satisfied to be told simply that he was mistaken in some of his empirical assumptions, which he thought precluded the occurrence of *X*; he will also want to be given a set of alternative, and presumably true, assumptions which, in conjunction with the rest of his background beliefs, explain to him why *X* occurred. The case of the melting spoon illustrates this. But if our questioner should believe that spilling salt is always followed by bad luck within three days, and if he were to ask 'How possibly could I have escaped bad luck though I spilled some salt three days ago?', then the answer could hardly do more than point out that his general hypothesis was false and, perhaps, that in the vast majority of cases, spilling salt is not followed by bad luck; but no 'why-necessarily' explanation for the questioner's avoidance of bad luck will be available.

why *p* is the case. In order to allay his perplexity he may have to be shown that some of the assumptions underlying his contrary expectation were in error.[6]

5.3 Explanation vs. Reduction to the Familiar. A predominantly pragmatic conception of explanation as aimed at dispelling the questioner's puzzlement also underlies the widely held view that an explanation must somehow reduce or link the puzzling phenomenon to something with which the questioner is already familiar, and which he accepts as unproblematic. Thus, Bridgman, for example, holds that "the essence of an explanation consists in reducing a situation to elements with which we are so familiar that we accept them as a matter of course, so that our curiousity rests."[7] An examination of this explicitly pragmatic characterization may serve further to clarify and support the case for constructing a nonpragmatic concept of scientific explanation.

Undeniably, many scientific explanations effect, in a sense, a "reduction to the familiar." This might be said, for example, of the wave-theoretical explanation of optical refraction and interference, and of at least some of the explanations achieved by the kinetic theory of heat. In cases of this kind, the concepts and principles invoked in the explanans bear a more or less close resemblance to concepts and principles that have long been used in the description and explanation of some familiar type of phenomenon, such as the propagation of wave motions on the surface of water or the motion of billiard balls.

Concerning the general view of explanation as a reduction to the familiar, let us note first that what is familiar to one person may not be so to another, and that, therefore, this view conceives of explanation as something relative to a questioner. But, as we noted earlier, explanations of the kind empirical science seeks are intended to exhibit objective relationships.

Secondly, the view here under discussion suggests that what is familiar requires no explanation. But this notion does not accord with the fact that scientists have gone to great lengths in an effort to explain "familiar" phenomena, such as the changes of the tides; lightning, thunder, rain, and snow; the blue color of the sky; similarities between parents and their offspring; the fact that the moon appears much larger when it is near the horizon than when it is high in the sky; the fact that certain diseases are "catching," while others are not; and even the familiar fact that it is dark at night. Indeed, the darkness of

6. This aspect of explanation, and various related ones, have been perceptively and lucidly examined by S. Bromberger (1960). For suggestive observations on the pragmatic aspects of explanation, see also Passmore (1962).

7. Bridgman (1927), p. 37. The pragmatic character of this conception is clearly reflected in Bridgman's remark that "an explanation is not an absolute sort of thing, but what is satisfactory for one man will not be for another." *Loc. cit.*, p. 38.

the night sky appears as a phenomenon much in need of explanation, in view of Olbers' paradox. This argument, put forward in 1826 by the German astronomer Heinrich Olbers, rests on a few simple assumptions, roughly to the effect that the distances and the intrinsic luminosities of the stars have about the same frequency distribution throughout the universe in the past as well as at present; that the basic laws for the propagation of light hold true in all spatio-temporal areas of the universe, and that the universe at large is static, i.e., that no large-scale systematic motions take place in it. From these assumptions it follows that the sky, in all directions and at all times, should be of enormous uniform brightness, and that the energy thus streaming in upon the surface of the earth should correspond to a temperature of more than 10,000 degrees Fahrenheit.[8]

Olbers' paradox thus raises a 'how-possibly?' question. An answer to it is suggested by the recent theory that the universe is steadily expanding. This theory implies, first, that Olbers' assumption of a static universe is in error, and it supplies, secondly, a positive explanation of the dark night sky by showing that the energy of the radiation received from very distant stars is enormously reduced by the high velocities of their recession.

This example also illustrates a further point, namely, that instead of reducing the unfamiliar to the familiar, a scientific explanation will often do the opposite: it will explain familiar phenomena with the help of theoretical conceptions which may seem unfamiliar and even counter-intuitive, but which account for a wide variety of facts and are well supported by the results of scientific tests.[9]

These observations are applicable also outside the domain of the natural sciences. Their relevance to sociology, for example, is suggested in the opening passage of a book by Homans: "My subject is a familiar chaos. Nothing is more familiar to men than their ordinary, everyday social behavior . . . every man makes his own generalizations about his own social experience, but uses them *ad hoc* within the range of situations to which each applies, dropping them as soon as their immediate relevance is at an end and never asking how they are related to one another . . . the purpose of this book is to bring out of the familiar chaos some intellectual order."[10] Incidentally, Homans goes on to say that the requisite ordering of a body of empirically established sociological facts, represented by low-level generalizatons, calls for an *explanation* of those facts; and that such explanation is achieved by means of a "set of more general propositions, still of the same form as the empirical ones, from which you can

8. For a fuller presentation of the paradox, and a critical analysis in the light of current cosmological theorizing, see, for example, Bondi (1961), chapter 2, and Sciama (1961), chapter 6.

9. This point is stressed also in Feigl's concise and illuminating article (1948); and it is lucidly illustrated by reference to the theory of relativity in Frank (1957), pp. 133-34.

10. Homans (1961), pp. 1-2.

logically deduce the latter under specified given conditions. To deduce them successfully is to explain them."[11]

To this emphasis on the sociologist's interest in the theoretical explanation of "familiar" generalizations about social behavior, there should be added a reminder that has been stressed by Lazarsfeld, among others; namely, that what are widely regarded as obvious and familiar facts of everyday psychological and sociological experience are sometimes not facts at all but popular stereotypes. This is true—to mention but one of Lazarsfeld's interesting illustrations—of the idea that the intellectual is emotionally less stable than the psychologically more impassive man-in-the-street, and that therefore it was to be expected that among the U.S. soldiers in the Second World War, better educated men showed more psychoneurotic symptoms than those with less education. In fact, the opposite was found to be the case.[12] Thus an explanation of some particular case by reference to the low-level generalization of this sterotype is simply false even though it might be said to effect a reduction to the familiar.

Such reduction, then, as has now been argued at some length, is surely not a necessary condition for an acceptable scientific explanation. But neither is it a sufficient condition; for a request for an explanation is sometimes answered in a way which puts the questioner's curiosity to rest by giving him a sense of familiarity or at-homeness with an initially puzzling phenomenon, without conveying a scientifically acceptable explanation. In this case, one might say, familiarity breeds content, but no insight. For example, as we have just seen, the proffered explanation might be based on a familiar and yet mistaken belief, and will then be false. Or the proposed account might rely on untestable metaphorical or metaphysical ideas rather than on general empirical hypotheses, and then would not afford even a potential scientific explanation. Take for example the "hypothesis of a common subconscious," which has been propounded to explain presumptive telepathic phenomena.[13] It asserts that while in their conscious domains human minds are separate entities, they are connected by a common subconscious, from which the individual consciousnesses emerge like mountainous islands joined by a submarine continent. The suggestive imagery of this account may well evoke a sense of intuitive understanding of telepathic phenomena; the latter seem to have been explained by reduction to ideas with which we are quite familiar. Yet we have been given a simile rather than a scientific explanation. The account offers us no grounds

11. Homans (1961), pp. 9-10, italics the author's.

12. See Lazarsfeld (1949), pp. 379-80.

13. See the critical reference in Price (1945) and *cf.* Carington's use of the idea as "a simile" (1949, pp. 223ff.), as well as his more specific account of the conception of a common subconscious, *loc. cit.*, pp. 208ff.

on which it would be reasonable to expect the occurrence of telepathic phenomena, nor does it give us any clues as to the conditions under which such phenomena are likely to occur. Indeed, in the form here outlined the notion of a common subconscious has no clear implications concerning empirical phenomena and is not amenable, therefore, to objective test or to significant explanatory or predictive use.

A similar critique applies to neovitalistic explanations of certain biological phenomena in terms of entelechies or vital forces. Such accounts do not specify under what conditions a vital force will exert its influence and what specific form its manifestations will take, nor, in the case of external interference with an organism, to what extent an entelechy will compensate for the resulting disturbance. By contrast, an explanation of planetary motions in terms of the Newtonian theory of gravitation specifies what gravitational forces will be exerted upon a given planet by the sun and by other planets, given their masses and distances, and it specifies further what changes in motion are to be expected as a result of those forces. Both accounts invoke certain "forces" that cannot be directly observed—one of them, vital forces, the other, gravitational ones; yet the latter account has explanatory status while the former does not. This is a consequence of the fact that the Newtonian theory offers specific laws governing gravitational forces, whereas neovitalism specifies no laws governing vital forces and is, in effect, only metaphorical. Thus, it is covering laws or theoretical principles that are crucial to a scientific explanation, rather than the sense of familiarity that its wording may impart.

The laws invoked in a proposed scientific explanation are of course capable of test; and adverse test results may lead to their rejection. No such fate threatens explanations in terms of similes or metaphors: since they do not specify what to expect under any empirical conditions, no empirical test can possibly discredit them. But absolute immunity to disconfirmation is not an asset but a fatal defect when we are concerned, as is scientific research, to arrive at an objectively testable and empirically well-supported body of empirical knowledge. An account that has no implications concerning empirical phenomena cannot serve this purpose, however strong its intuitive appeal: from the point of view of science, it is a *pseudo-explanation*, an explanation in appearance only.

In sum then, it is neither necessary nor sufficient for the scientific adequacy of an explanation that it should reduce the explanandum to ideas with which we are already familiar.

6. MODELS AND ANALOGIES IN SCIENTIFIC EXPLANATION

Explanatory accounts offered in empirical science are sometimes formulated in terms of a "model" of the phenomena to be explained, or in terms of analogies between those phenomena and others that have been previously explored. In the present section I propose to examine some forms of this procedure and to appraise their explanatory significance.

Let us consider first the use—quite widespread in the nineteenth and early twentieth centuries—of more or less complex mechanical systems as models of electric, magnetic, and optical phenomena, of the luminiferous ether, and so forth. The importance that some eminent scientists attributed to such representations is reflected in the famous pronouncement of Sir William Thomson (later Lord Kelvin):

> I never satisfy myself until I can make a mechanical model of a thing. If Ican make a mechanical model I can understand it. As long as I cannot make a mechanical model all the way through I cannot understand. . . .[1]
>
> My object is to show how to make a mechanical model which shall fulfill the conditions required in the physical phenomena that we are considering, whatever they may be. At the time when we are considering the phenomenon of elasticity in solids, I want to show a model of that. At another time, when we have vibrations of light to consider, I want to show a model of the action exhibited in that phenomenon. . . . It seems to me that the test of "Do we or not understand a particular subject in physics?" is, "Can we make a mechanical model of it?"[2]

Sir Oliver Lodge, whose book on electricity presents a multitude of mechanical models, says in a similar vein:

> Think of electrical phenomena as produced by an all-permeating liquid embedded in a jelly; think of conductors as holes and pipes in this jelly, of an electrical machine as a pump, of charge as excess or defect, of attraction as due to strain, of discharge as bursting. . . . By thus thinking you will get a more real grasp of the subject and insight into the actual processes occurring in Nature—unknown though these may still strictly be—than if you employed the old ideas of action at a distance, or contented yourselves with no theory at all on which to link the facts. . . . I am also convinced that it is unwise to drift along among a host of complicated phenomena without guide other than that afforded by hard and rigid mathematical equations.[3]

These pronouncements reflect variants of the idea that explanation in science must involve a reduction to the familiar. What this variant demands is not simply that an explanation somehow render a phenomenon plausible or familiar, but more specifically that it provide a model governed by the laws of mechanics, which in this context are accorded the status of familiar principles.

But just what does the construction of a mechanical model accomplish? It is not intended, of course, to identify the modeled phenomenon with the

1. Thomson (1884), pp. 270-71.
2. Thomson (1884), pp. 131-32.
3. Lodge (1889), pp. 60-61.

model. An electric current maintained in a wire by means of a battery is not claimed to be the same thing as the flow of a liquid through pipes, maintained by means of a pump, nor the same thing as an inextensible loop of cord kept circulating over pulleys by means of a sinking weight.[4] The claim is merely that there obtains an analogy between the model and the phenomenon it represents. And the relevant analogy lies in a formal similarity between certain laws governing the mechanical system and corresponding laws for the modeled phenomenon.

Consider, for example, the often cited analogy between the flow of an electric current in a wire and the flow of a fluid in a pipe. If the fluid flows with moderate speed through a fairly narrow pipe with circular inner cross section then according to Poiseulle's law the volume V of fluid flowing through a fixed cross-section per second is proportional to the difference in pressure between the ends of the pipe:

(6.1a) $V = c \cdot (p_1 - p_2)$

This law has the same form as Ohm's law for the flow of electricity in a metallic conductor:

(6.1b) $I = k \cdot (v_1 - v_2)$

Here the strength of the current, I, may be said to represent the amount of electric charge flowing through a fixed cross-section of the wire per second; $v_1 - v_2$ is the potential difference maintained between the ends of the wire; and k is the reciprocal of its resistance.

The analogy goes further. The factor c in (6.1a) is inversely proportional to the length l_1 of the pipe:

(6.2a) $c = \dfrac{c'}{l_1}$

and similarly, the factor k in (6.1b) is inversely proportional to the length, l_2, of the wire:

(6.2b) $k = \dfrac{k'}{l_2}$

Thus, the analogy in virtue of which the flow of a fluid here constitutes a model of the flow of a current may be characterized as follows: A certain set of laws governing the former phenomenon has the same syntactical structure as a corresponding set of laws for the latter phenomenon; or, more explicitly,

4. A profusion of such models can be found in Lodge (1889) and in Thomson (1884). A hydrodynamic model that represents in quite a similar manner certain aspects of the behavior of nervous systems is described in S. B. Russell (1913).

the empirical terms (i.e., those which are not logical or mathematical)[5] occurring in the first set of laws can be matched, one by one, with those of the second set in such a way that if in one of the laws of the first set each term is replaced by its counterpart, a law of the second set is obtained; and vice versa. Two sets of laws of this kind will be said to be syntactically isomorphic. Briefly, then, the relevant similarity or "analogy" between a model of the kind here considered and the modeled type of a phenomenon consists in a *nomic isomorphism*, i.e., a *syntactic isomorphism between two corresponding sets of laws*. The notion of model thus obtained is not limited to mechanical systems, of course; we can speak, in the same sense, also of electrical, chemical, and still other kinds of "analogical models."

But in our illustration, as in other cases of analogical modeling, the isomorphism has its limits: some laws for the flow of a fluid in pipes do not carry over to electric currents in wires. For example, if the length of the pipe and the pressure difference between its ends are fixed, V is proportional to the fourth power of the radius of the cross sections, whereas under corresponding circumstances, the current is proportional to the square of the wire's cross section:

$$(6.3a)\quad V = \frac{\pi r_1^4}{8 l_1 s}\,(p_1 - p_2)$$

$$(6.3b)\quad I = \frac{\pi r_2^2}{l_2 q}\,(v_1 - v_2)$$

Here, s is the viscosity of the fluid and q the specific resistance of the metal of which the wire is made; r_1 is the radius of the inner cross section of the pipe; and r_2 is the radius of the wire.

Thus, the statement that a system S_1 is an analogical model of a system S_2 is elliptical. A complete sentence expressing the relationship would have to take the form: 'S_1 is an analogical model of S_2 with respect to the sets of laws L_1, L_2'. This sentence is true if the laws in L_1 apply to S_1 snd those in L_2 to S_2, and if L_1 and L_2 are syntactically isomorphic.[6]

The concept of analogy as a nomic isomorphism plays an important role in Maxwell's essay on Faraday's lines of force. Maxwell here says: "By a physical analogy I mean that partial similarity between the laws of one science

5. Physical constants such as 'S' and 'q' in (6.3a) and (6.3b) count here as empirical terms.

6. This characterization of analogical models accords with Maxwell's and Duhem's conceptions of analogy in physics, about which more will be said presently. It is also supported by the way in which Boltzmann (1891) uses mechanical models to represent the Carnot cycle in the theory of heat (1891, chapter 2) and various electric phenomena. Heinrich Hertz's general concept of a "dynamic model" reflects the same basic idea; *cf.* Hertz (1894), p. 197.

and those of another which makes each of them illustrate the other." He notes, concerning the analogy between light and the vibrations of an elastic medium, that "though its importance and fruitfulness cannot be overestimated, we must recollect that it is founded only on a resemblance *in form* between the laws of light and those of vibrations."[7] Maxwell continues: "It is by the use of analogies of this kind that I have attempted to bring before the mind, in a convenient and manageable form, those mathematical ideas which are necessary to the study of the phenomena of electricity. . . . I am not attempting to establish any physical theory . . . , and . . . the limit of my design is to shew how, by a strict application of the ideas and methods of Faraday, the connexion of the very different orders of phenomena which he has discovered may be clearly placed before the mathematical mind."[8] The analogy Maxwell then develops in detail rests on a representation of Faraday's lines of force by tubes through which an incompressible liquid flows. It is of interest to note that while Maxwell is able to give an analogical representation of a great many electric and magnetic phenomena, he finds himself unable to extend the analogy when he comes to the discussion of what Faraday had called the electro-tonic state; here, he resorts to the formulation of a theory in purely mathematical form.[9]

The views of men like Kelvin and Lodge concerning the importance of analogical models for explanation in physics were severely criticized by Duhem. Duhem sees the aim of physics in the construction of theories couched in precise mathematical terms, from which empirically established laws can be deduced, and he argues that mechanical models contribute nothing to that objective. In reference to Lodge's book Duhem comments: "Here is a book meant to expound the modern theories of electricity . . .; it talks only of cords that move over pulleys, that wind themselves up on drums, that traverse beads, that carry weights; of tubes that pump water and of others that expand and contract; of cog wheels that mesh with each other and drive toothed racks; we thought we were entering the peaceful and carefully ordered abode of reason, and we find ourselves in a factory."[10] Duhem goes on to complain that far from facilitating the understanding of a theory "for a French reader," the use of such mechanical models requires of him a serious effort just to understand the working of the complicated apparatus and to recognize analogies between the properties of the model and the theory that is being illustrated.

Although Duhem rejects the explanatory use of mechanical models, he

7. Maxwell (1864), p. 28, italics the author's.

8. Maxwell (1864), p. 29.

9. Maxwell (1864), pp. 51ff. For a fuller discussion of Maxwell's views on the importance of analogies for physical theorizing, see Turner's studies (1955), (1956).

10. Translated from Duhem (1906), p. 111.

stresses that, by contrast, analogies may prove very fruitful in physical research. The analogies he has in mind are those based on what we have called nomic isomorphisms. He mentions, for example, Ohm's transfer of the laws of heat conduction to electric conduction, and he stresses the importance of those cases in which extensive theories for two distinct and dissimilar categories of phenomena have the same algebraic form.[11]

However, if our characterization is correct, then the mechanical models scorned by Duhem exhibit nomic isomorphisms of basically the same kind as those scientific analogies in Duhem's sense which are not specifically formulated in the parlance of models. Duhem's distinction between models and analogies, for which he states no precise criteria, then reflects not a difference in logical status, but rather a difference in the precision and the scope of the isomorphic sets of laws. Among the laws governing a mechanical model, those which carry over isomorphically to the modeled phenomenon are usually few in number and limited in scope, so that sometimes several different models are used to represent different aspects of one kind of physical entity or phenomenon. For example, Kelvin offers quite different mechanical models of molecules to represent elasticity in crystals, the dispersion of light, and the rotation of the plane of polarization of a light beam;[12] and Lodge designs entirely different mechanical systems, of the sort referred to by Duhem in the passage quoted earlier, to represent various electrostatic, electrodynamic, and electromagnetic phenomena. In the case of fruitful analogies of the kind envisaged by Duhem, on the other hand, the isomorphic laws or theoretical principles are stated in precise mathematical terms and are strong enough to permit the deduction of a great variety of consequences which themselves constitute important laws. This is illustrated by the extensive nomic isomorphisms that permit the application of the mathematical theory of wave motions to certain parts of mechanics, optics, and quantum mechanics.[13]

In order to appraise the explanatory significance of analogical models, and more generally of analogies based on nomic isomorphisms, let us suppose that some "new" field of inquiry is being explored, and that we try to explain the phenomena encountered in it by analogical reference to some "old," previously explored domain of inquiry. This calls for the establishment of an isomorphism

11. Duhem (1906), pp. 152-54. Boltzmann characterizes physical analogies in a similar manner: ". . . Nature seemed, as it were, to have built the most diverse things exactly according to the same plan, or, as the analytic mathematician says dryly, the same differential equations hold for the most diverse phenomena." Translated from Boltzmann (1905), p. 7.

12. *Cf.* Thomson (1884).

13. Further examples of analogies based on nomic isomorphisms in physics will be found in Seeliger's article (1948); for an illuminating discussion, well illustrated by examples, of the significance of nomic isomorphisms in physics, see also Watkins (1938), chapter 3.

between a set of laws, say L_1, pertaining to the old field and a corresponding set, say L_2, in the new. To that end, we obviously must first discover a suitable set L_2 of laws in the new field. But once this has been done, those laws can be used directly for the explanation of the "new" phenomena, without any reference to their structural isomorphism with the set L_1. For the systematic purposes of scientific explanation, reliance on analogies is thus inessential and can always be dispensed with.

This observation applies equally to analogical models of a nonmechanical sort, such as the physico-chemical systems which have been used to imitate phenomena that are often considered as specifically biological. Leduc, for example,[14] was able to produce by purely chemical means a large variety of osmotic growths whose highly diversified forms strikingly resemble those of familiar plants and animals, and which, in their development, exhibit remarkable analogies to organic growths. The analogical models thus obtained are based on an isomorphism of non-quantitative laws:

> An osmotic growth has an evolutionary existence; it is nourished by osmosis and intussusception; it exercises a selective choice on the substances offered to it; it changes the chemical constitution of its nutriment before assimilating it. Like a living thing it ejects into its environment the waste products of its function. Moreover, it grows and develops structures like those of living organisms, and it is sensitive to many exterior changes, which influence its form and development. But these very phenomena—nutrition, assimilation, sensibility, growth, and organization—are generally asserted to be the sole characteristics of life.[15]

These analogies, and various others, between organisms and physico-chemical systems have often been used to answer the vitalistic claim that growth, metabolism, regeneration, and the like are phenomena that cannot be exhibited by a "machine" or by a system governed exclusively by physico-chemical laws.[16] But, while the models can refute that contention, they do not provide a positive theoretical explanation of the biological phenomena in question. In fact, Leduc does not even state any physico-chemical laws that would explain the striking plantlike shapes exhibited by some of the osmotic growths he produces by chemical means; even less, therefore, does he establish that the

14. See Leduc's profusely illustrated books (1911), (1912).
15. Leduc (1911), p. 159.
16. *Cf.*, for example, the crystal analogy, which is discussed in Bertalanffy (1933), pp. 100-102; and see also the instructive discussion of physico-chemical models of biological phenomena in Bonhoeffer (1948), where the motivating consideration here referred to is explicitly suggested. In this context, we might mention also some more recent physical models of certain aspects of learning, whose construction, again, is prompted at least in part by the desire to counter vitalistic and similar claims: such models are presented in Baernstein and Hull (1931) and Krueger and Hull (1931).

same laws also account for the shapes of the "natural" plants modeled by those artificial growths. Similar comments apply to "metabolism," "regeneration," and so forth in osmotic and in organic growths.

Besides, the isomorphisms exhibited by Leduc's and similar models concern only regularities of a vague qualitative kind illustrated by the passage quoted above: organisms grow and decay, and so do their osmotic counterparts; there is an exchange of materials between organism and environment, and an exchange of materials between each of the models and its environment; there is some measure of repair of injuries in organisms and in their physico-chemical models, and so on. Because of their lack of specificity, generalizations of this kind do not have much explanatory force. In this respect, the analogies here exhibited are vastly inferior to those between water waves and electromagnetic waves, for example, which rest on a syntactical isomorphism of two extensive theories formulated in mathematical terms.

As we noted, all references to analogies or analogical models can be dispensed with in the systematic statement of scientific explanations. But the discovery of an isomorphism between different sets of laws or theoretical principles may prove useful in other respects.

First, it may make for "intellectual economy":[17] If certain laws governing a "new" class of phenomena are isomorphic with those for another class, which have already been studied in detail, then all the logical consequences of the latter can be transferred to the new domain by simply replacing all extra-logical terms by their counterparts. An important study by Gauss[18] takes as its point of departure the observation that the forces of gravitational attraction and of electric and magnetic attraction and repulsion between any two "elements" are all inversely proportional to the square of their distance and directly proportional to the product of their masses or electric charges, or magnetic strengths, respectively. On the basis of this nomic isomorphism, Gauss develops a general mathematical theory for all forces governed by a law of the specified form, and especially for the corresponding potentials, without distinguishing between the different subject matters to which the resulting theory can be applied.[19] This aspect of nomic isomorphisms has recently found important practical applications in the construction of analogue computers and similar devices. For example, the isomorphism underlying the analogy between the flow of a

17. Duhem (1906), p. 154.

18. Gauss (1840.

19. The discovery and utilization of nomic isomorphisms between different fields of inquiry is one of the objectives of "general system theory" as conceived by Bertalanffy; see his brief statements (1951) and (1956), where many further references will be found. Some comments on the program of exploring isomorphisms in the manner envisaged by Bertalanffy are included in Hempel (1951a).

liquid through a pipe and the flow of an electric current through a wire enables the designer of a large and costly water-pumping system to determine the optimal characteristics of the pumps and the network of pipes by means of small and inexpensive electric analogues.

Analogies and models based on nomic isomorphisms may also facilitate one's grasp of a set of explanatory laws or theoretical principles for a new domain of inquiry by exhibiting a parallel with explanatory principles for a more familiar domain: in this manner, they can contribute to the pragmatic effectiveness of an explanation.

More important, well-chosen analogies or models may prove useful "in the context of discovery," i.e., they may provide effective heuristic guidance in the search for new explanatory principles. Thus, while an analogical model itself explains nothing, it may suggest extensions of the analogy on which it was originally based. Norbert Wiener mentions a case of this kind. An analogy he and Bigelow had envisaged between certain types of voluntary human behavior and the behavior of a machine governed by a negative feedback system suggested to them that there might exist, for purposive behavior, an analogue to the conditions, which are theoretically well understood, in which a feedback system breaks down through a series of wild oscillations. Such an analogue was indeed found in the pathological condition of purpose tremor, in which a patient trying to pick up an object overshoots the mark and then goes into uncontrollable oscillations.[20] To give another example: Maxwell appears to have arrived at his equations for the electromagnetic field by judicious use of mechanical analogies of electromagnetic phenomena. This led Boltzmann to say that the high praise Heinrich Hertz had bestowed on Maxwell's theoretical accomplishment was earned primarily by Maxwell's ingenuity in devising fruitful mechanical analogies rather than by his mathematical analysis.[21]

Analogies may prove useful in devising, and in expanding, microstructure theories such as the kinetic theory of heat or the theory accounting for the coding and transmission of genetic information in terms of specific hypotheses about the molecular structure of the genes. It should be noted, however, that such theories are intended to explain observable macrophysical uniformities by suitable assumptions about the underlying microphysical structures and processes and that the latter are not, as a rule, presented as analogical models only. When Lord Kelvin sought to account for uniformities in the absorption

20. See Wiener (1948), pp. 13-15 and chapter 4.

21. Boltzmann (1905), p. 8; also (1891), p. iii. For various other illustrations and an illuminating general discussion of the role of analogies in physical theorizing, see Nagel (1961), pp. 107-17.

and dispersion of light by construing each of the material molecules involved in these processes on the model of a set of nested rigid metal spheres separated from each other by springs, he did not, of course, claim to describe the actual microstructure of matter, and it would have been beside the point to request evidence in support of the assumption that molecules consist of nested metal spheres and springs. However, the kinetic theory of heat does assert, among other things, that a gas consists of molecules in rapid motion; it specifies the numbers and masses of the particles involved, the distribution of their velocities and its dependence on the temperature, the mean free paths of the molecules and the mean time interval between successive collisions, and so forth; and in regard to these and many other specific implications, supporting evidence can be significantly asked for and can indeed be supplied.

Similarly, theories about the elementary particles constituting the atomic nuclei of various elements, or about the molecular structure of the genes, are presented as accounts of the actual structure of the systems in question, and not just as analogical models. Like any other theory in empirical science, such microstructure theories are put forward "until further notice," i.e., with the understanding that they may have to be modified or completely withdrawn in the light of subsequently discovered unfavorable evidence; and often they are offered only as approximations. Nevertheless, they differ in the respect just indicated from accounts formulated in terms of analogical models.

In some microstructure theories, the basic constituents of the macrophenomena under study are assumed to be governed by laws that are identical or syntactically isomorphic with a set of laws governing an already well-explored field of inquiry. A characteristic example is the assumption that the motions and collisions of gas molecules conform to the laws for the motions and collisions of elastic billiard balls. Indeed, some writers have insisted that the basic assumptions or equations of any good scientific theory must exhibit that kind of analogy. One eloquent proponent of this view is the physicist N. R. Campbell.

Campbell considers it the principal function of theories to provide deductive explanations of laws, i.e., of "propositions which assert uniformities discovered by experiment or observation."[22] He characterizes a theory as consisting of two sets of propositions, which he calls the hypothesis and the dictionary. The hypothesis is formulated in terms of "ideas which are characteristic of the theory," or in terms of theoretical concepts, as we might say. The dictionary provides a physical interpretation of the hypothesis by translating some but not necessarily all of its propositions into others which involve no theoretical

22. Campbell (1920), p. 71.

concepts and which can be verified or falsified, without any reference to the theory, by suitable experiments or observations.[23]

Campbell demands of a scientific theory that it be capable of explaining empirically established laws: such explanation consists in deducing the laws from the hypothesis in conjunction with the dictionary. "But," he insists, "in order that a theory may be valuable it must have a second characteristic; it must display an analogy. The propositions of the hypothesis must be analogous to some known laws." He adds: "analogies are not 'aids' to the establishment of theories; they are an utterly essential part of theories, without which theories would be completely valueless and unworthy of the name."[24] In support of this contention, Campbell constructs a small quasi-theoretical system which does deductively imply an empirical law, but which clearly is not an acceptable scientific theory; and this, in Campbell's opinion, because its hypothesis lacks the requisite analogy to known laws. Let us briefly consider that system, which I will call S.[25]

The hypothesis of S is expressed in terms of four quantitative theoretical concepts a, b, c, d, which are functions of certain "independent variables" u, v, w,.... The hypothesis states that a and b are constant functions, and that c is identical with d.

The dictionary of S consists of the following two specifications: the statement that $(c^2 + d^2)\, a = R$, where R is a positive rational number, implies that the resistance of some particular piece of pure metal is R; and the statement that $cd/b = T$ implies that the temperature of the same piece of metal is T.

Now, the hypothesis of S deductively implies that

$$(c^2 + d^2)\, a \bigg/ \frac{cd}{b} = 2\, ab = \text{constant}$$

Interpreting the quotient on the left by means of the dictionary we obtain, according to Campbell, the following law: "The ratio of the resistance of a piece of pure metal to its absolute temperature is constant." (Actually, this

23. Campbell (1920), pp. 122, states: "The dictionary relates some of these propositions of which the truth or falsity is known to certain propositions involving the hypothetical ideas by stating that if the first set of propositions is true then the second set is true and *vice versa*; this relation may be expressed by the statement that the first set *implies* the second." (Italics supplied.) This is clearly a nonstandard use of the word 'implies'; in the following discussion, I will therefore use the phrase 'deductively implies' to refer to the nonsymmetrical logical relation, in contradistinction to the symmetrical relation which Campbell has in mind, and which I suggested by saying that according to Campbell the dictionary *translates* certain theoretical propositions into empirical ones.

24. Campbell (1920), p. 129.

25. See Campbell (1920), pp. 123-24.

proposition follows only for the particular piece of metal referred to in the dictionary; but let us waive this point as inessential for the idea under consideration).

This law, then, is logically deducible from the system S and is in this sense explained by S. But Campbell argues: "If nothing but this were required we should never lack theories to explain our laws; a schoolboy in a day's work could solve the problems at which generations have laboured in vain by the most trivial process of trial and error. What is wrong with the theory . . . , what makes it absurd and unworthy of a single moment's consideration, is that it does not display any analogy."[26]

Campbell is certainly right in rejecting the "theory" S, but his diagnosis of its shortcomings seems to me incorrect. What is wrong with the theory, so it seems to me, is that it has no empirically testable consequences other than the law in question (and whatever is logically implied by it alone); whereas a worthwhile scientific theory explains an empirical law by exhibiting it as one aspect of more comprehensive underlying regularities, which have a variety of other testable aspects as well, i.e., which also imply various other empirical laws. Such a theory thus provides a systematically unified account of many different empirical laws. Besides, as was noted in section 2, a theory will normally imply refinements and modifications of previously established empirical laws rather than deductively imply the laws as originally formulated.

The diagnosis that it is this defect rather than the absence of analogy which disqualifies S can be further supported by the observation that systems can readily be constructed which do display some analogy to known laws and which are nevertheless worthless for science because they suffer from the same defect as S. For example, let the hypothesis of a system S' assert of four theoretical quantities a, b, c, d that for any object u,

$$c(u) = \frac{k_1 a(u)}{b(u)}; \quad d(u) = \frac{k_2 b(u)}{a(u)}$$

where k_1 and k_2 are numerical constants; and let the dictionary of S' specify that for any piece u of pure metal, $c(u)$ is its resistance and $d(u)$ the reciprocal of its absolute temperature. Then S', too, deductively implies the law cited above, and, in addition, each of the two propositions in the hypothesis displays an analogy to a known law; for example, to Ohm's law. Yet, S' does

26. Campbell (1920), pp. 129-30. Campbell allows, however, that there is a type of theory, illustrated by Fourier's theory of heat conduction, for which analogy may play a less important role (pp. 140-44). For the purposes of the present discussion, those theories clearly need not be considered.

not qualify as a scientific theory any more than does *S*, and clearly for the same reason.

While thus, in my judgment, Campell fails to establish that analogy plays an essential logic-systematic role in scientific theorizing and theoretical explaining, some of his pronouncements squarely place his requirement of analogy within the domain of the pragmatic-psychological aspects of explanation. This is illustrated by his statement that "an analogy is a function of the contemplating mind; when we say that one set of propositions is analogous to another we are saying something about its effect on our minds; whether or no it produces that effect on the minds of others, it will still have that effect on our own."[27] Surely, analogy thus subjectively conceived cannot be an indispensible aspect of objective scientific theories.

Considering the great heuristic value of structural analogies, it is natural that a scientist attempting to frame a new theory should let himself be guided by concepts and laws that have proved fruitful in previously explored areas. But if these should fail, he will have to resort to ideas that depart more and more from the familiar ones. In Bohr's early theory of the atom, for example, the assumption of electrons orbiting around the nucleus without radiating energy violates the principles of classical electrodynamics; and in the subsequent development of quantum theory, the analogy of the basic theoretical principles to "known laws" has been reduced considerably further in return for increased scope and greater explanatory and predictive power.

What remains as the principal requirement for scientific explanation is thus the inferential subsumption of the explanandum under comprehensive general principles, irrespective of the analogies these may display to previously established laws.

There is yet another kind of model, often referred to as theoretical or mathematical model, which is widely used for explanatory purposes, for example in psychology, sociology, and economics. It is exemplified by the numerous mathematical models of learning, by theoretical models of attitude change and of conflict behavior, and by a great variety of models for social, political, and economic phenomena.[28]

27. Campbell (1920), p. 144. For further light on these issues see Hesse (1963); chapter 2 of this book has the form of a dialogue between a "Campbellian" and a "Duhemian", in which various arguments concerning the significance of models and analogies for scientific theorizing are surveyed and suggestively appraised.

28. The relevant literature is vast, and only a very few specific references can be given here. A particularly lucid general discussion of theoretical models in psychology, together with a specific model of conflict behavior, is presented in Miller (1951). On models for learning, see for example Bush and Mosteller (1955); the introduction of this book lucidly formulates the methodology of the authors' procedure. The collective volume Lazarsfeld (1954)

(continued overleaf)

Broadly speaking, and disregarding many differences in detail, a theoretical model of this kind has the character of a theory with a more or less limited scope of application. Its basic assumptions concern interdependencies of different characteristics of the subject matter in question. Those characteristics are often, but not always, represented by quantitative parameters or "variables"; these may be more or less directly observable or measurable, or they may have the status of theoretical concepts with at least a partial empirical interpretation, effected, perhaps, by "operational definition." This is true, for example, of those parameters which represent statistical probabilities for certain kinds of behavior. The basic hypotheses of the model often construe some of the parameters as mathematical functions of others, but they do not always have this quantitative character.[29] From the basic hypotheses, in conjunction with the interpretation, specific consequences can be inferred concerning the empirical phenomena to which the model pertains: thus, it becomes possible to test the model and to put it to explanatory and predictive use. The resulting explanations and predictions may be deductive-nomological or inductive-statistical, depending on the form of the hypotheses included in the model.

The use of the term 'theoretical model' rather than 'theory' is perhaps meant to indicate that the systems in question have distinct limitations, especially when compared with advanced physical theories. To begin with, their basic assumptions are often known to be idealizations or oversimplifications. For example, they may disregard certain factors that are known to be of some relevance to the given subject matter; this would be true, e.g., of a theoretical model for economic behavior based on the assumption of strict economic rationality of the agents concerned. Next, the formulation of the interrelations between different factors may be deliberately oversimplified, perhaps in order to make the application of the model to particular cases mathematically manageable. In addition, the class of phenomena with which the model is concerned may be quite limited; for example, a theoretical model of decision making under risk might be restricted to decisions which are made under rather arti-

29. This is true, for example, of Miller's theoretical model of conflict behavior, which is formulated in terms of comparative hypotheses such as "The tendencey to approach a goal is stronger the nearer the subject is to it." Miller (1951), p. 90.

includes presentations of mathematical models for various aspects of social behavior as well as essays devoted to the analysis of particular models or to general problems concerning the methodology of model construction. An excellent general account of the role of mathematical models in the social sciences is given in Arrow (1951), and the symposia Society for Experimental Biology (1960) and International Union of History and Philosophy of Sciences (1961) contain some interesting papers on the role of models in empirical science. The essay Brodbeck (1959) includes illuminating observations on the character and the function of theoretical models.

ficial experimentally controlled conditions, and which are limited to a small number of rather trivial options.

But such peculiarities can also be found in the field of physical theorizing, and they do not bar the systems in question from the status of potentially explanatory theories. However, a limited scope and only approximate validity within that scope may severely restrict the actual explanatory and predictive value of a theoretical model.

7. GENETIC EXPLANATION AND COVERING LAWS

The covering-law models have often been criticized on the ground that while they may correctly represent the structure and the import of some of the explanations put forward by empirical science, they fail to do justice to many others. In the present section and in those that follow I propose to examine some important modes and aspects of scientific explanation that have been cited in support of this contention, and I will attempt to indicate what light the covering-law conception can shed upon their logic and their force.

One explanatory procedure, which is widely used in history, though not in history alone, is that of genetic explanation; it presents the phenomenon under study as the final stage of a developmental sequence, and accordingly accounts for the phenomenon by describing the successive stages of that sequence.

Consider, for example, the practice of selling indulgences, in the form it had taken when Luther was a young man. The ecclesiastic historian H. Boehmer tells us that until the beginning of the twentieth century, "the indulgence was in fact still a great unknown quantity, at sight of which the scholar would ask himself with a sigh: 'Where did it come from?'" An answer was suggested by Adolf Gottlob, who tackled the problem by asking himself what led the popes and bishops to offer indulgences. As a result, ". . . origin and development of the unknown quantity appeared clearly in the light, and doubts as to its original meaning came to an end. It revealed itself as a true descendant of the time of the great struggle between Christianity and Islam, and at the same time a highly characteristic product of Germanic Christianity."[1]

According to this conception,[2] the origins of the indulgence date back to the ninth century, when the popes were strongly concerned with the fight against Islam. The Mohammedan fighter was assured by the teachings of his

1. Boehmer (1930), p. 91. Gottlob's study, *Kreuzablass und Almosenablass*, was published in 1906; *cf.* the references to the work of Gottlob and other investigators in Schwiebert (1950), notes to chapter 10.

2. I am here following the accounts in Boehmer (1930), chapter 3 and in Schwiebert (1950), chapter 10.

religion that if he were to be killed in battle his soul would immediately go to heaven, but the Christian had to fear that he might still be lost if he had not done the regular penance for his sins. To allay these doubts, John VII, in 877, promised absolution for their sins to crusaders who should be killed in battle. "Once the crusade was so highly thought of, it was an easy transition to regard participation in a crusade as equivalent to the performance of atonement . . . and to promise remission of . . . penances in return for expeditions against the Church's enemies."[3] Thus, there was introduced the indulgence of the Cross, which granted complete remission of the penitential punishment to all those who participated in a religious war. "If it is remembered what inconveniences, what ecclesiastical and civil disadvantages the ecclesiastical penances entailed, it is easy to understand that penitents flocked to obtain this indulgence."[4] A further strong incentive came from the belief that whoever obtained an indulgence secured liberation not only from the ecclesiastical penances, but also from the corresponding suffering in purgatory after death. The benefits of these indulgences were next extended to those who, being physically unfit to participate in a religious war, contributed the funds required to send a soldier on a crusade. In 1199, Pope Innocent III recognized the payment of money as adequate qualification for the benefits of a crusading indulgence.

When the crusades were on the decline, new ways were explored of raising funds through indulgences. Thus, there was instituted a "jubilee indulgence," to be celebrated every hundred years, for the benefit of pilgrims coming to Rome on that occasion. The first of these indulgences, in 1300, brought in huge sums of money, and the interval between successive jubilee indulgences was therefore reduced to 50, 33, and even 25 years. And from 1393 on, the jubilee indulgence was made available, not only in Rome, but everywhere in Europe, through special agents who were empowered to absolve penitent sinners upon receiving appropriate payment. The development went still further: in 1477, a dogmatic declaration by Sixtus IV attributed to the indulgence the power of delivering even the dead from purgatory.

Undeniably, a genetic account of this kind can enhance our understanding of a historical phenomenon. But its explanatory role seems to me basically nomological in character. For the successive stages singled out for consideration surely must be qualified for their function by more than the fact that they form a temporal sequence and that they all precede the final stage, which is to be explained: the mere enumeration in a yearbook of "the year's important events" in the order of their occurrence clearly is not a genetic explanation of the final event or of anything else. In a genetic explanation each stage must be

3. Boehmer (1930), p. 92.
4. Boehmer (1930), p. 93.

shown to "lead to" the next, and thus to be linked to its successor by virtue of some general principles which make the occurrence of the latter at least reasonably probable, given the former. But in this sense, even successive stages in a physical phenomenon such as the free fall of a stone may be regarded as forming a genetic sequence whose different stages—characterized, let us say, by the position and the velocity of the stone at different times—are interconnected by strictly universal laws; and the successive stages in the movement of a steel ball bouncing its zigzaggy way down a Galton Board[5] may be regarded as forming a genetic sequence with probabilistic connections.

The genetic accounts given by historians are not, of course, of the purely nomological kind suggested by these examples from physics. Rather, they combine a certain measure of nomological interconnecting with more or less large amounts of straight description. For consider an intermediate stage mentioned in a genetic account. Some aspects of it will be presented as having evolved from the preceding stages (in virtue of connecting laws, which often will be no more than hinted at); other aspects, which are not accounted for by information about the preceding development, will be descriptively added because they are relevant to an understanding of subsequent stages in the genetic sequence. Thus, schematically speaking, a genetic explanation will begin with a pure description of an initial stage; thence, it will proceed to an account of a second stage, part of which is nomologically linked to, and explained by, the characteristic features of the initial stage, while the balance is simply added descriptively because of its relevance for the explanation of some parts of the third stage, and so forth.[6]

The following diagram schematically represents the way nomological explanation is combined with straightforward description in a genetic account of this kind:

$$S_1 \left.\begin{matrix} \nearrow & S'_2 \\ & \\ & +D_2 \end{matrix}\right\} S_2 \left.\begin{matrix} \nearrow & S'_3 \\ & \\ & +D_3 \end{matrix}\right\} S_3 \nearrow \;.\;\;.\;\;. \left.\begin{matrix} \nearrow & S'_{n-1} \\ & \\ & +D_{n-1} \end{matrix}\right\} S_{n-1} \rightarrow S_n$$

Each arrow indicates a presumptive nomic connection between two successive stages; it presupposes uniformities which as a rule are not stated fully

5. For a description of the device and a probabilistic analysis of its workings, see, for example, Mises (1939), pp. 237-40.

6. This conception of the structure of genetic explanation in history is in basic accord with that set forth by Nagel (1961), pp. 564-68, in the context of a very substantial and comprehensive discussion of problems in the logic of historical inquiry. The presupposition of connecting generalizations in historic-genetaic explanations is emphasized also in Frankel (1959), p. 412 and in Goldstein (1958), pp. 475-79. On the role of "coherent narrative" *vs.* covering-law explanation in natural history, see also Goudge (1958).

and explicitly, and which may be of the strictly universal kind or—more likely—of a statistical kind. $S_1, S_2, \ldots, S_n$ are sets of sentences expressing all the information that the genetic account gives about the first, second, . . , nth stage. For each of these stages except the first and the last, the information thus provided falls into two parts: one—represented by $S'_2, S'_3, \ldots, S'_{n-1}$—describes those facts about the given stage which are explained by reference to the preceding stage; the other—represented by $D_2, D_3, \ldots, D_{n-1}$—constitutes information about further facts which are adduced without explanation, because of their explanatory significance for the next stage. It will hardly be necessary to re-emphasize that this characterization of genetic explanation is highly schematic; it is intended to exhibit the affinities which this procedure has to nomological explanation on one hand and to description on the other. In practice, these two components will often be hard to separate; instead of neatly presenting a set of interconnecting but distinct stages in temporal succession, a genetic account is likely to give descriptions of, and suggest connections between, a great variety of facts and events that are spread over a certain temporal range and are not easily grouped into clusters constituting successive stages.

In our illustration the assumption of some connecting laws or lawlike principles is indicated by the references to motivating factors; for example, the explanatory claims made for the popes' desire to secure a fighting force or to amass even larger funds clearly presupposes psychological assumptions about the manner in which an intelligent individual will tend to act, in the light of his factual beliefs, when he seeks to attain a given objective. Psychological uniformities are implicit also in the reference to the fear of purgatory as explaining the eagerness with which indulgences were bought. Again, when one historian observes that the huge financial success of the first jubilee indulgence "only whetted the insatiable appetite of the popes. The period of time was variously reduced from 100 to 50, to 33, to 25 years,"[7] the explanation thus suggested rests on a psychological assumption akin to the idea of reinforcement by rewards. But, of course, even if some formulation of this idea were explicitly adduced, the resulting account would provide at the very most a partial explanation; it could not show, for example, why the intervening intervals should have had the particular lengths here mentioned.

Those factors which, in our illustration, are simply described or tacitly presupposed as "brute facts," to use Nagel's phrase,[8] include, for example, the relevant doctrines, the organization, and the power of the Church; the occurrence of the crusades and the eventual decline of this movement; and a great

7. Schwiebert (1950), p. 304.
8. Nagel (1961), p. 566.

many additional factors which are not explicitly mentioned, but which have to be understood as background conditions if the genetic account is to serve its explanatory purpose.

Let us consider briefly another example of genetic explanation, taken from Toynbee. In 1839 the principal maternity hospital in the city of Alexandria was located on the grounds of the navy arsenal. "This sounds odd," Toynbee notes, "but we shall see that it was inevitable as soon as we retrace the sequence of events that led to this at first surprising result."[9] Toynbee's genetic account is, briefly, as follows. By 1839 Mehmed 'Ali Pasha, the Ottoman governor of Egypt, had been at work for more than thirty years to equip himself with effective armaments, and particularly with a fleet of warships in the Western style. He realized that his naval establishment would not be self-sufficient unless he was in a position to have his warships built in Egypt by native workers, and that a competent group of Egyptian naval technicians could be trained only by Western naval specialists, who would have to be hired for this purpose. The governor therefore advertised for Western experts, offering them very attractive salaries. But the specialists who applied for the positions were unwilling to come without their families, and they wanted to be sure of medical care that was adequate by Western standards. The governor therefore also hired Western physicians to attend the naval experts and their families. The doctors found, however, that they had time to do additional work; and, "being the energetic and public-spirited medical practitioners that they were, they resolved to do something for the local Egyptian population as well. . . . Maternity work was obviously the first call. So a maternity hospital arose within the precincts of the naval arsenal by a train of events which, as you will now recognize, was inevitable."[10]

Toynbee thus seeks to explain the initially odd fact in question by showing how it came about "inevitably," as the final stage of a sequence of interconnected events; and he refers to the case as an example of the "process of one thing leading to another"[11] in intercultural relations. But wherein lies the inevitability with which one thing leads to the next? At several points in Toynbee's account, the presumptive connection is suggested by the explanatory reference to the motivating reasons of the agents; but these provide explanatory grounds for the resulting actions only on the assumption that people motivated by such and such reasons will *generally* act, or will *tend* to act, in certain characteristic ways. Thus, the conception of one thing inevitably leading to another here presupposes a connection by lawlike principles that hold for certain kinds of

9. Toynbee (1953), p. 75.
10. *Ibid.*, p. 77.
11. *Ibid.*, p. 75.

human action. The character of such principles and the logic of the explanations based on them will be examined more closely in sections 9 and 10 of this essay.

I will now briefly consider some controversial issues concerning genetic explanation in history on which the preceding considerations might shed some light.

Dray has argued that genetic explanation in history has logical peculiarities which can be thrown into relief by a comparison with what he calls "the model of the continuous series."[12] He illustrates the model by an account that explains the stalling of an automobile engine by tracing it back to a leak in the oil reservoir: as a result of the leak, the oil drained out, which deprived the cylinders and pistons of lubrication, thus leading to frictional heating and expansion of the pistons and cylinder walls, so that the metals locked tightly and the engine stopped. Dray puts much emphasis on the claim that by revealing the mechanism of the failure, this stepwise account provides an understanding that would not be conveyed by citing a covering law linking the failure directly to the leak: "*Of course* the engine seized up—and I say this because I can now envisage a *continuous series of happenings* between the leak and the engine seizure which themselves are quite understandable—as the original sequence 'leak-to-seizure' was not."[13]

If I understand it correctly, Dray's defense of this claim rests to a considerable extent on undeniable pragmatic differences between the two accounts: the sequential account affords an insight that is not provided when the final stage is immediately linked to the initial one. But this pragmatic difference is associated, I think, with a non-pragmatic one which justifies the claim that the two accounts differ in explanatory power. To see this, let us, for the sake of the argument, grant nomological status to the statement, L, that whenever the oil reservoir of a properly built car develops a leak, its engine will fail. This law could then be invoked for a low-level explanation of certain particular cases of engine failure. The sequential account, on the other hand, traces the process through a sequence of stages and presents each of these as governed by certain "sub-laws", as Dray calls them, such as those connecting the friction between pistons and cylinder walls with heating and expansion of the metals. But an adequate set of such laws will enable us not only to account for particular cases of engine failure, but also to explain why the law L holds, i.e., why it is that an oil leak in a properly built car *generally* leads to engine failure.

In the case of genetic explanation in history, there is an additional reason for

12. Dray (1957), pp. 66 ff.

13. *Ibid.*, p. 68, italics the author's. For observations in a similar vein and further illustrations, see Danto (1956), pp. 23-25.

regarding an account by stages as essential for the achievement of understanding: here we have no overall law which, in analogy to the law L of the preceding example, links the final stage of the process immediately to the initial one. As our schematic characterization indicates, the particular data about the initial stage do not by themselves suffice to account for all specified aspects of the final stage. To explain the latter, we need further data, and these are provided in installments by the information about additional "brute facts" in the descriptions of the intervening stages.

Our construal of genetic explanation also does justice to the complaint that the laws we might actually be able to adduce in the context of historical explanation, including psychological and other laws of common experience, prove trivial and inadequate when we try to account for the rich and distinctive peculiarities which supposedly make historical events unique, and which are therefore of special interest to the historian. Considering, for example, the subtlety and complexity of some of the psychological explanations that have been proposed for the actions of historical figures, this charge may be somewhat overstated; but undeniably it has a good deal of merit. And the model just outlined makes allowance for the difficulty by providing for the introduction into a genetic account of a more or less extensive mass of details which are simply described, without being explained by reference to other particular facts and connecting uniformities.

8. EXPLANATION-BY-CONCEPT

Another mode of explanation which presumably presents difficulties for the covering-law conception has been pointed out by Dray, who considers its role in historical inquiry. Dray calls it "explaining what" or "explanation-by-concept," on the ground that a request for an account of this kind typically takes the form 'what was it that happened in this case?', and that the historian "deals with it by offering an explanation of the form 'it was a so-and-so'."[1] Dray illustrates the idea by a passage from Ramsey Muir's *Short History of the British Commonwealth*. It describes certain changes that took place in late eighteenth century England—such as the enclosure of agricultural lands, the beginnings of industrial production, and the improvement of communication—and then continues: "It was not merely an economic change that was thus beginning; it was a social revolution." Dray argues that though the historian does not attempt to tell us here why or how the events under investigation came about, his "assertion, 'it was a social revolution', is an explanation never-

1. Dray (1959), p. 403, italics the author's.

theless. It explains what happened *as* a social revolution."[2] Dray characterizes this kind of account as "explanation by means of a general concept rather than a general law. For the explanation is given by finding a satisfactory *classification* of what seems to require explanation."[3] Dray adds that if any generalization is essential to this kind of explanation, then it does not take the form of a general law; for "what is to be explained is a collection of happenings or conditions, x, y and z; and the relevant generalization would be of the form: 'x, y and z amount to a Q'. Such an explanatory generalization is summative; it allows us to refer to x, y and z collectively as 'a so-and-so'. And historians find it intellectually satisfying to be able to represent the events and conditions they study as related in this way."[4]

But surely not every such representation can be regarded as explanatory: the particular occurrences referred to by Muir, for example, might be truthfully but unilluminatingly classified also as changes involving more than 1000 persons and affecting an area of over 100 square miles. If there is explanatory significance to characterizing x, y, and z collectively as a Q, it is because the characterization implies that the particular cases fit into, or conform to, some general pattern that is characteristic of Q.

I will illustrate this first by some examples which show, at the same time, that the procedure in question is also used outside the domain of historiography.

Torricelli's explanation of why a simple suction pump can raise water by no more than 34 feet has been said to rest on the "conceptual scheme" of a "sea of air" surrounding the earth.[5] But clearly that scheme has explanatory force only because it assumes a nomic analogy between the sea of air and a sea of water, namely, that "there would be an air pressure on all objects submerged in this sea of air exactly as there is water pressure below the surface of the ocean,"[6] and that the pressure is determined by the weight of the column of air above the object in question: this is indeed how Torricelli reasoned. Thus the explanation by means of his conceptual scheme effects a subsumption of the explanandum phenomenon under general hypotheses.

Next, as an example that shows a clear similarity to that cited by Dray, consider the statement: 'Otto's running nose and inflamed eyes, and the red spots surrounded by white areas that have just appeared on the mucous linings of his cheeks are not just isolated occurrences: they are, all of them, symptoms marking the onset of a full-blown case of the measles'. This diagnostic classifi-

2. *Ibid.*, italics the author's.
3. Dray (1959), p. 404, italics the author's.
4. *Ibid.*, p. 406.
5. Conant (1951), p. 69.
6. *Ibid.*

cation accounts for the particular complaints cited by pointing out that they jointly conform to the clinical pattern of the measles; i.e., that they are of certain characteristic kinds and occur in a characteristic temporal order, that they will be followed by further specific symptoms, and that the illness will tend to take a certain characteristic course. To interpret a set of complaints as manifestations of the measles is surely to claim that they fit into a certain pattern of regularities (which will be of statistical rather than of strictly universal form); and such an account accords with the covering-law conception of explanation.

Or consider the "classification" of a particular sequence of lightning and thunder as a case of a powerful electric discharge generating a violent disturbance of the air. This does indeed have explanatory import, but clearly by virtue of pointing out that the particular set of events showed the characteristics generally exhibited by powerful discharges and by the disturbances they create in the air; or, more precisely, that they conform to the laws characteristic of the sort of phenomenon as an instance of which the particular case is classified or interpreted.

In Dray's quotation from Muir, the pronouncement "it was a social revolution" similarly carries the suggestion that an explanatory diagnosis is being offered—a suggestion that is reinforced by the following amplificatory passage, which directly follows the sentence quoted by Dray: "The old, settled, stable order which we described as existing in Britain in the middle of the eighteenth century was being wholly transformed. . . . But the full significance of this change was as yet quite unrealized. Securely enthroned, the old governing classes were wholly blind to the forces that were at work beneath their feet, undermining the very foundations of their power, and making it inevitable that sooner or later the political system should be readjusted to accord with the change in the social order."[7] We have here the suggestion of a diagnosis or interpretation to the effect that the particular changes in agriculture, industrial production, and communications that Muir had described before were early manifestations of a larger process whose different phases are not associated coincidentally, but with some inevitability. Thus again—if only very vaguely and sketchily—the particular cases are assigned a place in a comprehensive pattern of connections. Whatever explanatory significance Muir's statement may have—and to me, it seems rather slight—surely lies in the suggestion of a diagnosis of the sort that is more plainly illustrated by our preceding two illustrations, which conform, in broad outline, to the covering-law conception.

7. Muir (1922), p. 123.

Other examples of what Dray calls explanation-by-concept are provided by the various interpretations of the American Civil War as the result of a conspiracy by some Northern—or Southern—groups of "wicked men"; as a quarrel between two rival regions; as a contest over types of government; as an outgrowth of the "irrepressible conflict" between freedom and slavery; as a basically economic contest; and so forth.[8] Each of these explanations of the Civil War "as a so-and-so" attributes special or overriding causal significance to factors of some special type and accordingly presupposes suitable nomic connections in support of those assumptions.[9]

Dray explicitly acknowledges that "explanation-by-concept may sometimes *in fact* subsume the explicandum under law,"[10] but holds that this is not generally the case. Specifically, he takes issue with an earlier statement of mine that "what is sometimes, misleadingly, called an explanation by means of a certain *concept* is, in empirical science, actually an explanation in terms of *universal hypotheses* containing that concept."[11] Against this view, Dray argues as follows: "Presumably the law which lurks in the background when something is explained 'as a revolution' is one which would contain the concept in its apodosis. . . . But to explain, say, what happened in France in 1789 'as a revolution' would surely not be equivalent to bringing it under any law of the form, 'Whenever $C_1, C_2, \ldots, C_n$ then a revolution'."[12] But my earlier remark does not limit an explanation-by-concept to one general hypothesis, nor does it limit the explanatory hypotheses to the type envisaged by Dray. It applies as well, for example, to the explanation of certain complaints "as symptoms of the measles," which rests on general hypotheses to the effect that if a person suffers from the measles, then he will exhibit symptoms of such a kind; here, the explanatory concept is referred to in the protasis rather than in the apodosis.

Or consider what might loosely be called "explaining the glow of a falling meteorite as a case of intense heat generated by friction." Here several laws are involved, among them two to the effect that a body moving through air

8. On these different interpretations see, for example, Beale (1946).

9. The problem of weighting causal factors according to their relative importance in a historical explanation is lucidly dealt with in Nagel (1961), pp. 582-88.

10. Dray (1959), p. 405, italics the author's.

11. Hempel (1942), footnote 3, italics in the original. Homans has recently stressed the same point in reference to sociology. He holds that much of modern sociological theory fails to explain anything, partly because "much of it consists of systems of categories, or pigeonholes, into which the theorist fits different aspects of social behavior. . . . but this in itself is not enough to give it explanatory power. . . . The science also needs a set of general propositions about the relations between the categories, for without such propositions explanation is impossible." Homans (1961), p. 10.

12. Dray (1959), p. 404.

encounters friction and that friction generates heat; so that the explanatory concepts might be said to figure partly in the protasis, partly in the apodosis of the corresponding general laws.

Dray's own example is stated so sketchily that it is difficult to appraise the explanation supposedly achieved. A statement characterizing what happened in France in 1789 as a revolution would seem to provide a very vague description rather than any explanation of those events. Some explanatory import might be claimed if the concept of revolution were understood in a restricted technical sense implying perhaps a sequence of characteristic stages in the process, or certain characteristic changes in the structure of political power, or the like; then some of the particular events of 1789 might be shown to conform to the patterns implied by the given concept of revolution and might thus be regarded as partly explained by it. But in this case, the explanation would evidently be achieved by reference to the implied uniformities.

In sum, then, an explanatory use of concepts must always rely on corresponding general hypotheses.

9. DISPOSITIONAL EXPLANATION

Another kind of explanation that has been held to defy a covering-law analysis invokes in a characteristic manner certain dispositional properties of the objects or agents whose "behavior" is to be accounted for; I will refer to this procedure as dispositional explanation.

The familiar method of explaining human decisions and actions in terms of purposes, beliefs, character traits, and the like is basically of this kind; for to ascribe to an agent such motivating factors is to assign to him certain more or less complex dispositional characteristics: this has been argued in detail by Ryle[1], whose ideas have had great influence on the discussion of the subject. Explanations by motivating reasons will be examined in some detail in section 10. In the present section we will consider the logical structure of some dispositional explanations in physics and compare it with that of explanations by covering laws.

Consider first an example discussed by Ryle. When a window pane shatters upon being struck by a stone, the breaking of the glass can be causally explained, according to Ryle, by pointing out that a stone hit it; but we often seek an explanation in a different sense: "We ask why the glass shivered when struck by the stone and we get the answer that it was because the glass was brittle."[2]

1. See especially Ryle (1949).
2. Ryle (1949), p. 88.

Here the explanation is achieved, not by specifying an independent event "which stood to the fracture of the glass as cause to effect",[3] but by attributing to the glass a certain dispositional property, brittleness. To ascribe this property to a particular window pane is, at least by implication, to assert a *general* hypothesis, roughly to the effect that if at any time the pane is sharply struck by any physical body, or is sharply twisted by any agent, it will fly into fragments. But while thus being general in character, a dispositional statement nevertheless also mentions a particular individual, such as the window pane. In this respect, dispositional statements differ from general laws, which Ryle construes as making no mention of individuals at all. To indicate their resemblance to general laws and also their difference from them, Ryle calls dispositional statements "law-like."[4]

It should be noted, however, that neither of the two kinds of explanation here distinguished by Ryle is sufficient by itself to account for the given event. The report that the pane was struck by a stone explains its being broken only in conjunction with the additional information that the pane was brittle: it is in virtue of the general hypothesis implied by this dispositional attribution that being hit by the stone becomes a cause rather than an accidental antecedent in regard to the breaking of the pane. Similarly, the dispositional statement can explain the breaking of the glass only when taken in conjunction with the report that the glass was sharply struck; and indeed, as we saw, Ryle himself describes the dispositional statement as explaining "why the glass shivered when struck by the stone," and not simply why the glass shivered. Thus either of the two explanations here distinguished is incomplete and requires complementa-

3. *Ibid.*

4. For details, see Ryle (1949), pp. 43-44, 89, 120-25. Strictly speaking, the intended distinction between law-like sentences and general laws cannot be satisfactorily explicated in terms of whether or not the sentences in question "mention particular things or persons," as Ryle (*loc. cit.*, p. 123) puts it; for explicit mention of an individual can be circumvented by rephrasing. For example, the general sentence 'All places on the surface of the earth within 100 miles of the North Pole are cold' would count as law-like because it mentions the North Pole. Yet it can be rephrased as 'All polar places are cold', where 'polar' is used as synonymous with 'lying on the surface of the earth within 100 miles of the North Pole'; and under the contemplated criterion, the rephrasal would have to be counted as a general law because it does not mention (i.e., it does not contain a designation of) any partciular person, place, or thing. For a fuller discussion of the issue *cf.* Hempel and Oppenheim (1948), section 6 and Goodman (1955), especially chapters 1 and 3. Note, incidentally, that Goodman uses the term 'lawlike' in a sense quite different from Ryle's, namely, to refer to sentences having all the characteristics of a law, except for possibly being false (*loc. cit.*, p. 27). To avoid a lengthy digression, we will here forego an attempt to offer a more adequate explication of the important distinction made by Ryle, and will consider the idea as intuitively sufficiently clear for our present purposes.

tion by the other. Jointly, they provide an adequate account, which might be schematically formulated as follows:

(9.1)

(C_1) The pane was sharply struck by a stone at time t_1

(L_1) For any time t it is the case that if the pane is sharply struck at t, then it breaks at t

---

(E_1) The pane broke at t_1

This account is a deductive-nomological explanation except for invoking a law-like statement instead of a completely general law. In this latter respect, the argument is in good company: Galileo's and Kepler's laws, for example, surely are used for explanatory purposes; and yet the former, when fully stated, specifies that its formula applies to free fall near the surface of the earth, and it thus mentions an individual object; while Kepler's laws, as originally conceived, refer to the motions of the planets of one particular object, the Sun. To be sure, these laws have since been subsumed under the Newtonian laws of motion and of gravitation, which are of completely general form. A similar step is possible in the example of the broken window, where the statement 'the pane was brittle' may be replaced in the explanatory argument by a completely general hypothesis, 'All glass is brittle (under standard conditions)', and the singular statement 'The pane was made of glass (and was under standard conditions)'.

However, currently available theories do not enable us to perform this sort of subsumption under strictly general laws or theoretical principles for all law-like statements, and especially for all statements ascribing psychological dispositions to individuals. But one other step can always be taken even in these cases: instead of putting the explanatory dispositional statement into the form of a generalization mentioning a particular individual in the manner of L_1 in (9.1), we can express it by two separate statements: a singular one, asserting that the given individual has the dispositional property in question, say, D; and a completely general one characterizing the disposition D. In the case of (9.1), this would amount to replacing the sentence L_1 by the following two:

(C_2) The pane was brittle at time t_1.

(L_2) Any brittle object, if sharply struck at any time, breaks at that time.

It might be objected that the only general statement which occurs in the resulting modification of (9.1), namely L_2, does not have the character of an *empirical law* about brittle objects, but rather that of a *definition* of brittleness; and that accordingly, the explanatory force of the argument continues to reside in the attribution of brittleness to a particular pane, and thus in the law-like statement L_1 rather than in a general law about all brittle objects.

This objection carries some weight when a dispositional characteristic represents just one kind of law-like behavior, such perhaps as breaking under specified impact. But a dispositional characteristic, say M, of the kind invoked for explanatory purposes can usually manifest itself in a variety of symptomatic ways, depending on the circumstances.[5] For example, magnetization of an iron bar can manifest itself by the fact that iron filings will cling to its ends; but also by the fact that one of its ends will attract the north pole, the other one the south pole of a compass needle; and no less by the fact that if the bar is broken in two, each of the parts will display the two kinds of disposition just described for the whole bar. Many of the "symptom statements" thus characterizing some peculiar way in which M may manifest itself might be regarded as expressing either a necessary or a sufficient condition for the presence of M, and M itself might be referred to as a broadly dispositional characteristic. To such characteristics the objection at hand does not apply, as I will now try to show.

Symptom sentences expressing necessary conditions for M might take the following form:

(9.2a) If an object or individual x has the property M, then under test conditions, or stimulus conditions, of kind S_1, x will regularly respond in manner R_1; under conditions S_2, in manner R_2; and so on.

Symptom sentences expressing sufficient conditions for M might correspondingly take the form:

(9.2b) If x is in conditions of kind S^1, then if x responds in manner R^1, x has the property M; if x is in S^2, then if x responds in manner R^2, x has the property M; and so on.[6]

Each symptom sentence of either type may be regarded as expressing a partial criterion of application for the term 'M'.

The construal of symptom statements as expressing strictly necessary or strictly sufficient conditions for M is an oversimplification in many cases. For example, in medical symptom statements and in the formulation of partial criteria for character traits, beliefs, desires, etc., the connection between

5. That the attribution of a disposition usually implies many hypothetical propositions has been stressed by Ryle (1949), pp. 43-44. Earlier, a much fuller formal study of the logic of such broadly dispositional concepts had been carried out by Carnap in his essay "Testability and Meaning" (1936-37), esp. Part 2, which specifically provided for the possibility of introducing a scientific term by means of a set of reduction sentences, each of which is a symptom sentence in our sense. For a more recent discussion, which sheds further light on the issues here considered, see also Carnap (1956).

6. The two types of symptom sentences, or partial criteria of application, here considered correspond to the two basic types of "reduction sentences" in Carnap's study (1936-37); see especially section 8, "Reduction Sentences."

M and its symptomatic manifestations will often have to be conceived as probabilistic in character. In this case, the symptom sentences might take the following statistical forms, which are counterparts to (9.2a) and (9.2b) above:

(9.3a) For objects or individuals that have the property M and are under test conditions of kind S_1 ($S_2, \ldots$), the statistical probability of responding in manner R_1 ($R_2, \ldots$) is r_1 ($r_2, \ldots$).

(9.3b) For objects or individuals that are under test conditions of kind S^1 ($S^2, \ldots$) and respond in manner R^1 ($R^2, \ldots$), the statistical probability of possessing the property M is r' ($r'', \ldots$).

For the sake of full concentration on the basic issues presently under discussion, however, we will limit our attention, for the time being, to broadly dispositional traits M characterized by non-probabilistic symptom sentences of the forms (9.2a) and (9.2b).

Let U be the set of all symptom sentences for M. This set evidently implies a sentence, expressible in terms of 'R_1','S_1','R_2','S_2', . . . , 'R^1','S^1','R^2','S^2', . . . to the effect that any x satisfying some one of the sufficient conditions for M as specified in U also satisfies any one of the necessary conditions for M as specified in U.[7] As will be shown presently, this statement normally has the character of a general empirical law: and if the symptom statements for M thus jointly have empirical implications, they clearly cannot all be claimed to hold true simply by definitional fiat.[8]

To illustrate by reference to an earlier example: one of the necessary conditions for an iron bar being magnetic might be:

(9.4a) If an iron bar x is magnetic then if iron filings are placed close to x (condition S_1), the filings will cling to its ends (response R_1).

And one of the sufficient conditions might be:

(9.4b) If an iron bar x is in the vicinity of a compass needle (condition S^1) then if one of its ends attracts the north pole of the needle and repels the south pole, whereas the other end shows the opposite behavior (response R^1), then x is magnetic (has property M).

But jointly, these two symptom sentences imply the general statement that any iron bar which satisfies the compass needle condition also satisfies the iron filings condition: and this surely is not a definitional truth, but a statement that has the character of an empirical law.

Thus, as a rule, the set U of symptom statements for a broadly dispositional

7. This statement is equivalent to what Carnap calls the "representative sentence" of the set U of reduction sentences for M; for it "represents, so to speak, the factual content" of U. See Carnap (1936-37), pp. 451.

8. This point is lucidly argued and illustrated, by reference to the broadly dispositional concept of a person *wanting* a certain state of affairs, in Brandt and Kim (1963), pp. 428-29.

term has empirical consequences. But then it would be quite arbitrary to construe some of those symptom statements as analytic-definitional and to assign to others the status of empirical laws.[9] For this would amount to decreeing that the former were not liable to modification if empirical evidence should be found to conflict with the laws implied by the set U; but in empirical science no statements other than logical and mathematical truths can be regarded as enjoying such unqualified immunity. Accordingly, the total set of symptom statements is more appropriately regarded as part of the system of general laws governing the concept in question.

Suppose, now, that in order to explain why a given particular object or individual i behaved in a certain manner, say R_3, it is pointed out that i was in a situation of kind S_3, and that i has a broadly dispositional property M whose presence is characterized by the disposition to respond to S_1 in manner R_1, to S_2 in manner R_2, to S_3 in manner R_3, and so on. This explanatory argument may then be schematized as follows:

(C_1) i was in a situation of kind S_3
(C_2) i has the property M
(9.5) (L) Any x with the property M will, in a situation of kind S_3, behave in manner R_3

(E) i behaved in manner R_3

This account is clearly of deductive-nomological form; for the general statement L, as we have just noted, has to be accorded the status of an empirical law rather than that of a "mere definition."

But the preceding account of "dispositional explanation" calls for some further qualification. What has been said so far might suggest, for example, that to ascribe to an iron bar the "broadly dispositional property" of being magnetic is tantamount to attributing to it a set of simple dispositions, each of them characterized, in the sense reflected by our symptom statements, by the association of some specific kind of manifest "response" with certain manifest "stimulus conditions." This would be too simple a conception, however. For the general physical statements pertaining to the property of being magnetic include, besides such symptom statements, also certain general laws which represent no dispositional tendencies, and which are no less characteristic of the concept of being magnetic than are the pertinent symptom statements. Among them is the law that a moving magnetic field will produce an electric field, which implies that in a closed wire loop near a moving magnet an electric current will be induced, which in turn implies a general statement concerning

9. On this point, see also pp. 113–115 in this volume.

the response made by an ammeter which is put into a closed wire loop near a moving magnet. This last statement may be regarded as a further symptom statement for the property of being magnetic, but it should be noted that the symptom here specified is associated with the property of being magnetic by virtue of theoretical principles connecting the given characteristic with other theoretical concepts, such as that of electric and magnetic fields and their interrelations. Thus, when a concept like that of a magnet functions in a theory, then, in applying it to some particular object, we are not simply attributing to this object a set, however extensive, of dispositions to display certain kinds of observable response under given, observable stimulus conditions: the assignment also has various theoretical implications, including the attribution of other "broadly dispositional" characteristics.

These observations concerning the theoretical aspects of broadly dispositional concepts also will be found relevant to an analysis of the explanatory role of motivating reasons, which forms the subject of the next section.

10. THE CONCEPT OF RATIONALITY AND THE LOGIC OF EXPLANATION BY REASONS

10.1 Two Aspects of the Concept of Rationality. In the present section, I propose to examine the logic of the familiar method of accounting for human decisions and actions in terms of motivating reasons—a method widely held to be entirely different from the explanatory procedures of the natural sciences and to defy analysis by means of the covering-law models.

In an explanation by motivating reasons the idea of rationality usually plays an important role; and I will therefore begin with some remarks on this concept. To qualify a given action as rational is to put forward an *empirical hypothesis* and a *critical appraisal*. The hypothesis is to the effect that the action was done for certain reasons, that it can be *explained* as having been motivated by them. The reasons will include the ends that the agent presumably sought to attain, and the beliefs he presumably entertained concerning the availability, propriety, and probable effectiveness of alternative means of attaining those ends. The critical appraisal implied by the attribution of rationality is to the effect that, judged in the light of the agent's beliefs, the action he decided upon constituted a *reasonable* or *appropriate* choice of means for achieving his end. These two aspects of the concept of rational action will now be examined in turn.

10.2 Rationality as a Normative–Critical Concept. The clarification of the critical, or normative, idea of rational action calls for the statement of clear criteria of rationality which might provide us with standards for appraising the rationality of particular actions, and which might thus also guide us in making rational decisions.

Rationality in this sense is obviously a relative concept. Whether a given action—or the decision to perform it—is rational will depend on the objectives the action is meant to achieve and on the relevant empirical information available at the time of the decision. Broadly speaking, an action will qualify if, on the given information, it offers optimal prospects of achieving its objectives. Let us now consider more closely the key concepts invoked in this characterization: the concepts of the information basis and of the objectives of an action, and finally that of rationality relative to a given basis and given objectives.

If we are to choose a rational course of action in pursuit of given ends, we will have to take into account all available information concerning such matters as the particular circumstances in which the action is to be taken; the different means by which, in these circumstances, the given ends might be attained; and the side-effects and aftereffects that may be expected from the use of different available means.

The total empirical information available for a given decision may be thought of as represented by a set of sentences, which I will call the *information-basis* of the decision or of the corresponding action. This construal of the empirical basis for a decision takes account of an obvious but important point: to judge the rationality of a decision, we have to consider, not what empirical facts—particular facts as well as general laws—are actually relevant to the success or failure of the action decided upon, but what information concerning such facts is available to the decision-maker. Indeed, a decision may clearly qualify as rational even though it is based on incomplete or false empirical assumptions. For example, the historian, precisely in order to present an action by a historical figure as rational, will often have to assume—and may well be able to show on independent grounds—that the agent was incompletely informed, or that he entertained false beliefs concerning relevant empirical matters.

But while the information basis of a rational action thus need not be true, should there not at least be good reasons for believing it true? Should not the basis satisfy a requirement of adequate evidential support? Some writers do consider this a necessary condition of rational action, and this view is indeed quite plausible. For example, as one of its recent advocates, Quentin Gibson, points out: "If someone were, carefully and deliberately, to walk round a ladder because he believed, without evidence, that walking under it would bring him bad luck, we would not hesitate to say that he acted irrationally."[1]

1. Gibson (1960), p. 43. Chapters 4 and 14 of Gibson's work include many illuminating observations on the questions examined in this section.

No doubt we often understand rationality in this restricted sense. But if we wish to construct a concept of rational action that might later prove useful in explaining certain types of human behavior, then it seems preferable not to impose on it a requirement of evidential support; for in order to explain an action in terms of the agent's reasons, we need to know what the agent believed, but not necessarily on what grounds. For example, an explanation of the behavior of Gibson's ladder-shunner in terms of motivating reasons would have to invoke the man's superstitious beliefs, but not necessarily the grounds on which he holds them; and the man may well be said to be acting quite reasonably, given his beliefs.

From the information basis of a decision I now turn to its objectives. In very simple cases, an action might be construed as intended to bring about a particular state of affairs, which I will call the end state. But even in such simple cases, some of the courses of action which, according to the information basis, are available and are likely to bring about the end state, may nevertheless be ruled out because they violate certain general constraining principles, such as moral or legal norms, contractual commitments, social conventions, the rules of the game being played, or the like. Accordingly, the contemplated action will be aimed at achieving the end state without such violation. What I will call its *total objective* may then be characterized by a set E of sentences describing the intended end state, in conjunction with a set N of constraining norms.

Again, as in the case of the empirical basis, I will not impose the requirement that there must be "good reasons" for adopting the given ends and norms: rationality of an action will be understood in a strictly relative sense, as its suitability, judged by the given information, for achieving the specified objective.

How can such suitability be defined? For decision situations of the simple kind just contemplated, a characterization can readily be given: if the information basis contains general laws by virtue of which certain of the available actions would be bound to achieve the total objective, then, clearly, any one of those actions will count as rational in the given context. If the information basis does not single out any available action as a sufficient means for attaining the objective, it may yet assign a numerical probability of success to each of the different available actions; in this case, any action will count as rational whose probability of success is not exceeded by that of any available alternative.

For many problems of rational decision, however, the available information, the objectives, and the criteria of rationality cannot be construed in this simple manner. Our construal becomes inapplicable, in particular, when the objective of a proposed action does not consist in attaining a specified end state. This is quite frequently the case, as we will now see.

To begin with, even when a particular end state is aimed at, the available information will often indicate that there are several alternative ways of definitely or probably attaining it, each attended by a different set of side-effects and aftereffects which are not part of it. Some of these anticipated incidental consequences will be regarded as more or less desirable, others as undesirable. In a theoretical model of such decision situations the total goal must accordingly be indicated, not simply by describing the desired end state, but by specifying the relative desirability of the different total outcomes that may result from the available courses of action.

In the mathematical theory of decision-making, various models of rational choice have been constructed in which those desirabilities are assumed to be specified in numerical terms, as the so-called utilities of the different total outcomes.

The case in which the given information basis also specifies the 'probabilities[2] of the different outcomes is called *decision under risk*. For this case, one criterion of rationality has gained wide acceptance, namely that of *maximizing expected utility*. The expected utility which, on the given information, is associated with a contemplated course of action is determined by multiplying, for each possible outcome of the action, its probability with its utility, and adding the products. An action, or the decision to perform it, then qualifies as rational if its expected utility is maximal in the sense of not being exceeded by the expected utility of any alternative action.

Another decision problem which has been the subject of mathematical study, and which is of considerable philosophic interest, is that of *decision under uncertainty*. Here it is assumed that the given information basis indicates the different available courses of action and specifies for each a set of mutually exclusive and jointly exhaustive possible outcomes, without, however, assigning probabilities to them;[3] finally, each of the possible outcomes is assumed to have

2. The probabilities and utilities here referred to are subject to certain mathematical requirements which cannot be discussed in the context of the present paper. The classical statement is given in von Neumann and Morgenstern (1947); lucid presentations of the requirements, and of the reasons underlying them, will be found in Luce and Raiffa (1957), chaps. 1-4 and in Baumol (1961), chaps. 17 and 18. Among the questions passed over here is the very important one of how the concept of the probability of outcomes should be understood in the context of decision theory. For a large class of problems the familiar statistical construal of probability as a long-run relative frequency will be practically sufficient, and the current mathematical theory of games and decisions does rely on it to a large extent. Alternative conceptions have been proposed, however, Among them are Carnap's concept of inductive or logical probability (*cf.* Carnap (1950), (1962)) and the concept of personal probability (*cf.* Savage (1954), especially chaps. 3 and 4).

3. Strictly speaking, this situation cannot arise on a theory of inductive logic, such as Carnap's, according to which the given empirical information, whatever it may be, always assigns a definite logical probability to each of the statements describing one of the possible outcomes.

been assigned a utility. By way of illustration, suppose that you are offered as a present the metal ball that you will obtain by one single drawing made, at your option, from one of two urns. You are given the information that the metal balls are of the same size; that the first urn contains platinum balls and lead balls in an unspecified proportion; and the second urn, gold and silver balls in an unspecified proportion. Suppose that the utilities you assign to platinum, gold, silver, and lead are in the ratio of 1000:100:10:1. From which urn is it rational to draw? Several quite different criteria of rational choice under uncertainty have been set forth in recent decision theory. Perhaps the best-known of them is the *maximin rule*; it directs us to maximize the minimum utility, i.e., to choose an action whose worst possible outcome is at least as good as the worst possible outcome of any alternative. In our example, this calls for a drawing from the second urn; for at worst, it will give you a silver ball, whereas the worst outcome of a drawing from the first urn would give you a lead ball. This rule clearly represents a policy of extreme caution, reflecting the pessimistic maxim: act on the assumption that the worst possible outcome will result from your action.

An alternative policy, expressed by the so-called *maximax rule*, reflects the optimistic expectation that our action will lead to the best possible outcome; it directs us to choose a course of action whose best possible outcome is at least as good as the best possible outcome of any alternative action open to us. In our example, the proper decision under this rule would be to draw from the first urn; for at best this will give us a platinum ball, whereas a drawing from the second urn can at best yield a gold ball.

Various interesting alternative rules have been proposed for the case of decision under uncertainty, but for our purposes it is not necessary to consider them here.[4]

The mathematical models here briefly characterized do not offer us much help for a rational solution of the grave and complex decision problems that confront us in our daily affairs. For in these cases, we are usually far from having the data required by our models: we often have no clear idea of the available courses of action, nor can we specify the possible outcomes, let alone their probabilities and utilities. In contexts, however, where such information is available, mathematical decision theory has been applied quite successfully even to rather complicated problems, for example, in industrial quality control and some phases of strategic planning.

But whatever their practical promise, these models contribute, I think, to the analytic clarification of the concept of rational action. In particular, they

4. Accounts of those rules can be found, for example, in Luce and Raiffa (1957), chap. 13 and in Baumol, chap. 19.

throw into relief the complex, multiply relative, character of this concept; and they show that some of the characterizations of rational action which have been put forward in the philosophical literature are of a deceptive neatness and simplicity. For example, Gibson, in his careful and illuminating study, remarks: "there may be various alternative ways of achieving an end. To act rationally. . . is to select what on the evidence is *the best* way of achieving it";[5] and he refers to "an elementary logical point—namely, that, given certain evidence, there can only be one correct solution to the problem as to the best way of achieving a given end."[6] Gibson offers no criterion for what constitutes the best solution; but surely, what he asserts here is not an elementary logical point, and indeed it is not true. For, first, even when the decision situation is of a kind for which one definite criterion of rational choice may be assumed to be available and agreed upon—for example, the principle of maximizing expected utilities—then that criterion may qualify several different courses of action as equally rational. Secondly, and more importantly, there are various kinds of decision, such as decision under uncertainty, for which there is not even agreement on a criterion of rationality, where maximin opposes maximax and both are opposed by various alternative rules.

It is important to bear in mind that the different competing criteria of rationality do not reflect differences in the evaluation of the various ends which, on the given information, are attainable: all the competing rules here referred to presuppose that the utilities of those ends have been antecedently fixed. Rather, the different decision rules or criteria of rationality reflect different inductive attitudes, and in some cases, as we saw, different degrees of optimism or pessimism as to what to expect of the world, and correspondingly different degrees of boldness or caution in the choice of a course of action.

The diversity of conflicting rules proposed for decision under uncertainty suggests the question whether it might not be possible to specify some unique sense of rationality which is independent of such differences of outlook, and which can be shown to be more adequate than the conceptions of rationality reflected by the competing criteria we have mentioned. The prospects of specifying such a sense are dim indeed, and this again is indicated by some results of mathematical decision theory. Specifically, it is possible to formulate a set of general desiderata, or conditions of adequacy, for any proposed decision rule, and to show that though each of the desiderata appears perfectly reasonable and, so to speak, "essential" to rational choice, nevertheless (i) every decision rule that has been proposed in the literature violates one or more of the desiderata, and, indeed (ii) despite their intuitive plausibility, the desiderata are

5. Gibson (1960), p. 160, italics the author's.

6. Gibson (1960), p. 162.

logically incompatible.[7] This result certainly must serve as a warning against the assumption that the idea of rationality, or of the best way to act in a given situation, is reasonably clear, and that the formulation of criteria which make the notion explicit is a basically trivial, though perhaps tedious, explicatory task.

The considerations here outlined concerning the critical or normative notion of rationality have important implications for the explanatory use of the idea of rational action, as we will now see.

10.3 RATIONALITY AS AN EXPLANATORY CONCEPT. Human actions are often explained in terms of motivating reasons. The preceding considerations suggest that a full statement of those reasons will have to indicate the agent's objectives as well as his beliefs about the means available to him and their probable consequences. And the explanation will aim at showing that the action was to be expected in view of those objectives and beliefs. Such explanatory accounts rest therefore, as Peters has put it, on the "concealed assumption" that "*men are rational* in that they will take means which lead to ends if they have the information and want the ends."[8] Here, then, the concept of rationality is used in an explanatory hypothesis. Let us now examine the logic of such explanations.

10.3.1 *Dray's Concept of Rational Explanation.* As our point of departure let us choose Dray's stimulating and suggestive study of such explanations and particularly of their role in historical inquiry[9]—a study which led him to conclude that "the explanation of individual human behavior as it is usually given in history has features which make the covering law model peculiarly inept."[10] Dray refers to the kind of explanation here referred to, namely, explanation by motivating reasons, as *rational explanation* because, as he says, it "displays the *rationale* of what was done" by offering "a reconstruction of the agent's *calculation* of means to be adopted toward his chosen end in the light of the circumstances in which he found himself. To explain the action we need to know what considerations convinced him that he should act as he did."[11] But Dray attributes to rational explanation a further characteristic, which clearly assigns an essential role to the evaluative or critical concept of rationality. According to him, the "goal of such explanation is to show that what was done was the thing to have done for the reasons given, rather than merely the thing

7. For details see Luce and Raiffa (1957), chap. 13, especially sections 3 and 4.

8. Peters (1958), p. 4, italics supplied. For another statement concerning the explanatory and predictive use of the assumption of rationality, *cf.* Gibson (1960), p. 164.

9. See especially Dray (1957), chap. 5 and Dray (1963).

10. Dray (1957), p. 118.

11. Dray (1957), pp. 124 and 122, italics the author's.

that is done on such occasions, perhaps in accordance with certain laws."[12] Hence, "Reported reasons, if they are to be explanatory in the rational way, must be *good* reasons at least in the sense that *if* the situation had been as the agent envisaged it. . . then what was done would have been the thing to have done."[13] To show that the agent had good reasons for his action, a rational explanation must therefore invoke, not a general empirical law, but a "*principle of action,*" which expresses "a judgment of the form: 'When in a situation of type $C_1 \ldots C_n$ the thing to do is x'."[14] Thus, such explanations contain "an element of *appraisal* of what was done."[15] And it is precisely in this reliance on a principle of action expressing a standard of appropriateness or rationality that Dray sees the essential difference between rational explanations and those accounts which explain a phenomenon by subsuming it under covering general laws that describe certain uniformities but do not appraise.

Dray does not further specify the character of the "situations" referred to in his principles of action; but in order to do justice to his intent, those situations must surely be taken to include such items as (i) the end the agent sought to attain, (ii) the agent's beliefs concerning the empirical circumstances in which he had to act and concerning the means available to him for the attainment of his objective, (iii) moral, religious, or other norms to which the agent was committed. For only when these items are specified does it make sense to raise the question of the appropriateness of what the agent did in the given situation.

It seems fair, then, to say that according to Dray's conception, a rational explanation answers a question of the form 'why did agent A do X?' by offering an explanans of the following type (instead of Dray's '$C_1 \ldots C_n$', we write 'C' for short, bearing in mind that the situation thus referred to may be very complex):

A was in a situation of type C

In a situation of type C the appropriate thing to do is X

But this construal of rational explanation presupposes a criterion of rationality which, for the given kind of situation, singles out one particular course of action as *the* thing to do: and as we saw earlier this presupposition is highly questionable.

More importantly however, even if such a criterion were available, an account of the form here considered cannot possibly explain why A did X. For according to the requirement of adequacy set forth in section 2.4 of this essay, any adequate answer to the question why a given event occurred will

12. Dray (1957), p. 124.
13. Dray (1957), p. 126, italics the author's.
14. Dray (1957), p. 132, italics the author's.
15. Dray (1957), p. 124, italics the author's.

have to provide information which, if accepted as true, would afford good grounds for believing that the event did occur. Now, the information that agent A was in a situation of kind C and that in such a situation the rational thing to do is x, affords grounds for believing that *it would have been rational for A to do x*, but no grounds for believing that A did in fact do x.[16] To justify this latter belief, we clearly need a further explanatory assumption, namely that—at least at the time in question—A was a *rational agent* and thus was *disposed* to do whatever was rational under the circumstances.

But when this assumption is added, the answer to the question 'Why did A do x?' takes on the following form:

A was in a situation of type C
A was a rational agent
(Schema R) In a situation of type C, any rational agent will do x

Therefore, A did x

This schema of rational explanation differs in two respects from what I take to be Dray's construal: first, the assumption that A was a rational agent is explicitly added; and second, the evaluative or appraising principle of action, which specifies the thing to do in situation C, is replaced by an empirical generalization stating how rational agents will act in situations of that kind. Thus, Dray's construal fails just at the point where it purports to exhibit a logical difference between explanations by reference to underlying reasons and explanations by subsumption under general laws, for in order to ensure the explanatory efficacy of a rational explanation, we found it necessary to replace Dray's normative principle of action by a statement that has the character of a general law. But this restores the covering-law form to the explanatory account.

That the appraising function which Dray considers essential for rational explanation has no explanatory import is shown also by this consideration: Doubts concerning a given explanation in terms of a specified rationale could not significantly be expressed in the form 'Was X actually the thing to do under the circumstances?', but they might well take the form 'Was A actually inclined to regard X as the thing to do?'. Accordingly, it would be irrelevant to argue, in defense of a proposed explanation, that X was indeed (by some theoretical standard of rationality) "the thing to do," whereas it would be distinctly relevant to show that A was generally disposed to do X under circumstances of the specified kind. And the explanatory import of this

16. The same objection has been raised, in effect, by Passmore, in the following comment on Dray's conception: ". . . explanation by reference to a 'principle of action' or a 'good reason' is not, by itself, explanation at all. . . . For a reason may be a 'good reason'—in the sense of being a principle to which one *could* appeal in justification of one's action—without having in fact the slightest influence on us." Passmore (1958), p. 275, italics the author's.

latter information would be completely independent of whether the contemplated action did or did not conform to the explainer's—or the questioner's—standards of rationality.

In thus disagreeing with Dray's analysis of rational explanation, I do not wish to deny that an explanatory account in terms of motivating reasons may well have evaluative overtones: what I maintain is only that whether a critical appraisal is included in, or suggested by, a given account, is irrelevant to its explanatory force; and that an appraisal alone, by means of what Dray calls a principle of action, cannot explain at all why A did in fact do x.

10.3.2 *Explanation by Reasons as Broadly Dispositional.* The notion of rational agent invoked in Schema R above must of course be conceived as a descriptive-psychological concept governed by objective criteria of application; any normative or evaluative connotations it may carry with it are inessential for its explanatory use. To be sure, normative preconceptions as to how a truly rational person ought to behave may well influence the choice of descriptive criteria for a rational agent—just as the construction of tests, and thus the selection of objective criteria, for intelligence, verbal aptitude, mathematical aptitude, and the like will be influenced by pre-systematic conceptions and norms. But the descriptive-psychological use of the term 'rational agent' (just like that of the terms 'IQ', 'verbal aptitude', 'mathematical aptitude', and the like) must then be governed by the objective empirical rules of application that have been adopted, irrespective of whether this or that person (e.g., the proponent of a rational explanation or the person to whom it is addressed) happens to regard those objective rules as conformable to his own normative standards of rationality.

By whatever specific empirical criteria it may be characterized, rationality in the descriptive-psychological sense is a *broadly dispositional trait;* to say of someone that he is a rational agent is to attribute to him, by implication, a complex bundle of dispositions. Each of these may be thought of as a tendency to behave—uniformly or with a certain probability—in a characteristic way under conditions of a given kind, whose full specifications may have to include information about the agent's objectives and beliefs, about other aspects of the psychological and biological state he is in, and about his environment. To explain an action in terms of the agent's reasons and his rationality is thus to present the action as conforming to those general tendencies, or as being a manifestation of them.[17] According as the sentences expressing the tendencies

17. This construal is in basic agreement, of course, with the general conception set forth in Ryle (1949). For a lucid characterization, in accordance with Ryle's ideas, of the force of explanations referring to an agent's wants, intentions, and plans, see Gardiner (1952), Part IV, section 3; and *cf.* also the expository and critical discussion in Dray (1957), pp. 144 and *passim*.

in question are of strictly universal form or of a statistical form such as (9.3a), or (9.3b), the resulting dispositional explanation will be deductive or inductive-probabilistic in character. But in any event it will subsume the given particular case under a general uniformity. However, this brief general characterization must now be amplified and must also be qualified in certain points of detail.

To begin with, the dispositions implied by the psychological concept of rational agent are not simply dispositions to respond to specifiable external stimuli with certain characteristic modes of overt behavior. They differ in this respect from at least some of the dispositions implied when we say of a person that he is allergic to ragweed pollen; for to say this is to imply, among other things, that he will exhibit the symptoms of a head cold when exposed to the pollen. When we call someone a rational agent, we assert by implication that he will behave in characteristic ways if he finds himself in certain kinds of situation; but such situations cannot be described simply in terms of environmental conditions and external stimuli; for characteristically they include the agent's objectives and his relevant beliefs. To mark this difference, we might say that the dispositions implied by attributing rationality to a person are *higher-order-dispositions*; for the beliefs and ends-in-view in reponse to which, as it were, a rational agent acts in a characteristic way are not manifest external stimuli but rather, in turn, broadly dispositional features of the agent. Indeed, to attribute to someone a particular belief or end-in-view is to imply that in certain circumstances he will tend to behave in certain ways which are indicative or symptomatic of his belief or his end-in-view.

There is yet another reason why we must avoid an overly narrow dispositional construal of an agent's beliefs, objectives, and rationality; and the qualified phrase 'broadly dispositional' is meant to serve as a reminder of this point as well: a statement attributing to a person certain objectives or beliefs or the property of being a rational agent, implies, *but is not equivalent to*, a set of other statements attributing to the person certain clusters of dispositions.

To elucidate and support this view, I will first adduce an analogous case from physics. To say of a body that it is electrically charged or that it is magnetic is to attribute to it, *by implication*, bundles of dispositions to respond in characteristic or symptomatic ways to various testing procedures. But this does not exhaust what is being asserted; for the concepts of electric charge, magnetization, and so on are governed by a network of theoretical principles interconnecting a large number of physical concepts. Conjointly, these theoretical principles determine an indefinitely large set of empirically testable consequences, among them various dispositional statements which provide operational criteria for ascertaining whether a given body is electrically charged, magnetic, and the like. Thus, the underlying theoretical assumptions contribute essentially to what is being asserted by the attribution of those physical properties. Indeed, it is only in conjunction with such theoretical background assumptions that a statement attributing an electric charge to a given body implies a set of dispositional statements; whereas the whole set of dispositional statements does not imply the statement about the charge, let alone the theoretical background principles.

Now, to be sure, the psychological concepts that serve to indicate a person's beliefs, objectives, moral standards, rationality, and so forth, do not function in a theoretical network comparable in scope or explicitness to that of electromagnetic theory. Nevertheless, we use those psychological concepts in a manner that clearly presupposes certain similar connections—we might call them *quasi-theoretical connections.*[18] For example, we assume that the overt behavior shown by a person pursuing a certain objective will depend on his beliefs; and conversely. Thus the attribution to Henry of the belief that the streets are slushy will be taken to imply that he will put on galoshes only on suitable assumptions about his objectives and indeed about his further beliefs,[19] such as that he wants to go out, wants to keep his feet dry, believes that his galoshes will serve the purpose, is not in too much of a hurry to put them on, and so on. This plainly reflects the assumptions of many complex interdependencies among the psychological concepts in question. And it is these assumptions which determine our expectations as to what behavioral manifestations, including overt action, a psychological trait will have in a particular case.

To reject the construal of those characteristics as simply bundles of behavioral dispositions is not to conjure up again the ghost in the machine, so deftly and subtly exorcised by Ryle and earlier—more summarily, but on basically similar grounds—by the logical behaviorism of Carnap.[20] The point is rather that in order to characterize the psychological features in question, we have to consider not only their dispositional implications, which provide operational criteria for attributing certain beliefs, objectives, and the like; we must also take account of the quasi-theoretical assumptions connecting them. For these, too, govern the use of those concepts, and they are not logically implied by the sets of dispositional statements associated with them.

18. Some plausible quasi-theoretical principles for the concept of an agent having a certain objective, or "wanting" a certain state of affairs, are set forth by Brandt and Kim (1963), p. 427, who suggest that the concept "wants" might helpfully be viewed as a theoretical construct. Tolman (1951) presents, in somewhat schematic and programmatic outline, a psychological model theory of action which includes among its "intervening variables" the "Belief-Value Matrix" as well as the "Need System" of the agent, but which also, quite rightly, considers the external conditions in which the action takes place.

19. On this point, *cf.* Chisholm (1962), pp. 513 ff. and especially p. 517.

20. See Ryle (1949); Carnap (1938) and, for a more technical account, Carnap (1936-37).

10.3.3 *Epistemic Interdependence of Belief Attributions and Goal Attributions.* The quasi-theoretical connections just referred to give rise to a problem that requires at least brief consideration. For our purposes it will suffice to examine one form of it, which is of fundamental importance to the idea of rational explanation. What sorts of dispositions do we attribute to a person by implication when we assert that he has certain specified objectives or beliefs? The statement that Henry *wants* a drink of water implies, among other things, that Henry is disposed to drink a liquid offered him—provided that he *believes* it to be potable water (and provided that he has no overriding reasons for refusing it). Thus, ascription of an objective here has implications concerning characteristic overt behavior only when taken in conjunction with ascriptions of appropriate beliefs. Similarly, in our earlier example, the hypothesis that Henry *believes* the streets to be slushy implies the occurrence of characteristic overt behavior only when taken in conjunction with suitable hypotheses about Henry's *objectives.*

Indeed, it seems that a hypothesis about an agent's objectives generally can be taken to imply the occurrence of specific overt action only when conjoined with appropriate hypotheses about his beliefs, and *vice versa.* Hence, strictly speaking, an examination of an agent's behavior can serve to test assumptions about his beliefs or about his objectives, not separately, but only in suitable pairs. That is, belief attributions and goal attributions are *epistemically interdependent.*

This fact does not make it impossible, however, to ascertain a person's beliefs or his objectives. For often we have good antecedent information about one of the interpendent items, and then a hypothesis about the other may be tested by ascertaining how the person acts in certain situations. For example, if we have good grounds for the assumption that our man is subjectively honest, that he endeavors to "tell the truth", then his answers to our questions may afford a reliable indication of his beliefs. Conversely, we are often able to test a hypothesis about a person's objectives by examining his behavior in certain critical situations because we have good reason to assume that he has certain relevant beliefs.

But the epistemic interdependence here referred to does raise the question whether an explanation by motivating reasons ever requires the explanatory assumption that the acting person was, at least at the time in question, a rational agent. How this question arises can be seen by taking a closer look at the test criteria for belief attributions and for goal attributions.

Suppose we know an agent's beliefs and wish to test the hypothesis that he wants to attain goal *G*. Just what sort of action is implied by this hypothesis? The criterion used in such cases seems to be roughly this: if *A* actually wants to attain G then he will follow a course of action which, in the light of his beliefs, offers him the best chance of success. In the parlance of our earlier discussion, therefore, the test of our goal attribution appears to presuppose the assumption that A will choose an action that is rational relative to his objectives and beliefs. This would mean that the way in which we use a person's actions as evidence in ascertaining his goals has the assumption of rationality built right into it. An analogous comment applies to the way in which we normally use the actions of a person whose objectives we know as evidence in ascertaining his beliefs.[21] But this seems to discredit the construal of rational explanation as involving, in the manner suggested in Schema R, an explanatory hypothesis to the effect that the person in question was a rational agent. For the considerations just outlined suggest that this hypothesis is always made true by a tacit convention implicit in our test criteria for the attribution of motivating objectives and beliefs to the agent. If this is generally the case, then the assumption of rationality could not possibly be violated; any apparent violation would be taken to show only that our conjectures about the agent's beliefs, or those about his objectives, or both, were mistaken. And, undeniably, such will in fact often be our verdict.

But will it always be so? I think there are various kinds of circumstances in which we might well retain our assumptions about the agent's beliefs and objectives and abandon instead the assumption of rationality. First of all, in deciding upon his action, a person may well overlook certain relevant items of information which he clearly believes to be true and which, if properly taken into account, would have called for a different course of action. Second, the agent may overlook certain aspects of the total goal he is seeking to attain, and may thus decide upon an action that is not rational as judged by his objectives and beliefs. Third, even if the agent were to take into account all aspects of his total goal as well as all the relevant information at his disposal, and even if he should go through a deliberate "calculation of means to be adopted toward his chosen end" (to repeat an earlier quotation from Dray), the result may still fail to be a rational decision because of some logical flaw in his calculation. Clearly there could be strong evidence, in certain cases, that an agent had fallen short of rationality in one of the ways here suggested; and indeed, if his decision had been made under pressure of time or under emotional strain, fatigue, or other disturbing influences, such deviations from rationality would be regarded as quite likely. (This reflects another one of the quasi-theoretical connections among the various psychological concepts that play a role in explanations by reasons or by motives.)

21. *Cf.*, for example, the discussion in Churchman (1961), pp. 288-91, which illustrates this point.

In sum then, rationality of human actions is not guaranteed by conventions implicit in the criteria governing the attribution of goals and beliefs to human agents; there may be good grounds for ascribing to an agent certain goals and beliefs and yet acknowledging that his action was not rationally called for by those goals and beliefs.

10.3.4 *Rational Action as an Explanatory Model Concept.* For further clarification of the role that the assumption of rationality plays in explanations by motivating reasons, it may be illuminating to ask whether the concept of rational agent might not be viewed as an idealized explanatory model comparable to the explanatory concept of an ideal gas, that is, a gas conforming exactly to Boyle's and Charles's laws. No actual gas strictly satisfies those laws; but there is a wide range of conditions within which many gases conform very closely to the account the model gives of the interrelations between temperature, pressure, and volume. Moreover, there are more general, but less simple laws—such as van der Waals', Clausius', and others—which explain to a large extent the deviations from the ideal model that are exhibited by actual gases.

Perhaps the concept of a rational agent can be similarly regarded as an explanatory model characterized by an "ideal law," to the effect that the agent's actions are strictly rational (in the sense of some specific criterion) relative to his objectives and beliefs. How could this programmatic conception be implemented? How could an explanatory model of rational action be precisely characterized, and how could it be applied and tested?

As noted earlier, the concept of rationality is by on means as clear and unequivocal as is sometimes implied in the literature on rational explanation. But let us assume that the proposed explanatory use of the concept is limited, to begin with, to cases of a relatively simple type for which some precise criterion of rationality can be formulated and incorporated into our model.

Then there is still the question of how to apply the model to particular instances, how to test whether a given action does in fact conform to the criterion of rationality the model incorporates. And this raises a perplexing problem. The problem is not just the practical one of how to *ascertain* an agent's beliefs and actions in a given case, but the conceptual one of what is to be *understood* by the beliefs and objectives of an agent at a given time, and by what logical means they might be properly characterized. Let me amplify this briefly.

A person must surely be taken to hold many beliefs of which he is not conscious at the time, but which could be elicited by various means. Indeed, a person may be held to believe many things he has never thought of and perhaps never will think of as long as he lives. If he believes that seven and five are twelve we would surely take him to believe also that seven speckled hens and five more make twelve speckled hens, although he may never consciously entertain this particular belief. Generally, a man will be taken to believe certain things that are consequences of other things he believes; but surely he cannot be taken to believe *all* those consequences since, to mention but one reason, his logical perspicacity is limited.

Hence, while in a theoretical model of the normative or critical concept of rational decision the information basis may be construed as a set of statements that is closed under an appropriate relation of logical derivability, this assumption cannot be transferred to an explanatory model of rational decision. In particular, a person may well give his believing assent to *one* of a pair of logically equivalent statements but withhold it from the other—although both express the same proposition. It seems clear, therefore, that the objects of a person's beliefs cannot be construed to be propositions each of which may be represented by any one of an infinite set of equivalent statements: in specifying an agent's beliefs, the mode of its formulation is essential. (This peculiarity seems closely akin to what Quine has called the referential opacity of belief sentences.)[22]

Presumably, then, in an explanatory model conception of rational action, the agent's beliefs should be represented by some set of sentences that is not closed under logical derivability. But what set? For example, should a person's belief-set be taken to include all sentences to which he could be induced to assent by pertinent questions and arguments, no matter how numerous or complex? Clearly such construal is unwarranted if we are interested in specifying a set of beliefs that can be regarded as motivating factors in explaining an action done by the agent. Where the boundary line of the belief-set is to be drawn—conceptually, not just practically—is a puzzling and obscure question.

Similar observations apply to the problem of how to characterize an agent's total objectives in a given decision situation.

Consequently, though in a normative-critical model of decision, rationality is always judged by reference to the total information basis and the total objective specified, it would be self-defeating to incorporate into an explanatory model of rational action the principle that a rational agent acts optimally, as judged by specified criteria, on the basis of his total set of objectives and beliefs: this notion is simply too obscure.

10.3.5 *The Model of a Consciously Rational Agent.* A way out seems to be suggested by the observation that many explanations present an action as determined

22. *Cf.* Quine (1960), section 30; and see also sections 35, 44, 45, which deal further with the problems of a logically adequate construal of belief-attributions. Several of these problems, and similar ones concerning the construal of goal-attributions, are searchingly examined in Scheffler (1963), Part I, section 8.

by reasons which presumably the agent took consciously into account in making his decision. Let us say that a person is a *consciously rational agent* (at a certain time) if (at that time) his actions are rational (in the sense of some clearly specified criterion) relative to those of his objectives and beliefs which he consciously takes into account in arriving at his decision.

By way of exploring the potential applicability of this model of a consciously rational agent, let us consider Bismarck's editing of the so-called Ems telegram, which played a crucial role in touching off the war between France and Prussia in 1870. Political relations between the two nations had been strained by France's strong opposition to the prospect, which for some time seemed likely, of a Hohenzollern prince accepting the throne of Spain. Bismarck had hoped that this issue might provide Prussia with a *casus belli* against France; but the prince resigned his candidacy, and the prospect of a military conflict with France seemed to vanish. At this juncture a French emissary approached King William of Prussia, who was staying at the spa of Ems, with the request that the king rule out resumption of the candidacy for all future times. The king declined this and informed Bismarck of the incident in a telegram in which he indicated no ruffled feelings but simply sought to convey his reasons for refusing the request. The king explicitly left it to Bismarck to decide whether to publish the content of the telegram. Bismarck seized the opportunity to edit the text for publication in a manner calculated to induce France to go to war. The reasons behind this action have been discussed by many writers, including Bismarck himself.

In his memoirs,[23] Bismarck states, first of all, his reasons for seeking war against France. Among these are his concern to preserve Prussia's national honor; his belief that otherwise the resulting loss of prestige would gravely interfere with the development of a German Reich under Prussian leadership; the expectation that a national war against France would serve to bridge the differences between many of the German nations Bismarck sought to unite; and the information, provided by the chief of the General Staff, that in view of Prussia's state of military preparedness no advantage was to be expected from deferring the outbreak of war. Bismarck concludes this part of his account with the words: "All these considerations, conscious and unconscious, strengthened my opinion that war could be avoided only at the cost of the honour of Prussia and of the national confidence in it. Under this conviction I made use of the royal authorization . . . to publish the contents of the telegram; and . . . I reduced the telegram by striking out words, but without adding or altering."[24]

23. Bismarck (1899), pp. 97 ff. The text of the King's telegram is quoted on p. 97, that of the edited version on pp. 100–101.

24. Bismarck (1899), p. 100.

The edited version of the Ems telegram created the impression that the king had treated the French emissary in an insulting manner. In his memoirs, Bismarck candidly states his reasons for this choice of means toward his end: he expected that the edited text would "have the effect of a red rag upon the Gallic bull. Fight we must. . . . Success, however, essentially depends upon the impression which the origination of the war makes upon us and others; it is important that we should be the party attacked, and this Gallic overweening and touchiness will make us if we announce in the face of Europe . . . that we fearlessly meet the public threats of France."[25] The publication of the edited text had the effect Bismarck had expected: in Paris it was taken as a national insult, and the French Cabinet decreed mobilization.

As for the explanatory force of Bismarck's own account or of those given by various historians, let us note first that no matter how illuminating a statement of motivating reasons may be, it cannot, and does not purport to, shed light on one very important aspect of Bismarck's action, namely, why the thought of editing the text occurred to him in the first place. In the context of our explanation by reasons, the statement that it did occur to him is simply offered as an explanatory datum, as part of the requisite specification of what courses of action the agent believed were open to him. Thus the explanatory account we have surveyed can claim at most to answer the question: given that the possibility occurred to Bismarck, why did he choose that course of action?

Let us consider now to what extent the explanation here outlined conforms to the model of a consciously rational action. First of all, it does represent Bismarck as having arrived at his decision as a result of a careful deliberation concerning the best available means toward his end of provoking France into going to war. The account indicates further that in the given situation, Bismarck believed several courses of action open to him: publication of an edited version of the telegram; publication of the original text; and no publication at all. In his estimate the first alternative, and it alone, was likely to have the desired effect. Hence if the list of motivating considerations is factually correct and complete in the sense of omitting none of the possibilities actually contemplated by Bismarck, then the account shows that his action was that of a consciously rational agent, and that relative to his beliefs and objectives it was rational in the sense of one of the simplest criteria mentioned in section 10.2.

Actually, however, the account is not likely to be strictly complete. For example, Bismarck must have considered, however briefly, some alternative courses of action—among them, different ways of editing the text—which are

25. Bismarck (1899), p. 101.

not mentioned in his own statement nor in the accounts given by various other writers who have dealt with the matter. The available studies suggest that Bismarck may have fleetingly entertained the possibility of releasing the relevant information to all Prussian embassies but not to the press for publication. Thus, there are good reasons to doubt that the available accounts are actually as complete as would be necessary to exhibit Bismarck's action as consciously rational. In defense of the presumptive omissions, it might be argued that greater completeness would have been pedantic and gratuitous, for does not the very fact that Bismarck chose to publish an edited version suffice to show that even if he should have entertained alternatives other than those explicitly mentioned, he dismissed them as less promising? This is indeed quite a plausible way of defending the claim that among all the possible actions he considered, Bismarck chose what in his estimate was the optimal one; but as far as this argument is relied on, the rationality of Bismarck's decision is safeguarded by tacitly building it into our construal of Bismarck's expectations: he could not have expected much of the alternatives or else he would have acted differently.

Thus, though in the case of the Ems telegram an unusually large amount of apparently reliable information on the motivating reasons is available, and though Bismarck's decision seems to have been arrived at by cool and careful deliberation, the rigorous requirements of the model of consciously rational action are not completely satisfied.

There are other cases which perhaps come even closer to the "ideal" of the model. Consider, for example, a competent engineer who seeks an optimal solution to a problem of design for which the range of permissible solutions is clearly delimited, the relevant probabilities and utilities are precisely specified, and even the criterion of rationality to be employed (e.g., maximization of expected utilities) is explicitly stated. In this case, the objectives and beliefs that determine the engineer's decision may be taken to be fully indicated by the specification of the problem; and by applying to the engineer the explanatory model of a consciously rational agent (whose standard of rationality is that specified in the given problem) we can explain—or predict—his arriving at a solution, or set of solutions, which is identical with the theoretically optimal one.

The broadly dispositional property of conscious rationality need not, and indeed cannot, be conceived as an enduring trait. A man may be disposed to act with conscious rationality at some times, when psychological and environmental conditions are favourable, yet fail to do so at other times, when disturbing external circumstances or such factors as fatigue, pain, or preoccupation with other matters prevent strictly rational deliberation. But similarly, a given body of gas may behave "ideally" at certain times, when it is at high temperature and under low pressure, yet nonideally at other times, when the circumstances are reversed.

However, while for a given body of gas the conditions of near-ideal behavior can be stated with considerable precision in terms of just a few quantitative parameters, the conditions under which a given individual will come very close to acting with conscious rationality can be indicated only vaguely and by means of a long, and open-ended, list of items which includes environmental as well as physiological and psychological factors. Very broadly speaking, the explanatory model concept of consciously rational action will be applicable in those cases where the decision problem the agent seeks to solve is clearly structured and permits of a relatively simple solution, where the agent is sufficiently intelligent to find the solution, and where circumstances permit careful deliberation free from disturbing influences.[26]

The idea of a consciously rational agent, with its very limited scope of application, does not offer the only way in which a model concept of rational decision might be put to explanatory and predictive use. One interesting alternative has been put forward in a study by Davidson, Suppes, and Siegel.[27] These investigators present an empirical theory of human choice which is modeled on the mathematical model of decision under risk and incorporates the hypothesis that the choices made by human subjects will be rational in the precise sense of maximizing expected utilities.

As might be anticipated, the rigorously quantitative character of the theory has the price of limiting its applicability to decisions of a rather simple type, which permit of strict experimental control. In the authors' test of the theory, the subjects had to make a series of decisions each of which called for a choice between two options. Each option offered the prospect of either gaining a specified small amount of money or losing some other specified small amount, depending on the outcome of a certain random experiment, such as rolling a regular die with peculiar markings on its faces. The random experiments, their possible outcomes, and the corresponding gains or losses were carefully described to the subject, who then made his choice.

The results of this experiment conformed quite well to the hypothesis that the subjects would choose the option with the greater *expected utility*, where the expected utility of an option is computed on the basis of theoretically postulated *subjective* probabilities and utilities which the different outcomes have for the choosing individual. The theory proposed by the authors provides an objective, if indirect, method for the simultaneous and independent measurement of such subjective probabilities and utilities for a given agent. Experimental

26. *Cf.* also the observations in Gibson (1960), pp. 165-68, which bear on this point.

27. Davidson, Suppes, and Siegel (1957).

study shows that the subjective probability which a specified outcome possesses for a given subject is not, in general, equal to its objective probability, even though the subject may know the latter; nor are the subjective utilities proportional to the corresponding monetary gains or losses. Indeed, a person normally will be entirely unaware of the subjective probabilities and utilities which, on the theory under consideration, the possible outcomes possess for him.

Thus, as far as the theory is correct, it gives a quite peculiar twist to the idea of rational action: though the subjects make their choices in clearly structured decision situations, with full opportunity for antecedent deliberation and even calculation, they act rationally (in a precisely defined quantitative sense) relative to subjective probabilities and utilities which they do not know, and which, therefore, they cannot take into account in their deliberations. They act rationally in the sense of acting *as if* they were trying to maximize expected utilities. Here, then, we seem to have a type of conscious decision which is *nonconsciously rational* with quantitative precision.

10.3.6 *The "Rationality" of Nondeliberative Actions. Explanation by Unconscious Motives.* Many purposive actions are taken without prior conscious deliberation, without any calculation of means to be chosen toward the attainment of an envisaged end; and yet such actions are often accounted for in terms of motivating reasons. Dray, who specifically includes such accounts in the scope of his analysis, argues that his conception of rational explanation is applicable to any purposive action, on the ground that "in so far as we say an action is purposive at all, no matter at what level of conscious deliberation, there is a calculation which could be constructed for it: the one the agent would have gone through if he had had time, if he had not seen what to do in a flash, if he had been called upon to account for what he did after the event, etc. And it is by eliciting some such calculation that we explain the action".[28]

But the explanatory significance of reasons or calculations constructed in this manner is certainly puzzling. If an agent arrives at his decision "in a flash" rather than by deliberation then it seems false to say that the decision can be accounted for by some argument which the agent might have gone through under more propitious circumstances, or which he might produce later if called upon to account for his action; for, by hypothesis, no such argument was in fact gone through by the agent at the crucial time; considerations of appropriateness or rationality played no part in shaping his decision, and an explanation in terms of such deliberations or calculations is simply fictitious.

Nevertheless I think Dray has a point in viewing some nondeliberative

28. Dray (1957), p. 123.

actions as akin to those which are decided upon by careful deliberation. For "rational explanations" of such actions may be viewed as broadly dispositional accounts invoking certain behavior patterns which the agent acquired by a learning process whose initial phases did involve conscious reflection and deliberation. Consider, for example, the complex set of maneuvers required in driving a car through heavy traffic, in using a sewing machine, or in performing a surgical operation: all these are learned by training processes which initially involve more or less complex deliberation, but which eventually come to be performed automatically, with little or no conscious reflection, yet often in a manner that the agent would have chosen if he had given the matter adequate thought. Accordingly, a particular action of this kind might be explained, not by a constructed calculation which in fact the agent did not carry out, but by exhibiting it as a manifestation of a general behavioral disposition which the agent has learned in the manner just suggested.[29]

The attempt to explain a given action by means of motivating reasons faces another well-known difficulty: it will frequently result in a rationalization rather than an explanation, especially when it relies on the reasons adduced by the agent himself. As G. Watson remarks, "Motivation, as presented in the perspective of history, is often too simple and straightforward, reflecting the psychology of the Age of Reason. . . . Psychology has come . . . to recognize the enormous weight of irrational and intimately personal impulses in conduct. In history, biography, and in autobiography, especially of public characters, the tendency is strong to present 'good' reasons instead of 'real' reasons."[30] Accordingly, as Watson goes on to point out, it is important, in examining the motivation of historical figures, to take into account the significance of such psychological mechanisms as reaction formation, "the dialectic dynamic by which stinginess cloaks itself in generosity, or rabid pacifism arises from the attempt to repress strong aggressive impulses."[31]

Increasing awareness that actions may be prompted to a considerable extent by motivating factors of which the agent is not conscious has prompted some historians to place strong emphasis on a more systematic use of the ideas of psychoanalysis or related depth-psychological theories in the context of his-

29. Scheffler (1963), pp. 115-16, has suggested in a similar fashion that an interpretation in terms of learning may illuminate some types of teleological statements about human behavior. On this point, see also the highly relevant article Suppes (1961); and *cf.* Gibson (1960), pp. 157-58, where a dispositional construal of nondeliberately rational acts is presented.

30. Watson (1940), p. 36.

31. *Ibid.* For some suggestive observations from a psychoanalytic point of view on the notion of "rationalization" in specifying the motives for an action, *cf.* F. Alexander (1940).

torical explanation. W. L. Langer's presidential address before the American Historical Association in 1957,[32] is a forceful statement of and plea for, this program.

Similar considerations have led some philosophical writers on motivation to distinguish, in explanations of a person's action, between "his reasons" for doing what he did and "the reasons" or "the real reasons" for his action.[33] In his illuminating study of historical explanation, Gardiner makes this observation on the latter notion: "In general, it appears safe to say that by a man's 'real reasons' we mean those reasons he would be prepared to give under circumstances where his confession would not entail adverse consequences to himself. An exception to this is the psycho-analyst's usage of the expression where different criteria are adopted."[34] But if Gardiner is right in his characterization of what is ordinarily understood by a man's real reasons for acting the way he did, then surely the historian in search of reasons that will correctly explain human actions will have to forego reliance on "real reasons" in the ordinary sense if psychological and other investigations show that they do not yield as adequate an understanding of human actions as does an interpretation in terms of less familiar conceptions, including perhaps a theory of subconscious motivation. That such a reorientation is in fact needed has been strongly urged by Langer: "Viewed in the light of modern depth psychology, the homespun, commonsense psychological interpretations of past historians, even some of the greatest, seem woefully inadequate, not to say naive. Clearly the time has come for us to reckon with a doctrine that strikes so close to the heart of our own discipline."[35]

As for the notion of the "real reasons" for a given action, I would say then, first, that psychological or historical explanation cannot be bound by the use of that notion in everyday discourse. But secondly, I doubt that the characterization which Gardiner suggests in an expressly tentative fashion does full justice even to what we mean in ordinary language when we speak of the real reasons that prompted a given action. For the idea of subconscious motives is quite familiar in our time, and we are therefore prepared to say in ordinary discourse that the reasons given by an agent may not be the "real reasons" behind his action, even if his statement is subjectively honest and he has no grounds to expect adverse consequences. And no matter whether an expla-

32. Langer (1958). For observations in a similar vein, see chap. 3 of Hughes (1964) and Mazlish's Introduction to the anthology, Mazlish (1963), which includes a number of specific examples of psychoanalytically inspired interpretations of historical materials.

33. See, for example, Peters (1958), pp. 3-9 and *passim*.

34. Gardiner (1952), p. 136.

35. Langer (1958), p. 90. Peters (1958), p. 63, explicitly notes that an unconscious wish might constitute "the reason" for a man's action.

nation of human actions is attempted in ordinary language or in the technical terms of some theory, the overriding criterion for what—if anything—should count as a "real," and thus explanatory, reason for a given action is surely not to be found by examining the way in which the term 'real reason' has thus far been used, but by investigating what conception of real reasons would yield the most satisfactory explanation of human conduct. Ordinary usage gradually changes accordingly.

The logical structure of explanations in terms of subconscious motives and processes is again broadly dispositional in the sense we considered earlier: the ascription of such motives amounts to attributing to the agent certain broadly dispositional characteristics, and the reference to subconscious mechanisms or to psychodynamic processes reflects the assumption of laws or theoretical principles involving those characteristics. To say this is not, however, to imply that all psychoanalytic interpretations that have actually been offered meet the basic requirements for scientifically adequate dispositional explanations. In fact, the empirical or operational criteria of application for psychoanalytic concepts, and the theoretical principles in which these concepts function, are often not nearly as clear as is desirable in the interest of objective applicability and testability.[36] But it should not be forgotten that in this respect common-sense motivational explanations, too, often leave much to be desired, and furthermore, that efforts are being made to put psychoanalytic and similar conceptions into a methodologically more satisfactory form.

10.3.7 *A Note on Causal Aspects of Dispositional Explanations.* It is often held that explanations in terms of motivating reasons, learned skills, personality traits, and the like, being dispositional in character, are for this reason noncausal. But this thesis seems to me misleading. For, first of all, as is shown by schemata (9.1) and (9.5), a dispositional explanation invokes, in addition to the appropriate dispositional property M, also the presence of circumstances, say S, in which the property M will manifest itself by the symptom—say, behavior of the kind R—whose occurrence is to be explained. For example, the attribution of venality to an agent will explain his having committed treason only in conjunction with suitable further assumptions, such as that he was offered a large bribe, which in virtue of his venal propensity led to the act in question. Here the offer of a bribe, in analogy to the impact of the stone in (9.1), may be said, in everyday parlance, to have caused the explanandum event. Dispositional explanations of this kind, therefore, cannot be said to be noncausal.

36. On this point, see, for example, the critique presented in Nagel (1959); and *cf.* also the critique and the defense of psychoanalytic conceptions in various other essays included in Hook (1959).

To be sure, possession of the dispositional property M would not ordinarily count as a cause: but then, the possession of M alone does not explain the given event.

Thus when Gardiner remarks that an explanation of the form 'x did y because he wanted z' does not refer to a causal relation between two events,[37] he is right in the sense that the statement 'x wanted z' does not describe an event, but ascribes to x a broadly dispositional property. But a because-sentence of the specified form surely affords an explanation only on the further assumption that x was in circumstances in which, at least by his lights, doing y could be expected to lead to z; and when supplemented by this further statement, the account takes on the form (9.5), which cannot be said to be noncausal. Gardiner's insistence that "motivational explanations . . . are not causal at all"[38] may serve a good purpose in cautioning—as it is intended to do—against the conception of motives as ghostly causes of overt behavior, and against the notion that "in history we have to do with a world of 'mental agencies', mysteriously lying behind the world of physical bodies and actions, separate from it and yet controlling it";[39] but it runs the risk of obscuring the close similarities here noted between motivational explanations and certain other accounts generally considered as causal.[40]

11. CONCLUDING REMARKS

At the beginning of this essay we contrasted reason-seeking and explanation-seeking why-questions. The former solicit grounds that will make empirical

37. Gardiner (1952), p. 124.

38. Gardiner (1952), pp. 133-34. *Cf.* also Ryle's view that "to explain an action as done from a specified motive or inclination is not to describe the action as the effect of a specified cause. Motives are not happenings and are not therefore of the right type to be causes." (1949, p. 113).

39. Gardiner (1952), p. 51.

40. In this context, see also the suggestive discussion of dispositions, reasons, and causes in Dray (1957), pp. 150-55. In contrast to the view that "only events and processes can be causes" (p. 151), Dray holds that a dispositional characteristic "is a type of 'standing condition'; and standing conditions, as well as precipitating ones, can be causes." (p. 152). The thesis that explanation by reasons is "a species of ordinary causal explanation" is interestingly argued, on rather different grounds than those here presented, in Davidson (1963), where also a number of further objections are examined. It should also be borne in mind that the everyday conception of causal explanation is rather narrow and vague and that at least in physics it has been replaced by the more general and precise conception of an explanation by means of a deterministic theory. It is illustrated by the case, considered in section 2, of the Newtonian theory of motion and of gravitation: given the "state" of a closed system of point masses at some time, the theory determines the state of the system at any other time and thus permits the explanation of a particular state of the system by reference to an earlier one. The terms of the causal relation consist here, not in events, but in momentary *states* of the system, as represented by the masses, positions, and velocities of the constituent particles at the moment in question.

statements *credible*; the latter solicit information that will explain empirical facts and thus render them *intelligible*. Our main concern has been to examine the ways in which science answers why-questions of the latter type and to characterize the kind of understanding it thereby affords.

We noted that scientific explanation is not aimed at creating a sense of familiarity with the explanandum; "reduction to the familiar" is at best an incidental aspect of it. The understanding it conveys lies rather in the insight that the explanandum fits into, or can be subsumed under, a system of uniformities represented by empirical laws or theoretical principles. Depending on the logical character of the uniformities, such subsumption will be deductive or inductive in a sense which our two basic models are intended to make explicit.

I would like to stress here once more that there are profound logical differences between those two modes of explanation. Not that in a statistical account the explanandum sentence is qualified by a modal clause such as 'probably' or 'almost certainly'; the explanandum is a nonmodal sentence in probabilistic no less than in deductive-nomological explanation and prediction. But in inductive-statistical explanation in contrast to its deductive counterpart, the explanans makes the explanandum only more or less probable and does not imply it with deductive certainty. Another difference, which so far does not seem to have received attention, lies in what I called the epistemic relativity of probabilistic explanation, i.e., the fact that we can significantly speak of a probabilistic explanation, even a potential one, only relative to some class K of statements representing a particular knowledge situation. The concept of deductive-nomological explanation requires no such relativization.

The explanatory role of presumptive laws and theoretical principles was illustrated and made explicit by an analysis of various kinds of explanation offered in different fields of empirical science. That survey does not claim completeness; it could have been expanded by examining the explanatory use of typological concepts and theories, of functional analysis, of psychoanalytic ideas, and so forth.[1]

The central theme of this essay has been, briefly, that all scientific explanation involves, explicitly or by implication, a subsumption of its subject matter under general regularities; that it seeks to provide a systematic understanding of empirical phenomena by showing that they fit into a nomic nexus. This construal, which has been set forth in detail in the preceding sections, does

1. The first two of these further topics are dealt with in two other essays in this volume: "Typological Methods in the Natural and the Social Sciences" and "The Logic of Functional Analysis." An interesting and useful collection of explanatory accounts from physics, biology, psychology, and history is offered in Kahl (1963).

not claim simply to be descriptive of the explanations actually offered in empirical science; for—to mention but one reason—there is no sufficiently clear generally accepted understanding as to what counts as a scientific explanation. The construal here set forth is, rather, in the nature of an *explication*, which is intended to replace a familiar but vague and ambiguous notion by a more precisely characterized and systematically fruitful and illuminating one. Actually, our explicatory analysis has not even led to a full definition of a precise "explicatum"-concept of scientific explanation; it purports only to make explicit some especially important aspects of such a concept.[2]

Like any other explication, the construal here put forward has to be justified by appropriate arguments. In our case, these have to show that the proposed construal does justice to such accounts as are generally agreed to be instances of scientific explanation, and that it affords a basis for a systematically fruitful logical and methodological analysis of the explanatory procedures used in empirical science. It is hoped that the arguments presented in this essay have achieved that objective.

BIBLIOGRAPHY

Alexander, F. "Psychology and the Interpretation of Historical Events." In Ware (1940), pp. 48-57.

Alexander, H. G. "General Statements as Rules of Inference?" In Feigl, Scriven, and Maxwell (1958), pp. 309-29.

Arrow, K. J. "Mathematical Models in the Social Sciences." In Lerner, D. and H. D. Laswell (eds.) *The Policy Sciences*. Stanford: Stanford University Press, 1951, pp. 129-54.

Baernstein, H. D. and Hull, C. L. "A Mechanical Model of the Conditioned Reflex." *The Journal of General Psychology* 5:99-106 (1931).

Barker, S. F. *Induction and Hypothesis*. Ithaca, N.Y.: Cornell University Press, 1957.

Barker, S. F. "The Role of Simplicity in Explanation." In Feigl and Maxwell (1961), pp. 265-74.

Bartley, W. W. "Achilles, the Tortoise, and Explanation in Science and History." *The British Journal for the Philosophy of Science* 13:15-33 (1962).

Baumol, William J. *Economic Theory and Operations Analysis*. Englewood Cliffs, N. J.: Prentice-Hall, 1961.

Baumrin, B. (ed.) *Philosophy of Science. The Delaware Seminar*. Volume I, 1961-62. New York: John Wiley & Sons, 1963.

2. That a fuller characterization of this concept, and *a fortiori* a complete explicative definition, poses further problems is made clear in section 6 of the essay "Studies in the Logic of Explanation" and in the Postscript to it. Another question that arises here is mentioned in footnote 33 of that essay as reprinted in the present volume.

Beale, H. K. "What Historians Have said About the Causes of the Civil War." In *Theory and Practice in Historical Study: A Report of the Committee on Historiography*, Social Science Research Council, Bulletin 54; New York: 1946, pp. 53-92.

Bertalanffy, L. von. *Modern Theories of Development*. London: Oxford University Press, 1933.

Bertalanffy, L. von. "Problems of General System Theory." *Human Biology* 23:302-12 (1951).

Bertalanffy, L. von. "General System Theory." In Bertalanffy, L. von, and A. Rapoport, (eds.) *General Systems. Yearbook of the Society for the Advancement of General Systems Theory*. Volume I, 1956.

Bismarck, Otto von. *Bismarck. The Man and the Statesman: Being the Reflections and Reminiscences of Otto, Prince von Bismarck*. Translated from the German under the supervision of A. J. Butler. Volume II. New York: Harper and Row, 1899.

Boehmer, H. *Luther and the Reformation in the Light of Modern Research*. Translated by E. S. G. Potter. New York: The Dial Press, 1930.

Boltzmann, L. *Vorlesungen über Maxwells Theorie der Elektrizität und des Lichtes*. I. Theil. Leipzig: Barth, 1891.

Boltzmann, L. *Populäre Schriften*. Leipzig: Barth, 1905.

Bonhoeffer, K. F. "Über physikalisch-chemische Modelle von Lebensvorgängen." *Studium Generale* 1:137-43 (1948).

Bondi, H. *The Universe at Large*. London: Heinemann, 1961.

Braithwaite, R. B. *Scientific Explanation*. Cambridge, England: Cambridge University Press, 1953.

Brandt, R. and J. Kim. "Wants as Explanations of Actions." *The Journal of Philosophy* 60:425-35 (1963).

Bridgman, P. W. *The Logic of Modern Physics*. New York: Macmillan, 1927.

Brodbeck, May. "Models, Meaning, and Theories." In Gross (1959), pp. 373-403.

Brodbeck, May. "Explanations, Predictions, and 'Imperfect' Knowledge." In Feigl and Maxwell (1962), pp. 231-72.

Bromberger, S. "The Concept of Explanation." Ph.D. thesis, Harvard University, 1960.

Bromberger, S. "An Approach to Explanation." In Butler, R. (ed.) *Studies in Analytical Philosophy*. Oxford: Blackwell, forthcoming.

Bush, R. R. and F. Mosteller. *Stochastic Models for Learning*. New York: John Wiley & Sons, 1955.

Campbell, N. R. *Physics: The Elements*. Cambridge, England: Cambridge University Press, 1920.

Campbell, N. R. *What is Science?* New York: Dover, 1952. (First published in 1921.)

Carington, W. *Matter, Mind and Meaning*. London: Methuen, 1949.

Carnap, R. "Testability and Meaning." *Philosophy of Science* 3, 1936 and 4, 1937. Reprinted in part in Feigl and Brodbeck (1953).

Carnap, R. *The Logical Syntax of Language*. New York: Harcourt, Brace and World, 1937.

Carnap, R. "Logical Foundations of the Unity of Science." In *International Encyclopedia of Unified Science*, Volume I, Number 1. Chicago: University of Chicago Press, 1938. Reprinted in Feigl and Sellars (1949), pp. 408-23.

Carnap, R. "On Inductive Logic." *Philosophy of Science* 12:72-97 (1945).

Carnap, R. *Logical Foundations of Probability*. Chicago: University of Chicago Press 1950; second, revised, edition 1962. Cited in this essay as Carnap (1950).

Carnap, R. "Inductive Logic and Science." *Proceedings of the American Academy of Arts and Sciences*, volume 80:187-97 (1951-54).

Carnap, R. *The Continuum of Inductive Methods.* Chicago: University of Chicago Press, 1952.

Carnap, R. "The Methodological Character of Theoretical Terms." In Feigl and Scriven (1956), 38-76.

Carnap, R. "The Aim of Inductive Logic." In Nagel, Suppes, and Tarski (1962), pp. 303-18.

Chisholm, R. "Sentences about Believing." In Feigl, Scriven, and Maxwell (1958), pp. 510-20.

Churchman, C. W. *Prediction and Optimal Decision.* Englewood Cliffs, N. J.: Prentice-Hall, 1961.

Cohen, M. R. and E. Nagel. *An Introduction to Logic and Scientific Method.* New York: Harcourt, Brace & World, 1934.

Conant, James B. *Science and Common Sense.* New Haven: Yale University Press, 1951.

Craig, W. "Replacement of Auxiliary Expressions." *Philosophical Review* 65:38-55 (1956).

Cramér, H. *Mathematical Methods of Statistics.* Princeton: Princeton University Press, 1946.

Danto, A. C. "On Explanations in History." *Philosophy of Science* 23:15-30 (1956).

Davidson, D. "Actions, Reasons, and Causes." *The Journal of Philosophy* 60:685-700 (1963).

Davidson, D., P. Suppes, and S. Siegel. *Decision Making: An Experimental Approach.* Stanford: Stanford University Press, 1957.

Dewey, John. *How We Think.* Boston: D. C. Heath & Co., 1910.

Donagan, A. "Explanation in History." *Mind* 66:145-64 (1957). Reprinted in Gardiner (1959), pp. 428-43.

Dray, W. "Explanatory Narrative in History." *Philosophical Quarterly* 4:15-27 (1954).

Dray, W. *Laws and Explanation in History.* Oxford: Oxford University Press, 1957.

Dray, W. "'Explaining What' in History." In Gardiner (1959), pp. 403-08.

Dray, W. "The Historical Explanation of Actions Reconsidered." In Hook (1963), pp. 105-35.

Duhem, P. *La Théorie Physique. Son Objet et Sa Structure.* Paris: Chevalier et Rivière, 1906. (Also translated by P. P. Wiener, under the title *The Aim and Structure of Physical Theory.* Princeton: Princeton University Press, 1954).

Fain, H. "Some Problems of Causal Explanation." *Mind* 72:519-32 (1963).

Feigl, H. "Some Remarks on the Meaning of Scientific Explanation." In Feigl and Sellars (1949), pp. 510-14.

Feigl, H. "Notes on Causality." In Feigl and Brodbeck (1953), pp. 408-18.

Feigl, H. and M. Brodbeck (eds.) *Readings in the Philosophy of Science.* New York: Appleton-Century-Crofts, 1953.

Feigl, H. and G. Maxwell (eds.) *Current Issues in the Philosophy of Science.* New York: Holt, Rinehart & Winston, 1961.

Feigl, H. and G. Maxwell (eds.) *Minnesota Studies in the Philosophy of Science*, Volume III. Minneapolis: University of Minnesota Press, 1962.

Feigl, H. and M. Scriven (eds.) *Minnesota Studies in the Philosophy of Science*, Volume I. Minneapolis: University of Minnesota Press, 1956.

Feigl, H., M. Scriven, and G. Maxwell (eds.) *Minnesota Studies in the Philosophy of Science*, Volume II. Minneapolis: University of Minnesota Press, 1958.

Feigl, H. and W. Sellars (eds.) *Readings in Philosophical Analysis.* New York: Appleton-Century-Crofts, 1949.

Feyerabend, P. K. "Explanation, Reduction, and Empiricism." In Feigl and Maxwell (1962), pp. 28-97.

Feyerabend, P. K. Review of Hanson (1963) in *Philosophical Review* 73:264-66 (1964).

Frank, P. *Philosophy of Science.* Englewood Cliffs, N.J.: Prentice-Hall, 1957.

Frankel, C. "Explanation and Interpretation in History." In Gardiner (1959), pp. 408-27. Reprinted from *Philosophy of Science* 24:137-55 (1957).

French, T. M. *The Integration of Behavior.* Volume I. *Basic Postulates.* Chicago: University of Chicago Press, 1952.

Freud, S. *Psychopathology of Everyday Life.* Translated by A. A. Brill. New York: The New American Library (Mentor Book Series), 1951.

Galilei, Galileo. *Dialogues Concerning Two New Sciences.* Translated by H. Crew and A. de Salvio. Evanston: Northwestern University, 1946.

Gallie, W. B. "Explanation in History and the Genetic Sciences." *Mind* 64:1955. Reprinted in Gardiner (1959), pp. 386-402.

Gardiner, P. *The Nature of Historical Explanation.* Oxford: Oxford University Press, 1952.

Gardiner, P. (ed.). *Theories of History.* New York: The Free Press, 1959.

Gasking, D. "Causation and Recipes." *Mind* 64:479-87 (1955).

Gauss, C. F. "Allgemeine Lehrsaetze in Beziehung auf die im verkehrten Verhaeltnisse des Quadrats der Entfernung wirkenden Anziehungs- und Abstossungs-Kraefte." (Published 1840) Reprinted in *Ostwalds Klassiker der exacten Wissenschaften*, No. 2, Leipzig: Wilhelm Engelmann, 1889.

Gibson, Q. *The Logic of Social Enquiry.* London: Routledge and Kegan Paul; New York: Humanities Press, 1960.

Goldstein, L. J. "A Note on the Status of Historical Reconstructions." *The Journal of Philosophy* 55:473-79 (1958).

Goodman, Nelson. "The Problem of Counterfactual Conditionals." *The Journal of Philosophy* 44:113-28 (1947). Reprinted, with minor changes, as the first chapter of Goodman (1955).

Goodman, Nelson. *Fact, Fiction, and Forecast.* Cambridge, Mass.: Harvard University Press, 1955.

Goudge, T. A. "Causal Explanation in Natural History." *The British Journal for the Philosophy of Science* 9:194-202 (1958).

Gross, L. (ed.) *Symposium on Sociological Theory.* New York: Harper & Row, 1959.

Grünbaum, A. "Temporally Asymmetric Principles, Parity between Explanation and Prediction, and Mechanism vs. Teleology." In Baumrin (1963), pp. 57-96.

Grünbaum, A. *Philosophical Problems of Space and Time.* New York: Knopf, 1963a.

Hanson, N. "On the Symmetry between Explanation and Prediction." *The Philosophical Review* 68:349-58 (1959).

Hanson, N. R. *The Concept of the Positron. A Philosophical Analysis.* Cambridge, England: Cambridge University Press, 1963.

Helmer, O. and P. Oppenheim. "A Syntactical Definition of Probability and of Degree of Confirmation." *The Journal of Symbolic Logic* 10:25-60 (1945).

Helmer, O. and N. Rescher. "On the Epistemology of the Inexact Sciences." *Management Science* 6:1959.

Hempel, C. G. "The Function of General Laws in History." *The Journal of Philosophy* 39:35-48, 1942. Reprinted in this volume.

Hempel, C. G. "Studies in the Logic of Confirmation." *Mind* 54:1-26 and 97-121 (1945). Reprinted in this volume.

Hempel, C. G. "A Note on the Paradoxes of Confirmation." *Mind* 55:79-82 (1946).

Hempel, C. G. "Problems and Changes in the Empiricist Criterion of Meaning." *Revue Internationale de Philosophie*, No. 11:41-63 (1950).

Hempel, C. G. "The Concept of Cognitive Significance: A Reconsideration." *Proceedings of the American Academy of Arts and Sciences*, Vol. 80, No. 1:61-77 (1951).

Hempel, C. G. "General System Theory and the Unity of Science." *Human Biology* 23:313-27 (1951a).

Hempel, C. G. "The Theoretician's Dilemma." In Feigl, Scriven, and Maxwell (1958), pp. 37-98. Reprinted in this volume.

Hempel, C. G. "Empirical Statements and Falsifiability." *Philosophy* 33:342-48 (1958a).

Hempel, C. G. "The Logic of Functional Analysis." In Gross (1959), pp. 271-307. Reprinted in this volume.

Hempel, C. G. "Inductive Inconsistencies." *Synthese* 12: 439-69 (1960). Reprinted in this volume.

Hempel, C. G. "Deductive-Nomological *vs.* Statistical Explanation." In Feigl and Maxwell (1962), pp. 98-169.

Hempel, C. G. and P. Oppenheim. "A Definition of 'Degree of Confirmation'." *Philosophy of Science* 12: 98-115 (1945).

Hempel, C. G., and P. Oppenheim. "Studies in the Logic of Explanation." *Philosophy of Science* 15:135-75 (1948). Reprinted in this volume.

Henson, R. B. "Mr. Hanson on the Symmetry of Explanation and Prediction." *Philosophy of Science* 30:60-61 (1963).

Hertz, H. *Die Prinzipien der Mechanik*. Leipzig: Johann Ambrosius Barth, 1894.

Hesse, Mary, B. *Models and Analogies in Science*. London and New York: Sheed and Ward, 1963.

Homans, George C. *Social Behavior. Its Elementary Forms*. New York: Harcourt, Brace & World, 1961.

Hook, S. (ed.). *Psychoanalysis, Scientific Method, and Philosophy*. New York: New York University Press, 1959.

Hook, S. (ed.). *Philosophy and History*. New York: New York University Press, 1963.

Hughes, H. S. *History as Art and Science*. New York: Harper & Row, 1964.

International Union of History and Philosophy of Sciences. *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*. Dordrecht, Holland: D. Reidel, 1961.

Kahl, R. (ed.). *Studies in Explanation. A Reader in the Philosophy of Science*. Englewood Cliffs, N. J.: Prentice-Hall, 1963.

Kemeny, J. G. and P. Oppenheim. "Degree of Factual Support." *Philosophy of Science* 19:307-24 (1952).

Kemeny, J. G., and P. Oppenheim. "On Reduction." *Philosophical Studies* 7:6-19 (1956).

Keynes, J. M. *A Treatise on Probability*. London: Macmillan, 1921.

Kim, J. "Explanation, Prediction, and Retrodiction: Some Logical and Pragmatic Considerations." Ph.D. thesis, Princeton University, 1962.

Körner, S. (ed.). *Observation and Interpretation: Proceedings of the Ninth Symposium of the Colston Research Society*. New York: Academic Press, and London: Butterworths Scientific Publications, 1957.

Kolmogoroff, A. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer, 1933.

Krueger, R. G. and Hull, C. L. "An Electro-Chemical Parallel to the Conditioned Reflex." *The Journal of General Psychology* 5:262-69 (1931).

Langer, W. L. "The Next Assignment." *The American Historical Review* 63: 283-304 (1958). Reprinted in Mazlish (1963), pp. 87-107. Page references are to reprinted text.

Lazarsfeld, P. F. "The American Soldier—An Expository Review." *Public Opinion Quarterly* 13:377-404 (1949).

Lazarsfeld, P. F. *Mathematical Thinking in the Social Sciences*. New York: The Free Press, 1954.

Leduc, S. *The Mechanism of Life*. Translated by W. D. Butcher, New York: Rebman Co., 1911.

Leduc, S. *La Biologie Synthétique*. Paris, 1912.

Lewis, C. I. *An Analysis of Knowledge and Valuation*. La Salle, Ill.: Open Court Publishing Co., 1946.

Lodge, Sir O. *Modern Views of Electricity*. London: Macmillan, 1889.

Luce, R. D. and H. Raiffa. *Games and Decisions*. New York: John Wiley, 1957.

Mandelbaum, M. "Historical Explanation: The Problem of 'Covering Laws'." *History and Theory* 1:229-42 (1961).

Mandler, G. and W. Kessen. *The Language of Psychology*. New York: John Wiley & Sons, 1959.

Margenau, H. *The Nature of Physical Reality*. New York: McGraw-Hill, 1950.

Mazlish, B. (ed.). *Psychoanalysis and History*. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Maxwell, J. C. "On Faraday's Lines of Force." *Transactions of the Cambridge Philosophical Society*, 10:27-83 (1864).

Mendel, A. "Evidence and Explanation." In *Report of the Eighth Congress of the International Musicological Society, New York, 1961*. La Rue, Jan (ed.). Kassel: Bärenreiter-Verlag, 1962. Volume II, pp. 3-18.

Miller, N. E. "Comments on Theoretical Models. Illustrated by the Development of a Theory of Conflict Behavior." *Journal of Personality* 20:82-190 (1951).

Mises, R. von. *Wahrscheinlichkeitsrechnung und ihre Anwendungen in der Statistik und theoretischen Physik*. Wien, 1931. Republished New York: M. S. Rosenberg, 1945.

Mises, R. von. *Probability, Statistics and Truth*. London: William Hodge & Co., 1939.

Mises, R. von. *Positivism. A Study in Human Understanding*. Cambridge, Mass.: Harvard University Press, 1951.

Moulton, F. R. and J. R. Schifferes. *The Autobiography of Science*. Garden City, N.Y.: Doubleday & Co., 1945.

Muir, R. *A Short History of the British Commonwealth*. Volume II. London: George Philip and Son, 1922.

Nagel, E. *Principles of the Theory of Probability*. Chicago: University of Chicago Press, 1939.

Nagel, E. *Logic without Metaphysics*. New York: The Free Press, 1956.

Nagel, E. "Methodological Issues in Psychoanalytic Theory." In Hook (1959), pp. 38-56.

Nagel, E. *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World, Inc., 1961.

Nagel, E., P. Suppes, and A. Tarski (eds.). *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress.* Stanford: Stanford University Press, 1962.

Neumann, J. von and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton: Princeton University Press, 2d. ed., 1947.

Passmore, J. "Law and Explanation in History." *The Australian Journal of Politics and History* 4-269:76 (1958).

Passmore, J. "Explanation in Everyday Life, in Science, and in History." *History and Theory* 2-105:23 (1962).

Peters, R. S. *The Concept of Motivation.* London: Routledge and Kegan Paul; New York: Humanities Press, 1958.

Pitt, J. "Generalizations in Historical Explanation." *The Journal of Philosophy* 56:578-86 (1959).

Popper, K. R. *Logik der Forschung.* Vienna: Springer, 1935.

Popper, K. R. "The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory." In Körner (1957), pp. 65-70.

Popper, K. R. "The Aim of Science." *Ratio* 1:24-35 (1957a).

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

Popper, K. R. *Conjectures and Refutations.* New York: Basic Books (1962).

Price, H. H. "The Theory of Telepathy." *Horizon* 12: 45-63 (1945).

Quine, W. V. O. *Word and Object.* Published jointly by Technology Press of the Massachusetts Institute of Technology and John Wiley and Sons, New York. 1960.

Ramsey, F. P. *The Foundations of Mathematics and Other Logical Essays.* London: Routledge and Kegan Paul; New York: Harcourt, Brace & World, 1931.

Reichenbach, H. *The Theory of Probability.* Berkeley and Los Angeles: The University of California Press, 1949.

Rescher, N. "A Theory of Evidence." *Philosophy of Science* 25:83-94 (1958).

Rescher, N. "Discrete State Systems, Markov Chains, and Problems in the Theory of Scientific Explanation and Prediction." *Philosophy of Science* 30:325-45 (1963).

Russell, S. B. "A Practical Device to Simulate the Working of Nervous Discharges." *The Journal of Animal Behavior* 3:15-35 (1913.)

Ryle, G. *The Concept of Mind.* London: Hutchinson, 1949.

Ryle, G. "'If', 'So', and 'Because'." In Black, M. (ed.). *Philosophical Analysis.* Ithaca, N.Y.: Cornell University Press, 1950.

Savage, L. J. *The Foundations of Statistics.* New York: John Wiley & Sons, 1954.

Scheffler, I. "Explanation, Prediction, and Abstraction." *The British Journal for the Philosophy of Science* 7:293-309 (1957).

Scheffler, I. *The Anatomy of Inquiry: Philosophical Studies in the Theory of Science.* New York: Alfred A. Knopf, 1963.

Schlick, M. "Die Kausalität in der gegenwärtigen Physik." *Die Naturwissenschaften* 19 (1931). Translated by D. Rynin, "Causality in Contemporary Physics." *The British Journal for the Philosophy of Science* 12:177-93 and 281-98 (1962).

Schwiebert, E. G. *Luther and His Times.* St. Louis: Concordia Publishing House, 1950.

Sciama, D. W. *The Unity of the Universe.* Garden City, N.Y.: Doubleday and Co. (Anchor Books), 1961.

Scriven, M. "Definitions, Explanations, and Theories." In Feigl, Scriven, and Maxwell (1958), pp. 99-195.

Scriven, M. "Truisms as the Grounds for Historical Explanations." In Gardiner (1959), pp. 443-75.
Scriven, M. "Explanation and Prediction in Evolutionary Theory." *Science* 130:477-82 (1959a).
Scriven, M. "Explanations, Predictions, and Laws." In Feigl and Maxwell (1962), pp. 170-230.
Scriven, M. "The Temporal Asymmetry between Explanations and Predictions." In Baumrin (1963), pp. 97-105.
Scriven, M. "New Issues in the Logic of Explanation." In Hook (1963a), pp. 339-61.
Seeliger, R. "Analogien und Modelle in der Physik." *Studium Generale* 1:125-37 (1948).
Sellars, W. "Inference and Meaning." *Mind* 62:313-38 (1953).
Sellars, W. "Counterfactuals, Dispositions, and the Causal Modalities." In Feigl, Scriven, and Maxwell (1958), pp. 225-308.
Society for Experimental Biology. *Models and Analogues in Biology: Symposia of the Society for Experimental Biology, Number XIV.* Cambridge, England: Cambridge University Press, 1960.
Suppes, P. "The Philosophical Relevance of Decision Theory." *The Journal of Philosophy* 58:605-14 (1961).
Svedberg, T. *Die Existenz der Moleküle.* Leipzig: Akademische Verlagsgesellschaft, 1912.
Thomson, Sir William. *Notes of Lectures on Molecular Dynamics and the Wave Theory of Light.* Baltimore: The Johns Hopkins University, 1884.
Tolman, E. C. "A Psychological Model." In Parsons, T. and E. A. Shils (eds.) *Toward a General Theory of Action.* Cambridge, Mass.: Harvard University Press, 1951; pp. 277-361.
Toulmin, S. *The Philosophy of Science.* London: Hutchinson, 1953.
Toulmin, S. *The Uses of Argument.* Cambridge, England: Cambridge University Press, 1958.
Toulmin, S. *Foresight and Understanding.* London: Hutchinson, 1961; New York: Harper & Row (Torchbook), 1963.
Toynbee, A. *The World and the West.* London: Oxford University Press, 1953.
Turner, J. "Maxwell on the Method of Physical Analogy." *The British Journal for the Philosophy of Science* 6:226-38 (1955).
Turner, J. "Maxwell on the Logic of Dynamical Explanation." *Philosophy of Science* 23:36-47 (1956).
Ware, C. F. (ed.). *The Cultural Approach to History.* New York: Columbia University Press, 1940.
Watkins, W. H. *On Understanding Physics.* Cambridge, England: Cambridge University Press, 1938.
Watson, G. "Clio and Psyche: Some Interrelations of Psychology and History." In Ware (1940), pp. 34-47.
Weber, Max. *On the Methodology of the Social Sciences.* Translated and edited by Shils, E. A. and H. A. Finch. New York: The Free Press, 1949.
Weingartner, R. H. "The Quarrel about Historical Explanation." *The Journal of Philosophy* 58:29-45 (1961).
Wiener, N. *Cybernetics.* New York: John Wiley & Sons, 1948.
Williams, D. C. *The Ground of Induction.* Cambridge, Mass.: Harvard University Press, 1947.