3

Carl Hempel, "Two Models of Scientific Explanation"

Carl Gustav Hempel's (1905–1997) classic account of explanation is widely known as the "covering-law model." It includes two species: the Deductive-Nomological Explanation and the Probabilistic-Statistical Explanation. Since its introduction in the seminal paper co-authored by Hempel and Paul Oppenheim (1948), the model has been intensely debated and criticized in the philosophy of science literature.

For although certain experiments are always necessary to serve as a basis for reasoning, nevertheless, once these experiments are given, we should derive from them everything which anyone at all could possibly derive; and we should even discover what experiments remain to be done for the clarification of all further doubts. That would be an admirable help, even in political science and medicine, to steady and perfect reasoning concerning given symptoms and circumstances. For even while there will not be enough given circumstances to form an infallible judgment, we shall always be able to determine what is most probable on the data given. And that is all that reason can do.

(*The General Science*, Leibniz, 1677)

1. Introduction

Among the divers factors that have encouraged and sustained scientific inquiry through its long history are two pervasive human concerns which provide, I think, the basic motivation for all scientific research. One of these is man's persistent desire to improve his strategic position in the world by means of dependable methods for predicting and, whenever possible, controlling the events that occur in it. The extent to which science has been able to satisfy this urge is reflected impressively in the vast and steadily widening

range of its technological applications. But besides this practical concern, there is a second basic motivation for the scientific quest, namely, man's insatiable intellectual curiosity, his deep concern to *know* the world he lives in, and to *explain*, and thus to *understand*, the unending flow of phenomena it presents to him.

In times past questions as to the *what* and the *why* of the empirical world were often answered by myths; and to some extent, this is so even in our time. But gradually, the myths are displaced by the concepts, hypotheses, and theories developed in the various branches of empirical science, including the natural sciences, psychology, and sociological as well as historical inquiry. What is the general character of the understanding attainable by these means, and what is its potential scope? In this paper I will try to shed some light on these questions by examining what seem to me the two basic types of explanation offered by the natural sciences, and then comparing them with some modes of explanation and understanding that are found in historical studies.

First, then, a look at explanation in the natural sciences.

2. Two Basic Types of Scientific Explanation

2.1. Deductive-Nomological Explanation

In his book, *How We Think*,[1] John Dewey describes an observation he made one day when, washing dishes, he took some glass tumblers out of the hot soap suds and put them upside down on a plate: he noticed that soap bubbles emerged from under the tumblers' rims, grew for a while, came to a standstill, and finally receded inside the tumblers. Why did this happen? The explanation Dewey outlines comes to this: In transferring a tumbler to the plate, cool air is caught in it; this air is gradually warmed by the glass, which initially has the temperature of the hot suds. The warming of the air is accompanied by an increase in its pressure, which in turn produces an expansion of the soap film between the plate and the rim. Gradually, the glass cools off, and so does the air inside, with the result that the soap bubbles recede.

This explanatory account may be regarded as an argument to the effect that the event to be explained (let me call it the explanandum-event) was to be expected by reason of certain explanatory facts. These may be divided into two groups: (i) particular facts and (ii) uniformities expressed by general laws. The first group includes facts such as these: the tumblers had been immersed, for some time, in soap suds of a temperature considerably higher than that of the surrounding air; they were put, upside down, on a plate on which a puddle of soapy water had formed, providing a connecting soap

film, etc. The second group of items presupposed in the argument includes the gas laws and various other laws that have not been explicitly suggested concerning the exchange of heat between bodies of different temperature, the elastic behavior of soap bubbles, etc. If we imagine these various presuppositions explicitly spelled out, the idea suggests itself of construing the explanation as a deductive argument of this form:

$$\text{(D)} \qquad \frac{\begin{array}{c} C_1, C_2, \ldots, C_k \\ L_1, L_2, \ldots, L_r \end{array}}{E}$$

Here, $C_1, C_2, \ldots, C_k$ are statements describing the particular facts invoked; $L_1, L_2, \ldots, L_r$ are general laws: jointly, these statements will be said to form the explanans. The conclusion E is a statement describing the explanandum-event; let me call it the explanandum-statement, and let me use the word "explanandum" to refer to either E or to the event described by it.

The kind of explanation thus characterized I will call *deductive-nomological explanation*; for it amounts to a deductive subsumption of the explanandum under principles which have the character of general laws: it answers the question "*Why* did the explanandum event occur?" by showing that the event resulted from the particular circumstances specified in C_1, $C_2, \ldots, C_k$ in accordance with the laws $L_1, L_2, \ldots, L_r$. This conception of explanation, as exhibited in schema (D), has therefore been referred to as the covering law model, or as the deductive model, of explanation.[2]

A good many scientific explanations can be regarded as deductive-nomological in character. Consider, for example, the explanation of mirror-images, of rainbows, or of the appearance that a spoon handle is bent at the point where it emerges from a glass of water: in all these cases, the explanandum is deductively subsumed under the laws of reflection and refraction. Similarly, certain aspects of free fall and of planetary motion can be accounted for by deductive subsumption under Galileo's or Kepler's laws.

In the illustrations given so far the explanatory laws had, by and large, the character of empirical generalizations connecting different observable aspects of the phenomena under scrutiny: angle of incidence with angle of reflection or refraction, distance covered with falling time, etc. But science raises the question "why?" also with respect to the uniformities expressed by such laws, and often answers it in basically the same manner, namely, by subsuming the uniformities under more inclusive laws, and eventually under comprehensive theories. For example, the question, "Why do Galileo's and Kepler's laws hold?" is answered by showing that these laws are but special consequences of the Newtonian laws of motion and of gravitation; and

these, in turn, may be explained by subsumption under the more comprehensive general theory of relativity. Such subsumption under broader laws or theories usually increases both the breadth and the depth of our scientific understanding. There is an increase in breadth, or scope, because the new explanatory principles cover a broader range of phenomena; for example, Newton's principles govern free fall on the earth and on other celestial bodies, as well as the motions of planets, comets, and artificial satellites, the movements of pendulums, tidal changes, and various other phenomena. And the increase thus effected in the depth of our understanding is strikingly reflected in the fact that, in the light of more advanced explanatory principles, the original empirical laws are usually seen to hold only approximately, or within certain limits. For example, Newton's theory implies that the factor g in Galileo's law, $s = \frac{1}{2} gt^2$, is not strictly a constant for free fall near the surface of the earth; and that, since every planet undergoes gravitational attraction not only from the sun, but also from the other planets, the planetary orbits are not strictly ellipses, as stated in Kepler's laws.

One further point deserves brief mention here. An explanation of a particular event is often conceived as specifying its *cause*, or causes. Thus, the account outlined in our first illustration might be held to explain the growth and the recession of the soap bubbles by showing that the phenomenon was *caused* by a rise and a subsequent drop of the temperature of the air trapped in the tumblers. Clearly, however, these temperature changes provide the requisite explanation only in conjunction with certain other conditions, such as the presence of a soap film, practically constant pressure of the air surrounding the glasses, etc. Accordingly, in the context of explanation, a cause must be allowed to consist in a more or less complex set of particular circumstances; these might be described by a set of sentences: $C_1, C_2, \ldots, C_k$. And, as suggested by the principle "Same cause, same effect," the assertion that those circumstances jointly caused a given event—described, let us say, by a sentence E—implies that whenever and wherever circumstances of the kind in question occur, an event of the kind to be explained comes about. Hence, the given causal explanation implicitly claims that there are general laws—such as $L_1, L_2, \ldots, L_r$, in schema (D)—by virtue of which the occurrence of the causal antecedents mentioned in $C_1, C_2, \ldots, C_k$ is a sufficient condition for the occurrence of the event to be explained. Thus, the relation between causal factors and effect is reflected in schema (D): causal explanation is deductive-nomological in character. (However, the customary formulations of causal and other explanations often do not explicitly specify all the relevant laws and particular facts: to this point, we will return later.)

The converse does not hold: there are deductive-nomological explanations which would not normally be counted as causal. For one thing, the

subsumption of laws, such as Galileo's or Kepler's laws, under more comprehensive principles is clearly not causal in character: we speak of causes only in reference to *particular* facts or events, and not in reference to *universal facts* as expressed by general laws. But not even all deductive-nomological explanations of particular facts or events will qualify as causal; for in a causal explanation some of the explanatory circumstances will temporally precede the effect to be explained: and there are explanations of type (D) which lack this characteristic. For example, the pressure which a gas of specified mass possesses at a given time might be explained by reference to its temperature and its volume at the same time, in conjunction with the gas law which connects simultaneous values of the three parameters.[3]

In conclusion, let me stress once more the important role of laws in deductive-nomological explanation: the laws connect the explanandum event with the particular conditions cited in the explanans, and this is what confers upon the latter the status of explanatory (and, in some cases, causal) factors in regard to the phenomenon to be explained.

2.2. *Probabilistic Explanation*

In deductive-nomological explanation as schematized in (D), the laws and theoretical principles involved are of *strictly universal form*: they assert that in *all* cases in which certain specified conditions are realized an occurrence of such and such a kind will result; the law that any metal, when heated under constant pressure, will increase in volume, is a typical example; Galileo's, Kepler's, Newton's, Boyle's, and Snell's laws, and many others, are of the same character.

Now let me turn next to a second basic type of scientific explanation. This kind of explanation, too, is nomological, i.e., it accounts for a given phenomenon by reference to general laws or theoretical principles; but some or all of these are of *probabilistic-statistical form*, i.e., they are, generally speaking, assertions to the effect that if certain specified conditions are realized, then an occurrence of such and such a kind will come about with such and such a statistical probability.

For example, the subsiding of a violent attack of hay fever in a given case might well be attributed to, and thus explained by reference to, the administration of 8 milligrams of chlor-trimeton. But if we wish to connect this antecedent event with the explanandum, and thus to establish its explanatory significance for the latter, we cannot invoke a universal law to the effect that the administration of 8 milligrams of that antihistamine will invariably terminate a hay fever attack: this simply is not so. What can be asserted is only a generalization to the effect that administration of the drug will be followed by relief with high statistical probability, i.e., roughly speaking,

with a high relative frequency in the long run. The resulting explanans will thus be of the following type:

> John Doe had a hay fever attack and took 8 milligrams of chlor-trimeton.
> The probability for subsidence of a hay fever attack upon administration of 8 milligrams of chlor-trimeton is high.

Clearly, this explanans does not deductively imply the explanandum, "John Doe's hay fever attack subsided"; the truth of the explanans makes the truth of the explanandum not certain (as it does in a deductive-nomological explanation) but only more or less likely or, perhaps "practically" certain.

Reduced to its simplest essentials, a probabilistic explanation thus takes the following form:

$$(P) \qquad \left. \begin{array}{c} Fi \\ \hline p(O, F) \text{ is very high} \\ \hline\hline Oi \end{array} \right\} \text{makes very likely}$$

The explanandum, expressed by the statement "Oi," consists in the fact that in the particular instance under consideration, here called i (e.g., John Doe's allergic attack), an outcome of kind O (subsidence) occurred. This is explained by means of two explanans-statements. The first of these, "Fi," corresponds to C_1, C_2, . . ., C_k in (D); it states that in case i, the factors F (which may be more or less complex) were realized. The second expresses a law of probabilistic form, to the effect that the statistical probability for outcome O to occur in cases where F is realized is very high (close to 1). The double line separating explanandum from explanans is to indicate that, in contrast to the case of deductive-nomological explanation, the explanans does not logically imply the explanandum, but only confers a high likelihood upon it. The concept of likelihood here referred to must be clearly distinguished from that of statistical probability, symbolized by "p" in our schema. A statistical probability is, roughly speaking, the long-run relative frequency with which an occurrence of a given kind (say, F) is accompanied by an "outcome" of a specified kind (say, O). Our likelihood, on the other hand, is a relation (capable of gradations) not between kinds of occurrences, but between statements. The likelihood referred to in (P) may be characterized as the strength of the inductive support, or the degree of rational credibility, which the explanans confers upon the explanandum; or, in Carnap's terminology, as the *logical*, or *inductive*, (in contrast to statistical) *probability* which the explanandum possesses relative to the explanans.

Thus, probabilistic explanation, just like explanation in the manner of schema (D), is nomological in that it presupposes general laws; but because these laws are of statistical rather than of strictly universal form, the resulting explanatory arguments are inductive rather than deductive in character. An inductive argument of this kind *explains* a given phenomenon by showing that, in view of certain particular events and certain statistical laws, its occurrence was to be expected with high logical, or inductive, probability.

By reason of its inductive character, probabilistic explanation differs from its deductive-nomological counterpart in several other important respects; for example, its explanans may confer upon the explanandum a more or less high degree of inductive support; in this sense, probabilistic explanation admits of degrees, whereas deductive-nomological explanation appears as an either-or affair: a given set of universal laws and particular statements either does or does not imply a given explanandum statement. A fuller examination of these differences, however, would lead us far afield and is not required for the purposes of this paper.[4]

One final point: the distinction here suggested between deductive-nomological and probabilistic explanation might be questioned on the ground that, after all, the universal laws invoked in a deductive explanation can have been established only on the basis of a finite body of evidence, which surely affords no exhaustive verification, but only more or less strong probability for it; and that, therefore, all scientific laws have to be regarded as probabilistic. This argument, however, confounds a logical issue with an epistemological one: it fails to distinguish properly between the *claim* made by a given law-statement and the *degree of confirmation*, or *probability*, which it possesses on the available evidence. It is quite true that statements expressing laws of either kind can be only incompletely confirmed by any given finite set—however large—of data about particular facts; but law-statements of the two different types make claims of different kind, which are reflected in their logical forms: roughly, a universal law-statement of the simplest kind asserts that *all* elements of an indefinitely large reference class (e.g., copper objects) have a certain characteristic (e.g., that of being good conductors of electricity); while statistical law-statements assert that in the long run, a specified proportion of the members of the reference class have some specified property. And our distinction of two types of law and, concomitantly, of two types of scientific explanation, is based on this difference in claim as reflected in the difference of form.

The great scientific importance of probabilistic explanation is eloquently attested to by the extensive and highly successful explanatory use that has been made of fundamental laws of statistical form in genetics, statistical mechanics, and quantum theory.

3. Elliptic and Partial Explanations: Explanation Sketches

As I mentioned earlier, the conception of deductive-nomological explanation reflected in our schema (D) is often referred to as the covering law model, or the deductive model, of explanation: similarly, the conception underlying schema (P) might be called the probabilistic or the inductive-statistical, model of explanation. The term "model" can serve as a useful reminder that the two types of explanation as characterized above constitute ideal types or theoretical idealizations and are not intended to reflect the manner in which working scientists actually formulate their explanatory accounts. Rather, they are meant to provide explications, or rational reconstructions, or theoretical models, of certain modes of scientific explanation.

In this respect our models might be compared to the concept of mathematical proof (within a given theory) as construed in meta-mathematics. This concept, too, may be regarded as a theoretical model: it is not intended to provide a descriptive account of how proofs are formulated in the writings of mathematicians: most of these actual formulations fall short of rigorous and, as it were, ideal, meta-mathematical standards. But the theoretical model has certain other functions: it exhibits the rationale of mathematical proofs by revealing the logical connections underlying the successive steps; it provides standards for a critical appraisal of any proposed proof constructed within the mathematical system to which the model refers; and it affords a basis for a precise and far-reaching theory of proof, provability, decidability, and related concepts. I think the two models of explanation can fulfill the same functions, if only on a much more modest scale. For example, the arguments presented in constructing the models give an indication of the sense in which the models exhibit the rationale and the logical structure of the explanations they are intended to represent.

I now want to add a few words concerning the second of the functions just mentioned; but I will have to forgo a discussion of the third.

When a mathematician proves a theorem, he will often omit mention of certain propositions which he presupposes in his argument and which he is in fact entitled to presuppose because, for example, they follow readily from the postulates of his system or from previously established theorems or perhaps from the hypothesis of his theorem, if the latter is in hypothetical form; he then simply assumes that his readers or listeners will be able to supply the missing items if they so desire. If judged by ideal standards, the given formulation of the proof is elliptic or incomplete; but the departure from the ideal is harmless: the gaps can readily be filled in. Similarly, explanations put forward in everyday discourse and also in scientific contexts are often *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 the lawn sprinkler because the sunlight was reflected and refracted by the water droplets, we may be said to offer elliptic formulations of deductive-nomological explanations; an account of this kind omits mention of certain laws or particular facts which it tacitly takes for granted, and whose explicit citation would yield a complete deductive-nomological argument.

In addition to elliptic formulation, there is another, quite important, respect in which many explanatory arguments deviate from the theoretical model. It often happens that the statement actually included in the explanans, together with those which may reasonably be assumed to have been taken for granted in the context at hand, explain the given explanandum only *partially*, in a sense which I will try to indicate by an example. In his *Psychopathology of Everyday Life*, Freud offers the following explanation of a slip of the pen that occurred to him: "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 20th 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, the formulation of the intended explanation is *at least incomplete* in the sense considered a moment ago. In particular, it fails to mention any laws or theoretical principles in virtue of which the subconscious wish, and the other antecedent circumstances referred to, could be held to explain Freud's slip of the pen. However, the general theoretical considerations Freud presents here and elsewhere in his writings suggests strongly that his explanatory account relies on a hypothesis to the effect that when a person has a strong, though perhaps unconscious, desire, then if he commits a slip of pen, tongue, memory, or the like, the slip will take a form in which it expresses, and perhaps symbolically fulfills, the given desire.

Even this rather vague hypothesis is probably more definite than what Freud would have been willing to assert. But for the sake of the argument let us accept it and include it in the explanans, together with the particular statements that Freud did have the subconscious wish he mentions, and that he was going to commit a slip of the pen. Even then, the resulting explanans permits us to deduce only that the slip made by Freud would, *in some way or other*, express and perhaps symbolically fulfill Freud's subconscious wish.

But clearly, such expression and fulfillment might have been achieved by many other kinds of slip of the pen than the one actually committed.

In other words, the explanans does not imply, and thus fully explain, that the particular slip, say s, which Freud committed on this occasion, would fall within the narrow class, say W, of acts which consist in writing the words "Thursday, October 20th"; rather, the explanans implies only that s would fall into a wider class, say F, which includes W as a proper subclass, and which consists of all acts which would express and symbolically fulfill Freud's subconscious wish *in some way or other*.

The argument under consideration might be called a *partial explanation*: it provides complete, or conclusive, grounds for expecting s to be a member of F, and since W is a subclass of F, it thus shows that the explanandum, i.e., s falling within W, accords with, or bears out, what is to be expected in consideration of the explanans. By contrast, a deductive-nomological explanation of the form (D) might then be called *complete* since the explanans here does imply the explanandum.

Clearly, the question whether a given explanatory argument is complete or partial can be significantly raised only if the explanandum sentence is fully specified; only then can we ask whether the explanandum does or does not follow from the explanans. Completeness of explanation, in this sense, is relative to our explanandum sentence. Now, it might seem much more important and interesting to consider instead the notion of a complete explanation of some *concrete event*, such as the destruction of Pompeii, or the death of Adolf Hitler, or the launching of the first artificial satellite: we might want to regard a particular event as completely explained only if an explanatory account of deductive or of inductive form had been provided for all of its aspects. This notion, however, is self-defeating; for any particular event may be regarded as having infinitely many different aspects or characteristics, which cannot all be accounted for by a finite set, however large, of explanatory statements.

In some cases, what is intended as an explanatory account will depart even further from the standards reflected in the model schemata (D) and (P) above. An explanatory account, 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*: it may suggest, perhaps quite vividly and persuasively, the general outlines of what, it is hoped, can eventually be supplemented so as to yield a more closely reasoned argument based on explanatory hypotheses which are indicated more fully, and which more readily permit of critical appraisal by reference to empirical evidence.

The decision whether a proposed explanatory account is to be qualified as an elliptically formulated deductive or probabilistic 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 argument and of the background assumptions that may be assumed to have been tacitly taken for granted, or at least to be available, in the given context. Unequivocal decision rules cannot be set down for this purpose any more than for determining whether a given informally stated inference which is not deductively valid by reasonably strict standards is to count nevertheless as valid but enthymematically formulated, or as fallacious, or as an instance of sound inductive reasoning, or perhaps, for lack of clarity, as none of these.

. . .

Notes

1 See Dewey, John. *How We Think*. Boston, New York, Chicago, 1910; Chapter VI.

2 For a fuller presentation of the model and for further references, see, for example, Hempel, C. G. and P. Oppenheim, "Studies in the Logic of Explanation," *Philosophy of Science* 15: 135–175 (1948). (Secs. 1–7 of this article, which contain all the fundamentals of the presentation, are reprinted in Feigl, H. and M. Brodbeck (eds.), *Readings in the Philosophy of Science*. New York, 1953.)—The suggestive term "covering law model" is W. Dray's; cf. his *Laws and Explanation in History*. Oxford, 1957; Chapter I. Dray characterizes this type of explanation as "subsuming what is to be explained under a general law" (*loc. cit.*, p. 1), and then rightly urges, in the name of methodological realism, that "the requirement of a *single* law be dropped" (*loc. cit.*, p. 24; italics, the author's): it should be noted, however, that, like the schema (D) above, several earlier publications on the subject (among them the article mentioned at the beginning of this note) make explicit provision for the inclusion of more laws than one in the explanans.

3 The relevance of the covering-law model to causal explanation is examined more fully in sec. 4 of Hempel, C. G., "Deductive-Nomological *vs.* Statistical Explanation." In Feigl, H., et al. (eds.), *Minnesota Studies in the Philosophy of Science*, vol. III. Minneapolis, 1962.

4 The concept of probabilistic explanation, and some of the peculiar logical and methodological problems engendered by it, are examined in some detail in Part II of the essay cited in note 3.

5 Freud, S. *Psychopathology of Everyday Life*. Translated by A. A. Brill. New York (Mentor Books) 1951; p. 64.