Chapter 16

Laws and Their Role in Scientific Explanation

Carl Hempel

Two Basic Requirements for Scientific Explanations

To explain the phenomena of the physical world is one of the primary objectives of the natural sciences. Indeed, almost all of the scientific investigations that served as illustrations in the preceding chapters were aimed not at ascertaining some particular fact but at achieving some explanatory insight; they were concerned with questions such as how puerperal fever is contracted, why the water-lifting capacity of pumps has its characteristic limitation, why the transmission of light conforms to the laws of geometrical optics, and so forth. In this chapter ... we will examine in some detail the character of scientific explanations and the kind of insight they afford.

That man has long and persistently been concerned to achieve some understanding of the enormously diverse, often perplexing, and sometimes threatening occurrences in the world around him is shown by the manifold myths and metaphors he has devised in an effort to account for the very existence of the world and of himself, for life and death, for the motions of the heavenly bodies, for the regular sequence of day and night, for the changing seasons, for thunder and lightning, sunshine and rain. Some of these explanatory ideas are based on anthropomorphic conceptions of the forces of nature, others invoke hidden powers or agents, still others refer to God's inscrutable plans or to fate.

Accounts of this kind undeniably may give the questioner a sense of having attained some understanding; they may resolve his perplexity and in this sense "answer" his question. But however satisfactory these answers may be psychologically, they are not adequate for the purposes of science, which, after all, is concerned to develop a conception of the world that has a clear, logical bearing on our experience and is thus capable of objective test. Scientific explanations must, for this reason, meet two systematic requirements, which will be called the requirement of explanatory relevance and the requirement of testability.

The astronomer Francesco Sizi offered the following argument to show why, contrary to what his contemporary, Galileo, claimed to have seen through his telescope, there could be no satellites circling around Jupiter:

> There are seven windows in the head, two nostrils, two ears, two eyes and a mouth; so in the heavens there are two favorable stars, two unpropitious, two luminaries, and Mercury alone undecided and indifferent. From which and many other similar phenomena of nature such as the seven metals, etc., which it were tedious to enumerate, we gather that the number of planets is necessarily seven.... Moreover, the satellites are invisible to the naked eye and therefore can have no influence on the earth and therefore would be useless and therefore do not exist.[1]

Carl G. Hempel, *Philosophy of Natural Science*, copyright 1966, Chapter 5, pp. 47–69. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

The crucial defect of this argument is evident: the "facts" it adduces, even if accepted without question, are entirely irrelevant to the point at issue; they do not afford the slightest reason for the assumption that Jupiter has no satellites; the claim of relevance suggested by the barrage of words like 'therefore', 'it follows', and 'necessarily' is entirely spurious.

Consider by contrast the physical explanation of a rainbow. It shows that the phenomenon comes about as a result of the reflection and refraction of the white light of the sun in spherical droplets of water such as those that occur in a cloud. By reference to the relevant optical laws, this account shows that the appearance of a rainbow is to be expected whenever a spray or mist of water droplets is illuminated by a strong white light behind the observer. Thus, even if we happened never to have seen a rainbow, the explanatory information provided by the physical account would constitute good grounds for expecting or believing that a rainbow will appear under the specified circumstances. We will refer to this characteristic by saying that the physical explanation meets the *requirement of explanatory relevance*: the explanatory information adduced affords good grounds for believing that the phenomenon to be explained did, or does, indeed occur. This condition must be met if we are to be entitled to say: "That explains it—the phenomenon in question was indeed to be expected under the circumstances!"

The requirement represents a necessary condition for an adequate explanation, but not a sufficient one. For example, a large body of data showing a red-shift in the spectra of distant galaxies provides strong grounds for believing *that* those galaxies recede from our local one at enormous speeds, yet it does not explain *why*.

To introduce the second basic requirement for scientific explanations, let us consider once more the conception of gravitational attraction as manifesting a natural tendency akin to love. As we noted earlier, this conception has no test implications whatever. Hence, no empirical finding could possibly bear it out or disconfirm it. Being thus devoid of empirical content, the conception surely affords no grounds for expecting the characteristic phenomena of gravitational attraction: it lacks objective explanatory power. Similar comments apply to explanations in terms of an inscrutable fate: to invoke such an idea is not to achieve an especially profound insight, but to give up the attempt at explanation altogether. By contrast, the statements on which the physical explanation of a rainbow is based do have various test implications; these concern, for example, the conditions under which a rainbow will be seen in the sky, and the order of the colors in it; the appearance of rainbow phenomena in the spray of a wave breaking on the rocks and in the mist of a lawn sprinkler; and so forth. These examples illustrate a second condition for scientific explanations, which we will call the *requirement of testability*: the statements constituting a scientific explanation must be capable of empirical test.

It has already been suggested that since the conception of gravitation in terms of an underlying universal affinity has no test implications, it can have no explanatory power: it cannot provide grounds for expecting that universal gravitation will occur, nor that gravitational attraction will show such and such characteristic features; for if it did imply such consequences either deductively or even in a weaker, inductive-probabilistic sense, then it would be testable by reference to those consequences. As this example shows, the two requirements just considered are interrelated: a proposed explanation that meets the requirement of relevance also meets the requirement of testability. (The converse clearly does not hold.)

Now let us see what forms scientific explanations take, and how they meet the two basic requirements.

Deductive-Nomological Explanation

Consider once more Périer's finding in the Puy-de-Dôme experiment, that the length of the mercury column in a Torricelli barometer decreased with increasing altitude. Torricelli's and Pascal's ideas on atmospheric pressure provided an explanation for this phenomenon; somewhat pedantically, it can be spelled out as follows:

(*a*) At any location, the pressure that the mercury column in the closed branch of the Torricelli apparatus exerts upon the mercury below equals the pressure exerted on the surface of the mercury in the open vessel by the column of air above it.
(*b*) The pressures exerted by the columns of mercury and of air are proportional to their weights; and the shorter the columns, the smaller their weights.
(*c*) As Périer carried the apparatus to the top of the mountain, the column of air above the open vessel became steadily shorter.
(*d*) (Therefore,) the mercury column in the closed vessel grew steadily shorter during the ascent.

Thus formulated, the explanation is an argument to the effect that the phenomenon to be explained, as described by the sentence (*d*), is just what is to be expected in view of the explanatory facts cited in (*a*), (*b*), and (*c*); and that, indeed, (*d*) follows deductively from the explanatory statements. The latter are of two kinds; (*a*) and (*b*) have the character of general laws expressing uniform empirical connections; whereas (*c*) describes certain particular facts. Thus, the shortening of the mercury column is here explained by showing that it occurred in accordance with certain laws of nature, as a result of certain particular circumstances. The explanation fits the phenomenon to be explained into a pattern of uniformities and shows that its occurrence was to be expected, given the specified laws and the pertinent particular circumstances.

The phenomenon to be accounted for by an explanation will henceforth also be referred to as the *explanandum phenomenon*; the sentence describing it, as the *explanandum sentence*. When the context shows which is meant, either of them will simply be called the explanandum. The sentences specifying the explanatory information—(*a*), (*b*), (*c*) in our example—will be called the *explanans sentences*; jointly they will be said to form the *explanans*.

As a second example, consider the explanation of a characteristic of image formation by reflection in a spherical mirror; namely, that generally $1/u + 1/v = 2/r$, where u and v are the distances of object-point and image-point from the mirror, and r is the mirror's radius of curvature. In geometrical optics, this uniformity is explained with the help of the basic law of reflection in a plane mirror, by treating the reflection of a beam of light at any one point of a spherical mirror as a case of reflection in a plane tangential to the spherical surface. The resulting explanation can be formulated as a deductive argument whose conclusion is the explanandum sentence, and whose premisses include the basic laws of reflection and of rectilinear propagation, as well as the statement that the surface of the mirror forms a segment of a sphere.[2]

A similar argument, whose premisses again include the law for reflection in a plane mirror, offers an explanation of why the light of a small light source placed at the focus of a paraboloidal mirror is reflected in a beam parallel to the axis of the paraboloid (a principle technologically applied in the construction of automobile headlights, searchlights, and other devices).

The explanations just considered may be conceived, then, as deductive arguments whose conclusion is the explanandum sentence, *E*, and whose premiss-set, the explanans, consists of general laws, $L_1, L_2, \ldots, L_r$ and of other statements, $C_1, C_2, \ldots, C_k$, which make assertions about particular facts. The form of such arguments, which thus constitute one type of scientific explanation, can be represented by the following schema:

$$\text{D-N]} \quad \frac{\left.\begin{matrix} L_1, L_2, \ldots, L_r \\ C_1, C_2, \ldots, C_k \end{matrix}\right\} \text{ Explanans sentences}}{E \qquad \text{Explanandum sentence}}$$

Explanatory accounts of this kind will be called explanations by deductive subsumption under general laws, or *deductive-nomological explanation.* (The root of the term 'nomological' is the Greek word 'nomos', for law.) The laws invoked in a scientific explanation will also be called *covering laws* for the explanandum phenomenon, and the explanatory argument will be said to subsume the explanandum under those laws.

The explanandum phenomenon in a deductive-nomological explanation may be an event occurring at a particular place and time, such as the outcome of Périer's experiment. Or it may be some regularity found in nature, such as certain characteristics generally displayed by rainbows; or a uniformity expressed by an empirical law such as Galileo's or Kepler's laws. Deductive explanations of such uniformities will then invoke laws of broader scope, such as the laws of reflection and refraction, or Newton's laws of motion and of gravitation. As this use of Newton's laws illustrates, empirical laws are often explained by means of theoretical principles that refer to structures and processes underlying the uniformities in question. ...

Deductive-nomological explanations satisfy the requirement of explanatory relevance in the strongest possible sense: the explanatory information they provide implies the explanandum sentence deductively and thus offers logically conclusive grounds why the explanandum phenomenon is to be expected. (We will soon encounter other scientific explanations, which fulfill the requirement only in a weaker, inductive, sense.) And the testability requirement is met as well, since the explanans implies among other things that under the specified conditions, the explanandum phenomenon occurs.

Some scientific explanations conform to the pattern (D-N) quite closely. This is so, particularly, when certain quantitative features of a phenomenon are explained by mathematical derivation from covering general laws, as in the case of reflection in spherical and paraboloidal mirrors. Or take the celebrated explanation, propounded by Leverrier (and independently by Adams), of peculiar irregularities in the motion of the planet Uranus, which on the current Newtonian theory could not be accounted for by the gravitational attraction of the other planets then known. Leverrier conjectured that they resulted from the gravitational pull of an as yet undetected outer planet, and he computed the position, mass, and other characteristics which that planet would have to possess to account in quantitative detail for the observed irregularities. His explanation was strikingly confirmed by the discovery, at the predicted location, of a new planet, Neptune, which had the quantitative characteristics attributed to it by Leverrier. Here again, the explanation has the character of a deductive argument whose premisses include general laws—specifically Newton's laws of gravitation and of motion—as well as statements specifying various quantitative particulars about the disturbing planet.

Not infrequently, however, deductive-nomological explanations are stated in an elliptical form: they omit mention of certain assumptions that are presupposed by the explanation but are simply taken for granted in the given context. Such explanations are sometimes expressed in the form '*E* because *C*', where *E* is the event to be explained and *C* is some antecedent or concomitant event or state of affairs. Take, for example, the statement: 'The slush on the sidewalk remained liquid during the frost because it had been sprinkled with salt'. This explanation does not explicitly mention any laws, but it tacitly presupposes at least one: that the freezing point of water is lowered whenever salt is dissolved in it. Indeed, it is precisely by virtue of this law that the sprinkling of salt acquires the explanatory, and specifically causative, role that the elliptical because-statement ascribes to it. That statement, incidentally, is elliptical also in other respects; for example, it tacitly takes for granted, and leaves unmentioned, certain assumptions about the prevailing physical conditions, such as the temperature's not dropping to a very low point. And if nomic and other assumptions thus omitted are added to the statement that salt had been sprinkled on the slush, we obtain the premisses for a deductive-nomological explanation of the fact that the slush remained liquid.

Similar comments apply to Semmelweis's explanation that childbed fever was caused by decomposed animal matter introduced into the bloodstream through open wound surfaces. Thus formulated, the explanation makes no mention of general laws; but it presupposes that such contamination of the bloodstream generally leads to blood poisoning attended by the characteristic symptoms of childbed fever, for this is implied by the assertion that the contamination *causes* puerperal fever. The generalization was no doubt taken for granted by Semmelweis, to whom the cause of Kolletschka's fatal illness presented no etiological problem: given that infectious matter was introduced into the bloodstream, blood poisoning would result. (Kolletschka was by no means the first one to die of blood poisoning resulting from a cut with an infected scalpel. And by a tragic irony, Semmelweis himself was to suffer the same fate.) But once the tacit premise is made explicit, the explanation is seen to involve reference to general laws.

As the preceding examples illustrate, corresponding general laws are always presupposed by an explanatory statement to the effect that a particular event of a certain kind *G* (e.g., expansion of a gas under constant pressure; flow of a current in a wire loop) was *caused* by an event of another kind *F* (e.g., heating of the gas; motion of the loop across a magnetic field). To see this, we need not enter into the complex ramifications of the notion of cause; it suffices to note that the general maxim "Same cause, same effect", when applied to such explanatory statements, yields the implied claim that whenever an event of kind *F* occurs, it is accompanied by an event of kind *G*.

To say that an explanation rests on general laws is not to say that its discovery required the discovery of the laws. The crucial new insight achieved by an explanation will sometimes lie in the discovery of some particular fact (e.g., the presence of an undetected outer planet; infectious matter adhering to the hands of examining physicians) which, by virtue of antecedently accepted general laws, accounts for the explanandum phenomenon. In other cases, such as that of the lines in the hydrogen spectrum, the explanatory achievement does lie in the discovery of a covering law (Balmer's) and eventually of an explanatory theory (such as Bohr's); in yet other cases, the major accomplishment of an explanation may lie in showing that, and exactly how, the explanandum phenomenon can be accounted for by reference to laws and data about particular facts that are already available: this is illustrated by the explanatory derivation of the reflection laws for spherical and paraboloidal mirrors from the basic law of geometrical optics in conjunction with statements about the geometrical characteristics of the mirrors.

An explanatory problem does not by itself determine what kind of discovery is required for its solution. Thus, Leverrier discovered deviations from the theoretically expected course also in the motion of the planet Mercury; and as in the case of Uranus, he tried to explain these as resulting from the gravitational pull of an as yet undetected planet, Vulcan, which would have to be a very dense and very small object between the sun and Mercury. But no such planet was found, and a satisfactory explanation was provided only much later by the general theory of relativity, which accounted for the irregularities not by reference to some disturbing particular factor, but by means of a new system of laws.

Universal Laws and Accidental Generalizations

As we have seen, laws play an essential role in deductive-nomological explanations. They provide the link by reason of which particular circumstances (described by C_1, C_2, …, C_k) can serve to explain the occurrence of a given event. And when the explanandum is not a particular event, but a uniformity such as those represented by characteristics mentioned earlier of spherical and paraboloidal mirrors, the explanatory laws exhibit a system of more comprehensive uniformities, of which the given one is but a special case.

The laws required for deductive-nomological explanations share a basic characteristic: they are, as we shall say, statements of universal form. Broadly speaking, a statement of this kind asserts a uniform connection between different empirical phenomena or between different aspects of an empirical phenomenon. It is a statement to the effect that whenever and wherever conditions of a specified kind *F* occur, then so will, always and without exception, certain conditions of another kind, *G*. (Not all scientific laws are of this type. In the sections that follow, we will encounter laws of probabilistic form, and explanations based on them.)

Here are some examples of statements of universal form: whenever the temperature of a gas increases while its pressure remains constant, its volume increases; whenever a solid is dissolved in a liquid, the boiling point of the liquid is raised; whenever a ray of light is reflected at a plane surface, the angle of reflection equals the angle of incidence; whenever a magnetic iron rod is broken in two, the pieces are magnets again; whenever a body falls freely from rest in a vacuum near the surface of the earth, the distance it covers in t seconds is $16t^2$ feet. Most of the laws of the natural sciences are quantitative: they assert specific mathematical connections between different quantitative characteristics of physical systems (e.g., between volume, temperature, and pressure of a gas), or of processes (e.g., between time and distance in free fall in Galileo's law; between the period of revolution of a planet and its mean distance from the sun, in Kepler's third law; between the angles of incidence and refraction in Snell's law).

Strictly speaking, a statement asserting some uniform connection will be considered a law only if there are reasons to assume it is true: we would not normally speak of false laws of nature. But if this requirement were rigidly observed, then the statements commonly referred to as Galileo's and Kepler's laws would not qualify as laws; for according to current physical knowledge, they hold only approximately; and as we shall see later, physical theory explains why this is so. Analogous remarks apply to the laws of geometrical optics. For example, even in a homogeneous medium, light does not move strictly in straight lines: it can bend around corners. We shall therefore use the word 'law' somewhat liberally, applying the term also to certain statements of the kind here referred to, which, on theoretical grounds, are known to hold only approximately and with certain qualifications. ...

We saw that the laws invoked in deductive-nomological explanations have the basic form: 'In all cases when conditions of kind *F* are realized, conditions of kind *G* are realized as well'. But, interestingly, not all statements of this universal form, even if true, can qualify as laws of nature. For example, the sentence 'All rocks in this box contain iron' is of universal form (*F* is the condition of being a rock in the box, *G* that of containing iron); yet even if true, it would not be regarded as a law, but as an assertion of something that "happens to be the case", as an "accidental generalization". Or consider the statement: 'All bodies consisting of pure gold have a mass of less than 100,000 kilograms'. No doubt all bodies of gold ever examined by man conform to it; thus, there is considerable confirmatory evidence for it and no disconfirming instances are known. Indeed, it is quite possible that never in the history of the universe has there been or will there be a body of pure gold with a mass of 100,000 kilograms or more. In this case, the proposed generalization would not only be well confirmed, but true. And yet, we would presumably regard its truth as accidental, on the ground that nothing in the basic laws of nature as conceived in contemporary science precludes the possibility of there being—or even the possibility of our producing—a solid gold object with a mass exceeding 100,000 kilograms.

Thus, a scientific law cannot be adequately defined as a true statement of universal form: this characterization expresses a necessary, but not a sufficient, condition for laws of the kind here under discussion.

What distinguishes genuine laws from accidental generalizations? This intriguing problem has been intensively discussed in recent years. Let us look briefly at some of the principal ideas that have emerged from the debate, which is still continuing.

One telling and suggestive difference, noted by Nelson Goodman,[3] is this: a law can, whereas an accidental generalization cannot, serve to support *counterfactual conditionals*, i.e., statements of the form 'If *A* were (had been) the case, then *B* would be (would have been) the case', where in fact *A* is not (has not been) the case. Thus, the assertion 'If this paraffin candle had been put into a kettle of boiling water, it would have melted' could be supported by adducing the law that paraffin is liquid above 60 degrees centigrade (and the fact that the boiling point of water is 100 degrees centigrade). But the statement 'All rocks in this box contain iron' could not be used similarly to support the counterfactual statement 'If this pebble had been put into the box, it would contain iron'. Similarly, a law, in contrast to an accidentally true generalization, can support *subjunctive conditionals*, i.e., sentences of the type 'If *A* should come to pass, then so would *B*', where it is left open whether or not *A* will in fact come to pass. The statement 'If this paraffin candle should be put into boiling water then it would melt' is an example.

Closely related to this difference is another one, which is of special interest to us: a law can, whereas an accidental generalization cannot, serve as a basis for an explanation. Thus, the melting of a particular paraffin candle that was put into boiling water can be explained, in conformity with the schema (D-N), by reference to the particular facts just mentioned and to the law that paraffin melts when its temperature is raised above 60 degrees centigrade. But the fact that a particular rock in the box contains iron cannot be analogously explained by reference to the general statement that all rocks in the box contain iron.

It might seem plausible to say, by way of a further distinction, that the latter statement simply serves as a conveniently brief formulation of a finite conjunction of this kind: 'Rock r_1 contains iron, and rock r_2 contains iron, ..., and rock r_{63} contains iron'; whereas the generalization about paraffin refers to a potentially infinite set of particular cases and therefore cannot be paraphrased by a finite conjunction of statements describing individual instances. This distinction is suggestive, but it is overstated. For to begin with, the generalization 'All rocks in this box contain iron' does not in fact tell us how many rocks there are in the box, nor does it name any particular rocks r_1, r_2, etc. Hence, the general sentence is not logically equivalent to a finite conjunction of the kind just mentioned. To formulate a suitable conjunction, we need additional information, which might be obtained by counting and labeling the rocks in the box. Besides, our generalization 'All bodies of pure gold have a mass of less than 100,000 kilograms' would not count as a law even if there were infinitely many bodies of gold in the world. Thus, the criterion we have under consideration fails on several grounds.

Finally, let us note that a statement of universal form may qualify as a law even if it actually has no instances whatever. As an example, consider the sentence: 'On any celestial body that has the same radius as the earth but twice its mass, free fall from rest conforms to the formula $s = 32t^2$'. There might well be no celestial object in the entire universe that has the specified size and mass, and yet the statement has the character of a law. For it (or rather, a close approximation of it, as in the case of Galileo's law) follows from the Newtonian theory of gravitation and of motion in conjunction with the statement that the acceleration of free fall on the earth is 32 feet per second per second; thus, it has strong theoretical support, just like our earlier law for free fall on the moon.

A law, we noted, can support subjunctive and counterfactual conditional statements about potential instances, i.e., about particular cases that might occur, or that might have occurred but did not. In similar fashion, Newton's theory supports our general statement in a subjunctive version that suggests its lawlike status, namely: 'On any celestial body that there may be which has the same size as the earth but twice its mass, free fall would conform to the formula $s = 32t^2$'. By contrast, the generalization about the rocks cannot be paraphrased as asserting that any rock that might be in this box would contain iron, nor of course would this latter claim have any theoretical support.

Similarly, we would not use our generalization about the mass of gold bodies—let us call it H—to support statements such as this: 'Two bodies of pure gold whose individual masses add up to more than 100,000 kilograms cannot be fused to form one body; or if fusion should be possible, then the mass of the resulting body will be less than 100,000 kg', for the basic physical and chemical theories of matter that are currently accepted do not preclude the kind of fusion here considered, and they do not imply that there would be a mass loss of the sort here referred to. Hence, even if the generalization H should be true, i.e., if no exceptions to it should ever occur, this would constitute a mere accident or coincidence as judged by current theory, which permits the occurrence of exceptions to H.

Thus, whether a statement of universal form counts as a law will depend in part upon the scientific theories accepted at the time. This is not to say that "empirical generalizations"—statements of universal form that are empirically well confirmed but have no basis in theory—never qualify as laws: Galileo's, Kepler's, and Boyle's laws, for example, were accepted as such before they received theoretical grounding. The relevance of theory is rather this: a statement of universal form, whether empirically confirmed or as yet untested, will qualify as a law if it is implied by an accepted theory (statements of this kind are often referred to as theoretical laws); but even if it is empirically well confirmed and presumably true in fact, it will not qualify as a law if it rules out certain hypothetical occurrences (such as the fusion of two gold bodies with a resulting mass of more than 100,000 kilograms, in the case of our generalization *H*) which an accepted theory qualifies as possible.[4]

Probabilistic Explanation: Fundamentals

Not all scientific explanations are based on laws of strictly universal form. Thus, little Jim's getting the measles might be explained by saying that he caught the disease from his brother, who had a bad case of the measles some days earlier. This account again links the explanandum event to an earlier occurrence, Jim's exposure to the measles; the latter is said to provide an explanation because there is a connection between exposure to the measles and contracting the disease. That connection cannot be expressed by a law of universal form, however; for not every case of exposure to the measles produces contagion. What can be claimed is only that persons exposed to the measles will contract the disease with high probability, i.e., in a high percentage of all cases. General statements of this type, which we shall soon examine more closely, will be called *laws of probabilistic form* or *probabilistic laws*, for short.

In our illustration, then, the explanans consists of the probabilistic law just mentioned and the statement that Jim was exposed to the measles. In contrast to the case of deductive-nomological explanation, these explanans statements do not deductively imply the explanandum statement that Jim got the measles; for in deductive inferences from true premises, the conclusion is invariably true, whereas in our example, it is clearly possible that the explanans statements might be true and yet the explanandum statement false. We will say, for short, that the explanans implies the explanandum, not with "deductive certainty", but only with near-certainty or with high probability.

The resulting explanatory argument may be schematized as follows:

The probability for persons exposed to the measles
to catch the disease is high.

Jim was exposed to the measles.
══════════════════ [makes highly probable]
Jim caught the measles.

In the customary presentation of a deductive argument, which was used, for example, in the schema (D-N) above, the conclusion is separated from the premises by a single line, which serves to indicate that the premises logically imply the conclusion. The double line used in our latest schema is meant to indicate analogously that the "premises" (the explanans) make the "conclusion" (the explanandum sentence) more or less probable; the degree of probability is suggested by the notation in brackets.

Arguments of this kind will be called *probabilistic explanations*. As our discussion shows, a probabilistic explanation of a particular event shares certain basic characteristics with the corresponding deductive-nomological type of explanation. In both cases, the given event is explained by reference to others, with which the explanandum event is connected by laws. But in one case, the laws are of universal form; in the other, of probabilistic form. And while a deductive explanation shows that, on the information contained in the explanans, the explanandum was to be expected with "deductive certainty", an inductive explanation shows only that, on the information contained in the explanans, the explanandum was to be expected with high probability, and perhaps with "practical certainty"; it is in this manner that the latter argument meets the requirement of explanatory relevance.

Statistical Probabilities and Probabilistic Laws

We must now consider more closely the two differentiating features of probabilistic explanation that have just been noted: the probabilistic laws they invoke and the peculiar kind of probabilistic implication that connects the explanans with the explanandum.

Suppose that from an urn containing many balls of the same size and mass, but not necessarily of the same color, successive drawings are made. At each drawing, one ball is removed, and its color is noted. Then the ball is returned to the urn, whose contents are thoroughly mixed before the next drawing takes place. This is an example of a so-called random process or random experiment, a concept that will soon be characterized in more detail. Let us refer to the procedure just described as experiment *U*, to each drawing as one performance of *U*, and to the color of the ball produced by a given drawing as the result, or the outcome, of that performance.

If all the balls in an urn are white, then a statement of strictly universal form holds true of the results produced by the performance of *U*: every drawing from the urn yields a white ball, or yields the result *W*, for short. If only some of the balls—say, 600 of them—are white, whereas the others—say 400—are red, then a general statement of probabilistic form holds true of the experiment: the probability for a performance of *U* to produce a white ball, or outcome *W*, is .6; in symbols:

$$P(W, U) = .6. \tag{5a}$$

Similarly, the probability of obtaining heads as a result of the random experiment *C* of flipping a fair coin is given by

$$P(H, C) = .5, \tag{5b}$$

and the probability of obtaining an ace as a result of the random experiment *D* of rolling a regular die is

$$P(A, D) = 1/6. \tag{5c}$$

What do such probability statements mean? According to one familiar view, sometimes called the "classical" conception of probability, the statement (*5a*) would have to be interpreted as follows: each performance of the experiment *U* effects a choice of one from among 1,000 basic possibilities, or basic alternatives, each represented by one of the balls in the urn; of these possible choices, 600 are "favorable" to the outcome *W*; and the probability of drawing a white ball is simply the ratio of the number of favorable choices available to the number of all possible choices, i.e., 600/1,000. The classical interpretation of the probability statements (*5b*) and (*5c*) follows similar lines.

Yet this characterization is inadequate, for if before each drawing, the 400 red balls in the urn were placed on top of the white ones, then in this new kind of urn experiment—let us call it *U'*—the ratio of favorable to possible basic alternatives would remain the same, but the probability of drawing a white ball would be smaller than in the experiment *U*, in which the balls are thoroughly mixed before each drawing. The classical conception takes account of this difficulty by requiring that the basic alternatives referred to in its definition of probability must be "equipossible" or "equiprobable"—a requirement presumably violated in the case of experiment *U'*.

This added proviso raises the question of how to define equipossibility or equiprobability. We will pass over this notoriously troublesome and controversial issue, because—even assuming that equiprobability can be satisfactorily characterized—the classical conception would still be inadequate, since probabilities are assigned also to the outcomes of random experiments for which no plausible way is known of marking off equiprobable basic alternatives. Thus, for the random experiment D of rolling a regular die, the six faces might be regarded as representing such equiprobable alternatives; but we attribute probabilities to such results as rolling an ace, or an odd number of points, etc., also in the case of a loaded die, even though no equiprobable basic outcomes can be marked off here.

Similarly—and this is particularly important—science assigns probabilities to the outcomes of certain random experiments or random processes encountered in nature, such as the step-by-step decay of the atoms of radioactive substances, or the transition of atoms from one energy state to another. Here again, we find no equiprobable basic alternatives in terms of which such probabilities might be classically defined and computed.

To arrive at a more satisfactory construal of our probability statements, let us consider how one would ascertain the probability of the rolling of an ace with a given die that is not known to be regular. This would obviously be done by making a large number of throws with the die and ascertaining the *relative frequency*, i.e., the proportion, of those cases in which an ace turns up. If, for example, the experiment D' of rolling the given die is performed 300 times and an ace turns up in 62 cases, then the relative frequency, 62/300, would be regarded as an approximate value of the probability $p(A, D')$ of rolling an ace with the given die. Analogous procedures would be used to estimate the probabilities associated with the flipping of a given coin, the spinning of a roulette wheel, and so on. Similarly, the probabilities associated with radioactive decay, with the transitions between different atomic energy states, with genetic processes, etc., are determined by ascertaining the corresponding relative frequencies; however, this is often done in highly indirect ways rather than by simply counting individual atomic or other events of the relevant kinds.

The interpretation in terms of relative frequencies applies also to probability statements such as (*5b*) and (*5c*), which concern the results of flipping a fair (i.e., homogeneous and strictly cylindrical) coin or tossing a regular (homogeneous and strictly cubical) die: what the scientist (or the gambler, for that matter) is concerned with in making a probability statement is the relative frequency with which a certain outcome O can be expected in long series of repetitions of some random experiment R. The counting of "equiprobable" basic alternatives and of those among them which are "favorable" to O may be regarded as a heuristic device for guessing at the relative frequency of O. And indeed when a regular die or a fair coin is tossed a large number of times, the different faces tend to come up with equal frequency. One might expect this on the basis of symmetry considerations of the kind frequently used in forming physical hypotheses, for our empirical knowledge affords no grounds on which to expect any of the faces to be favored over any other. But while such considerations often are heuristically useful, they must not be regarded as certain or as self-evident truths: some very plausible symmetry assumptions, such as the principle of parity, have been found not to be generally satisfied at the subatomic level. Assumptions about equiprobabilities are therefore always subject to correction in the light of empirical data concerning the actual relative frequencies of the phenomena in question. This point is illustrated also by the statistical theories of gases developed by Bose and Einstein and by Fermi and Dirac, respectively, which rest on different assumptions concerning what distributions of particles over a phase space are equiprobable.

The probabilities specified in the probabilistic laws, then, represent relative frequencies. They cannot, however, be strictly defined as relative frequencies in long series of repetitions of the relevant random experiment. For the proportion, say, of aces obtained in throwing a given die will change, if perhaps only slightly, as the series of throws is extended; and even in two series of exactly the same length, the number of aces will usually differ. We do find, however, that as the number of throws increases, the relative frequency of each of the different outcomes tends to change less and less, even though the results of successive throws continue to vary in an irregular and practically unpredictable fashion. This is what generally characterizes a random experiment R with outcomes $O_1, O_2, \ldots, O_n$: successive performances of R yield one or another of those outcomes in an irregular manner; but the relative frequencies of the outcomes tend to become stable as the number of performances increases. And the probabilities of the outcomes, $p(O_1, R), p(O_2, R), \ldots, p(O_n, R)$, may be regarded as ideal values that the actual frequencies tend to assume as they become increasingly stable. For mathematical convenience, the probabilities are sometimes defined as the mathematical *limits* toward which the relative frequencies converge as the number of performances increases indefinitely. But this definition has certain conceptual shortcomings, and in some more recent mathematical studies of the subject, the intended empirical meaning of the concept of probability is deliberately, and for good reasons, characterized more vaguely by means of the following so-called *statistical interpretation of probability*:[5]

The statement

$$p(O, R) = r$$

means that in a long series of performances of random experiment R, the proportion of cases with outcome O is almost certain to be close to r.

The concept of *statistical probability* thus characterized must be carefully distinguished from the concept of *inductive or logical probability*. Logical probability is a quantitative logical relation between definite *statements*; the sentence

$$c(H, K) = r$$

asserts that the hypothesis H is supported, or made probable, to degree r by the evidence formulated in statement K. Statistical probability is a quantitative relation between repeatable *kinds of events*: a certain kind of outcome, O, and a certain kind of random process, R; it represents, roughly speaking, the relative frequency with which the result O tends to occur in a long series of performances of R.

What the two concepts have in common are their mathematical characteristics: both satisfy the basic principles of mathematical probability theory:

a) The possible numerical values of both probabilities range from 0 to 1:

$$0 \leq p(O, R) \leq 1,$$

$$0 \leq c(H, K) \leq 1.$$

b) The probability for one of two mutually exclusive outcomes of R to occur is the sum of the probabilities of the outcomes taken separately; the probability, on any evidence K, for one or the other of two mutually exclusive hypotheses to hold is the sum of their respective probabilities:

If O_1, O_2 are mutually exclusive, then

$$p(O_1 \text{ or } O_2, R) = p(O_1, R) + p(O_2, R).$$

If H_1, H_2 are logically exclusive hypotheses, then

$$c(H_1 \text{ or } H_2, K) = c(H_1, K) + c(H_2, K).$$

c) The probability of an outcome that necessarily occurs in all cases—such as O or not O—is 1; the probability, on any evidence, of a hypothesis that is logically (and in this sense necessarily) true, such as H or not H, is 1:

$$p(O \text{ or not } O, R) = 1,$$

$$c(H \text{ or not } H, K) = 1.$$

Scientific hypotheses in the form of statistical probability statements can be, and are, tested by examining the long-run relative frequencies of the outcomes concerned; and the confirmation of such hypotheses is then judged, broadly speaking, in terms of the closeness of the agreement between hypothetical probabilities and observed frequencies. The logic of such tests, however, presents some intriguing special problems, which call for at least brief examination.

Consider the hypothesis, H, that the probability of rolling an ace with a certain die is .15; or briefly, that $p(A, D) = .15$, where D is the random experiment of rolling the given die. The hypothesis H does not deductively imply any test implications specifying how many aces will occur in a finite series of throws of the die. It does not imply, for example, that exactly 75 among the first 500 throws will yield an ace, nor even that the number of aces will lie between 50 and 100, say. Hence, if the proportion of aces actually obtained in a large number of throws differs considerably from .15, this does not refute H in the sense in which a hypothesis of strictly universal form, such as 'All swans are white', can be refuted, in virtue of the *modus tollens* argument, by reference to one counter-instance, such as a black swan. Similarly, if a long run of throws of the given die yields a proportion of aces very close to .15, this does not confirm H in the sense in which a hypothesis is confirmed by the finding that a test sentence I that it logically implies is in fact true. For in this latter case, the hypothesis asserts I by logical implication, and the test result is thus confirmatory in the sense of showing that a certain part of what the hypothesis asserts is indeed true; but nothing strictly analogous is shown for H by confirmatory frequency data; for H does *not* assert by implication that the frequency of aces in some long run will definitely be very close to .15.

But while H does not logically preclude the possibility that the proportion of aces obtained in a long series of throws of the given die may depart widely from .15, it does logically imply that such departures are highly improbable in the statistical sense; i.e., that if the experiment of performing a long series of throws (say, 1,000 of them per series) is repeated a large number of times, then only a tiny proportion of those long series will yield a proportion of aces that differs considerably from .15. For the case of rolling a die, it is usually assumed that the results of successive throws are "statistically independent"; this means roughly that the probability of obtaining an ace in a throw of the die does not depend on the result of the preceding throw. Mathematical analysis shows that in conjunction with this independence assumption, our hypothesis H deductively determines the statistical probability for the proportion of aces obtained in n throws to differ from .15 by no more than a specified amount. For example, H implies that for a series of 1,000 throws of the die here considered, the probability is about .976 that the proportion of aces will lie between .125 and .175; and similarly, that for a run of 10,000 throws the probability is about .995 that the proportion of aces will be between .14 and .16. Thus, we may say that if H is true, then it is practically certain that in a long trial run the observed proportion of aces will differ by very little from the hypothetical probability value .15. Hence, if the observed long-run frequency of an outcome is not close to the probability assigned to it by a given probabilistic hypothesis, then that hypothesis is very likely to be false. In this case, the frequency data count as disconfirming the hypothesis, or as reducing its credibility; and if sufficiently strong disconfirming evidence is found, the hypothesis will be considered as practically, though not logically, refuted and will accordingly be rejected. Similarly, close agreement between hypothetical probabilities and observed frequencies will tend to confirm a probabilistic hypothesis and may lead to its acceptance.

If probabilistic hypotheses are to be accepted or rejected on the basis of statistical evidence concerning observed frequencies, then appropriate standards are called for. These will have to determine (*a*) what deviations of observed frequencies from the probability stated by a hypothesis are to count as grounds for rejecting the hypothesis, and (*b*) how close an agreement between observed frequencies and hypothetical probability is to be required as a condition for accepting the hypothesis. The requirements in question can be made more or less strict, and their specification is a matter of choice. The stringency of the chosen standards will normally vary with the context and the objectives of the research in question. Broadly speaking, it will depend on the importance that is attached, in the given context, to avoiding two kinds of error that might be made: rejecting the hypothesis under test although it is true, and accepting it although it is false. The importance of this point is particularly clear when acceptance or rejection of the hypothesis is to serve as a basis for practical action. Thus, if the hypothesis concerns the probable effectiveness and safety of a new vaccine, then the decision about its acceptance will have to take into account not only how well the statistical test results accord with the probabilities specified by the hypothesis, but also how serious would be the consequences of accepting the hypothesis and acting on it (e.g., by inoculating children with the vaccine) when in fact it is false, and of rejecting the hypothesis and acting accordingly (e.g., by destroying the vaccine and modifying or discontinuing the process of manufacture) when in fact the hypothesis is true. The complex problems that arise in this context form the subject matter of the theory of statistical tests and decisions, which has been developed in recent decades on the basis of the mathematical theory of probability and statistics[6]

Many important laws and theoretical principles in the natural sciences are of probabilistic character, though they are often of more complicated form than the simple probability statements we have discussed. For example, according to current physical theory, radioactive decay is a random phenomenon in which the atoms of each radioactive element possess a characteristic probability of disintegrating during a specified period of time. The corresponding probabilistic laws are usually formulated as statements giving the "half-life" of the element concerned. Thus, the statements that the half-life of radium^{226} is 1,620 years and that of polonium^{218} is 3.05 minutes are laws to the effect that the probability for a radium^{226} atom to decay within 1,620 years, and for an atom of polonium^{218} to decay within 3.05 minutes, are both one-half. According to the statistical interpretation cited earlier, these laws imply that of a large number of radium^{226} atoms or of polonium^{218} atoms given at a certain time, very close to one-half will still exist 1,620 years, or 3.05 minutes, later; the others having disintegrated by radioactive decay.

Again, in the kinetic theory various uniformities in the behavior of gases, including the laws of classical thermodynamics, are explained by means of certain assumptions about the constituent molecules; and some of these are probabilistic hypotheses concerning statistical regularities in the motions and collisions of those molecules.

A few additional remarks concerning the notion of a probabilistic law are indicated. It might seem that all scientific laws should be qualified as probabilistic since the supporting evidence we have for them is always a finite and logically inconclusive body of findings, which can confer upon them only a more or less high probability. But this argument misses the point that the distinction between laws of universal form and laws of probabilistic form does not refer to the strength of the evidential support for the two kinds of statements, but to their form, which reflects the logical character of the claim they make. A law of universal form is basically a statement to the effect that in *all* cases where conditions of kind F are realized, conditions of kind G are realized as well; a law of probabilistic form asserts, basically, that under certain conditions, constituting the performance of a random experiment R, a certain kind of outcome will occur in a specified percentage of cases. No matter whether true or false, well supported or poorly supported, these two types of claims are of a logically different character, and it is on this difference that our distinction is based.

As we saw earlier, a law of the universal form 'Whenever F then G' is by no means a brief, telescoped equivalent of a report stating for each occurrence of F so far examined that it was associated with an occurrence of G. Rather, it implies assertions also for all unexamined cases of F, past as well as present and future; also, it implies counterfactual and hypothetical conditionals which concern, so to speak "possible occurrences" of F: and it is just this characteristic that gives such laws their explanatory power. Laws of probabilistic form have an analogous status. The law stating that the radioactive decay of radium^{226} is a random process with an associated half-life of 1,620 years is plainly not tantamount to a report about decay rates that have been observed in certain samples of radium^{226}. It concerns the decaying process of any body of radium^{226}—past, present, or future; and it implies subjunctive and counterfactual conditionals, such as: if two particular lumps of radium^{226} were to be combined into one, the decay rates would remain the same as if the lumps had remained separate. Again, it is this characteristic that gives probabilistic laws their predictive and their explanatory force.

The Inductive Character of Probabilistic Explanation

One of the simplest kinds of probabilistic explanation is illustrated by our earlier example of Jim's catching the measles. The general form of that explanatory argument may be stated thus:

$p(O, R)$ is close to 1
i is a case of R
════════ [makes highly probable]
i is a case of O

Now the high probability which, as indicated in brackets, the explanans confers upon the explanandum is surely not a statistical probability, for it characterizes a relation between sentences, not between (kinds of) events. We might say that the probability in question represents the rational credibility of the explanandum, given the information provided by the explanans; and as we noted earlier, in so far as this notion can be construed as a probability, it represents a logical or inductive probability.

In some simple cases, there is a natural and obvious way of expressing that probability in numerical terms. In an argument of the kind just considered, if the numerical value of $p(O, R)$ is specified, then it is reasonable to say that the inductive probability that the explanans confers upon the explanandum has the same numerical value. The resulting probabilistic explanation has the form:

$$\begin{array}{l} p(O,R) = r \\ i \text{ is a case of } R \\ \hline\hline i \text{ is a case of } O \end{array} \quad [r]$$

If the explanans is more complex, the determination of corresponding inductive probabilities for the explanandum raises difficult problems, which in part are still unsettled. But whether or not it is possible to assign definite numerical probabilities to all such explanations, the preceding considerations show that when an event is explained by reference to probabilistic laws, the explanans confers upon the explanandum only more or less strong inductive support. Thus, we may distinguish deductive-nomological from probabilistic explanations by saying that the former effect a deductive subsumption under laws of universal form, the latter an inductive subsumption under laws of probabilistic form.

It is sometimes said that precisely because of its inductive character, a probabilistic account does not explain the occurrence of an event, since the explanans does not logically preclude its nonoccurrence. But the important, steadily expanding role that probabilistic laws and theories play in science and its applications, makes it preferable to view accounts based on such principles as affording explanations as well, though of a less stringent kind than those of deductive-nomological form. Take, for example, the radioactive decay of a sample of one milligram of polonium^{218}. Suppose that what is left of this initial amount after 3.05 minutes is found to have a mass that falls within the interval from .499 to .501 milligrams. This finding can be explained by the probabilistic law of decay for polonium^{218}; for that law, in combination with the principles of mathematical probability, deductively implies that given the huge number of atoms in a milligram of polonium^{218}, the probability of the specified outcome is overwhelmingly large, so that in a particular case its occurrence may be expected with "practical certainty".

Or consider the explanation offered by the kinetic theory of gases for an empirically established generalization called Graham's law of diffusion. The law states that at fixed temperature and pressure, the rates at which different gases in a container escape, or diffuse, through a thin porous wall are inversely proportional to the square roots of their molecular weights; so that the amount of a gas that diffuses through the wall per second will be the greater, the lighter its molecules. The explanation rests on the consideration that the mass of a given gas that diffuses through the wall per second will be proportional to the average velocity of its molecules, and that Graham's law will therefore have been explained if it can be shown that the average molecular velocities of different pure gases are inversely proportional to the square roots of their molecular weights. To show this, the theory makes certain assumptions broadly to the effect that a gas consists of a very large number of molecules moving in random fashion at different speeds that frequently change as a result of collisions, and that this random behavior shows certain probabilistic uniformities—in particular, that among the molecules of a given gas at specified temperature and pressure, different velocities will occur with definite, and different, probabilities. These assumptions make it possible to compute the probabilistically expected values—or, as we might briefly say, the "most probable" values—that the average velocities of different gases will possess at equal temperatures and pressures. These most probable average values, the theory shows, are indeed inversely proportional to the square roots of the molecular weights of the gases. But the actual diffusion rates, which are measured experimentally and are the subject of Graham's law, will depend on the actual values that the average velocities have in the large but finite swarms of molecules constituting the given bodies of gas. And the actual average values are related to the corresponding probabilistically estimated, or "most probable", values in a manner that is basically analogous to the relation between the proportion of aces occurring in a large but finite series of tossings of a given die and the corresponding probability of rolling an ace with that die. From the theoretically derived conclusion concerning the probabilistic estimates, it follows only that in view of the very large number of molecules involved, it is overwhelmingly *probable* that at any given time the actual average speeds will have values very close to their probability estimates and that, therefore, it is *practically certain* that they will be, like the latter, inversely proportional to the square roots of their molecular masses, thus satisfying Graham's law.[7]

It seems reasonable to say that this account affords an explanation, even though "only" with very high associated probability, of why gases display the uniformity expressed by Graham's law; and in physical texts and treatises, theoretical accounts of this probabilistic kind are indeed very widely referred to as explanations.

Notes

1. From Holton and Roller, *Foundations of Modern Physical Science,* p. 160.
2. The derivation of the laws of reflection for the curved surfaces referred to in this example and in the next one is simply and lucidly set forth in Chap. 17 of Morris Kline, *Mathematics and the Physical World* (New York: Thomas Y. Crowell Company, 1959).
3. In his essay, "The Problem of Counterfactual Conditionals," reprinted as the first chapter of his book, *Fact, Fiction, and Forecast,* 2nd ed. (Indianapolis: The Bobbs-Merrill Co., Inc., 1965). This work raises fascinating basic problems concerning laws, counterfactual statements, and inductive reasoning, and examines them from an advanced analytic point of view.
4. For a fuller analysis of the concept of law, and for further bibliographic references, see E. Nagel, *The Structure of Science* (New York: Harcourt, Brace & World, Inc., 1961), Chap. 4.
5. Further details on the concept of statistical probability and on the limit-definition and its shortcomings will be found in E. Nagel's monograph, *Principles of the Theory of Probability* (Chicago: University of Chicago Press, 1939). Our version of the statistical interpretation follows that given by H. Cramér on pp. 148–49 of his book, *Mathematical Methods of Statistics* (Princeton: Princeton University Press, 1946).
6. On this subject, see R. D. Luce and H. Raiffa, *Games and Decisions* (New York: John Wiley & Sons, Inc., 1957).
7. The "average" velocities here referred to are technically defined as root-mean-square velocities. Their values do not differ very much from those of average velocities in the usual sense of the arithmetic mean. A succinct outline of the theoretical explanation of Graham's law can be found in Chap. 25 of Holton and Roller, *Foundations of Modern Physical Science.* The distinction, not explicitly mentioned in that presentation, between the average value of a quantity for some finite number of cases and the probabilistically estimated or expected value of that quantity is briefly discussed in Chap. 6 (especially section 4) of R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Reading, Mass.: Addison-Wesley Publishing Co., 1963).