PART I

What you always wanted to know about the philosophy of science but were afraid to ask

1

From the received view to the views of Popper

The received view

Anyone consulting some current textbooks in the philosophy of science will soon discover that the philosophy of science is a very strange subject: it is not, as might be expected, a study of the psychological and sociological factors that promote and encourage the discovery of scientific hypotheses; it is not an examination of the philosophical views of the world that are implicit in leading scientific theories; it is not even a reflection on the principles, methods, and results of the physical and social sciences, describing at the highest level of generality the pinnacles of scientific achievement. Instead, it appears to consist largely of a purely logical analysis of the formal structure of scientific theories, which seem to be more concerned with prescribing good scientific practice than with describing what it is that has actually passed as science; and when it mentions the history of science at all, it is written as if classical physics were the prototype science to which all other disciplines must sooner or later conform if they are to justify the title of "science."

This characterization of the philosophy of science is now somewhat out of date, reflecting as it does the heyday of logical positivism in the interwar years. Between the 1920s and 1950s, philosophers of science did more or less agree with what Frederick Suppe (1974) has called "The Received View on Theories." But the works of Popper, Polanyi, Hanson, Toulmin, Kuhn, Lakatos, and Feyerabend, to mention only the leading names, have largely destroyed this *received view* without, however, putting any generally accepted alternative conception in its place. In short, the philosophy of science has been in something of a turmoil ever since the 1960s, which complicates the task of providing a simple guide to the subject in the space of two chapters. On balance, there is much to be said for beginning with some principal features of the received view and only then moving on to the new heterodoxy, using the work of Karl Popper as a watershed between the old and the new views of the philosophy of science.

The hypothetico-deductive model

The standard view of science in the middle of the nineteenth century was that scientific investigations begin in the free and unprejudiced observation of facts, proceed by inductive inference to the formulation of universal laws about these facts, and finally arrive by further induction at statements of still wider generality known as theories; both laws and theories are ultimately checked for their truth content by comparing their empirical consequences with all the observed facts, including those with which they began. This inductive view of science, perfectly summed up in John Stuart Mill's *System of Logic, Ratiocinative and Inductive* (1843) and remaining to this day the conception of science of the man-in-the-street, gradually began to break down in the last half of the nineteenth century under the influence of the writings of Ernst Mach, Henri Poincaré, and Pierre Duhem, and to be almost entirely reversed by *the hypothetico-deductive model of scientific explanation* that emerged after the turn of the century in the work of the Vienna Circle and the American pragmatists (see Alexander, 1964; Harré, 1967; and Losee, 1972, chaps. 10, 11).

Nevertheless, it was not until 1948 that the hypothetico-deductive model was written down in formal terms as the only valid type of explanation in science. This authorized version first appeared in a now famous paper by Carl Hempel and Peter Oppenheim (1965),[1] which argued that all truly scientific explanations have a common logical structure: they involve at least one universal law plus a statement of relevant initial or boundary conditions that together constitute the *explanans* or premises from which an *explanandum,* a statement about some event whose explanation we are seeking, is deduced with the aid of the rules of deductive logic. By a *universal law,* we mean some such proposition as "in all cases where events *A* occur, events *B* also occur," and such universal laws may be deterministic in form by referring to individual events *B* or statistical in form by referring to classes of events *B;* (thus, statistical laws take the form: "in all cases where events *A* occur, events *B* also occur with a probability of *p,* where $0<p<1$"). By the rules of deductive logic, we mean some sort of infallible syllogistic reasoning like "if *A*

[1] This was a more guarded version of the same thesis announced by Hempel (1942), which generated a great debate among historians about the meaning of historical explanations (see footnote 5). Earlier, less formally precise statements of the hypothetico-deductive model can be found in Popper's *The Logic of Scientific Discovery,* first published in German in 1934 and then in English in 1959 (1959, pp. 59, 68–9; also Popper, 1962, II, pp. 262–3, 362–4; Popper, 1976, p. 117), and indeed as early as 1843 in Mill (1973, 7, pp. 471–2).

is true, then *B* is true *A* is true; therefore *B* is true'' (this is an example of what logicians call a *hypothetical syllogism*). It need hardly be added that deductive logic is an abstract calculus and that the logical validity of deductive reasoning in no way depends on the material truth of either the major premise ''if *A* is true, then *B* is true,'' or the minor premise, ''*A* is true.''

It follows from the common logical structure of all truly scientific explanations, Hempel and Oppenheim went on to argue, that the operation called *explanation* involves the same rules of logical inference as the operation called *prediction*, the only difference being that explanations come after events and predictions before events. In the case of explanation, we start with an event to be explained and find at least one universal law plus a set of initial conditions that logically imply the statement of the event in question. In other words, to cite a particular cause as an explanation of an event is simply to subsume the event in question under some universal law or set of laws; for that reason one critic of the Hempel–Oppenheim thesis has called it ''the covering law model of explanation'' (Dray, 1957, chap. 1). In the case of prediction, on the other hand, we start with a universal law plus a set of initial conditions, and from them we deduce a statement about an unknown event; the prediction is typically used to see whether the universal law is in fact upheld. In short, explanation is simply ''prediction written backwards.''

This notion that there is a perfect, logical symmetry between the nature of explanation and the nature of prediction has been labeled the *symmetry thesis*. It constitutes the heart of the hypothetico-deductive or covering-law model of scientific explanation. The point of the model is that it employs no other rules of logical inference than that of deduction (the force of that remark will become clear in a moment). The universal laws that are involved in explanations are not derived by inductive generalization from individual instances; they are merely hypotheses, inspired conjectures if you like, that may be tested by using them to make predictions about particular events but which are not themselves reducible to observations about events.

The symmetry thesis

The covering-law model of scientific explanation has been attacked from a number of standpoints, and even Hempel himself, its most vigorous proponent, has retreated somewhat over the years in response to these attacks (Suppe, 1974, p. 28n). Most of the critics have seized on the symmetry thesis as the butt of all their objections. It has been argued that prediction need not imply explanation and even that explanation need not imply prediction. The former proposition, at any rate, is plain sailing: prediction only requires a correlation, whereas explanation cries out for something more. Thus, any linear extrapolation of an ordinary least squares regression is a prediction of sorts, and yet the regression itself may be based on no theory whatsoever of the relationship between the relevant variables, much less a notion of which are causes and which are effects. No economist needs to be told that accurate short-term economic forecasting, like accurate short-term weather forecasting, is perfectly possible with the aid of rules-of-thumb that yield satisfactory results, although we may have no idea why they do so. In short, it is only too obvious that it is perfectly possible to predict well without explaining anything.

This is not to say, however, that it is always easy to decide whether a particular scientific theory with an impressive predictive record achieves its results by fluke or by design. Some critics of the received view have argued that the covering-law model of scientific explanation is ultimately based on David Hume's analysis of causation. For Hume, what is called *causation* is nothing but the constant conjunction of two events that happen to be contінguous in time and space, the event that is prior in time being labeled the ''cause'' of the later event labeled ''effect,'' although there is actually no necessary connection between them (see Losee, 1972, pp. 104–6). In other words, we can never be sure that causation is not simply correlation between event at time t and event at time $t+1$. The critics have dismissed this Humean ''billiard ball model of causation'' and have instead insisted that genuine scientific explanation must involve an intervening mechanism connecting cause and effect, which guarantees that the relationship between the two events is indeed a ''necessary'' one (e.g., Harré, 1970, pp. 104–26, 1972, pp. 92–5, 114–32; and Harré and Secord, 1972, chap. 2).

The example of Newton's theory of gravitation, however, shows that the insistent demand for a truly causal mechanism in scientific explanation, if taken at face value, might well be harmful to scientific progress. Ignore everything about moving bodies, Newton said, except their positions, point masses, and velocities, and provide operational definitions for these terms; the resulting theory of gravity, incorporating the universal law that bodies attract each other with a force that varies directly with the product of their masses and inversely with the square of the distance between them, then enables us to predict the behavior of such diverse phenomena as the elliptical paths of planets, the phases of the moon, the occurrence of tides, the trajectory of missiles fired out of cannons, and even the rate at which apples fall from trees. Nevertheless, Newton provided no push-or-pull mechanism to account for the action of gravity – and none has ever been discovered – and he was unable to meet the objection of many of his contemporaries that the very notion of gravity acting instantaneously at a distance without any material medium to carry the force – ghostly fingers clutching through the void! – is utterly metaphysical.[2]

[2] We know that Newton was perfectly aware of this objection; as he wrote in a letter to a friend: ''Gravity must be caused by an agent acting constantly according to certain laws, but whether this agent be material or immaterial I have left to the consideration of my readers'' (quoted in Toulmin and Goodfield, 1963, pp. 281–2; see also Toulmin

Yet who could deny the extraordinary predictive power of Newtonian theory, particularly after the confirmation in 1758 of Edmond Halley's prediction of the return of ''Halley's comet,'' topped in 1846 by Leverrier's use of the inverse-square law to predict the existence of a hitherto unknown planet, Neptune, from the observed aberrations in the orbit of Uranus; the fact that Newtonian theory sometimes scored misses as well as hits (witness Leverrier's fruitless search for another unknown ''planet,'' Vulcan, to account for the irregularities in the motion of Mercury) was conveniently forgotten. In short, it can be argued that Newton's theory of gravity is merely a highly efficient instrument for generating predictions that are approximately correct for virtually all practical purposes within our solar system but which nevertheless fails really to ''explain'' the motion of bodies. Indeed, it was thoughts such as these that led Mach and Poincaré in the nineteenth century to assert that all scientific theories and hypotheses are merely condensed descriptions of natural events, neither true nor false in themselves but simply conventions for storing empirical information, whose value is to be determined exclusively by the principle of economy of thought – this is what is nowadays called the methodology of *conventionalism*.

Suffice it to say that prediction, even from a highly systematic and rigorously axiomatized theory, need not imply explanation. But what of the converse proposition: can we provide an explanation without making any prediction? The answer clearly depends on precisely what we mean by *explanation,* a question that we have so far carefully dodged. In the widest sense of the word, to explain is to answer a Why? question; it is to reduce the mysterious and unfamiliar to something known and familiar, thus producing the exclamation: ''Aha, so that is how it is!'' If this deliberately loose use of language is accepted, it would appear that there are scientific theories which generate a sense of Aha-ness and yet produce little or nothing in the way of prediction about the class of events with which they are concerned. A leading example, frequently cited by critics of the received view (e.g., Kaplan, 1964, pp. 346–51; Harré, 1972, pp. 56, 176–7), is Darwin's theory of evolution, which purports to explain how highly specialized biological forms develop from a succession of less specialized ones by a process of natural selection that acts to maximize reproductive capacity, without however being able to specify beforehand precisely what highly specialized forms will emerge under which particular environmental conditions.

Darwinian theory, say the critics, can tell us much about the evolutionary process once it has occurred, but very little about that process before it occurs.

and Goodfield, 1965, pp. 217–20; Hanson, 1965, pp. 90–1; Losee, 1972, pp. 90–3). Likewise, the history of the concept of hypnosis (through ''animal magnetism'' to ''mesmerism'' to ''hypnosis'') demonstrates that many well-attested natural phenomena, for example, the efficacious use of hypnosis as a medical anaesthetic, cannot be explained even now in terms of an intervening, causal mechanism.

It is not simply that Darwinian theory cannot spell out the initial conditions required for the operation of natural selection, but that it cannot provide definite universal laws about the survival rates of species under different environmental circumstances. Insofar as the theory predicts at all, it predicts the *possibility* of a certain outcome conditional on other events actually occurring and not the *likelihood* of that outcome if those events did occur. For example, it conjectures that a certain proportion of a species with the capacity to swim will survive the sudden inundation of its previously arid habitat, but it cannot predict what proportion will actually survive a real flooding and it cannot even predict whether this proportion will be larger than zero (Scriven, 1959).

It would be wrong to say that Darwinian theory commits the famous fallacy of *post hoc, ergo propter hoc,* that is, inferring causation from mere casual conjunction, because Darwin did provide a mechanism to account for the evolutionary process. The cause of the variation of species according to Darwin is natural selection, and natural selection expresses itself through a struggle for existence that operates via reproduction and chance variations on what he called ''gemmules,'' much like domestic selection by animal breeders. Darwin's mechanism of inheritance was essentially a system whereby the traits coming from each parent were blended in the offspring, the traits being steadily diluted in successive generations. Unfortunately, the specified mechanism is faulty: no new species could so arise because any mutation, or ''sport'' as Darwin used to say, would fade away by blending within several generations to the point where it would cease to have any selective value. Darwin himself came to appreciate this objection, and in the last edition of *The Origin of Species* he made increasing concessions to the discredited Lamarckian concept of the direct inheritance of acquired characteristics in the effort to provide something like a tenable explanation of evolution.[3]

For Lamarck, the giraffe grows a longer neck because it wants to get at leaves higher up the tree and this acquired characteristic is handed down to its progeny, who in turn stretch their necks still further. According to Darwin, giraffes have offspring with necks of different lengths and the scarcity of leaves gives young giraffes with longer necks a better chance to survive, mate, and thus produce more giraffes with long necks like themselves; over generations this same effect eventually produces the long-necked giraffe we

[3] It is with some satisfaction that we note that Darwin was inspired by one economist, Thomas Malthus, and decisively criticized by another, Fleeming Jenkin, a professor of engineering at the University of Edinburgh (Jenkin, incidentally, was the first British economist to draw demand and supply curves). It was Jenkin who first demonstrated in an 1867 review of *The Origin of Species* (1859) that Darwin's theory was incorrect as Darwin stated it. It was this objection which *may* have caused Darwin to insert a new chapter in the sixth edition of *The Origin of Species,* resuscitating the ideas of Lamarck (see Jenkin, 1973, particularly pp. 344–5; Toulmin and Goodfield, 1967, chap. 9: Ghiselin, 1969, pp. 173–4; Lee, 1969; Mayr, 1982, pp. 512–14).

know. The two evolutionary mechanisms are radically different and for Darwin to have conceded even a jot to Lamarck was a serious compromise of his fundamental argument.

The irony is that by 1872, unknown to Darwin or to anyone else, Mendel had already discovered the idea of genes, that is, discrete units of heredity that are transmitted from generation to generation without blending or dilution. Mendelian genetics provided Darwinian theory with a convincing causal mechanism, but from our point of view, it left the status of the theory of evolution essentially where it was before: Darwinian theory seems to explain what it cannot predict and offers few supports for its arguments except indirect ones after the fact. Darwin was himself a self-declared advocate of the hypothetico-deductive model of scientific explanation (Ghiselin, 1969, pp. 27–31, 59–76; George, 1982, pp. 140–50), but the fact remains that for some he provides to this day ''the paradigm of the explanatory but nonpredictive scientist'' (Scriven, 1959, p. 477).

This is perhaps to overstate the case because Darwinism does rest on a number of specific contingent claims about reality – for example, that offspring vary in phenotypes, that such variations are systematically related to the phenotypes of the parents, and that different phenotypes leave different numbers of offspring in remote generations. And Darwinism does imply some definite predictions, for example, that species never reappear; thus, if the dodo came back, Darwinism would be refuted (Mayr, 1982, chap. 10; Rosenberg, 1985, chaps. 5–7). Similarly, to say that Darwinian evolution can explain the modern giraffe's neck but could never have predicted it beforehand is really to misunderstand Darwinian theory, which predicts, if it predicts at all, not for individuals (like giraffes) or for organs (like necks) but rather for traits or sets of traits. Darwin himself was keenly aware that certain facts, such as the existence of neuter insects and sterile hybrids, appeared to contradict his theory: a whole chapter of the *Origin of Species* was devoted to ''miscellaneous objections to the theory of natural selection,'' that is, traits which could not have evolved by natural selection. In short, Darwinism is capable of being refuted by observations, quite apart from the fact that in recent times speciation à la Darwin has been directly observed (Ruse, 1982, pp. 97–108; Ruse, 1986, pp. 20–6). In that sense Darwinian evolution is not a logically different type of theory from, say, Newtonian mechanics or Einsteinian relativity (Williams, 1973; Flew, 1984, pp. 24–31; Caplan, 1985). Nevertheless, it may be granted that the covering-law model of scientific explanation with its corollary, the symmetry thesis, cannot easily accommodate the Darwinian theory of evolution.[4]

[4] Perhaps that is why Popper (1976, pp. 168, 171–80; also 1972a, pp. 69, 241–2, 267–8) once argued that the Darwinian theory of evolution is not a testable scientific theory

There are other examples of theories that appear to provide explanations without making definite predictions, such as Freudian depth psychology and Durkheim's theory of suicide, although any of them are susceptible to the retort that they are not truly scientific. But a still wider class of examples is furnished by all manner of historical explanations, which at best yield sufficient but not necessary conditions for a certain kind of event to occur or to have occurred; what historians explain is almost never strictly deducible from their *explanans* and hence does not result in anything like a strict prediction (or rather retrodiction). Historical explanations are indeed controlled by factual evidence like scientific explanations but the evidence is usually so sparse and so ambiguous as to be compatible with a large number of alternative and even conflicting explanations. It is difficult therefore to resist Hempel's (1942) argument that virtually all historical explanations are pseudo-explanations: they may be true or they may be false but we will rarely know which is the case and the historian is not typically prepared to help us to distinguish one from the other.

To sum up: we can make a case for the thesis of explanation-without-prediction but it is not a strong case and I myself remain persuaded that the covering-law model of scientific explanation survives all the criticisms it has received. This is clearly a controversial position but suffice it to say that we ought to be on our guard when offered an explanation that does not yield a prediction, that is, when instead of an explanation we are offered "understanding." "We understand the causes of earthquakes," Frank Hahn (1985, p. 10) tells us, "but we cannot at the moment predict them." On the contrary, however: geophysicists have made great progress in recent years in predicting earthquakes because they have come better to understand their precise causes. In any case, when understanding is not matched by predictability, we should ask, Is it because we cannot secure all the relevant information about the initial conditions, as with much of biological evolution, or is it because the explanation does not rest in any way on a universal law or at least a loose generalization of some kind, as with so many historical explanations? If the latter, I would argue that we are definitely being handed chaff for wheat because it is not possible to explain anything without reference to some larger set of things of which it is itself an element (see Elster, 1989).

Norms versus actual practice

We have seen that the covering-law model of scientific explanation excludes much of what at least some people have regarded as science. But

but rather "a metaphysical research programme – a possible framework for testable scientific theories."

this is precisely its aim: it seeks ''to tell it like it should be'' and not ''to tell it as it is.'' It is this prescriptive, normative function of the covering-law model of explanation that its critics find so objectionable. They argue that, instead of stating the logical requirements of a scientific explanation, or the minimum conditions that scientific theories ideally should satisfy, our time would be better spent in classifying and characterizing the theories that are actually employed in scientific discourse.[5] When we do so, they contend, what we shall find is that their diversity is more striking than their similarity; scientific theories seem to lack properties common to them all.

In addition to deductive, lawlike, statistical, and historical explanations, which we have already mentioned, biology and social science in general furnish numerous examples of functional or teleological explanations, which take the form of indicating either the instrumental role that a particular unit of an organism performs in maintaining a given state of the organism, or that individual human action plays in bringing about some collective goal (see Nagel, 1961, pp. 20–6). These four or five types of explanations appear in a variety of scientific theories, and the theories themselves may in turn be further classified along various dimensions (e.g., Suppe, 1974, pp. 120–5; Kaplan, 1964, pp. 298–302). But even such detailed typologies of scientific theories raise difficulties because many theories combine different modes of explanation, so that it is not even true that all the scientific theories classed together under some common heading will reveal the same structural properties. In other words, as soon as we take a comprehensive view of scientific practice, there is simply too much material to permit a single ''rational reconstruction'' of theories from which we might derive methodological norms that all proper scientific theories are supposed to obey.

The tension between description and prescription in the philosophy of science, between the history of science and the methodology of science, has been a leading factor in the virtual overthrow of the received view in the 1960s (see Toulmin, 1977). That tension also makes itself felt in Karl Popper's treatment of the role of falsifiability in scientific progress, which has

5 In the same way, historians have argued that the covering-law model of historical explanation misrepresents what historians actually do: history is an ''idiographic'' not a ''nomothetic'' subject, being concerned with the study of particular events and particular persons, not with general laws of development (see Dray, 1957; 1966). But the essence of Hempel's original argument was that even individual events cannot be explained except by invoking generalizations of some kind, however trivial, and that historians typically provide no more than an ''explanation sketch'' because they either fail to specify their generalizations, or else imply without warrant that they are well attested. The debate about the received view among philosophers of science is thus perfectly duplicated by the Hempel–Dray debate among philosophers of history (see McClelland, 1975, chap. 2, for a judicious and pointed summary).