Explanation in the Biological Sciences

MICHAEL SCRIVEN
Department of Philosophy
University of California, Berkeley

The changes in biology and in other sciences since the eighteenth century have made biology an increasingly typical science. If the historians of science can forgive a caricature, I would say that biology was once a mixture of magic and mechanism, in which the latter element gradually gained the upper hand; that it came to be dominated by a relatively novel mode of explanation which served as a paradigm for the burgeoning social sciences; and that finally it was invaded by the reductionist forces of molecular biology as were classical thermodynamics and behaviorist psychology among the other "pure" macro-subjects. Today it contains, and for obvious reasons will always retain, substantial commitments to both the ecological-evolutionary domain and the molecular-chemical domain; in the jargon—an idiographic-historical aspect, as well as a nomothetic-dynamic one. The peculiar problems of uniqueness and predictability arise most clearly in connection with the former aspect and hence it is to this that my remarks are most immediately applicable. But this does not imply that I think it is more important than the other; no such comparative evaluation seems useful. Nor do I think the methodological phenomena to which I attend are only to be found in evolutionary biology; they most certainly occur in molecular biology, as indeed in classical astrophysics. But they are most clearly seen in the historical side of biology or its short-term sidekick, ecology, precisely duplicating—it appears to me—the situation in geology, geography, government, and psychiatry, in all of which there is currently sharp professional competition between the historical-idiographic enthusiasts and the nomothetic-dynamic. I shall treat briefly several issues that bear on the methodology of these double-aspect subjects, beginning with a discussion of the nature of philosophy of science and proceeding to a general sketch of what one might call macro-philosophy of science.

THE NATURE OF PHILOSOPHY OF SCIENCE

In my view, philosophy of science is best thought of as simply a meta-science, like the sociology of science, the economics of science, the psychology of science, and so on. Philosophy is not *exactly* a science, any more than history is, but it is best to think of it this way because such a conception puts its descriptive and nomothetic features into correct perspective. Just as a physicist will unhesitatingly evaluate explanations on the basis of his physical knowledge, so a philosopher of science is entitled to evaluate *accounts* of explanations, *theories* of explanation (descriptions of their general logical features), on the basis of his knowledge of the *logical* properties of good individual scientific explanations. Of course, it is some set of scientists' judgments that determines the philosopher's data, but that does not mean the individual scientist or most scientists at one time cannot be corrected by the philosopher, since he may have extracted from his research certain key features of explanation the necessity of which has been overlooked by scientists engaged in the thrills of scientific conflict. This situation rarely persists for long since many scientists are also part-time philosophers of science and usually appeal to logical considerations quite quickly, often prematurely, in their debates with other scientists. But there is nothing more surprising about this possibility of normative philosophy of science than there is about normative science, that is, about one scientist correcting another on the ground that this theory does not fit in well with well-established physical principles. Metascience is empirical in the same sense as anthropology, though it studies verbal behavior rather than other quasi-arbitrary social customs. It is certainly closely analogous to the lexicographer's activity, which is surely empirically based, and just as clearly it forms a basis for normative and corrective claims. Frederick Churchill has kindly defended philosophers of science on the grounds that they are not really telling scientists how to do science. But sometimes they are, and certainly they are involved in a subject which entitles them to do just this at times. It is really much better to think of philosophy of science as an activity defined by its relation to "straight" science, or "science proper"—in other words, science as it is usually conceived, rather than as what people called philosophers of science do. For what we do is not distinguishable from what some scientists do some of the time and very properly. We are simply specialists in the logical analysis of scientific concepts, just as we might be specialists in the economic effects of scientific research. And some scientists achieve their most striking results from consideration of what are simply logical points about science, whether raised by a colleague with the same union card or a philosopher, just as some philosophers of science have made significant contributions to straight science, from Aristotle to Kant's hypothesis about the creation of the solar system and Russell's hypotheses about the fundamentals of mathematics. Scientists who do not like philosophers often say that such contributions are not made by a philosopher but by a man who was partly a scientist, or even "really" a scientist. But the other side of that coin is that many scientists, including Einstein, must then be regarded as partly philosophers or even "really" philosophers. The payoff is scientific progress, and that requires some logical analyses, for example of the concepts of distant simultaneity, or of species, or of information-content, or of causation, whoever does them. The real question is whether any straight philosophers of science ever say anything that leads their listeners to improve their practice as scientists. And that challenge I regard as a legitimate, indeed as the principal, criterion for criticizing my own work. I think there is another defense for the philosophy of science, viz. as a "pure" subject. But I find that a less challenging and, it is obvious, socially a less valuable goal. In either case, the *task* of philosophy of science is the same: to produce good analyses of scientific concepts. And the task of the history of science is similar with respect to the time dimension.

We have to tell it like it is, and that means both philosophers and historians provide data in terms of which it will be possible to criticize current practice destructively and decisively. To give two examples: it is inconceivable that any historian of psychology could fail to point out some howlers in recent psychology of perception that have simply been repeated because their perpetrators were ignorant of their earlier resolution. And I do not think a student of the Whorfian hypothesis or of Talcott Parsons' work could have failed to benefit from the trenchant criticism philosopher Max Black has directed to those topics.

So we philosophers have swords in our sheathes, but so do you as scientists. Just as we can appraise one aspect of your work, so you can do the same to us. Our task is to benefit each other's main enterprise as well as we can.

MACRO- AND MICRO-PHILOSOPHY OF SCIENCE

The debate between the reductionists and the historical biologists is simply one example of a debate that has occurred in many subjects at many times in their history—one thinks of Ostwald going to the grave (or *nearly* that far) denying the reality of the atom and of Skinner condemning the psychophysiological approach as irrevelant to psychology. On the basis of my analogy between philosophy of science and science, I wish to propose the existence of a radical difference of approach within the philosophy of science which corresponds to that difference in science. The terminology which the discipline of economics uses to mark the distinctions is the one I shall use, the difference between the macro and the micro approach, partly because it already has a more general currency and partly perhaps because in economics the two approaches have been reconciled, and I would hope for the same outcome in my own subject. Let me first mention one or two alternative terminologies that I considered, for this is as good a way as any to illuminate the meaning I wish to attach to the terms I do use.

In cryogenics about twenty years ago an ingenious hypothesis called the two-fluid theory was proposed to explain the bizarre behavior of liquid helium. It was widely described as a phenomenological theory, meaning one which (allegedly) had no ontological commitment as to the *real* nature of liquid helium but only proposed an "as if" account: if one regarded this substance as if it were a mixture of two fluids of very different properties, one could account for the fountain effect, and so on. I wish to introduce what might, in this sense, be called a "phenomenological" philosophy of science. The reason I do not call it so is a problem of confusion with the branch of contemporary philosophy called phenomenology, which may well have its own philosophy of science (though I know of no substantial body of such) and which most certainly would be very different from what I am talking about. Incidentally, the sense in which the two-fluid theory could actually explain phenomena while denying reality-content is quite obscure, and that same conclusion applies to my own proposal; the situation reminds one of Galileo and his "calculational device" defense against the heresy charges. Real explanations are reality-based.

One might also talk of a philosophy of science-in-practice as opposed to a philosophy of science-in-principle, and that distinction conveys much of my intent. Or one might talk of a phenotypical philosophy of science by contrast with a genotypical one, and convey a similar point; or of a philosophy of the approximational sciences, and so on.

Professor Simpson has said that most philosophers of science are, as everyone knows, really philosophers of physics. Indeed, it could be said with some plausibility that most philosophers of biology are really philosophers of physics; the same, indeed, could be said of most philosophers of the macro-sciences in general. It is of great importance, therefore, that one distinguish macro-philosophy of science from philosophy of the macro-sciences. The latter might be simply, in one worker's view, the same as philosophy of physics. But macro-philosophy of science, like macro-economics, is by definition different; it may be ultimately reducible to the other but it is the phenomenological account of the logic of the macro-sciences, which is obviously different with respect to many considerations which we shall shortly itemize. In saying that philosophers of macro-subjects are usually philosophers of physics, one does not just mean—and I believe that Professor Simpson did not just mean—that these philosophers do not know (or care, which may be more important) as much about biology as they do about physics—though that is often true —but rather than they are meta-reductionists who wish to reduce, not one science to another, but the philosophy of one science to the philosophy of another. I think someone should be trying to do that, just as reductionists should be working on biology and psychology and chemistry and engineering. But I doubt if it is the only useful approach in the philosophy of science, any more than in the substantive sciences. Let us set out some propositions in macro-philosophy of science and see whether it may not be a more useful instrument for the macro-scientists, regardless of whether one claims it is different for logical or merely factual reasons.

In macro-philosophy of science:

1. Determinism, with respect to macro-state variables, is often an excellent belief, especially in limited areas. That is, determinism is true, or true in very nearly all cases, within certain domains; it covers *much* of the action, including voluntary human behavior, in other areas (for example, spread of phenotypes in large populations). Of course, in the ultimate micro-science of particle physics it is false and very significantly so; its substitute there is statistical determinism, not limited determinism (each is both stronger and weaker than the other). The function of macro-determinism is thus twofold: as a methodological inspiration to further macro- and micro-research, and as a descriptive formula in certain restricted areas (see 8 and 9 below).

2. Emergence is obviously a common phenomenon. (In the microsubject, the issue is lost in morasses of competing definitions and interpretations.)

3. Definition of key concepts like species and function is

chiefly by giving examples, contrasts, and indicators rather than necessary and sufficient conditions or essences. (The micro-theorist may and often does continue to believe that we shall one day be able to identify the truly defining properties of each concept in some ideal reconstruction.)

4. Randomness is common in the macro-*environment,* since it means the absence of regularities which have any causal significance at the macro-level. (A roulette wheel is macro-random.) Randomness usually occurs in the behavior of macro-systems as the result of random input from some environmental system. We might describe this by saying that it is usually antecedent conditions and not the laws which have an irreducibly stochastic element (*macro*-irreducible of course).

5. Reducibility of the macro-subjects (like biology) in the sense of their eliminability in favor of micro-subjects (like physics) is pragmatically absurd, since the macro-subjects are identified in terms of macro-problems which are to be solved using the available macro-data. Of course, some macro-phenomena can not be *explained* at the macro-level, for example, the law of effect or the connection between the curvature of space and the distribution of matter—and these impose some of the limits on macro-determinism and provide an important reason for continuing some reducibility research. But the macro-subjects exist in response to an environmental demand which frequently allows far too little time and/or opportunity for access to micro-data, so their macro-defined concepts are necessary, which means their autonomy as subjects is assured. Often this practical demand is not only pragmatically but *definitionally* connected with macro-variables, as in the request for the historical account of the development of the mammals.

6. Reducibility of macro-phenomena in the sense of their explicability in terms of micro-variables is simply the last resort of whatever truth there is in micro-determinism. It has in fact been achieved on some occasions and may be achieved on many more, but it is chiefly a project for the micro-subjects with the above exceptions. When achieved, however, two points should be insisted on: (i) *an* insight has been achieved, for we do then understand something in *a* respect we did not before; (ii) *no* insight has been achieved with respect to possibly important questions at the macro-level couched in terms that require a macro-answer—an example is any request for historical background, but there are other totally different kinds, such as requests to fit a macro-event into some meaningful macro-pattern. ("That movement is part of a released escape mechanism"—M. B.'s example.) (iii) Both (i) and (ii) are in-

dependent of the question whether new and valuable macro-research has been suggested by the micro-theory which gives insight.

7. Reducibility in the sense of conceptual eliminability or the "defining-away" or "translating" of macro-concepts is nearly always impossible (except in the almost irrelevant sense of material equivalence); but then it has, I think, never been achieved nor would it had been very valuable in any field of science.

8. The most important explanatory concept at the macro-level is that of cause, and it is a completely autonomous macro-concept. That is, it has no logically necessary micro-commitment at all (for example, to the existence of intervening links between spatially or temporally separated causes and effects). Of course, the success of micro-theories in general makes it plausible to look for micro-equivalents of macro-causal claims, in terms of which micro-equivalents it is sometimes possible to answer other and interesting questions (about the mechanism of causal influence for example). But this *in no way* alters the fact that the account in terms of macro-causes is (a) a complete and correct answer to important scientific questions, (b) discoverable without any micro-investigation. Before going into some detail about this and other points, let me complete my list of macro-claims by mentioning the most important consequences of this one.

9. a) The information which enables us to identify the cause of a macro-event beyond any doubt is often entirely inadequate for predicting that same macro-event prior to its occurrence. (Radiation sometimes produces a viable mutation, but you have to bet that it will not on *any* given occasion.)

b) The information which allows a prediction of a macro-event is often entirely inadequate for its explanation at the macro-level. (A mere correlation is not explanatory, but is predictive.)

c) The information which allows a micro-explanation of a macro-phenomenon is often entirely inadequate for its macro-explanation, as in the case where the required macro-explanation is historical and refers to the unique features of the event.

d) The same point as (c) applies to micro-prediction but for a relatively trivial reason and under restrictions.

To sum up point 9, explanation and prediction are essentially different processes practically as well as logically; and the macro- and micro-levels are essentially independent levels with respect to explanation, and to a lesser extent, prediction.

10. The conceptual apparatus of methodology—the concepts

of concept, law, model, theory, hypothesis, and so forth—must be defined in macro-appropriate terms for macro-use, and that means abandoning any Nervous Nellieism about their approximative status. Almost no physical laws are literally true, especially those which are precisely expressed, and any useful generalization at the macro-level which needs a convenient name-tag can justifiably be called a law or principle unless too weak ("hypothesis") or too strong ("central dogma").

To support points 8 and 9 requires a radical and very extensive reassessment of the whole logic of knowledge, understanding, and inference, which cannot be done adequately in the present paper. I shall try instead one simple line of argument, hoping this will give some assurance that an alternative line of approach is possible. I propose to argue that (a) explanation and prediction are *connected* in the sense that both processes *often* draw on *some* of the same data, although they use them in different ways and often use and need other data; (b) a cause is simply a single-factor explanation. The notion of "the cause" is therefore appropriate in just those cases where a simple, single-factor explanation is appropriate, that is, roughly those cases where the inquirer is looking for and there exists one missing link in the general picture that he has of the phenomenon in question.

The two main "axioms," of widely different kinds, on which I base the theory of explanation that leads to these conclusions, are as follows:

1) *An axiom from pragmatic logic.* Explanation is an essentially pragmatic rather than syntactical or semantic notion. In particular it is highly context-dependent, in the sense that a proper answer to a request for an explanation can only be identified by means of cues provided by the context, for example, the *level* of knowledge of the inquiry (or, in particular, that of an inquirer), and the *type* of interest that he has. The notion of "the scientific explanation of x" must be regarded as an abstraction from the basic notion of "explaining x to y" via the notion of "the explanation of x for y."

2) *An axiom from the evolutionary theory of cognitive processes.* The need for understanding and explanation transcends but springs from the need for knowledge. It is a further requirement for effective adaptation of an information-processing system in an environment providing it with more information than it can efficiently handle if it employs the simplest sequential-storage, total-search-and-retrieval system.

Let us begin by identifying the typical questions that call for explanations and predictions. It has been wrongly suggested that explanations are always or nearly always responses to "why" questions. They may in fact be responses to questions of almost any form, for example: What is the nature of this complex process? How did this result come about? Where and in what way do the waste products get treated in this system? Indeed, it is probably just as satisfactory to think that "why" questions can be translated into these terms as the reverse.

Requests for predictions, on the other hand, come from a completely different family; for example: What will be the future development of this life-form? Will it be 6 feet tall a year from now? Where will Mars be next April 9th? When will the next solar eclipse occur at San Francisco? How will the Fifth Republic end? To put the matter in its simplest terms, prediction require us to say something about the nature of the future, whereas explanations require us to identify a feature of the past that bears a special relationship to the present. Speaking atemporally, the simple fact is that any kind of empirically reliable correlation between events is an adequate basis for prediction, but explanations require a special kind of connection, such as causal.

Whereas *any* kind of nonsimultaneous correlation will give you *some* kind of prediction, if you want to predict where *Mars* will be next April *9th,* then you have to operate in a more directed manner. You work from the orbital equation for Mars and the antecedent condition of its position at a certain time to generate the required position. This is very like a theorem-proving program in the computer field, where we specify the task in a certain way, as, for instance, by asking the computer to prove a theorem about a certain relationship, to prove or disprove, for example, the collinearity of three specified points in a certain family of triangles. The program begins with certain axioms, perhaps with intermediate theorems, and tries to apply rules of inference to them according to certain heuristic instructions until it generates a result which has the required formal properties to within a single application of the negation operator.

On the other hand, the kind of activity involved in the search for an explanation varies a great deal more, depending upon the particular level and type of explanation that is being sought. The computer must first infer or be given data about level and type. Then, in response to a request to explain the nature of a complex process (such as a tribal activity to an anthropologist unfamiliar with the tribe in question, or the operation of a complicated distillation plant), the prime task of the explaining program is to identify in this process some already understood patterns, that is, patterns which it is reasonable to infer that the *inquirer* will understand, and attempt to tie together a set of these with understood links so that they completely describe the observed phenomenon in the required degree of detail. The simplest kind of request for an explanation—for example, that posed by the police to the coroner—calls for a single cause, e.g. of death, and here again a pattern recognition activity is involved. The explainer examines the circumstances and compares them with his repertoire of known causal patterns until he obtains the best match. An intermediate case is involved in responding to a question of the form, "How did this come about?", where one is expected to produce a rather long but simple chain, as in answering a request to explain the emergence of the black nations as a significant world force in the latter half of the twentieth century. Here we have a recognition and construction task, the over-all requirements of which are very similar to those of the complex prediction type, if described on a sufficiently general level. But the moment we get down to looking carefully at the two tasks, we notice that in the prediction task we are restricted to using patterns of a particular kind, those with "projective reliability," that is, those with highly stable later portions tied to a highly specific early portion. For we are trying to build up a bridge to a future event described in informative detail, with a degree of accuracy that mere guessing would not lead us to achieve. This is the sole dimension of merit in a prediction—the degree to which it is precise or informative with respect to the required prediction-request. But in building the connecting links between one or several causes and their effect, we are not at all concerned to show that the effect was the only one that could have come about—our task is rather to show how it did, in fact, come about, since we already know *that* it did. That is, we only need patterns with good fit to a *completed* process: no projective reliability is required. On the other hand, the acceptable patterns must meet another criterion; they must themselves be comprehensible. This does not just mean familiar, or true, though it is connected with these concepts; it means, roughly, reducible to certifiable models where "certifiable" means that their consequences are known (to at least a considerable depth) and true.

So we see that there are affinities between the prediction task and the explanation task which are related to those between theorem proving and pattern recognition tasks, but there are also crucial differences.

Without going into great detail, I shall now describe one particular case that bears on claim 8. This is what I call the basic statistical case of causation. In this case, the following conditions obtain:

a) The effect Y never occurs except in circumstances where the factor X is also to be found.
b) In those cases where the factor X is introduced randomly, as well as where it is spontaneous, a certain small percentage are observed to also include the factor Y.

No scientist would hesitate to say that on the occasions where both X and Y occur, X is the cause of Y. I say that a scientist would not hesitate to describe this as a case of causation with some confidence because it is a common case in quantum physics, and is there so described. But its importance has not been recognized by scientists in general, or philosophers of science in particular, who are still committed to the view that there cannot be any definite undeniable cases of causation in an individual situation if there is no underlying determinism.

The case just described can be weakened to allow a spontaneous base rate for Y, to cover the case of several independent possible causes of Y, etc. In all of these situations, a substantial amount of causal vocabulary can be retained if certain other experimental results are known.

Some philosophers faced with considerations such as the above have been led to suppose that there is an analysis of causation which would handle them. This would equate a cause with a probability-modifying operator. Unfortunately, this is both too restrictive and too inclusive a definition. It is too inclusive because the symptoms of a disease are probability-modifying operators indicating the presence of the disease but not causing it; it is too restrictive because of a family of cases which are also crucial for certain other analyses in the concept of cause, such as cases of over-determination. These are cases where there is present in a situation more than one factor or group of factors which are independently sufficient to bring about the effect. Since it is possible to predict the occurrence of the effect with probability from any such group, and since in such circumstances it is frequently true that only one of them is, in fact, the cause of the effect occurrence, it must be false that a cause is a probability modifier. These basic cases are, of course, crucial to one or more of the earlier claims, particularly claim 8.

To conclude, then, I have proposed that we should develop and employ a macro-philosophy of science just as long as we find it appropriate to employ macro-sciences. The day when chemistry, biology, and psychology are reduced to atomic physics it would be quite possible that we should reduce macro-philosophy of science to micro-philosophy of science. It might be appropriate at an earlier date, or a later one; but until it can be demonstrated that there are greater advantages to talking always in terms of what is "in principle possible," then I would be inclined to think that we should stick as closely as we can to a description of the methodology and logic of science which is as closely tailored to its reasoning and procedures as its own concepts are tailored to its phenomena.