Scientific Explanation

There are two main types, and their interrelationship has philosophical as well as scientific implications.

Warren Weaver

The explanations of science are often regarded as so complete, so precise, so irrefutable, and so ultimate in nature that other types of interpretation of experience are rudely crowded out. Numbers, charts, equations expressed in abstract variables—these are by some supposed to constitute the final, absolute, and complete explanation of all phenomena. And as science successfully moves from its major conquest of the physical world to equally promising attacks on the biological world, including all mental and emotional phenomena—indeed as science moves towards an analysis of the nature of life itself—this assumption that science "explains everything" becomes more and more formidable.

Thus it seems useful to examine the character of scientific explanation. Possessing not even the vocabulary of philosophy, I propose to phrase my comments in very simple language—what my friend Fred Mosteller calls "kitchen words." What is lost in the appearance of scholarship will, I hope, be at least somewhat compensated by clarity.

Consider, then, the person who is confronted by something which he does not understand. He goes through some process of talking, or listening, or reading, or thinking, or experimentation, or perhaps of all of these. It may take 10 minutes, or it may take years. Subsequent to that process, he says, "Well, at least I have made a start in a good direction, for now I understand better than I did."

What has happened to that person in the interval between his complaint that he does not understand and his later feeling that he now does partially or even fully understand? What, in other words, is the nature of this strange process we call explanation?

Perhaps it will be well, at the outset, to note that a person usually considers a statement as having been explained if, after the explanation, he feels intellectually comfortable about it. I am sure that this criterion is too vague to be approved by the philosophers; but I am also sure that it expresses something that is widely understandable and acceptable. The average person applies this criterion to the explanation of a machine, or a process, or any natural phenomenon. The scientist is influenced by similar considerations. After a good explanation he is likely, because of his special interest, not only to feel intellectually more comfortable but also to experience a very active satisfaction or a very real esthetic pleasure. And *scientific* explanations characteristically have two further very important aspects.

First, the scientist who "understands" a phenomenon is almost always in a position to *predict*. He can say with confidence, "under such and such conditions, such and such will happen." If he has rather full understanding, he may be able to add, "If you change the conditions in such and such a way, then the results will be altered in such and such a way."

The second aspect flows directly out of the last preceding remark. For if the scientist knows how to change the result by altering the attendant circumstances, then he is well along toward accomplishing *control*. And the control of natural phenomena, to bring them more effectively to the service of men, is obviously one of the major aims of science. Indeed a scientist is very likely to say that he "understands" a phenomenon if he can predict it and can control it. If he can express this in mathematical equations, and if he can thus relate the phenomenon in question to a wide range of other phenomena, then he is likely to consider the explanation satisfactory and complete.

The Two Types

With these preliminaries behind us, we can now begin a direct discussion of the nature of explanation. I suggest that there are two main types of explanation, very different in character, and useful in different circumstances.

The first, the more familiar, the older, and by far the more popular, consists of explaining something by *restating or describing the unfamiliar in terms of the familiar*.

This is the way the dictionary explains the meaning of a word. You may not understand the word *euphroe*; but when Webster tells you that it is a little block of wood with a hole in each end, used to cinch a tent rope, then the meaning of the word has been explained.

This is what happens when a person, completely mystified by the idea of electromagnetic waves spreading out from a radio station, is told: "You have, of course, seen the circular ripples expand on the surface of a still pond when you drop a pebble. Notice that the ripples get weaker as they get further and further from the center. (Do you live so far from your station that the signals are weak?) You have doubtless noticed that, just behind a rock which sticks up above the surface, there is calm water with no ripples; but a few feet beyond the rock the ripple patterns from either side join up and show little or no residual effect of the rock. (If you live just 'behind' a big steel-frame building you will have trouble, won't you, in getting the radio signals; but a half mile beyond, the building does not shield or interfere.) And remember that the water ripples are two-dimensional waves on the surface of the pond; but the radio waves are three-dimensional, going out in spherical form like the successive spherical shells of a magic onion which keeps growing larger and larger."

Quite apart from the obvious incompleteness of these remarks, this is a reasonable example of at least the early stages of an "explanation" of the sort that many persons find satis-

The author is vice president of the Alfred P. Sloan Foundation, New York. This article is adapted from the 4th Daniel Coit Gilman lecture, given at the Johns Hopkins University School of Medicine, Baltimore, Md., on 24 September 1963.

fying. Indeed, examples not very much more sophisticated than this have played exceedingly important roles in the development of scientific theories, particularly in physics.

We should note, parenthetically, that every poet should feel friendly toward this type of scientific explanation; for it is essentially equivalent to the similes and metaphors that, explicit or implied, are the very essence of poetry.

Indeed, the matter goes deeper than this. As Bronowski has said (*1*, p. 35): "The scientist or the artist takes two facts or experiences which are separate; he finds in them a likeness which had not been seen before; and he creates a unity by showing the likeness."

When one thinks a little it is promptly clear that this first type of explanation (which points out that the unfamiliar is in certain respects like the familiar), when considered simply *as* an explanation, is almost completely illusory. It restates in terms which are very familiar, to be sure, but which, upon really honest examination, are just exactly as little "understood" as the unfamiliar concept, procedure, or phenomenon for which explanation is sought. A person may be very familiar with expanding ripples on a pond without actually having any clear idea at all as to *how* these ripples propagate, interfere, or attenuate. Long familiarity has dulled penetrating curiosity, and one just unthinkingly accepts the familiar as understood.

It does not, however, at all follow that this type of explanation is silly or useless. For it is a rather remarkable fact that there is a tremendous amount of non-obvious isomorphism in the logical structures of natural phenomena, especially when all the phenomena in question are broadly macroscopic in scale—say, involving space dimensions ranging from those of optical microscopy up to planetary and very possibly to galactic dimensions. That is, there are very numerous pairs, *A* and *B*, of "things" in the physical world which in important respects behave similarly; so that it often constitutes a useful, illuminating, and suggestive explanation, when meeting a strange pair of related variables, to be reliably told that this unfamiliar pair is "like" a familiar pair. This turns out to be mentally satisfying; and it often suggests the application, to the new and initially strange pair, of a lot of procedures which have previously been found useful in the case of familiar pairs. Presently, of course, the new pair becomes a familiar pair.

This recognition of similarity of behavior is one of the major ways in which science moves forward in its great task of bringing its type of order and beauty out of confusion. Bronowski, in his superb little book *Science and Human Values* (*1*), says, "All science is the search for unity in hidden likenesses [p. 23]. . . . The scientist looks for order in the appearances of nature by exploring such likenesses [p. 24]. . . . The discoveries of science, the works of art, are explanation—more, are explosions, of a hidden likeness. The discoverer or the artist presents in them two aspects of nature and fuses them into one. This is the act of creation, in which an original thought is born, and it is the same act in original science and original art [p. 30]."

There are serious limitations to this first type of explanation, but before commenting on them, I will say a little about the second type of explanation.

The second type of explanation has no concern whatever with familiarity. On the contrary, it characteristically describes a phenomenon or a statement in terms which are almost indefinitely *less* familiar, or in any event more basic and more abstract, than the phenomenon or statement being explained.

A good example is furnished by mathematics. Suppose one is confronted by a theorem which is complicated, subtle, and wholly unfamiliar. The first stage of explanation (useful in the case of a person well trained in the mathematical field in question) consists of *proving* this theorem by showing that it logically follows from various theorems previously proved. At this stage the second type of explanation shares, to a mild degree, the essential feature of the first type. For the trained mathematician will in fact be familiar with the previously known theorems which he uses in proving the new theorem.

The essential difference, however, is that the mathematician, while he enjoys the fact that he is familiar with the older theorems, does not in the least base his just-gained confidence in the new equation upon that element of familiarity. For it is not at all familiarity which makes him trust and understand the older theorems. His trust and understanding rest solidly on the fact that these older theorems have, in their turn, been logically deduced from a still older, still more primitive, set of theorems.

One must not underestimate the power and excitement of this step-by-step procedure of proof. His friend John Aubrey described the reaction of Thomas Hobbes, who, about 1630 when he was 40 years old, accidentally looked at a copy of Euclid's *Elements* open at the pages containing the proof of the famous theorem of Pythagoras.

> "By G- -," sayd he (He would now and then sweare, by way of emphasis) "By G- -, sayd he, this is impossible!" So he read the demonstration of it, which referred him back to such a proposition; which proposition he also read. *Et sic deinceps,* that at last he was demonstratively convinced of that trueth. This made him in love with geometry.

Now this procedure of pushing the explanation down, step by step, to lower and lower, more and more primitive levels of explanation obviously requires examination. For where does this descent stop?

As an historical fact, there have been three answers to this question. The first and least satisfactory answer comes from those individuals who, so to speak, descend one or two or three steps and then get so bored or so confused that they are content to give up. For them this descending set of steps ends in a fog.

The second kind of answer is that which appealed to Euclid. One descends, step by step, until one reaches a "Bottom Step," on which are found statements (axioms) which supposedly are so obviously true, so clearly necessary, and so patently clear that all reasonable men are supposed to agree to these statements and to accept them without any further examination.

This kind of answer seemed satisfactory to many persons over hundreds of years. But, as every schoolboy now knows, this kind of answer is no good. For it turned out that the axioms on this supposedly bottom step simply were *not* obviously true, nor were they *necessary*. Euclid considered it unthinkable to question the statement "through a point not on a straight line it is possible to draw one and only one straight line parallel to the given line." But we now know that a person can perfectly well assume that through this external point one can draw *more* than one straight line parallel to the given line. The result of this second assumption is not logical chaos or contradiction but an alternative geometry, rich, consistent, beautiful, and useful.

So it is not tolerable to let these descending stair steps, on which are found successively more and more simple explanations, terminate in a fog, nor is it tolerable to let them terminate in a universally accepted bottom step. What then can be done? It was a triumph of 19th-century mathematics to see a respectable alternative. It is to descend down to a step which is *not* labeled "Unique Bottom Step" but which *is* labeled "This Is As Far As We Go." On this step one does not find axioms—statements which are supposed to be necessary and obviously acceptable to all. One finds postulates—statements which, for the purposes in hand, are simply assumed to be true. Two different mathematicians can perfectly properly, even when working in the same field, assume two quite different sets of postulates. If you do not like a given set of assumptions, there is no compulsion. You can just decide not to play and can take your doll rags and go home. Or you may decide to accept the set of postulates, this not at all meaning that you "believe" them but simply meaning that you adopt them, and that you will now, starting there, apply logical procedures and lift yourself, step by step, to ever higher levels of complicated and sophisticated deductions.

The "explanation" of a statement on a high-level step is now clearly to be obtained by tracing the relationship between that statement and statements on the next lower step, the second lower step, the third lower step . . . , until one reaches the step labeled "This Is As Far As We Go." On this step one finds nothing "obvious," nothing "true." One finds only statements which have a footnote, "It seems interesting to assume this set of remarks." And it seems to me that this type of explanation is precisely the process of descending to the step on which we agree to stop.

We may again parenthetically note that poets should find nothing alien or objectionable here, but rather ought to be enthusiastic about the element of ultimate mysticism that exists at the bottom of this type of scientific explanation.

Interrelationship

In terms of the stair-step metaphor we have been using, we can now describe the interrelationship between the two types of explanation. In the first type, being located for the moment on a step which contains strange and not-understood things, one looks horizontally about him and observes a neighboring set of steps. This second set—or at least a few steps of it—are friendly and familiar. One has been on them many times. And one notices (or is told) that the strange elements on the step one is now occupying are "like" elements over on a familiar step (presumably at about the same level) of the other set.

It is a curious fact that this observation is very comforting, whether or not one has ever visited more than one or two steps of the second set.

In terms of this metaphor, one can usefully refer to this first type as "horizontal explanation." The procedure does not move vertically downward to deeper levels of simplicity or abstraction but moves horizontally over to more familiarity. And although this is an interesting, pleasant, and clearly useful procedure, it seems clear that it does not, in any sophisticated sense, and certainly not in any ultimate sense, constitute "explanation" at all.

The second type is "vertical explanation." And it seems fair to say that it is deep and logical, but that, again, it certainly is not ultimate.

At this point it is useful to return to some remarks I made earlier, when referring to the non-obvious isomorphism of natural phenomena—namely, remarks concerning the importance of the scale of the events. For it seems at least generally true that horizontal explanation is useful when the scale of the events being explained is roughly the same as the scale of the more familiar events used in the explanation. The electrical oscillations in circuits are usefully discussed in terms of the oscillations of a taut string; and the sizes, masses, periodic times, and so on, in the two cases are, very roughly at least, of the same large-scale order of magnitude. This statement about spatial and temporal size is an exceedingly rough one. It can be expressed (and very probably this is not accidental) by saying that the distances and times and masses involved must not be too extremely small (or large) as compared with the dimensions and mass of a man, and as compared with the times (years, days, hours, minutes, and seconds) which enter directly into human experience. The horizontal-explanation method has worked surprisingly well on a planetary, or even larger, scale; and surprisingly well down to molecular dimensions. But when one tries to push this method down to atomic dimensions, and certainly when one tries to push it to nuclear dimensions, then the method of horizontal explanation (at least so I believe) collapses entirely.

For example, you will read in articles on modern physics that the density within the nucleus of an atom is of the order of ten thousand million tons per cubic inch. I refuse to gasp and say, "Isn't that amazing!" For it seems to me that such a statement is simply meaningless, and the collapse of meaning has resulted from trying to use man-sized language, and horizontal explanation, where they are totally inapplicable.

The situation at the cosmic scale does not seem so clear; but perhaps attempts to answer such questions as, "Is the universe expanding?" or "What is the age of the universe?" all experience difficulties which result from the inappropriateness of using "human-size" concepts for such problems.

If it is indeed entirely useless—as I here suggest it is—to try to apply horizontal explanation when the *terms of the explanation* are macroscopic but the *thing being explained* belongs to the submicroscopic world of events in which the elementary particles of physics are the actors, then, as far as I can see, we must cheerfully and completely accept the type of formalism which now characterizes the present-day theoretical physics of elementary events. According to this formalism, the scientist performs well-specified experiments and obtains numbers; he inserts these numbers into equations and gets new numbers; and he then consults nature (that is, performs further well-specified experiments) and obtains numbers which, to a satisfying approximation, discretely or statistically agree with those calculated from the equations. This agreement indicates that he has a good theory; and one simply has to accept the fact that the procedure is formal, and that no one has, or in the nature of things *can* have, a "physical interpretation" for the various quantities in the equations. For this desire for a "physical explanation" is, I believe, precisely a desire for a horizontal explanation when that procedure is inapplicable.

The yearning for "physical explanation" (which as far as I can see always means horizontal explanation) is an urgent one, which extends to all levels of sophistication in science. It is clear that Einstein never gave up the idea that physical interpretation of the unitary events of physics was both possible and desirable. There is a long list of earnest and able individuals who have been puzzled by "action at a distance," and who have sought some other model with macroscopic properties which would help them escape the, for them, intolerable fact that action at a distance is not "understandable" (although, curiously, action *not* at a distance presents equally grave difficulties). All of these persons have, in my judgment, not faced up to the nature of explanation. Vertical explanation has not been satisfying to them; and their concern has been with cases to which horizontal explanation is not applicable.

There remains, we must confess, an underlying mystery here. Why is it that the universe furnishes so many paired instances of useful isomorphism as long as the scale of events lies, roughly, between 10^{-8} centimeter and a few hundred thousand light years but recedes into completely special and unique abstractness when the scale is roughly 10^{-13} centimeter or smaller, or is as large as, say, a billion light years? Is this because our physical theories remain too anthropomorphic, influenced too much by the accident of our own size and by the illusion of continuity at macroscopic dimensions? Will we ever have the courage and imagination to leap over this barrier of smallness into the world of unitary events, and construct a theory which starts at the right place and with the right concepts? Such a theory will surely begin with no recognizable space and time variables, but will, at a much later stage, develop the traditional and continuous time and space measurements as statistical consequences, appropriate only on a macroscopic scale, of the discrete variables of the more basic theory. If and when such a theory is available, certain presently unsatisfactory aspects of the explanation of physical events will have disappeared.

I want to emphasize an aspect which the two types of explanation have in common. It is an aspect, moreover, which the scientist values very highly indeed, for both practical and esthetic reasons. Namely, either type of explanation addresses itself to an element of our experience and gives meaning to it, gives new significance and richness to it, suggests new usefulness for it, in short *explains* it, by placing it in a broader context. Horizontal explanation does this vividly, but narrowly. An electromagnetic wave is put into the context of more familiar mechanical or hydrodynamical waves on strings or ponds. Vertical explanation probes ever so much deeper into the isomorphism of phenomena and puts the case under study within the total context of all the possible phenomena which conform to all the relationships deducible from their common origin—namely, the postulates on the bottom step. The electromagnetic wave thus is placed within the broad context of all possible types of solutions of certain very general types of differential equations. All the practical and esthetic values which result from this recognition of relatedness constitute, I think, the important essence of explanation.

Reference

1. J. Bronowski, *Science and Human Values* (Harper and Row, New York, 1959).

Federal Support of Science: A Formula for Cooperation

National Academy Public Policy Committee offers suggestions for eliminating sources of difficulty.

The following are the formal conclusions contained in Federal Support of Basic Research in Institutions of Higher Learning, *a report issued this week by the National Academy of Sciences Committee on Science and Public Policy. A discussion of the report appears in News and Comment, page 1304.*

The commitment of large public funds for the support of basic research in universities has led not only to spectacular growth of the scope of scientific effort but also to advances in quality: American science has reached a position of world leadership. We attribute this in no small measure to enlightened policies of several federal agencies committed to furtherance of basic research; specifically to the current emphasis on support by research project grants and by fixed-price research contracts (not too unlike grants), coupled with an extensive use of advisory scientific bodies, such as panels or study sections, to select scientifically meritorious projects for support. We believe that research project grants and contracts should remain the backbone of federal policy in support of basic research in science in universities. The emphasis on large programmatic ventures and laboratories which has been manifest in recent times must not lead to a loss of emphasis on individual scientists: the individual investigator has been and will remain the source of strength in American science.

Concerning Federal Agencies

1. The criterion of selection for grant or contract support of basic research has