

Forms of Explanation

RETHINKING THE QUESTIONS
IN SOCIAL THEORY

ALAN GARFINKEL

New Haven and London
Yale University Press

1981

Introduction

We are surrounded by explanations. The child is failing at school because he or she has a low IQ. Prices are going up because of inflation. Inflation exists because "too many dollars are chasing too few goods." We do not normally stop to ask what these explanations mean or what they are supposed to be explaining. This can lead to problems. Often we first become aware of trouble when we encounter conflicts, when several explanations, coming from different angles and speaking to different aspects of a problem, crowd around a single object. We look at a body of theory and find a confusing patchwork of schools and approaches, and it is very hard to see how they fit together.

This happens all the time in the social sciences. In psychology alone we find Freudians (of various kinds), behaviorists, cognitive developmentalists, physiological psychologists, holistic psychologists, humanistic psychologists, transactionalists, gestaltists, social psychologists, symbolic interactionists, personality theorists, existential psychoanalysts, and perception theorists, to name just the best-known varieties. Add this to similar lists for economics, sociology, anthropology, and linguistics, and we have a bewildering variety of ways of approach and modes of explanation.

Faced with any such list, what strikes us is the difficulty of finding a coherent way of comparing the different theories. They seem to be different *sorts* of things. Some of the theories may address different phenomena or different realms of phenomena. Some are genuinely competing, others can be reconciled with one another, while still others pass one another by, answering different questions. They fit together only in a very complicated and overlapping geometry.

It will help to try to map out this geometry. As theoreticians we need to understand how the explanations relate to one another, and as consumers of explanations, people who are trying to deal with some concrete problem, we need to understand how explanations answer or fail to answer the questions we are asking. What is needed is an analysis of

explanation that will help us to sort out this variety. My aim here will not be to construct a general philosophy of explanation based on first principles. Instead I will look at a variety of examples and attempt by their means to develop some elements of a theory. The examples are chosen with an eye to the central questions a philosophy of explanation must deal with:

- When are two explanations inconsistent with each other?
- When are two explanations irrelevant to each other?
- When can two explanations from different theories be added or joined to each other?
- How does one explanation replace or supplant another?
- When does one explanation presuppose another?
- When are two explanations from different theories really explaining the same thing?
- What could make one explanation superior to another?

Explanations in Conflict

If a child is failing at school, there are an embarrassingly large number of potential explanations. If the child happens to be black, there are even more. A short list of explanations includes: low IQ (genetic or environmental causes), culture of poverty, lack of proper prenatal diet, institutional racism, bad teachers, "cultural remnants of slavery," biased educational standards, lack of appropriate role models, economic pressures, matriarchal families, minimal brain damage, and lack of future orientation.

Now what do all these explanations have to do with one another? Are they competing? Do they reinforce one another? Are they complementary to one another, or perhaps just irrelevant to one another, existing on different levels? The person who thinks that the explanation lies in brain damage clearly disagrees with the person who thinks that it is a matter of biased teachers, and both reject an explanation in terms of matriarchal family structure. Yet it is far from clear how we know that these are mutually exclusive.

This is an example of the most basic problem in sorting out a mass of explanations: Which of them are in conflict with which others? As the example illustrates, this can be a very hard question to answer.

On occasion it is easy, if the inconsistency is right on the surface of the explanations themselves. "There was a conspiracy to kill John F.

Kennedy" is inconsistent with "Oswald acted alone." But usually the conflict is not obvious from the formal structure of the statements, and then our analysis of it depends a great deal on background theory and assumptions.

We hear discussions, for example, of conflicts between genetic and environmental explanations of race differences. Someone says that something is "eighty percent explained by genetics"; someone else says, "No, it is eighty percent environmental." They seem to agree that the form of explanation is that any trait is $x\%$ due to genes and $100 - x\%$ ("the rest") due to environment. But are genetic explanations really in conflict with environmental ones? Are they jointly exhaustive? The answer is *no*, in both cases, although this is not understood by many of the participants in these arguments.

At the very least it seems that before we plunge into such debates we should try to sketch what the basic categories of explanation are. This is not usually done, in part because of the sheer difficulty of deciding when two explanations are really in conflict.

Neither is it clear what to do when faced with apparently conflicting explanations. Must we opt for one or the other? Or is it somehow possible to maintain both?

In quantum mechanics the principle of complementarity says that for certain purposes an electron can be viewed as a particle, while for other purposes it can be viewed as a wave. The two modes of explanation, particle theory and wave theory, attribute inconsistent properties to the electron and therefore cannot be applied simultaneously. Yet neither one is true to the exclusion of the other. I am inclined to think that this duality is intolerable: future science will have to eliminate it in favor of a single, coherent picture. Others welcome it as a paradigm in physics itself of the possibility of multiple "conceptual frameworks" or "points of view." And anytime two rival forms of explanation seem applicable to the same thing, it can be tempting to see a case of complementarity. For example, mind and body can be viewed as providing complementary modes of explanation of human action (psychology and physiology). But is there any validity to this view beyond the superficial similarities? We do not know what is going on in the quantum mechanics case, and even less whether there are any genuine examples of complementarity on the macroscopic level.

So we see that among the members of a collection of explanations there will be a number of distinct kinds of relations, straightforward and

complicated. Of course, a multiplicity of explanations does not necessarily mean that there is any internal contradiction at all. Consider this set of explanations of the death of Socrates:

Socrates died because:

- Athens feared his independence
- he drank hemlock
- he was tried and convicted of a capital offense
- he suffered cardiac/respiratory arrest secondary to ingestion of coniine alkaloid
- he was too closely linked to the antidemocratic forces
- he refused Crito's offer of escape.

Here the explanations are not mutually contradictory. Some are different parts of the story, others treat the event on different levels or from the standpoint of different kinds of inquiry. All of them can be maintained simultaneously.

These examples suggest that the first task we might set for a philosophy of explanation is that it give us some account of these conflicts, complementarities, overlaps, and displacements, that it give, as it were, an elementary algebra of explanations. Its purpose would be to tell us when they can be added together and when they must be subtracted from one another.

Whatever Happened to Neurasthenia?

The variety is further complicated by the fact that there are not only different explanations but different conceptions of what an explanation is. Perhaps the most important intellectual development of the twentieth century has been the recognition that there is a variety of conceptual frameworks, forms of understanding, or cognitive points of view. Like a Cubist painting, the contemporary world-picture features a simultaneous presentation of multiple perspectives.

We no longer understand the development of science as a smooth, linear growth of a monolithic entity, Knowledge. Rather, we see it as marked by discontinuities in conceptualization, by radical shifts in the very idea of what the problem is and of what a scientific explanation might look like.

The source of this understanding can ultimately be traced back to Kant's demonstration that the forms of empirical knowledge are subject to prior categories of the understanding. Once we see how concepts shape our knowledge and perception, we see how other categories and other concepts could produce radically different forms of knowledge and

explanation. For Kant, these categories were given once and for all; later came the realization that they are changing and developing, determined and conditioned by period, culture, and context. T. S. Eliot writes:

Even Kant, devoting a lifetime to the pursuit of categories, fixed only those which he believed, rightly or wrongly, to be permanent, and overlooked or neglected the fact that these are only the more stable of a vast system of categories in perpetual change.¹

Recent history and philosophy of science have stressed the idea that developments in knowledge often take the form, not of discoveries of new facts, but of shifts in the conception of what the phenomena to be explained are and of what counts as an explanation of them. The work of Bachelard on conceptions of fire and space, of Foucault on hospitals, madness, and prisons, and more recently the writings of such people as Toulmin and Kuhn have made people more aware of the ways in which the science of a particular period views the world through concepts very much its own.

Thus the prescientific view of the heavens was that everything revolved around man; the early scientific revolution inverted that to say that everything revolved around the sun. But the modern view calls into question the very concept of something "revolving around" something else. Strictly speaking, nothing "revolves around" anything else.

Consider some of the ways in which psychology has characterized the objects of its explanations and the styles of explanation appropriate to those objects. We are inclined, for instance, to think of physiological explanation in psychology as something recent, but in fact it has been very much in vogue at other times as well. Seventeenth-century psychology postulated physiological explanations of behavior in terms of airs, humors, and other material substances. Sir Robert Burton's classic treatise *The Anatomy of Melancholy* discusses melancholy as a pervasive and general condition and sees it as a fundamental psychological diagnostic category. He says that people become melancholic when a certain material humor in them changes from sweet to sour, a process he likens to wine turning to vinegar. They become melancholy "as vinegar out of the purest wine . . . becomes sour and sharp." He goes on to show how this explanation also accounts for other observed

1. In his introduction to Charlotte Eliot's *Savonarola: A Dramatic Poem* (London: R. Cobden-Sanderson, 1926), p. viii.

Introduction

phenomena of melancholy: "From the sharpness of this humour proceeds much waking, troublesome thoughts and dreams, etc."²

Now what is this? An explanation? An analogy? A metaphor? It is all of these really. The distinctions are not hard and fast to begin with, and some things can be substantive theories at one point and literary metaphors at another (e.g., "that makes my blood boil"). At least the *form* of Burton's explanation is more or less familiar to us, although we have some trouble understanding what exactly is the object being explained, this condition called "melancholy." Nowadays we do not use this as a general descriptive term, and so the object of explanation for us will be different. Part of what "melancholy" meant would be covered by our (that is, the current psychoanalytic) concept of "depression." But only part of it. The rest corresponds to other current concepts or to no concept at all. The recent edition of Burton has a jacket description that characterizes "melancholy" as "a term used in the seventeenth century to cover everything from schizophrenia to a lover's moping." "Schizophrenia," on the other hand, is a twentieth-century term used to cover everything from out-and-out madness to political dissent. It has been severely criticized as ideological by Foucault, Szasz, Laing, and others.

Earlier in this century, melancholy would have been diagnosed as "neurasthenia," a term then very much in fashion among psychologists. In fact it was one of their main diagnostic categories and was applied to everything from depression to shyness and anemia. It is no longer used at all, and it was formally dropped as a diagnostic category by the American Psychological Association some years ago. The epic account of the conquest of neurasthenia waits to be written.

Radical changes in styles and objects of explanation can be found in all the sciences, not just psychology. When Galileo reported seeing moons circling Jupiter, he was refuted by a priori arguments that there could be no such thing, since the number of planets (i.e., objects in the solar system) was necessarily seven. The explanation of why there had to be seven took the form of correlating them with the seven apertures of the human head!

There are seven windows given to animals in the domicile of the head, through which the air is admitted to the tabernacle of the body, to enlighten, to warm and to nourish it. What are the parts of the *microcosmos*? Two nostrils, two eyes, two ears and a mouth. So in the heavens, as in a *macrocosmos*, there are two favorable stars, two unpropri-

2. *The Anatomy of Melancholy* (1621) (New York, Vintage Books, 1977), p. 174.

Introduction

tious, two luminaries, and Mercury undecided and indifferent. From this and many other similarities in nature, such as the seven metals, etc., which it was tedious to enumerate, we gather that the number of planets is necessarily seven.³

This case is different from Burton's psychology, which we can vaguely assimilate to our own modes of explanation. First of all, the style of explanation, the microcosm-macrocosm analogy, is one that we cannot assimilate to any model we currently use. But there is more. The "fact" which is being explained, the existence of seven planets, is of course no fact at all. There are not seven, there are nine, or more, depending on what you count. But the real problem lies in the very idea that this is the kind of thing that can be explained at all.

After all, suppose there are nine planets. *Why* is this so? What explanation does modern astrophysics give us for the fact? It turns out that there is no nontrivial explanation. Modern science rejects the idea of explaining that sort of thing, except by the trivial statement that that is how many there turned out to be. Here the difference is not about facts but about what kinds of facts we can expect to explain.

We can distinguish two different issues. The first concerns changes in the general form of the explanation, while the second concerns changes in the object of explanation. In the first case we see claims that one form of explanation is to be rejected in favor of another, while in the second there are shifts and dislocations in the very nature of the phenomena being explained or even in what is held to be capable of explanation at all. We need a term to refer to these modes of explanation and associated objects; I propose to call them explanatory frames. An explanatory frame is therefore a model or paradigm of a form of explanation and an object to be explained.

Answers and Questions

Perhaps the most interesting cases of changes in explanatory frames are ones in which there is a shift in the nature of the *question* being asked. Explanations are sometimes answers to explicit questions. Why is the sky blue? Why do metals expand when heated?⁴ But often there is no

3. From Charles Taylor's *Hegel*, p. 4. Taylor is quoting from S. Warhaft, ed., *Francis Bacon: A selection of his works* (Toronto, 1965), p. 17.

4. For one treatment of explanations as answers to questions, see S. Bromberger, "Why Questions," in B. Brody, ed., *Readings in the Philosophy of Science* (Englewood Cliffs, N.J.: Prentice-Hall, 1970), and "An Approach to Explanation," in R. J. Butler, ed., *Studies in Analytical Philosophy*, 2nd ser. (Oxford: Blackwell, 1965).

Introduction

explicit question at hand, and in those cases it can be very instructive to perform a kind of diagnostic inference and ask what question the explanation is really answering.

The emphasis on questions, and on ferreting out the implicit question behind an explanation, is crucial to this entire work. Attending to the questions rather than the answers and looking for the implicit question hiding behind the answer are a useful device for analyzing explanations and understanding historical shifts. In general, epochs in history, the history of science or any other history, are marked as much by the questions they ask as by the answers that they give.⁵

The first example I want to examine, from physics, concerns the shift from medieval to Newtonian theories of motion.⁶ The medieval physicists looked at an object in motion and asked, Why does it keep moving? This seemed like a natural question, and there had to be some answer to it, some kind of force that keeps the object moving. They called it "impetus".

Newton rejected such forces. But he did not offer in their place an alternative explanation for why the object keeps moving. Instead, the "explanation" he did offer was peculiar: he said that things do not need anything to keep moving, and hence that the question was mistaken. An object in motion just tends to remain in motion unless acted on by an outside force. In a certain way, this is trivial. Not as a scientific advance, for it was a major scientific breakthrough, but trivial as an answer to the question "Why does the object keep moving?" For it says, in effect, "It just keeps moving." Newton rejected this question and by doing so rejected the forces that the medievals had postulated. Even though those forces were, in the current phrase, "inferences to the best explanation," the explanatory frame that required them was rejected.

The shift to the Newtonian explanatory frame is a shift to thinking that what stands in need of explanation is not why an object is moving but rather why the motion of an object *changes*. What stands in need of explanation is acceleration, change of motion, not motion itself.

Toulmin describes this by saying that when the body is in constant

5. Marx writes (in the *Grundrisse*, trans. M. Nicolaus [New York: Random House, 1973]): "Frequently the only possible answer is a critique of the question, and the only solution is to negate the question."

6. The facts of my account are drawn from the discussion in Toulmin's *Foresight and Understanding*, the chapter entitled "Ideals of Natural Order" (New York: Harper & Row, 1961).

Newton
stated
relativistic
vs
absolute

Introduction

motion, the "body's motion is treated as self-explanatory."⁷ The statement that the motion is self-explanatory and the statement that it is explained only trivially amount to the same thing, that the explanation takes the form of saying that something happened because nothing prevented it from happening. By themselves, such explanations tell us nothing. It does not help us to be told that Saturn has rings because nothing happened to prevent it from having rings.

The role of such vacuous explanations is not to stand on their own as independent, informative statements, but rather to signal us that we have reached the outline of the explanatory frame we are using. It tells us what sorts of things we try to explain, and in what ways. If an airplane crashes, we ask why and expect an answer. But suppose flight 123 is a normal, routine flight and arrives safely. If we ask, Why didn't flight 123 crash?, there is no answer except: because nothing happened to make it crash. What we are saying is that we do not explain safe flights the way we explain crashes.

The second example of a shift in explanatory frame is one from evolutionary biology. Aristotle wondered why we have the species that we do. That is, if we look at the species that exist, they are an odd lot. There are, for example, porcupines and giraffes but no unicorns. Why are there no unicorns? The set of actually existing species forms a hazardous subset of the set of all possible species. It becomes natural, in a certain frame of mind, to ask why this or that species was or was not actualized. Why these and not those?

Aristotle wanted a genuine answer to this. He rejected as unscientific the view that species are generated randomly ("by chance") and then either survive or do not. If a particular species exists, there has to be some nontrivial answer to the question of why it exists. This leads him to the conclusion that "it is plain then that nature is a cause, a cause that operates for a purpose."⁸

We are inclined to think Aristotle naive, or prescientific, and to feel self-congratulatory about the "the modern theory of evolution," but we should first ask what answer modern biology *does* give to Aristotle's question. Why are there no unicorns? It turns out that there is no real answer given, at least no nontrivial one. Mutation and natural selection does not tell us why there are no unicorns; it just says that there happen

7. *Ibid.*, p. 55.

8. *Physics* 199 b 32.

x cause = A from bios. 1.1.10 p. 8
defining value. Not v. 1.0.1
is Ho. def. v. 1.1.1

never to have been any. This is different from the case of dinosaurs, in which there is a nontrivial answer to the question of why they do not exist: the environment could not support them, or something like that. It is also different from the question of why there are no flying horses, for there is also a real answer to that: flying horses are mechanically impossible. But with unicorns there is no such answer.

Darwinian biology simply does not answer Aristotle's question. The scientific advance that Darwin made can partly be seen as a rejection of that question and the substitution of a different question, namely: *given* that a species comes to exist (however it does), why does it continue to exist or cease to exist? That is precisely not the question of the origin of species but rather why species survive. This question is given a non-trivial answer. And so, once again, the shift from one explanatory frame to another consists of a shift in the question.

Questions and Purposes

The examples given above should illustrate the importance of a sense of the question in understanding historical developments. Such a sense is also important for understanding explanations here and now, for they exhibit a similar kind of relativity. The variety of potential questions that can be asked produces a relativity of possible explanations. This can give rise to misunderstandings, cases where it *looks* like people are disagreeing about the correct explanation of something but where they are really answering different questions.

A couple was once discussing in my presence the reasons for the breakdown of their relationship. Various factors were offered as the explanation, and fairly soon it became obvious that there were a number of different questions that were being argued at cross purposes. The last straw for the couple had been a fight they had had after one of them was involved in an auto accident. There was mention of the accident itself as the cause of the crisis, but people have accidents all the time without causing breakdowns in their relationships. Therefore, we must distinguish the question

What brought on the crisis? (the auto accident)
from the question

What caused it to precipitate a crisis?

This is only the beginning. Other questions must be distinguished, the

answer to each of which could claim to be "the" explanation of the breakup:

Why did the fight over the accident lead to the breakup of the couple? *was due to*

Who started the fight? *was due to*

Whose fault is the general situation?

What could change it now?

What should these people learn for the future from this?

The answer to every one of these questions can, in one context or another, be called the explanation of the breakup.

As we begin to realize the multiplicity of questions that can be asked, it is natural to wonder how we could ever choose among them. Looking at this example, we can begin to see certain themes. The most basic differences among the questions are the differences in their practical point of view: they are oriented toward different purposes. For example, the answer to the first question, What brought it on?, may be "the auto accident." But it may be *pointless* to dwell on this fact; there may be no purpose served by that question, no future in it. Things like that happen, we would say; the question is, Why did it have that effect? Here the shift in question is being urged for a practical reason.

Sometimes there are whole classes of questions which are practically useless in the way that dwelling on the auto accident can be useless. Certain ways of questioning may focus on the wrong aspects of the situation or be the wrong questions to ask. This is one way in which value considerations enter into the choice of explanation (see chaps. 5 and 6 below). One explanation may be better than another because it lends itself to practical use better than the other.

Perhaps the simplest kind of case is the one where the requirement is simply that the explanation *be* pragmatic. We sometimes reject a particular form of explanation because it gives us no practical handle on the situation. This is the position that B. F. Skinner takes toward Freudian explanation in terms of an "unconscious." His claim is not that there is no such thing as the unconscious but rather that explanations in terms of it are *useless*: "The objection to inner states is not that they do not exist, but that they are not relevant in a functional analysis."⁹ The criteria for what goes into a "functional analysis" are basically practical.

9. *Science and Human Behavior* (New York: Macmillan, 1953), p. 35.

A of give 57102: paradigm 7.1.5.4
1.19.6.1.4.2

e.g. could in multi-tasking, was of definition

