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 conine 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.

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 melancholy 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

---

Introduction

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

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 priors 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-


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 were tedious to enumerate, we gather that the number of planets is necessarily seven.

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? But often there is no


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

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 haphazard 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 naïve, 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

\(^5\) 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).

\(^7\) Ibid., p. 55.

\(^8\) *Physics* 199 b 32.
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 nontrivial 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 breakups 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?
Who started the fight?
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.

Therefore he dismisses explanation in terms of inner mental states (as well as explanations in terms of physiological states of the nervous system) as “of limited usefulness in the prediction and control of specific behavior” (p. 29). This, in turn, is true because any explanatory factor is “useless in the control of behavior unless we can manipulate it” (p. 34).

This is a quite particular view of the relation between explanation and practical control, in a sense the most extreme view. There are several faults with it. First of all, the relation between practical control and goodness of explanation is not as straightforward as Skinner has it. The Copernican, heliocentric view of the solar system does not give us manipulable factors or practical control any more than the Ptolemaic system does. Explanations in geology—say, of the formation of continents—do not give us the ability to predict or control, but they are good explanations for all that.

The second fault with Skinner’s extreme pragmatism is that it is not at all clear that explanations in terms of the unconscious really do fail his practicality test. Such explanations are of practical help, at least sometimes. Classical psychoanalytic explanations sometimes help people change their behavior, and so it is simply a mistake to dismiss them on the grounds that they have no practical or therapeutic consequences.  

**Questions and Objects**

The example of the couple breaking up teaches a basic lesson: We need to pay more attention to what exactly is being explained by a given explanation. Too often, theories talk as if they are addressing some problem, though they are really addressing different problems or different aspects, interpretations, or readings of the problem. For when a theory talks about a phenomenon, it inevitably does so in terms of its own representation of it. The phenomenon gets incorporated into the theory in a particular way, structured by a definite set of assumptions and presuppositions about its nature. This makes it very important that we recognize those presuppositions and discover how the theory has represented a particular object of explanation.

---

10. It is a very interesting question to ask how this is so. How exactly does the psychoanalytic explanation of my behavior enable me to change it? How does finding out, for example, that something I am doing is “really” an expression of hostility toward my father help me to change what I am doing?

11. Goethe wrote, in *Maximen und Reflexionen*, “Mathematicians are like Frenchmen: whatever you say to them they translate into their own language and forthwith it is something completely different” (cited by Morris W. Hirsch, *Differential Topology* [New York, Springer-Verlag, 1976], p. 169).
people are having in them. We might be led to think about this as the problem of why some people “do well” in school and other people “do poorly.” We might then be led to think about it as a problem of “intelligence.” A bit later we would find ourselves in the middle of a discussion of IQ, talking about what factors influence it and what its distribution in the population is. From there it is a short step to talking about IQ rank correlations when controlled for SES, and to ask how “predictive” or how “heritable” they are.

The relation of any of these technical formulations to the original problems is far from clear. It is time to start reexamining the technical concepts of the social sciences to see what their presuppositions are. There is an ultimate sense in which the definition of the problem must be in pretechnical, human, terms. The spread of science, especially social science, has effected a revolution in which the influence of human concerns no longer shows itself in the shape of the theory. In a way it is a kind of Copernican revolution: a decentring of human concerns relative to the scientific scheme of things. But there is a sense in which the Copernican revolution was mistaken: people are the center of the universe, at least in the sense that ultimately human concerns shape physical theory. Physical theory ultimately revolves around us, even if the planets do not.

The sense of this has been largely lost in recent theorizing in the social sciences. It is very hard to recognize the objects of our concern in the technical terms of modern social science. It is time for a humanist counterrevolution, reasserting the primacy of our pretheoretical, ethical, concerns. What is needed is a critical philosophy of explanation. Its point would be to give us an understanding of what the objects of explanations are, what we want them to be, what forms of explanation are appropriate to those objects, and how various explanations fit together, excluding or requiring one another, supplanting one another historically, presupposing one another.

Reductionism
One of the deepest relations that one explanation can have toward another is that of reducibility. The reductionist claims that one class of phenomena, more or less well explained by some body of theory, is really explainable by some other theory, which is thought of as deeper or more basic. This, we might say, reduces the apparent complexity of the world.

Some of the most basic claims of science are to be found in examples of reduction. Is all human behavior reducible to the working out of unconscious sexuality? This is a simple example of a reduction. So is the claim to explain all human behavior in terms of stimulus conditioning. Many social theorists (I think wrongly) cite Marx as the source for their view that all social phenomena are reducible to economics. Regardless of who held it, it is an important reduction to understand. Other reductions, also influential, have been based on biology and seek to explain social phenomena as the working out of various biological imperatives. Examples of this range from the social Darwinist and Malthusian social theories of the 1800s to the contemporary discussion about biologically based “aggression” or “territoriality” or the recent sociobiology.

The pull of reductionist views is very strong. They give us a kind of understanding that we regard as profound. When Newton demonstrated that terrestrial phenomena, like falling bodies, and celestial phenomena, like planetary motions, could be brought under a single set of laws, the effect on the general world view was profound. For before Newton, no two things could be more different than leaden weights falling from earthy towers and the patterns of the heavens. Newton changed this. The same sort of conceptual power, the ability to change the way we see a large class of phenomena, makes reductions very attractive, be they physical, biological, economic, sexual, or any other kind.

In this work I examine reductionism from the point of view of the theory of explanation. Does the reducing theory in these various examples really give us explanations of the phenomena? In order to answer this, we will have to look more closely at the explanations that are being offered, but we shall also have to examine the notion of explanation itself. What exactly is it that we are looking for when we seek these kinds of reductive explanations?

One answer is that we are looking to go beyond the ordinary explanations we give of phenomena. The power of Marxist or Freudian or sociobiological explanations is precisely that they give us a radically new view of what is “really” going on in what we thought was a familiar realm. The ordinary phenomenon is displayed transformed by the reductionist explanation. This fact, which gives power to reductionist explanations, is also responsible for the most basic problem: Are the new phenomena explained by the reducing theory really the same phenomena as the familiar ones? The reducing or underlying theory is supposed to explain the same phenomena as the reduced or upper-level theory. This presupposes
that one explanation is an explanation of the same phenomenon as the other.

But is the "aggression" that sociobiology seeks to explain the same aggression we find in war or civil strife? Is it the same aggression that Freud talked about? Is the "social stratification" of the recent IQ theorists the same stratification that economics and sociology try to explain? Is it the same as the stratification Marx speaks of?

Sometimes, as we shall see, the answer is no, and when it is, a simple-minded reductionism will be untenable. But the surprising thing is not that the answer is sometimes negative but that this essential question is usually not even asked. Writers on this subject make claims which turn on such questions, yet they use the notions of explanation and its objects in an unreflective and uncritical way.

The question of when two explanations are explanations of the same phenomenon is another of the basic problems in sketching our algebra of explanations. It is fundamental for understanding such earlier questions as when two explanations are inconsistent with each other and when one explanation supplants another. Since reduction involves the notion of explanation across theories, the problem of the identity of the objects of explanation is crucial here.

Individualism
In one of the most basic forms of reduction a theory of one realm of phenomena is reduced to an underlying theory whose objects are the physical constituents of the objects of the first theory. Such a reduction tries to explain the phenomena of one "level" by appeal to the theory of what the things on that level are made of. Following standard practice, we shall call such reductions microreductions.

The paradigm of microreduction in physical science is the reduction of thermodynamics, the theory of the observable properties of gases (temperature, pressure, and so forth), to statistical mechanics, which postulates that the gas is made up of certain kinds of molecules. One then derives the higher-level laws from the lower-level mechanical assumptions. This is often taken as a paradigm of microreduction, not just in physical science, but in social theory as well. In general its strategy is to explain the upper-level phenomena by showing how they arise from the interaction of the atomic constituents. This is an extremely powerful form of explanation and has had an enormous influence.

When this paradigm is applied in social theory, we get the various forms of individualism. What they all have in common is the idea that the characteristics of society can be explained as arising from the characteristics of individuals, just as the characteristics of the gas can be explained as arising from the properties of its molecules. What the specific theory of the individuals is, of course, varies from case to case. It may be the psychoanalytic theory of individual psychology, as in Freud's Civilization and Its Discontents, or it may be a biological theory of the formation of individual characteristics, as in the recent sociobiology. It may be the commonsense explanation-by-reasons of the actions of individuals, as in traditional narrative history, or again it may be the explanation of individual choices of the kind represented by rational-choice theory in economics.

The essence of these is the methodological form of explaining social phenomena from individual phenomena. This is the guiding methodology of much contemporary social theory, the subject of much debate; under the name methodological individualism, it has been discussed both by philosophers and by social scientists. Clear endorsements of methodological individualism can be found in Hobbes, Locke, Adam Smith, Mill, and Weber among the classics, and in contemporary social scientists like Kenneth Arrow and George Homans. In contemporary social theory, it is the guiding methodology of such work as:

Theories of the market that seek to explain economic phenomena as arising from the sum of individual choices (Milton Friedman, F. A. Hayek)
The work of social-choice theorists in attempting to reduce the problem of collective choice to the theory of individual choice (K. Arrow, Mancur Olson, A. K. Sen)
Attempts to reduce the problem of distributive justice to the justification of individual "holdings" (Nozick)

12. Social theory is the term I am using indiscriminately to refer to social science, social philosophy, and their various mixtures. Obviously, the use of this term suggests that I do not think that a "normative vs. positive" distinction can be usefully drawn. In fact, all the examples partake of the nature of both.

13. There are, therefore, cases of microreduction that are not atomistic, for example, a microreduction where the underlying level is a continuous medium. Such cases will not concern us here.

Theories of political representation which seek to construct an overall political decision out of the individual preferences (Buchanan and Tullock, Downs)

The methodological issue has been the subject of much debate; yet, for all that, the debate is of little help in understanding the controversy. Partly, this is because the debate is muddled by an astonishing variation in the theses called "methodological individualism." We have, for example: theses about what kinds of entities are "real" (or "really real"), theses about how we know about social phenomena, theses about the "derivability" of certain kinds of laws from others, and theses about what kinds of explanations exist or are "ultimate" ones. Writers use these variants interchangeably and will often shift from one to another in the course of an argument.\(^{15}\)

This lack of clarity is especially alarming because if there is one thing that everyone agrees on, it is that the debate is of more than academic interest and that important moral and political issues are lurking in the background. Karl Popper's *The Open Society and Its Enemies* is a two-volume blast at all nonindividualistic social theory. He says that failure to grasp methodological individualism leads not only to philosophical and scientific error but to moral and political evil. Plato, Hegel, and most of all Marx are the examples he cites of what can happen if we are not individualistic; the consequences are totalitarianism. Nonindividualists, he suggests, believe in trampling, in thought and deed, on individual liberty. On the other hand, other writers have pointed out that individualism is itself not without ideological consequences and presuppositions.

The main focus of this work is the assessment of individualism, both as a general philosophy of social explanation and as a guiding methodology. I join the controversy because I think it is a chance for philosophy to be useful in the analysis of real scientific and social issues. The principal strategy will be to attack the problem via an examination of the concept of explanation itself. Methodological individualism, like all species of reductionism, consists in a claim that certain kinds of explanation are available, that the theory of individuals explains the phenomena which were previously the province of the upper-level social theory.

\(^{15}\) Lukes's survey article "Methodological Individualism Reconsidered," in A. Ryan, ed., *The Philosophy of Social Explanation* (Oxford: Oxford University Press, 1973), distinguishes eight importantly different theses, including some, but not all, of the above!


that did not conform to this model was either defective or "not really scientific."

This is not the place for a critique of this doctrine. I will have more to say about it in chapter 5. It is not my goal to offer a detailed criticism of the positivist model. Rather, I want to move beyond it, to take a position in post-positivist philosophy of explanation, via a consideration of the kinds of questions I have posed in this introduction. The formal, positivist model gives us either no answers at all to those questions or answers that are just false.

1 Explanatory Relativity

When Willie Sutton was in prison, a priest who was trying to reform him asked him why he robbed banks. "Well," Sutton replied, "that's where the money is."

There has been a failure to connect here, a failure of fit. Sutton and the priest are passing each other by. The problem is to say how, exactly, they differ. Clearly there are different values and purposes shaping the question and answer. They take different things to be problematic or stand in need of explanation. For the priest, what stands in need of explanation is the decision to rob at all. He does not really care what. But for Sutton, that is the whole question. What is problematic is the choice of what to rob.

We could say that Sutton and the priest have different notions of what the relevant alternatives to bank robbing are. For the priest, the relevant alternative to bank robbing is leading an honest life, not robbing anything. But for Sutton, the relevant alternatives to bank robbing are: robbing grocery stores, robbing gas stations, and so on. What Sutton is really explaining is why he robs banks rather than robbing grocery stores, etc. We could say that the priest has asked why Sutton robs banks and Sutton has answered why he robs banks!

The difference between them is that they have two different contrasts in mind, two different sets of alternatives to the problematic: Sutton robs banks. They are embedding the phenomenon to be explained in two different spaces of alternatives, which produces two different things-to-be-explained, two different objects of explanation.

The object of explanation here is therefore not a simple object, like an event or a state of affairs, but more like a state of affairs together with a definite space of alternatives to it. In the Sutton case the priest's object is

\[
\text{Sutton} \begin{cases} \text{does not rob} \\ \text{robs} \end{cases} \text{banks,}
\]
whereas Sutton’s object is

\[
\begin{align*}
\text{Sutton robs} & \quad \text{other things} \\
\text{banks} &
\end{align*}
\]

Clearly, if the same event is embedded in two different contrast spaces, the answers to the two different questions so generated will not necessarily be the same, and will often be different. Many jokes, like the Sutton joke, have as their structure a question and answer having different presuppositions, and often this will take the form of a dislocation from one contrast space to another. Children’s jokes make use of this device in such classics as: Why do firemen wear red suspenders? (To keep their pants up); and: Why do ducks fly south in the winter? (It’s too far to walk). Sometimes, as with Sutton, the answerer answers a question that is much narrower in scope than the intended question. Sometimes it is the reverse, as the answerer answers a very general interpretation of a narrow question. The detective, questioning a suspect about a murder, asks the suspect, Why did he die? The suspect tentatively suggests, Well, everyone has to go sometime, sir. Here, the suspect is dodging the detective’s “real” question. The explanation is formally an answer to the question, but what is really being answered is the very general question why

the victim \{ lived forever \}

\{ died \}

whereas the detective’s question was really why

the victim \{ died at some other time \}

\{ died \}

The effect of such differing spaces of alternatives is not always a joke; what aspect of a given state of affairs we take to be problematic radically affects the success or failure of potential explanations. For an explanation to be successful, it must speak to the question at hand, whether explicit or implicit, or else we will have failures of fit like Sutton and the priest. What we need, therefore, is some way of representing what is really getting explained in a given explanation, and what is not. The contrast spaces give us such a representation of one basic way in which explanation is “context relative.” My claim is that this relativity-to-a-contrast-space is quite general; I will call it explanatory relativity.

Once sensitized to this phenomenon, one can easily find examples of

it, for there are many cases in which explanations are rejected as belonging to the “wrong” contrast space. For example, at one point in his analysis of dreams, Jung is discussing a dream about an auto accident:

We reduce the dream-picture to its antecedents with the help of the dreamer’s recollections. He recognizes the street as one down which he had walked on the previous day. . . . The car accident reminds him of an accident that had actually occurred a few days before, but of which he had only read in a newspaper. As we know, most people are satisfied with a reduction of this kind. “Aha,” they say, “that’s why I had this dream.”

Obviously this reduction is quite unsatisfying from the scientific point of view. The dreamer had walked down many streets the previous day; why was this one selected? He had read about several accidents; why did he select this one?

The complaint is that a certain contrast is crucial for a successful explanation, and that some would-be explanation fails to account for that contrast. A good example is furnished by Meyer Shapiro’s “Nature of Abstract Art,” which considers and rejects the standard forms of explanation for the rise and fall of artistic styles: explanations that appeal to “the exhaustion of possibilities” in earlier styles and “pendulum swing” theories. His complaint is: “From the mechanical theories of exhaustion, boredom and reaction we could never explain why the reaction occurred when it did” (p. 190). Later he criticizes one such explanation for the rise of Futurism, saying that it “makes no effort to explain why this art should emerge in Italy rather than elsewhere” (p. 208).

On a more abstract level, Aristotle complains that atomists try to explain movement by postulating an eternal motion of the atoms. But, he says, this is a poor explanation because it does not explain why things move one way rather than another. And, at the other extreme, explanations in everyday life also reflect this relativity, as advertising slogans urge: “Don’t ask me why I smoke. Ask me why I smoke Winstons.”

3. “Why and what this movement is they do not say, nor, if the world moves one way rather than another, do they tell us the cause of its doing so” (Metaphysics 1071 b 33).
Yet despite the frequent occurrence of explanatory relativity, not much attention has been paid to the general phenomenon. There are several good examples in a recent paper by Fred Dretske. Here is one of them:

Suppose Alex, after being fired, needs some money to meet expenses until he finds another job. Clyde lends him $300. It seems fairly obvious that there are three different questions (at least) that we can ask with the words "Why did Clyde lend him $300?" and, accordingly, three different explanations one can give for Clyde's lending him $300. We may want to know why Clyde lent him $300. The answer might be that this is how much Alex thought he would need; or perhaps, though Alex wanted more, this is all the ready cash that Clyde had available. On the other hand, we may want to know why Clyde lent him $300—why didn't he just give it to him? . . . Finally, we may be interested in finding out why Clyde lent him $300. 4

Here, the three different explanations can be represented as the answers to the questions:

1. Why Clyde lent Alex
   \[
   \begin{cases}
   \{ \text{some other sum} \} \\
   \text{gave Alex $300 and} \\
   \text{ lent} \\
   \text{someone else lent Alex} \\
   \text{Clyde}
   \end{cases}
   \]

2. Why Clyde
   \[
   \begin{cases}
   \text{gave} \\
   \text{lent}
   \end{cases}
   \text{Alex $300 and}
   \]

3. Why
   \[
   \begin{cases}
   \text{someone else} \\
   \text{Clyde}
   \end{cases}
   \text{lent Alex $300.}
   \]

Dretske observes that the differences among various stresses is essentially of a pragmatic nature and says that examples like these "constitute serious obstacles to any attempt to formulate a purely syntactical characterization of explanation."

This is true. But then how should we characterize it? Dretske represents the variations by using contrastive stress, the linguists' term for the device of underlining (or vocally stressing) part of the sentence. But the voice or the underline is a symptom of whatever is going on here, not an analysis of it. And in addition to not being an analysis, it cannot represent the more general forms of explanatory relativity, since it is obviously limited to cases where the problematic consists of a variation in one of the explicit syntactic parts of the sentence. If the aspect being varied is not a syntactic part of the sentence, contrastive stress cannot represent the explanatory relativity. For example, take the case of Clyde lending Alex $300, and suppose that we know that Alex is the sort who, if he could not raise the money from friends, would take Willie Sutton's advice and rob a bank. Clyde knows this too and so lends his friend $300 to keep him from getting into trouble.

If we then asked, "Why did Clyde lend Alex $300?" and received the answer: to keep his friend out of jail, how are we to represent the "real" question? It is not:

1. why Clyde (rather than Bob) lent Alex $300
2. why Clyde lent (rather than gave) Alex $300
3. why Clyde lent Alex (rather than Phil) $300
4. why Clyde lent Alex $300 (rather than some other sum)

but rather why Clyde lent Alex $300 (without emphasis) rather than letting him rob a bank. The contrast space is the only possible representation in cases like these.

The Algebra of Explanations
Let us begin, then, with the simple idea that an explanation always takes place relative to a background space of alternatives. Then different spaces of alternatives may therefore require different explanations. And sometimes we can compare two explanations to see how their contrast spaces differ. This gives us a measure of the dislocation between two explanations.

Contrast spaces therefore give us a useful tool for comparing explanations with each other. In particular I want to go back to the examples of the introduction and show how sometimes a shift from one explanatory frame to another is just a shift in the relevant contrast space. Consider the case of Newton vs. the medieval physicists: The medievalists asked why something keeps moving. Their object can now be represented as asking why

\[
\text{the thing is} \begin{cases}
\text{moving} \\
\text{not moving}
\end{cases} \text{ at } t.
\]

In order to answer this, they had to postulate a force acting at each time.
Explanatory Relativity

We are now in a position to represent the epistemological break that Newton achieved.\(^5\) It is to reconstitute the object of explanation as asking why

- the thing has \{ given acceleration \}
- some other acceleration \}

at \( t \).

For this object the only nontrivial explanations are explanations of changes of motion, that is, of accelerations. So the break from the medievals to Newton can be represented as a shift in the contrast space of their explanations.

The same things can be said about the other example of that section, the shift from Aristotelian to evolutionary biology. Aristotle asked why there are the species that there are. We can now state the object of his question: he was asking why these species exist \( \text{rather than} \) the other possible species, that is, why

- \{ other possible species \}
- \{ these species \}

exist.

The shift to the Darwinian question is the shift to explaining \( \text{not the origin of species but rather their survival} \). In other words the object becomes why

- \{ these species \}
- \{ become extinct \}

exist.

This representation makes a little clearer what is and what is not getting explained in a given explanation. And this, in turn, means that the problems of explanation discussed in the introduction are sensitive to the contrast space phenomenon.

Consider, for example, the problem of deciding if pretheoretical questions are really being answered. Certainly a necessary condition for a theory to be a real answer to a pretheoretical question is that it embody

\[^5\text{The terminology ("coupure epistemologique") is taken from Bachelard. We could also have said "change in explanatory frame," or for that matter, "paradigm shift," "scientific revolution," "change in epistemology" or "aspect shift." There are embarrassingly many terms, all of vague meaning, and all meaning vaguely similar things. I choose Bachelard's term because it greatly predates the others. See G. Bachelard, The Psychoanalysis of Fire (Boston: Beacon Press, 1964), and The Poetics of Space (New York: Orion Press, 1964).}\]
each case there is more to be said about how contrast spaces enable us to represent these relations, and about how the various relations among explanations correlate with various relations among their contrast spaces.

In particular this is true of the problems of theoretical reduction. We said that one of the main requirements of a would-be reduction is that it enable us to explain “the same phenomena” as the preduction theory. But this notion of “the same phenomena” is clearly sensitive to explanatory relativity: two different contrast spaces may smuggle in a Suttoninesque ambiguity into the situation. Consequently, in evaluating all reductionist claims, we must be careful that the objects match up and that the reducing theory really does explain the same phenomena (with the same contrast space) as the reduced theory. This will be one of the main tools I will use in the next chapter, in the analysis of reductionism.

Presuppositions of Explanations
The general claim of explanatory relativity is the claim that explanation takes place relative to a contrast space. I mean this as a claim about how to explain explanations: the contrast space is a basic presupposition of the explanation context, an additional piece of structure necessary to explain how explanations function.

Without some such hypothesis, I do not think we can give an account of how explanations actually work, or fail to. The contrast space determines, in part, what counts as a successful explanation. I want to examine this more closely, especially the idea of what really does get explained, and what does not, in a typical explanation. For there is an important way in which certain things really do not get explained. Now of course any explanation leaves something unexplained, and in particular any part of the object of explanation that is outside the contrast space will not be explained.

But there is something more. Look at the explanation that Sutton gave. The question was: why do you rob banks? and the answer was: they have the most money. He is saying, in other words, that the fact that something has the most money

explains why

Sutton robs that thing.

Now we might, especially if we were sympathetic to the priest, ask why

the first thing explains the second thing. Why, after all, does the fact that something has the most money explain why Sutton robs it? Note that it follows from the form of his explanation that, since it is (in some sense) necessary that something (or other) has the most money, therefore, necessarily, Sutton robs something. The fact that Sutton robs something or other follows from the form of the explanation.

But this is just the thing that is bothering the priest. He wants to know why rob anything at all, a question Sutton has not answered. But it is not just that he has not answered it. Rather, Sutton’s answer and the fact that it is taken to be an answer (i.e., an explanation) indicate that Sutton is presupposing a satisfactory answer to the priest’s question. He is presupposing it in the simple sense that his answer does not even make sense unless one supposes the priest’s question already to have been answered satisfactorily. Someone for whom the priest’s question still lingers is not someone who can accept that Sutton’s answer is an answer at all, for the implicit question which Sutton is answering is

Given that you are going to rob something, why do you rob banks?

So the underlined phrase is a presupposition in the straightforward sense that whether or not one accepts it affects the success of the explanatory act.

Looking at how such “given” clauses function in explanations gives us another view of the phenomenon of explanatory relativity. The “given” clause often (but not always) functions to express the same presupposition as the contrast space. Roughly speaking, the question

Given A, why B?

is equivalent to the contrast

Why B rather than any of the other alternatives to B in which A is true?

The “given” clause tells us, at the very least, what the outer bound is on the variation in B: we are to consider only such alternatives to B as also satisfy A. Why must this be so? Why is it that explanations limit their alternatives in this way? Why do we have explanations of why X rather than Y, or why A, given B, rather than simply explaining why X or why A?6

6. Notice, incidentally, that this very question has the form “Why P rather than Q?”
Explanatory Relativity

The answer, I think, lies in our need to have a limited negation, a determinate sense of what will count as the consequent’s “not” happening. Lacking such a determinate sense of alternatives, one has difficulty seeing how we could give explanations at all; they would have to be so all encompassing as to be impossible.

Let me give an example. Suppose that I got up one day and went out for a drive. I was doing about 110 when I rounded a bend, around which a truck had stalled. Unable to stop in time, I crashed into the truck. Later, chastising me for the accident, you say, “If you hadn’t been speeding, you wouldn’t have had that accident.” I reply, “Yes, that’s true, but then if I hadn’t had breakfast, I would have gotten to that spot before the truck stalled, so if I hadn’t eaten breakfast, I wouldn’t have had the accident. Why don’t you blame me for having had breakfast?”

What’s wrong with my reply? It is based on the truth of a causal conditional: if I had not had breakfast, I would not have had that accident. But while it is true that I would not have had that accident, nevertheless, if I had been speeding then it is likely that I would have had another accident. My claim is based on the assertion that if something (eating breakfast) had not happened, the accident would not have happened. The problem is, what is going to count as that accident’s not happening? If “that accident” means, as it must if my statement is going to be true, “that very accident,” that concrete particular, then everything about the situation is going to be necessary for it: the shirt I was wearing, the kind of truck I hit, and so forth, since if any one of them had not occurred, it would not have been that accident.

But this is absurd, and in order to escape this absurdity and not have everything be necessary for the accident, we must recognize that the real object of explanation is not my having had that accident.

We need something in addition to explain what is really getting explained, something that will account for the fact that my objection somehow misses the point. For not any difference from that very accident is going to count as relevantly different, only certain ones will. And so we need, in addition to the event, a set of perturbations which will count as irrelevant or inessentially different. These irrelevant perturbations determine an equivalence relation, “differs inessentially from,” and the real object of explanation is an equivalence class under this relation. The equivalence relation determines what is going to count as the event’s not happening.

If we consider the auto accident we can see that, for the usual purposes, many “alternatives” will be considered as “not essentially different” under this equivalence relation. Having had a similar accident in a similar place driving a similar car but wearing a different colored shirt, for example, will count as “not essentially different.” The equivalence relation determines the object of explanation in the sense that the explanation we seek must not dwell on the factors which select among equivalent states of affairs but only on factors which are responsible for the accident’s happening rather than some inequivalent state of affairs.

Of course, the exact specification of this equivalence relation, the specification of what exactly is going to count as relevantly different, is not given in advance or once and for all. The relation can be drawn tightly or loosely. In the auto accident case I am trying to evade responsibility by drawing the relation very tightly, thus making a great deal necessary for it to have occurred. I am prepared to say that an accident in which I was wearing a different shirt counts as “the same,” and so I would not try to use the shirt as a necessary cause; I have some nontrivial conception of what counts as “the same” accident. But I have chosen (implicitly) a specific object of explanation. My detractors, blaming me for the accident, are insisting on a different equivalence relation. For them, the question is why I had that accident rather than not having had any accident. In other words they count my having had another accident just down the road as something which is essentially the same as what actually happened.

So each equivalence relation lays a grid or mesh over the possible phenomena, and corresponding to each mesh there is a conception of what the object of explanation is. We can, as in the auto accident, insist on one mesh or another. But the choice is not entirely arbitrary. There is a general fact which must be taken into account. As the mesh becomes finer and finer, that is, as the equivalence classes become smaller and more numerous, the resulting object, and hence the resulting explanation, becomes less and less stable. That is, the explanation:

7. Each equivalence class consists of the set of “inessentially different” objects, collapsed into one for this purpose.
Explanatory Relativity

reckless driving causes accidents (somewhere or other)
is highly stable under all sorts of perturbations of the underlying situation: the weather, road conditions, etc. We can perturb them almost at will, and the causal relation remains. On the other hand the explanation

reckless driving and breakfast and ... causes accident at x, t, ...
is highly unstable under these same perturbations. This is sometimes obscured by the fact that we call auto crashes "accidents." Sometimes a crash is really accidental, that is, is highly unstable with respect to its antecedents. But other times, as in this case, accidents are not accidental at all, even though we can (perversely) cast the object of explanation so as to make it seem accidental.

The general need to have an object of explanation which is somewhat stable under such perturbations means that the choice of object is not entirely arbitrary. Each equivalence relation determines the kind of object, and hence a theory, in much the same way as Felix Klein defined a geometry as arising when we specify an equivalence relation. Any equivalence relation, or sense of what is essentially the same as what, gives us a new set of objects. We study the features which are invariant under the various perturbations. Each equivalence relation therefore gives rise to a different "geometry." But which geometry is the "right" one? Clearly, there are some pragmatic, practical factors at work. Yet the situation is not completely determined by these factors, for these practical demands must be reconciled with the nature of the phenomena themselves and with the stability demands of good scientific explanation.

So the answer to the question, Are irrelevance-geometries stipulated, or are they "in the world"? is: both! We can stipulate equivalences at will, but the result will be a good explanation or a good piece of science only if the way we are treating things as inessentially different corresponds to the way nature treats things as inessentially different.

Let us grant, then, that when we explain an event not everything about it is essential to it. The explanation has to be stable in some neighborhood of the actual world. The problem now is: How large a neighborhood? What are its boundaries?

Clearly, the particulars of a specific explanation can be perturbed substantially while the explanation retains its force. If the explanation for why a leaf is green is that it contains chlorophyll, then the causal form:

contains chlorophyll causes green color

holds generally in a wide class of circumstances. But not in all circumstances: jars which contain chlorophyll are not caused thereby to be green. So obviously, there is a presupposition working here, to the effect that this case is the kind of case for which "contains chlorophyll" explains "greenness." Again, if you sought to explain why something moved as it did by citing Newton's laws of falling bodies, you are presupposing that the thing in question is the kind of thing for which those laws hold, namely, a physical object, etc. If it turns out that the moving thing was a shadow on the side of a building, you must withdraw or amend your explanation, for we have passed out of the realm for which such explanations hold.

Each time, there is a presupposition that this case lies inside the domain of validity of the explanatory form, that is, that it is the kind of thing for which such an explanation can hold. Now this gives us a kind of test for the presuppositions of an explanation: see how large a neighborhood of the actual situation will maintain the validity of the explanation. The outer boundaries of that neighborhood will represent the presuppositions of the explanation. In the simple cases above, those presuppositions are, respectively, that we are dealing with a plant, and that we are dealing with a physical object.

The presuppositions become much more complex when we pass to more difficult cases. Recall, for example, the explanation of why Nixon got the Republican nomination: all the other major candidates had alienated some faction or other in the party. Certainly this is only an explanation if we are presupposing that the four major candidates were the only possibilities; but something more is true. Suppose someone said, "Well, I understand that the other candidates had offended sections of the party. But Nixon is so awful; why did they nominate anyone at

8. Alternatively, it could be suggested that the problem turns on the ambiguity of the word "contains." If plants contained chlorophyll like jars do, they would not necessarily be green. But this is really no different from what I am saying, which can be put as the presupposition that we are talking about "contains" in the sense appropriate to plants, what the word "contains" means in the category of plants.
all? Why not just pass? Or why didn’t they compromise by nominating
the set of the four of them to run junta-style?”

Such possibilities are essentially beyond the frame of the explanation
given. In our language the form of the explanation was something
like: given that exactly one person receives the Republican nomination,
why was it Nixon rather than Goldwater, and so forth? Once we
wander outside the boundaries of the given clause, the explanation
collapses. The domain of validity of this explanation includes only those
situations in which exactly one person receives the Republican nomi-
nation.

The same sort of thing is true of the Sutton case. Since the form of
his explanation

X has the most money explains Sutton robs X,

it is clear that this explanation is valid only in situations in which Sutton
robs exactly one (kind of) thing. The fact that Sutton robs exactly one
thing, therefore, is true of every possibility envisaged by the presup posi-
tion. It is a pure consequence of the presupposition and is therefore not
itself explained. The same thing is true of the Republican nomination
case, in which the fact that the Republicans nominate exactly one person
is a pure consequence of the presupposition.

So every explanation must have some generality, yet it obviously can-
not have complete generality. Somewhere in the middle, then, are the
boundaries of the realm for which the explanation holds. This will vary
from case to case, and in each case the size and shape of the outer bound-
ary of contemplated possibility will reflect (part of) the presuppositions
of that explanation.

I will return immediately to these points, develop them further, and
then apply them to ferret out the presuppositions of various explana-
tions. First, I want to take a brief detour and make essentially the same
points again, from a slightly different standpoint: a consideration of the
role of laws in explanations.

Many philosophers think that laws play an essential role in explana-
tions. In fact the covering-law model of Hempel and Oppenheim says,
basically, that an explanation consists in subsuming the situation to be
explained under a law. The role of such laws in explanations is twofold.
First, the law is supposed to provide the necessary generality that an

explanation must have: to say that A causes B, Hume noted, is to say
that A and B are instances of some more general relation. Second, the
law is supposed to capture the idea of a connection between A and B,
the causal connection whose existence means that A and B are not mere-
ly accidentally co-occurring.

Now, I do not want to get into the issue of whether the kinds of laws
contemplated in the standard treatments really do capture these notions
or whether explanations really must contain laws in them at all. (Some
people think not: Scriven, Arronson, and Davidson, for example, think
that there must be a law but that the explanation need not cite it.)

All I want to do immediately is to follow the standard way of talking
about “deductive-nomological” explanations and to restate some of the
conclusions of the last few pages in that language. Let us start with a
simple example. When I explain why the plant is green by saying that it
contains chlorophyll, the law in this case is that all plants which contain
chlorophyll are green. So far, so good. Now consider the example of the
Republican nomination, which Nixon won by not alienating anyone. Is
the law here something like

Anyone who does not alienate anyone gets the Republican nomina-
tion?

There are several problems with this. First of all, it clearly applies only
to leading candidates. So we must add the qualification:

Anyone who (1) is a leading candidate

and (2) does not alienate anyone

gets the Republican nomination.

But this still will not do. Suppose two people had not alienated anyone?
What would have happened then? Here the law gives a confused answer.
On the one hand there is nothing in either of the conditions, or their
conjunction, to rule out the possibility of two people satisfying them
jointly. On the other hand it is a logical consequence of the law that it
can be true of exactly one person.

Let me take the second point first. “Gets the Republican nomination”
can be expressed more explicitly as

becomes the one and only person who gets the Republican nomina-
tion.

It is a logical consequence of the law that exactly one person gets the
Explanatory Relativity

nomination. This we have already seen. That exactly one person get the nomination is in some sense a necessary truth. Here this fact is expressed by the fact that it is a pure consequence of the law.

Why could not two people have the property of being leading candidates who had not alienated anyone? It was perfectly possible; it just did not happen to be. So the correct form of the law is therefore:

If there happens to be a unique person among the leading candidates who has not alienated anyone, then that person becomes the nominee.

And so we see that if there had been two such people, the whole explanation would have to be withdrawn because the very applicability of the law depends on the condition that there is only one such person. It must be withdrawn completely in the case where two people have the necessary properties, and we must go out to look for a whole different form of explanation in the case where that happens.

The case is somewhat different in the Willie Sutton example, where the law is something like:

If something has the (most) money, then Sutton will rob it.

What if several things had the most money? Several things cannot have the most money; it is part of the meaning of the word most that exactly one thing has the most money. In this case the fact that there is exactly one thing satisfying the requisite properties is a necessary truth, not an additional assumption about what just happened to be.

And so it does not really matter for the time being whether we speak of presuppositions of explanations, pure consequences of the law, or consequences of the meaning of the terms being used, for in any case the conclusion is that in cases like these the fact that there is a unique thing satisfying a certain property is not explained but, on the contrary, is a necessary truth in one way or another.

So much for the role of laws in explanations. I ventured into it in order to show, using concepts which are philosophically familiar, how presuppositions function to shape the space of contemplated alternatives, the phenomenon I am calling “explanatory relativity.”

9. Roughly speaking, and ignoring the possibility of ties.

Explanatory Relativity and the Philosophy of Explanation

Although, as we saw, there is no shortage of examples of explanatory relativity, there has not been very much discussion of it as a phenomenon in the philosophy of explanation. There are hints of it here and there.

Aristotle devotes the last section of book Z of the Metaphysics to a discussion of explanation. Book Z is concerned with expounding the nature of substance, the foundation concept of his metaphysics, and in the last section he says that we can “make another start” on that subject by considering the notion of explanation (attia), “for substance is a kind of principle and explanation.” He continues immediately:

Now to ask why is always why something belongs to something else (1041 a 12).

He goes on to provide a number of remarks about what various questions really mean. In each case the real meaning of the question is given by what amounts to a contrast. He says generally that to ask why is to ask “why something of something else”; he criticizes questions of the form Why A? and says that a true question has the form Why does A belong to X?

“What does it thunder?” means “why is a noise produced in the clouds?”

In each case he gives, like this one, the first element is the variable or problematic factor, and the second element is the unchanging substance. So we are asking of this substance, X, why it is A. We can add: that is, why it is A rather than not-A.

In adding this explicit contrast I am only filling in what Aristotle has said elsewhere: that substance terms do not admit contraries but that the attributes of substance do. Thus his dictum “Substance is that which bears contraries” means that the substance X can be A or not-A; the person can be musical or unmusical. The substance term itself, X, does not have a contrary. The general picture this gives us of explanation is that explanation is of some substance X and explains why X is A rather than not-A.

This brings Aristotle’s formula into conformity with what I have been

10. And it is essential that what it is is an X.
saying. The "substance" term X expresses the presupposition, that which is (taken as) fixed. The variation "A rather than not-A" expresses the limited set of alternatives. I say "limited" because the alternatives to the state of affairs

A belongs to X

are those in which

not-A belongs to X.

But these do not exhaust all of logical space because they share a common presupposition, X.

This general schema captures the ordinary cases of explanations. When we ask why the sky is blue, Aristotle would say we are really asking of the sky why it is blue rather than some other color. In contrast space language, we are asking why

the sky is \{ another color \} \{ blue \}.

This means, first of all, that the existence of the sky itself is not problematic for this explanation, and second, that what is problematic is (only) the color of the sky.

The two terms of the explanation, the substance term and the variation term, are also related in a very important, and very Aristotelian, way: the substance term tells us what the form of the explanation is going to be. It will be whatever the appropriate form of explanation is for things of that category.\textsuperscript{11} Consider the variable predicate "red," something which can be had or not had by various kinds of substances.

If we ask:

Why is X red?

even if we understand that to mean:

Why is X \{ another color \} \{ red \}?

we still do not have a determinate form of explanation until we know what kind of thing X is. There is not, in this view, one all-purpose form of explanation for why any old thing is red, be it a book or a person or a sky or a dream-image. Instead, there is one kind of explanation for each kind of thing that is red.

If we ask why a person is red, for example, the answer might be "from exertion" or "because of anger." But "anger" does not generally produce redness, only in people. Horses who are angry do not become red. Likewise if we ask why a piece of metal is red, the answer might be "because it was heated to a high temperature." But the general relation between being heated and becoming red is only a relation in the category of metals. If a fluid is heated, it does not become red.

In each case, then, when we ask why X is A, we get the answer that it is because it is B. But, in general, B-ness will explain A-ness only for the kind of thing that X is. The kind of thing that X represents is therefore a presupposition, in the sense of the previous section.

Outside the classical period\textsuperscript{12} there has not been much discussion of the forms of explanation until fairly recently, although there are some remarks in scattered places which talk about the need for contrasts of various kinds. For example, Wittgenstein speaks in the Tractatus about "logical spaces" whose structure is exactly that of a contrast space:

Each thing is, as it were, in a space of possible states of affairs. . . . A spatial object must be situated in infinite space. . . . A speck in the visual field, though it need not be red, must have some color; it is, so to speak, surrounded by color-space. Notes must have some pitch, objects of the sense of touch some degree of hardness, and so on.\textsuperscript{13}

But remarks like these are not, and they were not intended to be, part of a philosophy of explanation, and the role of such contrast spaces in explanation has not received much notice.

One interesting exception is Josiah Royce. His seminars were often devoted to the concept of explanation, with various people presenting papers on very modern-sounding topics in "comparative methodology."

\textsuperscript{11} More accurately, the substance term tells us which forms of explanation will be appropriate for which predicates. There is not just one kind of explanation for each kind of substance but several. The form of explanation of the color of the sky, for example, is different from the form for explaining why it is cloudy.

\textsuperscript{12} Cf. also Plato's discussion of the forms of explanation in the Phaedo, 97-107.

\textsuperscript{13} Tractatus Logico-Philosophicus (London: Kegan Paul, 1921), 2.013f.
(For example, the graduate student T. S. Eliot presented one on the relation between interpretation and explanation in comparative religion.) Florence Webster presented a paper on the notion of cause in biology. She argued for a notion which was "interestingly analogous to Mill's method of difference. It is, namely, that while you cannot find the cause for an event, you can find the cause for the difference between two events" (p. 133). The following interchange took place:

Costello: What is interest?
Miss Webster: Depends on choice of events you compare an event with.
Royce: Interest is objective in being determined by environment. You compare with certain other events.
Miss Webster: Depends on the context of the interesting objects.14

These contrast spaces are still not well understood objects. Their structure is not readily identifiable with any of the traditional objects of logic, for example. They have some similarities with "possible worlds," for instance, but they are not simply spaces of possible worlds. They are more like equivalence classes of possible worlds (under the relation "differs inessentially from") with almost all possible worlds excluded altogether from the space. (Contrast spaces are typically quite small.)

The basic structure of a contrast space is something like this: If \( Q \) is some state of affairs, a contrast space for \( Q \) is a set of states \( [Qa] \) such that:

1. \( Q \) is one of the \( Qa \).
2. Every \( Qa \) is incompatible with every other \( Qb \).
3. At least one element of the set must be true.
4. All of the \( Qa \) have a common presupposition (i.e., there is a \( P \) such that for every \( Qa \), \( Qa \) entails \( P \)).

Basically, these spaces are similar to what physicists call state spaces. A state space is a geometric representation of the possibilities of a system; a parametrization of its states, a display of its repertoire. In the case of a simple switch the state space has two elements:

\[
\text{switch} = \begin{cases} 
\text{on} \\
\text{off} 
\end{cases}
\]

If such a switch is connected in a circuit with another switch, one which has, say, three positions (on, off, reversed), then the total state space for the complex system is the six-element product of the two:

\[
\text{switch}_1 \text{ is } \begin{cases} 
\text{on} \\
\text{off} 
\end{cases} \quad \times \quad \text{switch}_2 \text{ is } \begin{cases} 
\text{on} \\
\text{off} 
\end{cases} \text{ and } \begin{cases} 
\text{rev} 
\end{cases}.
\]

Such a space also represents the presuppositions of the explanation, in the sense that it makes clear how much is not being explained. For example, when the object of explanation is why the

\[
\text{switch} = \begin{cases} 
\text{on} \\
\text{off} 
\end{cases}
\]

the unvaried part gives us the presupposition. In this case it would include: what a switch is, why there is a switch here at all, why the switch has exactly those two positions, and so on.

Structural Presuppositions
There is a certain kind of presupposition that arises when the explanations we seek deal with individuals who are related in a larger system. The theory of these presuppositions is the foundation for much of what I am saying in this work.

Let me begin with an example. Suppose that, in a class I am teaching, I announce that the course will be "graded on a curve," that is, that I have decided beforehand what the overall distribution of grades is going to be. Let us say, for the sake of the example, that I decide that there will be one A, 24 B's, and 25 C's. The finals come in, and let us say Mary gets the A. She wrote an original and thoughtful final.

Now, if someone asks me why Mary got an A, I would say exactly what I just said: she wrote an original and thoughtful final. Yet this is inadequate as it stands. Suppose two people had written finals that were well thought out and original. Would two people have received A's? Not if I am really grading on a curve. Because of this, it is misleading in a certain way to answer why Mary got an A by citing this simple fact about her—that she wrote a good final. It is misleading because it gives the impression that "writing a good final" is sufficient to explain "getting an A." But that is not true. Even if someone writes a good final, they may fail to get an A because someone else has written a better one. So it is more accurate to answer the question by pointing to the relative fact that Mary wrote the best paper in the class.

What is assumed here is something like:

Whoever writes the best final gets the A.

But if this is the general principle, we can see that a direct consequence of it is that exactly one person gets an A (since exactly one person can "write the best final"). There is, therefore, an unexplained presupposition that there is exactly one A in the class.

The nature of this presupposition distinguishes this case from cases in which there is no curve. In those cases, if Mary gets an A, even if she happens to be the only one to get one, we can answer the question why Mary got an A by citing factors that are purely about Mary and that contain no hidden presuppositions. In those sorts of situations the answer to the question about Mary would be something like:

She had taken a math course that was helpful

or

There was a big party the night before the final, but her phone was out of order,

or some answer from which it follows that had this been true of several people, then all of them could have received A's. In cases where there is no curve, we can explain each individual's fortune by appealing only to facts about that individual.

This fails, by definition, in the cases where there is a curve, for in those cases, there will be the unexplained presupposition, the presupposition of the grading structure, that in this case there will be exactly one A. This is also reflected in the fact that if we asked

Why didn't Bob get an A?

it would be legitimate to answer

because Mary's paper was better.

Here the presupposition is obvious.

Generally speaking, we can distinguish two different kinds of questions. The first is a question about the grade distribution or structure:

Why was there exactly one A?

The second kind of question presupposes the first and asks

What is it about Mary in virtue of which she got the A?

The first is a question about a distribution, the second is a question about an individual's place in that distribution.

The relation between these two kinds of questions is parallel to the Willie Sutton example. There we distinguished the two questions:

1. Why is there something (at all) which Sutton robs? (the priest's question)

and

2. Given that there is something which Sutton robs, why is it banks?

Similarly, in this case we can distinguish:

1. Why is there exactly one person getting an A?

and

2. Given that exactly one person is to get an A, why was it Mary?

Answers to the second question presuppose, and do not explain, answers to the first question. This is true even if we look at the individual explanations for everyone's performance. If we take each person in the class and ask why that person got the grade he or she did, we have fifty answers to the questions why

Mary got an A
Bob got a B
Harold got a C

but the answers to those fifty questions do not add up to an answer to the question of why there was this distribution of grades.

Perhaps the clearest way to put this point is in terms of the contrast spaces of the two different explanations. If we look at the individuals, one by one, each individual has three possibilities: getting an A, getting a B, or getting a C. We can therefore represent a typical question as

why Mary gets \( \{ \text{A} \} \).
Now the class is in some sense the sum of the individuals, and so it is natural to try, a priori, to represent the possibility-space of the whole class as the product of fifty copies of the individual possibility-space, one copy for each individual. This would give us a space for the whole class that was

\[ S = \text{Mary gets } \{ A \} \times \text{Bob gets } \{ A \} \times \ldots \times \text{Harold gets } \{ A \}. \]

But this a priori possibility-space is not the true possibility-space of the class because it fails to take into account that certain combinations of individual possibilities are not collectively possible.

This would be the right space if the class were not graded on a curve. In the case where each individual's outcome depends only on facts about that individual, this is the true state space of the whole class. But in our example, the true possibility-space has far fewer than \(3^{50}\) elements because a set of additional conditions has been imposed on the overall space. I will call these structural conditions. The effect of such conditions is to reduce, before any of the imagined contingencies, the number of possibilities (or "degrees of freedom") available to the system.

The true contrast space for the question about Mary consists of the number of ways a set of 50 people can be subdivided into a set of 1, a set of 24, and a set of 25, in other words, only those grade distributions consistent with my policy. The contrast is between those distributions in which Mary got the A and ones where someone else got the A. All other differences among distributions are irrelevant. Consequently, the true contrast space for that question is not:

\[ \begin{align*}
\text{Why Mary got } & \{ A \} \text{ but rather} \\
\{ B \} & \\
\{ C \} \\
\text{Why } & \\
\{ \text{Mary} \} & \\
\{ \text{Bob} \} & \text{got the A.} \\
\{ \ldots \} & \\
\{ \text{Harold} \} &
\end{align*} \]

Writing the contrast space in this way makes it clear what is being presupposed by the questions about the individual's place in the distribution: they take the distribution itself as "given."

In the case of the structural question the situation is completely different. There the contrast would be between this grade distribution and the other possible grade distributions. The question "Why is there this distribution of grades?" contrasts this distribution with all the other possible distributions. Indeed, not only do the alternative possibilities include other grading distributions (5 A's, 20 B's, and so on), but for certain purposes one would have to include all other possible grading policies, even the nonstructured ones: giving no grades at all, giving A's to my friends, and so on.

In cases in which there is no predetermined distribution, the structural question

Why is there this distribution of grades?

does not really have a distinct answer. The answer to it is just that that's what the distribution of individual performances happened to be. If there are many C's, it is because many individuals happened to write poor papers. The question of the distribution collapses to the questions about the individuals. But where there is a predetermined policy, there is a separate nontrivial question about why that distribution was the case. What we are asking in those cases is not why there happened to be this distribution (e.g., one A) but rather why there had to be this distribution.

In cases like these, the imposed structural conditions radically alter the kinds of explanations we give because they constrain and truncate the contrast spaces. There is some precedent for this way of talking, and some good examples are to be found, in the state spaces of physics.

In analytical dynamics, the mathematical study of the physics of motion, these imposed conditions are called kinematical conditions. Consider, for example, two mass points moving freely in a plane. The total state space of these two points has eight dimensions: two location coordinates and two velocity coordinates for each of the two particles. We

can then talk about their gravitational interactions by means of a differential equation in this eight-dimensional space. So far the particles are moving freely, and so there are as yet no kinematical conditions. But now suppose the two particles are joined together by a rigid rod. Then there are no longer eight degrees of freedom, for there is the restriction that the distance between the two particles is constant. This is a simple algebraic relation among some of the coordinates:

$$(x_1 - y_1)^2 + (x_2 - y_2)^2 = k^2,$$

where $(x_1, x_2)$ and $(y_1, y_2)$ are the position coordinates of the two particles. These four coordinates, or four dimensions, are not independent, since given any three we can compute the fourth. We could even use this equation to rewrite the basic dynamics by eliminating one of the variables, say $x_1$, and substituting its equivalent in terms of the other three position variables, thus leaving an equation which has explicitly only seven degrees of freedom.

One problem with doing this is that the kinematical condition is obviously symmetric in the four variables, and the choice of one of the variables to be replaced as a function of the others is therefore arbitrary and somewhat misleading. The analogue in the grading example would be to take one person in the class, say Harold, and to say that if we know the grades of the other 49 students we can determine Harold’s grade; therefore, Harold can be eliminated as an independent variable. Why Harold? It is a more faithful representation of the situation to see it as an imposed relation among symmetric variables. (Indeed, one of the points of Lanczos’s book is to argue that this way of treating it is much more natural than the asymmetric way and allows the use of powerful mathematical techniques.) Let me just list more examples of such kinematical conditions:

1. If the two particles are constrained to remain on the surface of a sphere, we have lost two degrees of freedom, since the four-dimensional space of possible positions has been reduced to the two-dimensional surface of the sphere.

2. Consider a lever. First view it as a system of material particles held together by a variety of intermolecular forces. Its state space is therefore of very high dimension, millions of degrees of freedom for the individual particles. But the fact that the lever is rigid means that a substantial kinematical condition has been imposed, one which has the effect of reducing the degrees of freedom of the overall system from millions to exactly two: the location of one end-point and the angle of orientation.

3. (An example from social science.) Consider two people involved in what is called a zero-sum game. In such a game, one player’s wins are at the expense of the other. The state space of such a game would involve points representing the moves available to each player, and the outcomes dependent on them. The competitive or zero-sum nature of the game is then expressed by the imposed kinematical condition:

$$x_1 + x_2 = 0,$$

where $x_1$ is the payoff to the first player, and $x_2$ is the payoff to the second player.

The important point is this: the existence of such kinematical conditions makes it possible to make explanations within the system a lot more simply than we might be able to do otherwise. This is because when such conditions exist, the complexity of the explanation can be greatly reduced. For example, in the case of the lever, suppose it is in equilibrium with certain weights at certain points on it. In order to explain this in the very high dimensional state space of its constituent particles, we would have to know the representation of the state of the lever and its weights in that multidimensional space. The explanation, having literally millions of dimensions, would be awesomely complicated. But the assumption of rigidity enables us to reduce this complexity to a manageable level. What is more, if we were to try to explain a particular equilibrium in the particle-space, we would even have to know the nature of the underlying intermolecular forces that are responsible for the rigidity. The kinematical approach enables us to finesse this problem.

Similarly, in the case of the zero-sum game, the kinematical condition enables us to pare down explanations. In order to explain why the payoff to the two players was $(a, b)$, it suffices to give a one-dimensional explanation because the thing to be explained is, appearances to the contrary, only one dimensional: we know “a priori” that $b = -a$.

We might say, in the case of the zero-sum game, that we can explain why, given that there is a zero sum, it is this pair of values and not that, e.g., why it is $(6, -6)$ rather than $(9, -9)$. But we cannot, on this possibility-space, explain why there is a zero sum at all, because every possible
causal antecedent produces some zero sum or other. Thus it is like the grading example.

To summarize, explanations have presuppositions which, among other things, limit drastically the alternatives to the thing being explained. These presuppositions radically affect the success and failure of potential explanations and the interrelation of various explanations. Call this explanatory relativity.

A perspicuous way to represent this phenomenon is the device of contrast spaces, or spaces of live alternatives. The structure of these spaces displays some of the presuppositions of a given explanation.

One particular class of examples of explanatory relativity is especially worth noting: cases where a system consists of a number of individuals, each with its own individual possibility-space, but where the true possibility-space of the total system is not the full product of the individual spaces. In such cases the presuppositions (analogous to kinematical conditions) establish internal relations among the individuals, and in such cases explanations of individual properties will take a very special form. Moreover, such explanations (e.g., why Mary got an A) always presuppose, and hence can never explain, the overall structure of the system.

I now want to go on to apply these remarks in a study of various kinds of reductionism.

2 Reductionism

Reduction

Reductionist claims are often expressed by saying that something "is just" (or "is really") something else.

The claim that psychology is reducible to physics or chemistry is expressed as the statement that people "are just" physical objects. The claim that actions are reducible to primitive drives is put as the statement that human behavior "is just" the expression of those drives. There are claims that everything "is just" economics, while others say that everything "is just" biology. The claim that thermodynamics is reducible to statistical mechanics is expressed as the claim that a gas "is just" a collection of molecules, and the claim that social laws are reducible to the actions of individuals is expressed as the claim that society "is just" individuals.

The first problem with such claims is understanding what they could possibly mean. What does it mean to say that something "is just" (or "is really") something else?

The examples suggest that what is being claimed is a certain fact about explanation, namely, that the phenomena of the first kind are explainable from the theory of the second kind. The reducibility of psychology to physics and chemistry amounts to the claim that conduct can be explained wholly in terms of physical and chemical phenomena. Similarly in each of the other cases, the claim is that the one theory explains the other phenomena.

So reduction, which is on its face an ontological question, is really a question about the possibility of explanation: to say that something is reducible to something else is to say that certain kinds of explanations exist. This can be reconciled with more traditional conceptions, perhaps the best known of which is Quine's "ontological reduction." On his view an ontological reduction has been effected when one realm of discourse has been shown to be eliminable in favor of another: