

5

Philip Kitcher,* “Explanatory Unification and the Causal Structure of the World”

Drawing on some earlier work (Friedman 1974), Philip Kitcher (1981) developed his unification theory of explanation. Central to his approach is the idea that science enhances understanding by unifying disparate phenomena. This feature is overlooked by the covering-law model and this creates a number of problems for it—the problems that, Kitcher argues, are nicely handled by the unification theory. The selection below is excerpted from a later and more comprehensive paper in which the author explores, among other things, the connections between his theory and other approaches to explanation, in particular, Salmon’s causal approach (see next selection).

1. Introduction

The modern study of scientific explanation dates from 1948, the year of the publication of the pioneering article by C. G. Hempel and Paul Oppenheim. Nearly forty years later, philosophers rightly continue to appreciate the accomplishments of the covering-law models of explanation and the classic sequence of papers in which Hempel articulated his view. Even though it has become clear that the Hempelian approach to explanation faces difficulties of a number of types, the main contemporary approaches to explanation attempt to incorporate what they see as Hempelian insights (with distinct facets of the covering-law models being preserved in different cases), and they usually portray themselves as designed to accommodate one or more of the main problems that doomed the older view. My aim in this essay is to compare what I see as the chief contemporary rivals in the theory of explanation, to understand their affiliations to the covering-law models and their

P. Kitcher, “Explanatory Unification and the Causal Structure of the World,” in *Scientific Explanation*, ed. P. Kitcher and W.C. Salmon, 1989, pp. 410–505 (excerpts). Minneapolis: University of Minnesota Press.

efforts to address the troubles of those models, and to evaluate their success in doing so. Ecumenical as this may sound, the reader should be forewarned that I shall also be interested in developing further, and defending, an approach to explanation that I have championed in previous essays (1981, 1985).

1.1. *Hempel's Accounts*

Let us start with Hempel. The principal features of Hempel's account of explanation are (i) that explanations are arguments, (ii) that the conclusion of an explanation is a sentence describing the phenomenon to be explained, and (iii) that among the premises of an explanation there must be at least one law of nature. Although the original treatment (1948) focused on cases in which the argument is deductive and the conclusion a singular sentence (a sentence in which no quantifiers occur), it was clear from the beginning that the account could be developed along two different dimensions. Thus there can be covering-law explanations in which the argument is nondeductive or in which the conclusion is general. *D-N* explanations are those explanations in which the argument is deductive and the conclusion is either a singular sentence or a nonstatistical generalization. Hempel assigned deductive explanations whose conclusion is a statistical generalization a special category—*D-S* explanations—but their kinship with the official cases of *D-N* explanation suggests that we should broaden the *D-N* category to include them (see Salmon 1984 . . .). Finally, *I-S* explanations are those explanations in which the argument is inductive and the conclusion a singular sentence to which the premises assign high probability.

The motivation for approaching explanation in this way stems from the character of the explanations given in scientific works, particularly in those texts that are intended to introduce students to the main ideas of various fields. Expository work in physics, chemistry, and genetics (and, to a less obvious extent, in other branches of science) often proceeds by deriving descriptions of particular events—or, more usually, descriptions of empirical regularities—from sets of premises in which statements identified as laws figure prominently. Among the paradigms, we may include: the demonstration that projectiles obtain maximum range on a flat plain when the angle of projection is 45°, the Newtonian derivation of Galileo's law of free fall, Bohr's argument to show that the frequencies of the lines in the hydrogen spectrum satisfy the formulas previously obtained by Balmer and others, the kinetic-theoretic deduction of the Boyle–Charles law, computations that reveal the energy required for particular chemical reactions, and the derivation of expected distributions of traits among peas from specifications of the crosses and Mendel's laws. In all these cases, we can find scientific texts

that contain arguments that come very close indeed to the ideal form of explanation that Hempel describes.

1.2. *Hempel's Problems*

There are four major types of objection to the Hempelian approach. The first is the obverse of the motivational point just canvassed. Although we can identify some instances in which full-dress covering-law explanations are developed, there seem to be many occasions on which we accept certain statements as explanatory without any ability to transform them into a cogent derivation of a sentence describing the phenomenon to be explained. This objection, made forcefully in a sequence of papers by Michael Scriven (1959, 1962, 1963), includes several different kinds of case, of which two are especially important for our purposes here. One source of trouble lies in our propensity to accept certain kinds of historical narrative—both in the major branches of human history and in evolutionary studies—as explaining why certain phenomena obtain, even though we are unable to construct any argument that subsumes the phenomena under general laws. Another results from the existence of examples in which we explain events that are very unlikely. Here the paradigm is Scriven's case (later elaborated by van Fraassen) of the mayor who contracts paresis. Allegedly, we hold that the question “Why did the mayor get paresis?” can be answered by pointing out that he had previously had untreated syphilis, despite the fact that the frequency of paresis among untreated syphilitics is low.

A second line of objection to the covering-law models is based on the difficulty in providing a satisfactory analysis of the notion of a scientific law. Hempel is especially forthright in acknowledging the problem (1965, 338). The challenge is to distinguish laws from mere accidental generalizations, not only by showing how to characterize the notion of a projectible predicate (and thus answer the questions raised by Goodman's seminal 1956) but also by diagnosing the feature that renders pathological some statements containing only predicates that are intuitively projectible (for example, “No emerald has a mass greater than 1000 kg.”).

The first objection questions the necessity of Hempel's conditions on explanation. The third is concerned with their sufficiency. As Sylvain Bromberger made plain in the early 1960s (see especially his 1966), there are numerous cases in which arguments fitting one of Hempel's preferred forms fail to explain their conclusions. One example will suffice for the present. We can explain the length of the shadow cast by a high object (a flagpole or a building, say) by deriving a statement identifying the length of the shadow from premises that include the height of the object, the elevation of the sun, and the laws of the propagation of light. That derivation fits Hempel's D–N

model and appears to explain its conclusion. But, equally, we can derive the height of the object from the length of the shadow, the elevation of the sun and the laws of the propagation of light, and the latter derivation intuitively *fails* to explain its conclusion. Bromberger's challenge is to account for the asymmetry.

A close cousin of the asymmetry problem is the difficulty of debarring Hempelian arguments that appeal to irrelevant factors. If a magician casts a spell over a sample of table salt, thereby "hexing" it, we can derive the statement that the salt dissolved on being placed in water from premises that include the (apparently lawlike) assertion that all hexed salt dissolves on being placed in water. (The example is from Wesley Salmon's seminal 1970; it originally comes from Henry Kyburg [1965]). But, it is suggested, the derivation does not explain why the salt dissolved.

Finally, Hempel's account of statistical explanation was also subject to special problems. One trouble, already glimpsed in the paresis example, concerns the requirement of high probability. Among the guiding ideas of Hempel's account of explanation is the proposal that explanation works by showing that the phenomenon to be explained was to be expected. In the context of the statistical explanation of individual events, it was natural to formulate the idea by demanding that explanatory arguments confer high probability on their conclusions. But, as was urged by both Richard Jeffrey (1969) and Wesley Salmon (1970), this entails a whole class of counterintuitive consequences, generated by apparently good explanations of improbable occurrences. Moreover, the high-probability requirement itself turns out to be extremely hard to formulate (see Hempel 1965 for the surmounting of preliminary difficulties, and Coffa 1974 for documentation of residual troubles). Indeed, critics of Hempel's I-S model have charged that the high-probability requirement can only be sustained by supposing that all explanation is fundamentally deductive (Coffa 1974; Salmon 1984, 52–53).

Even a whirlwind tour of that region of the philosophical landscape occupied by theories of explanation (a region thick with syphilitic mayors, flagpoles, barometers, and magicians) can help to fix our ideas about the problems that an adequate account of scientific explanation must overcome. Contemporary approaches to the subject rightly begin by emphasizing the virtues of Hempel's work, its clarity, its connection with parts of scientific practice, its attention to the subtleties of a broad range of cases. When we have assembled the familiar difficulties, it is appropriate to ask "What went wrong?" The main extant rivals can be viewed as searching for the missing ingredient in the Hempelian approach, that crucial factor whose absence allowed the well-known troubles that I have rehearsed. I shall try to use the four main problem-types to chart the relations among Hempel's successors, and to evaluate the relative merits of the main contemporary rivals.

2. The Pragmatics of Explanation

Not all of the problem-types need be viewed as equally fundamental. Perhaps there was a basic mistake in Hempel’s account, a defect that gave rise directly to one kind of difficulty. Solve that difficulty, and we may discover that the remaining troubles vanish. The suggestion is tantalizing, and it has encouraged some important proposals.

One approach is to regard the first type of problem as fundamental. Hempel clearly needed an account of the pragmatics of explanation. As his own detailed responses to the difficulties raised by Scriven (Hempel 1965, 359–64, 427) make entirely clear, he hoped to accommodate the plausible suggestion that narratives can serve an explanatory function even when we have no idea as to how to develop the narrative into an argument that would accord with one of the models. The strategy is to distinguish between what is said on an occasion in which explanatory information is given and the ideal underlying explanation.¹ Although the underlying explanation is to be an argument including laws among its premises, what is said need not be. Indeed, we can provide some information about the underlying argument without knowing all the details, and this accounts for the intuitions of those (like Scriven) who insist that we can sometimes say explanatory things without producing a fully approved Hempelian argument (or without knowing much about what the fully approved argument for the case at hand would be).

Instead of backing into the question of how to relate explanations to what is uttered in acts of explaining, we can take the characterization of explanatory acts as our fundamental problem. This strategy has been pursued in different ways by Peter Achinstein and Bas van Fraassen, both of whom believe that the main difficulties of the theory of explanation will be resolved by gaining a clear view of the pragmatics of explanation. Because van Fraassen’s account introduces concepts that I take to be valuable to any theory of explanation, I shall consider his version.²

2.1. *Van Fraassen’s Pragmatics*

Van Fraassen starts with the claim that explanations are answers to why-questions. He proposes that why-questions are essentially contrastive: the question “Why *P*?” is elliptical for “Why *P* rather than *P**, *P*** , . . .?” In this way he can account for the fact (first noted in Dretske 1973 and further elaborated in Garfinkel 1981) that the same form of words can pose different contrastive why-questions. When Willie Sutton told the priest that he robbed banks because that is where the money is, he was addressing one version of the question “Why do you rob banks?,” although not the one that the priest intended.

With this in mind, van Fraassen identifies a why-question as an ordered triple $\langle P_k, X, R \rangle$. P_k is the topic of the question, and an ordinary (elliptical) formulation of the question would be “Why P_k ?” X is the contrast class, a set of propositions including the topic P_k . Finally R is the relevance relation. Why-questions arise in contexts, where a context is defined by a body of background knowledge K . The questions have presuppositions: each why-question presupposes that its topic is the only true member of the contrast class (intuitively, the question “Why P_k in contrast to the rest of X ?” is inappropriate if P_k is false or if some other member of the contrast class is true), and also that there is at least one true proposition A that stands in the relation R to $\langle P_k, X \rangle$. A why-question arises in a context K provided that K entails that the topic is the only true member of the contrast class and does not entail that there is no answer to the question (more exactly, that there is no true A bearing R to $\langle P_k, X \rangle$).

Van Fraassen recognizes that the theory of explanation ought to tell us when we should reject questions rather than attempting to answer. His pragmatic machinery provides a convincing account. We reject the why-question Q in context K if the question does not arise in this context, and, instead of trying to answer the question, we offer corrections. If Q does arise in a context, then a direct answer to it takes the form “Because A ,” where A is a true proposition that bears R to $\langle P_k, X \rangle$. The proposition A is the core of the direct answer.

2.2. *Why Pragmatics Is Not Enough*

Because he hopes to avoid the tangles surrounding traditional approaches to explanation, van Fraassen places no constraints on the relations that can serve as relevance relations in why-questions. In consequence, his account of explanation is vulnerable to trivialization. The trouble can easily be appreciated by noting that it is *prima facie* possible for any true proposition to explain any other true proposition. Let A, B both be true. Then, given van Fraassen’s thesis that explanations are answers to why-questions, A will explain B in context K provided that there is a question “Why B ?” that arises in K for which A is the core of a direct answer. We construct an appropriate question as follows: let $X = \{B, -B\}$, $R = \{\langle A, \langle B, X \rangle \rangle\}$. Provided that K entails the truth of B and does not contain any false proposition entailing the nonexistence of any truth bearing R to $\langle B, X \rangle$, then the question $\langle B, X, R \rangle$ arises in K , its topic is B , and its only direct answer is A .

Wesley Salmon and I have argued (Kitcher and Salmon 1987) that van Fraassen’s account cannot avoid this type of trivialization. We diagnose the absence of constraints on the relevance relation as the source of the trouble. Intuitively, genuine why-questions are triples $\langle P_k, X, R \rangle$ where R is a genuine

relevance relation, and a large part of the task of a theory of explanation is to characterize the notion of a genuine relevance relation.

...

Van Fraassen contends that his pragmatic approach to explanation solves the problem of asymmetry that arises for the Hempelian account. His solution consists in showing that there is a context in which the question “Why is the height of the tower h ?” is answered by the proposition that the length of the shadow cast by the tower at a certain time of day is s . That proposition answers the question by providing information about the intentions of the builder of the tower. Thus it has seemed that van Fraassen does not touch the Hempelian problem of distinguishing the explanatory merits of two derivations (both of which satisfy the conditions of the D–N model), and that the claim to have solved the problem of asymmetry is incorrect (see Salmon 1984, 95, and Kitcher and Salmon 1987, for arguments to this effect).

...

I suggest that van Fraassen’s illuminating discussion of why-questions is best seen not as a solution to all the problems of the theory of explanation, but as a means of tackling the problems of the first type (see section 1). Given solutions to the difficulties with law, asymmetry, irrelevance, and statistical explanation, we could embed these solutions in van Fraassen’s framework, and thus handle the general topic of how to relate idealized accounts of explanation to the everyday practice of answering why-questions. This is no small contribution to a theory of explanation, but it is important to see that it cannot be the whole story.

2.3. Possible Goals for a Theory of Explanation

Van Fraassen’s work also enables us to see how to concentrate the three residual problems that arise for Hempel’s account into one fundamental issue. The central task of a theory of explanation must be to characterize the genuine relevance relations, and so delimit the class of genuine why-questions. To complete the task it will be necessary to tackle the problems of asymmetry and irrelevance, to understand the structure of statistical explanations, and, if we suppose that genuine relevance involves lawlike dependence, to clarify the concept of law.³ However, the formulation of the task is ambiguous in significant respects. Should we suppose that there is a single set of genuine relevance relations that holds for all sciences and for all times? If not, if the set of genuine relevance relations is different from science to science and from epoch to epoch, should we try to find some underlying characterization that determines how the different sets are generated, or should we rest content with studying a particular science at a particular time

and isolating the genuine relevance relations within this more restricted temporal and disciplinary area?⁴

It appears initially that Hempel sought a specification of the genuine relevance relations that was time-independent and independent also of the branch of science. However, in the light of our integration of Hempel's approach with van Fraassen's treatment of why-questions, I think we can achieve a more defensible view of the Hempelian task. Plainly, the set of ideal relevance relations (or of ideal why-questions) may be invariant across times and sciences, even though different actual questions become genuine in the light of changing beliefs. Thus one conception of the central problem of explanation—I shall call it the *Hempelian conception*—is the question of defining the class of genuine relevance relations that occur in the ideal why-questions of each and every science at each and every time. We can then suppose that variation in the why-questions arises partly from differing beliefs about which topics are appropriate, partly from differing views about the character of answers to underlying ideal why-questions, and partly from differing ideas about what would yield information about those answers.

...

Because philosophical attention to the history of science has exposed numerous important shifts in methodological ideals, the Hempelian conception of the theory of explanation may seem far too ambitious and optimistic. However, one way to respond to claims about shifting standards is to argue that there are overarching principles of *global* methodology that apply to all sciences at all times. As particular scientific fields evolve, the principles of global methodology are filled out in different ways, so that there are genuine modifications of *local* methodology.⁵ The version of the Hempelian conception that I have just sketched assigns to global methodology a characterization of ideal why-questions. Shifts in admissible why-questions, corresponding to changes in local methodology, can occur against the background of constancy in the underlying ideals—witness my brief discussion of functional/teleological questions.

Perhaps this picture makes the Hempelian conception somewhat less at odds with current thinking about the modification of methodology in the history of science. But can anything positive be said in favor of that conception? I believe it can. The search for understanding is, on many accounts of science, a fundamental goal of the enterprise. That quest may take different forms in different historical and disciplinary contexts, but it is tempting to think that there is something that underlies the various local endeavors, something that makes each of them properly be seen as a striving after the same goal. The Hempelian conception proposes that there is an abstract conception of human understanding, that it is important to the development of science, and that it is common to the variety of ways in which understand-

ing is sought and gained. Scientific explanations are intended to provide objective understanding of nature. The task of characterizing the ideal notions of explanation, why-question, and relevance is thus one of bringing into focus one of the basic aims of science.

I do not suppose that these remarks provide any strong reasons for thinking that the Hempelian conception is correct. It might turn out that there is nothing but ritual lip movements in the avowal of explanation as an aim of the sciences. Nonetheless, there is an obvious motivation for pursuing the Hempelian conception, for, if it is correct, then we can hope to obtain some insight into the rationality and progressiveness of science. Since I know of no conclusive reasons for abandoning my preferred version of the conception, I propose to consider theories of explanation that undertake the ambitious task of characterizing the ideal relevance relations. More modest projects can come once ambition has failed.

3. Explanation as Delineation of Causes

There are two main approaches to explanation that can be seen as undertaking the project just outlined. One of these can be motivated by considering the problems of asymmetry. Intuitively, the length of the shadow cast by a flagpole is causally dependent on the height of the flagpole, but the height is not causally dependent on the shadow-length. Thus we arrive at the straightforward proposal that Hempel's failure to solve problems of asymmetry (and irrelevance) stems from the fact that causal notions are avoided in his analyses. Diagnosis leads quickly to treatment: genuine relevance relations are causal relations, explanations identify causes.

Of course, the invocation of causal notions has its costs. Hempel's account of explanation was to be part of an empiricist philosophy of science, and it could therefore only draw on those concepts that are acceptable to empiricists. If causal concepts are not permissible as primitives in empiricist analyses, then either they must be given reductions to empiricist concepts or they must be avoided by empiricists. Hempel's work appears to stand in a distinguished tradition of thinking about explanation and causation, according to which causal notions are to be understood either in terms of the concept of explanation or in terms of concepts that are themselves sufficient for analyzing explanation. Empiricist concerns about the evidence that is available for certain kinds of propositions are frequently translated into claims about conceptual priority. Thus, the thesis that we can only gain evidence for causal judgments by identifying lawlike regularities generates the claim that the concept of law is prior to that of cause, with consequent dismissal of analyses that seek to ground the notion of law in that of cause.

One of Hume's legacies is that causal judgments are epistemologically

problematic. For those who inherit Hume's theses about causation (either his positive or his negative views) there are obvious attractions in seeking an account of explanation that does not take any causal concept for granted. A successful analysis of explanation might be used directly to offer an analysis of causation—most simply, by proposing that one event is causally dependent on another just in case there is an explanation of the former that includes a description of the latter. Alternatively, it might be suggested that the primitive concepts employed in providing an analysis of explanation are just those that should figure in an adequate account of causation.

Because the invocation of causal dependency is so obvious a response to the problems of asymmetry and irrelevance, it is useful to make explicit the kinds of considerations that made that response appear unavailable. One central theme of the present essay is that there is a tension between two attractive options. Either we can have a straightforward resolution of asymmetry problems, at the cost of coming to terms with epistemological problems that are central to the empiricist tradition, or we can honor the constraints that arise from empiricist worries about causation and struggle to find some alternative solution to the asymmetries. The two major approaches to explanation respond to this tension in diametrically opposite ways. As we may anticipate, the central issues that arise concern the adequacy of proposed epistemological accounts of causation and of suggestions for overcoming problems of asymmetry and irrelevance without appealing to causal concepts.

...

4. Explanation as Unification

On both the Hempelian and the causal approaches to explanation, the explanatory worth of candidates—whether derivations, narratives, or whatever—can be assessed individually. By contrast, the heart of the view that I shall develop in this section (and which I shall ultimately try to defend) is that successful explanations earn that title because they belong to a set of explanations, the *explanatory store*, and that the fundamental task of a theory of explanation is to specify the conditions on the explanatory store. Intuitively, the explanatory store associated with science at a particular time contains those derivations which collectively provide the best systematization of our beliefs. Science supplies us with explanations whose worth cannot be appreciated by considering them one-by-one but only by seeing how they form part of a systematic picture of the order of nature.

...

All this is abstract and somewhat metaphorical. To make it more precise, let us begin with the proposal that *ideal* explanations are derivations. Here

there is both agreement and disagreement with Hempel. An argument can be thought of as an ordered pair whose first member is a set of statements (the premises) and whose second member is a single statement (the conclusion). Hempel’s proposal that explanations are arguments appears to embody this conception of arguments as *premise–conclusion* pairs. But, on the systematization account, an argument is considered as a derivation, as a sequence of statements whose status (as a premise or as following from previous members in accordance with some specified rule) is clearly specified. An ideal explanation does not simply list the premises but shows how the premises yield the conclusion.

...

For a derivation to count as an *acceptable* ideal explanation of its conclusion in a context where the set of statements endorsed by the scientific community is K , that derivation must belong to the explanatory store over K , $E(K)$. At present, I shall assume that K is both consistent and deductively closed, and that the explanatory store over a set of beliefs is unique. $E(K)$ is to be the set of derivations that best systematizes K , and I shall suppose that the criterion for systematization is unification.⁶ $E(K)$, then, is the set of derivations that best unifies K . The challenge is to say as precisely as possible what this means.

We should be clear about just what is to be defined. The set of derivations we are to characterize is the set of explanations that would be acceptable to those whose beliefs comprised the members of K .

...

The idea that explanation is connected with unification has had some important advocates in the history of the philosophy of science. It appears to underlie Kant’s claims about scientific method⁷ and it surfaces in classic works in the logical empiricist tradition (see Hempel [1965] 345, 444; Feigl [1970] 12). Michael Friedman (1974) has provided the most important defense of the connection between explanation and unification. Friedman argues that a theory of explanation should show how explanation yields understanding, and he suggests that we achieve understanding of the world by reducing the number of facts we have to take as brute.⁸ Friedman’s motivational argument suggests a way of working out the notion of unification: characterize $E(K)$ as the set of arguments that achieves the best tradeoff between minimizing the number of premises used and maximizing the number of conclusions obtained.

Something like this is, I think, correct. Friedman’s own approach did not set up the problem in quite this way, and it proved vulnerable to technical difficulties (see Kitcher 1976 ...). I propose to amend the account of unification by starting from a slight modification of the motivational idea that Friedman shares with T. H. Huxley (see note 8). Understanding

the phenomena is not simply a matter of reducing the “fundamental incomprehensibilities” but of seeing connections, common patterns, in what initially appeared to be different situations. Here the switch in conception from premise-conclusion pairs to derivations proves vital. *Science advances our understanding of nature by showing us how to derive descriptions of many phenomena, using the same patterns of derivation again and again, and, in demonstrating this, it teaches us how to reduce the number of types of facts we have to accept as ultimate (or brute).*⁹ So the criterion of unification I shall try to articulate will be based on the idea that $E(K)$ is a set of derivations that makes the best tradeoff between minimizing the number of patterns of derivation employed and maximizing the number of conclusions generated.

...

5. Asymmetry and Irrelevance

The time has now come to put all this abstract machinery to work. Problems of asymmetry and irrelevance take the following general form. There are derivations employing premises which are (at least plausible candidates for) laws of nature and that fail to explain their conclusions. The task is to show that the unwanted derivations do not belong to the explanatory store over our current beliefs. To complete the task we need to argue that any systematization of our beliefs containing these derivations would have a basis that fares worse . . . than the basis of the systematization that we actually accept. In practice, this task will be accomplished by considering a small subset of the explanatory store, the derivations that explain conclusions akin to that of the unwanted derivation, and considering how we might replace this subset and include the unwanted derivation. I want to note explicitly that there is a risk that we shall overlook more radical modifications of the explanatory store which would incorporate the unwanted derivation. If there are such radical modifications that do as well by the criteria of unifying power as the systematization we actually accept, then my account is committed to claiming that we were wrong to treat the unwanted derivation as nonexplanatory.

5.1. *The “Hexed” Salt*

Let us start with the classic example of explanatory irrelevance. A magician waves his hands over some table salt, thereby “hexing” it. The salt is then thrown into water, where it promptly dissolves. We believe that it is not an acceptable explanation of the dissolving of the salt to point out that the salt was hexed and that all hexed salt dissolves in water. What is the basis of this belief?

“EXPLANATORY UNIFICATION AND CAUSAL STRUCTURE”

Suppose that $E(K)$ is the explanatory store over our current beliefs, K , and that S is some set of derivations, acceptable with respect to K , that has the unwanted derivation of the last paragraph as a member. One of the patterns used to generate $E(K)$ derives claims about the dissolving of salt in water from premises about the molecular composition of salt and water and about the forming and breaking of bonds. This pattern can be used to generate derivations whose conclusions describe the dissolving of hexed salt and the dissolving of unhexed salt. How does S provide similar derivations? Either the basis of S does not contain the standard pattern or it contains both the standard pattern and a nonstandard pattern that yields the unwanted derivation. In the former case, S fares less well than $E(K)$ because it has a more restricted consequence set, and, in the latter case, it has inferior unifying power because its basis employs all the patterns of the basis of $E(K)$ and one more besides.

It is obviously crucial to this argument that we exclude the gerrymandering of patterns. For otherwise the claim that the basis of S must contain either the nonstandard pattern alone or the nonstandard pattern plus the standard pattern would be suspect. The reason is that we could gerrymander a “pattern” by introducing some such Goodmanian predicate as “ x is either hexed, or is unhexed and has molecular structure $NaCl$.” Now we could recover derivations by starting from the claim that all table salt satisfies this predicate, by using the principle that all hexed table salt dissolves in water to generate the conclusion from one disjunct and by using the standard chemical derivation to generate the conclusion from the other disjunct. This maneuver is debarred by the requirement that the predicates used in patterns must be protectable from the perspective of K .

Consider next a refinement of the original example. Not all table salt is hexed, but presumably all of it is hexable. (For present purposes, we may assume that hexing requires only that an incantation be muttered with the magician’s thoughts directed at the hexed object; this will obviate any concerns that some samples of table salt might be too large or too inaccessible to have the magician wave a hand over them.) Suppose now that it is proposed to explain why a given sample of table salt dissolves in water by offering the following derivation:

a is a hexable sample of table salt.

a was placed in water.

Whenever a hexable sample of table salt is placed in water, it dissolves.

a dissolved.

I take it that this derivation strikes us as nonexplanatory (although it is useful to point out that it is not as badly nonexplanatory as the derivation in

the original example). Suppose that S is a systematization of K that contains the derivation. Can we show that S has less unifying power than $E(K)$?

Imagine that S had the same unifying power as $E(K)$. Now in $E(K)$ the mini-derivation that is most akin to the one we want to exclude derives the conclusion that a dissolved from the premise that a is a sample of table salt, the premise that a was placed in water, and the generalization that samples of table salt that are placed in water dissolve. Of course, this mini-derivation is embedded within a much more exciting chemical derivation whose conclusion is the generalization that samples of table salt dissolve when placed in water. That derivation instantiates a general pattern that generates claims about the dissolving (or failure to dissolve) of a wide variety of substances from premises about molecular structure. In its turn, that general pattern is a specification of an even more general pattern that derives conclusions about chemical reactions and state changes for all kinds of substances from premises about molecular structures and energy distributions. If S is to rival $E(K)$ then it must integrate the unwanted mini-derivation in analogous fashion.

That can be done. One way to proceed would be to use the standard chemical derivation to yield the conclusion that all samples of table salt dissolve when placed in water and then deduce that all hexable samples of table salt dissolve when placed in water. But now we can appeal to a principle of simplifying derivations to eliminate redundant premises or unnecessary steps. When embedded within the standard chemical derivation, the unwanted mini-derivation is inferior to its standard analog because the latter is obtainable more directly from the same premises. An alternative way of trying to save the unifying power of S would be to amend the standard chemical patterns to suppose that they apply only to hexable substances. But since it is supposed that *all* substances are hexable, and since this fact is used throughout S to generate derivations to rival those produced in $E(K)$, this option effectively generates a set of derivations that systematically contain idle clauses. Since it is believed that everything is hexable, the outcome is as if we added riders about objects being self-identical or being nameable to our explanations, and again a principle of simplification directs that the idle clauses be dropped.¹⁰

We can now achieve a diagnosis of the examples of explanatory irrelevance. Citation of irrelevant factors will either commit one to patterns of explanation that apply only to a restricted class of cases or the irrelevancies will be idle wheels that are found throughout the explanatory system. The initial hexing example illustrates the first possibility; the refinement shows the second.

5.2. *Towers and Shadows*

Let us now turn to the asymmetry problem, whose paradigm is the case of the tower and the shadow. Once again, let K be our current set of beliefs, and let us compare the unifying power of $E(K)$ with that of some systematization S containing a derivation that runs from the premises about shadow length and sun elevation to a conclusion about the tower’s height. As in the case of the irrelevance problem, there is a relatively simple argument for maintaining that S has less unifying power than $E(K)$. There are also some refinements of the original, troublesome story that attempt to evade this simple argument.

Within $E(K)$ there are derivations that yield conclusions about the heights of towers, the widths of windows, the dimensions of artifacts and natural objects alike, which instantiate a general pattern of tracing the present dimensions to the conditions in which the object originated and the modifications that it has since undergone. Sometimes, as with flagpoles and towers, the derivations can be relatively simple: we start with premises about the intentions of a designer and reason to an intermediate conclusion about the dimensions of the object at the time of its origin; using further premises about the conditions that have prevailed between the origin and the present, we reason that the object has persisted virtually unaltered and thus reach a conclusion about its present dimensions. With respect to some natural objects, such as organisms, stars, and mountain ranges, the derivation is much more complex because the objects have careers in which their sizes are substantially affected. However, in all these cases, there is a very general pattern that can be instantiated to explain current size, and I shall call derivations generated by this pattern *origin-and-development* explanations.

Now if S includes *origin-and-development* explanations, then the basis of S will include the pattern that gives rise to these derivations. To generate the unwanted derivation in S , the basis of S must also contain another pattern that derives conclusions about dimensions from premises about the characteristics of shadows (the *shadow* pattern). In consequence, S would fare worse than $E(K)$ according to our principles . . . because its basis would contain all the patterns in the basis of $E(K)$ and one more. Notice that, once again, the “no gerrymandering” requirement comes into play to block the device of fusing some doctored version of the pattern that generates *origin-and-development* explanations with the shadow pattern. So S must forswear *origin-and-development* explanations.

However, it now seems that S must have a consequence set that is more restricted than that of $E(K)$. The reason is that the *shadow* pattern cannot be instantiated in all the cases in which we provide *origin-and-development*

explanations. Take any unilluminated object. It casts no shadow. Hence we cannot instantiate the *shadow* pattern to explain its dimensions.

This is correct as far as it goes, but the asymmetry problem cuts deeper. Suppose that a tower is actually unilluminated. Nonetheless, it is possible that it should have been illuminated, and if a light source of a specified kind had been present and if there had been a certain type of surface, then the tower would have cast a shadow of certain definite dimensions. So the tower has a complex dispositional property, the disposition to cast a shadow of such-and-such a length on such-and-such a surface if illuminated by a light-source at such-and-such an elevation above the surface. From the attribution of this dispositional property and the laws of propagation of light we can derive a description of the dimensions of the tower. The derivation instantiates a pattern, call it the *dispositional-shadow* pattern, that is far more broadly applicable than the *shadow* pattern.

But can it be instantiated widely enough? To be sure it will provide surrogates for *origin-and-development* explanations in those cases in which we are concerned with ordinary middle-sized objects. But what about perfectly transparent objects (very thin pieces of glass, for example)? Well, they can be handled by amending the pattern slightly, supposing that such objects have a disposition to be coated with opaque material and then to cast a shadow. Objects that naturally emit light can be construed as having a disposition to have their own light blocked and then to cast a shadow. Objects that are so big that it is hard to find a surface on which their shadows could be cast (galaxies, for example) can be taken to have the dispositional property of casting a shadow on some hypothetical surface.

Yet more dispositional properties will be needed if we are to accommodate the full range of instances in which *origin-and-development* explanations are available. An embryologist might explain why the surface area of the primitive gut (archenteron) in an early embryo is of such-and-such a size by deriving a description of the gut from premises about how it is formed and how modified. To instantiate the *dispositional-shadow* pattern in such cases, we shall need to attribute to the gut-lining a dispositional property to be unrolled, illuminated, and thus to cast a shadow. A biochemist might explain the diameter in the double helix of a DNA molecule by identifying the constraints that the bonding pattern imposes on such molecules both as they are formed and as they persist.¹¹ Taking a clue from the principles of electronmicroscopy, the *dispositional-shadow* pattern can be instantiated by supposing that DNA molecules have a dispositional property to be coated and irradiated in specified ways and to produce absorption patterns on special surfaces. And so it goes.

Perhaps there are some objects that are too small, or too large, too light sensitive, or too energetic for us to attribute to them any disposition to cast

anything like a shadow. If so, then even with the struggling and straining of the last paragraph, the *dispositional-shadow* pattern will still fail to generate derivations to rival those present in $E(K)$. But I shall assume that this is not so, and that for any object whose dimensions we can explain using our accepted patterns of derivation, it is possible to find a dispositional property that has something to do with casting a shadow.

However, if we now consider the critical predicate that appears in the *dispositional-shadow* pattern, we find that it is something like the following: “ x has the disposition to cast a shadow if illuminated by a light source or x has the disposition to produce an absorption pattern if x is suitably coated and irradiated or x has the disposition to cast a shadow if x is covered with opaque material or x has the disposition to cast a shadow if x is sectioned and unrolled or x has the disposition to cast a shadow after x has been treated to block its own light sources or . . .” At this point it is surely plain that we are cutting across the distinctions drawn by the projectable predicates of our language. Any “pattern” that employs a predicate of the sort that I have (partially) specified is guilty of gerrymandering, for, from our view of the properties of things, the dispositions that are lumped together in the predicate are not homogeneous. I conclude that even if it is granted that we can find for each object some dispositional property that will enable us to derive a specification of dimensions from the ascription of the disposition, there is no *common* dispositional property that we can employ for *all* objects. To emulate the scope of $E(K)$, the basis of S would have to contain a multiplicity of patterns, and our requirement against gerrymandering prohibits the fusion of these into a single genuine pattern.

As in the case of the irrelevance problem, there is a natural diagnosis of the trouble that brings out the central features of the foregoing arguments. Explanation proceeds by tracing the less fundamental properties of things to more fundamental features, and the criterion for distinguishing the less from the more fundamental is that appeal to the latter can be made on a broader scale. Thus an attempt to subvert the order of explanation shows up in the provision of an impoverished set of derivations (as in our original example of the tower and the shadow) or in the attempt to disguise an artificial congeries of properties as a single characteristic (as in our more recent reflections).

...

6. Conclusions

As Railton clearly recognizes, . . . differences in views about scientific explanation connect to differences in metaphysics. The causal approach is wedded to a strong version of realism in which the world is seen as having a

structure independent of our efforts to systematize it. It should be no surprise that the metaphysical extras bring epistemological problems in their train. . . . I have been trying to show that we can make sense of scientific explanation and our view of the causal structure of nature without indulging in the metaphysics. The aim has been to develop a simple, and, I think, very powerful idea. The growth of science is driven in part by the desire for explanation, and to explain is to fit the phenomena into a unified picture insofar as we can. What emerges in the limit of this process is nothing less than the causal structure of the world.

Notes

- * I owe a long-standing debt to Peter Hempel, who first inspired my interest in the study of scientific explanation and whose writings on the topic seem to me paradigms of what is best in twentieth-century philosophy. My own thinking about explanation was redirected by Michael Friedman's seminal essay on explanation and scientific understanding, and I have also learned much from the comments, encouragement, and advice of Paul Churchland, Paul Humphreys, David Papineau, Kenneth Schaffner, and Stephen Stich. Above all I am deeply grateful to Wesley Salmon, for the depth and lucidity of his ideas and the kindness and patience of his conversation. The present essay continues a long dialogue, and, because that dialogue has been so pleasant and so instructive, I trust that it is not yet over.
- 1 This approach to pragmatic issues has been articulated with considerable sophistication by Peter Railton. See his (1981) and his unpublished doctoral dissertation.
- 2 Achinstein's theory of explanation, as presented in his (1983), is extremely complex. I believe that it ultimately suffers from the same general difficulty that I present below for van Fraassen. However, it is eminently possible that I have overlooked some subtle refinement that makes for a disanalogy between the two versions.
- 3 As should now be evident, the second of the four problem-types that beset the Hempelian account assumes a derivative status. We may be able to manage the theory of explanation without a characterization of laws if we can distinguish the genuine relevance relations without invoking the notion of lawlike dependence. I shall articulate an approach below on which this possible strategy is attempted.
- 4 I am extremely grateful to Isaac Levi for raising the issue of the goals of a theory of explanation, by inquiring whether we can expect there to be a single set of relevance relations that applies for all sciences at all times.
- 5 The distinction between global and local methodology is drawn in more detail in chapter 7 of (Kitcher 1983). It is only right to note that some

“EXPLANATORY UNIFICATION AND CAUSAL STRUCTURE”

scholars have challenged the idea that there is any very substantive global methodology. See, for example, Laudan 1984.

- 6 We might think of the systematization approach as covering an entire family of proposals among which is that based on the view of systematization as unification. Since it appears to me that the latter view provides the best chances of success, I shall concentrate on it and ignore alternative possible lines of development.
- 7 See my (1986) for a reconstruction of Kant’s views that tries to defend this attribution.
- 8 “. . . our total picture of nature is simplified via a reduction in the number of independent phenomena that we have to accept as ultimate” (Friedman 1974, 18). There is an interesting recapitulation here of T. H. Huxley’s summary of Darwin’s achievement. “In ultimate analysis everything is incomprehensible, and the whole object of science is simply to reduce the fundamental incomprehensibilities to the smallest possible number” (Huxley 1896, 165).
- 9 I think it entirely possible that a different system of representation might articulate the idea of explanatory unification by employing the “same way of thinking again and again” in quite a different—and possibly more revealing—way than the notions from logic that I draw on here. Kenneth Schaffner has suggested to me that there is work in AI that can be deployed to provide the type of account I wish to give, and Paul Churchland has urged on me the advantages of connectionist approaches. I want to acknowledge explicitly that the adaptation of ideas about logical derivation may prove to be a ham-fisted way of developing the idea of explanatory unification. But, with a relatively developed account of a number of facets of explanation available, others may see how to streamline the machinery.
- 10 Notice that derivations that systematically contain idle clauses are not so clearly nonexplanatory as the kind of irrelevant derivation with which we began. It seems to me that this is because the unwanted mini-derivations are viewed as giving us information about the structure of a full, ideal, derivation, and the natural implication is that the properties picked out in the premises will play a key role. Once we see that these properties are inessential, and that predicates referring to them figure throughout all our derivations, then we may feel that cluttering up the explanatory store does nothing more than add a harmless irrelevancy. The resultant derivations are untidy, but I think that there is reason to argue about whether they should be counted as nonexplanatory.
- 11 For this example, it is important to recognize that the *origin-and-development* pattern must allow for explanatory derivations in which we appeal to general constraints that keep a system close to an equilibrium state throughout its career. . . .

References

- Achinstein, Peter. 1983. *The Nature of Explanation*. New York: Oxford University Press.
- Bromberger, Sylvain. 1963. A Theory about the Theory of Theory and about the Theory of Theories. In *Philosophy of Science: The Delaware Seminar*, ed. W. L. Reese. New York: John Wiley.
- . 1966. Why-Questions. In *Mind and Cosmos*, ed. R. Colodny. Pittsburgh: University of Pittsburgh Press.
- Coffa, J. Alberto. 1974. Hempel's Ambiguity, *Synthese* 28: 141–64.
- Dretske, Fred. 1973. Contrastive Statements, *Philosophical Review*, 82: 411–37.
- Feigl, Herbert. 1970. The "Orthodox" View of Theories: Remarks in Defense as well as Critique. In *Minnesota Studies in the Philosophy of Science*, Volume IV, eds. M. Radner and S. Winokur. Minneapolis: University of Minnesota Press.
- Friedman, Michael. 1974. Explanation and Scientific Understanding, *Journal of Philosophy*, 71: 5–19.
- Garfinkel, Alan. 1981. *Forms of Explanation*. New Haven: Yale University Press.
- Goodman, Nelson. 1956. *Fact, Fiction, and Forecast*. Indianapolis: Bobbs-Merrill.
- Hempel, C. G. 1965. *Aspects of Scientific Explanation*. New York: Free Press.
- Hempel, C. G., and Oppenheim, P. 1948. Studies in the Logic of Explanation. Chapter 9 of Hempel (1965).
- Huxley, Thomas Henry. 1896. *Darwiniana*. New York: Appleton.
- Jeffrey, Richard. 1969. Statistical Explanation vs. Statistical Inference. In *Essays in Honor of Carl G. Hempel*, ed. Nicholas Rescher. Dordrecht, Holland: D. Reidel.
- Kitcher, Philip. 1976. Explanation, Conjunction, and Unification, *Journal of Philosophy* 73: 207–12.
- . 1981. Explanatory Unification, *Philosophy of Science*, 48: 507–31.
- . 1983. *The Nature of Mathematical Knowledge*. New York: Oxford University Press.
- . 1985. Two Approaches to Explanation, *Journal of Philosophy*, 82: 632–39.
- . 1986. Projecting the Order of Nature. In *Kant's Philosophy of Physical Science*, ed. Robert Butts. Dordrecht: D. Reidel.
- , and Salmon, Wesley. 1987. Van Fraassen on Explanation, *Journal of Philosophy*, 84: 315–30.
- Kyburg, Henry. 1965. Comment, *Philosophy of Science*, 35: 147–51.
- Laudan, Larry. 1984. *Science and Values*. Berkeley: University of California Press.
- Railton, Peter. 1981. Probability, Explanation, and Information, *Synthese*, 48: 233–56.
- Salmon, Wesley. 1970. Statistical Explanation. In *The Nature and Function of Scientific Theories*, ed. R. Colodny. Pittsburgh: University of Pittsburgh Press.
- . 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Scriven, Michael. 1959. Definitions, Explanations, and Theories. In *Minnesota Studies in the Philosophy of Science*, Volume II, eds. H. Feigl, M. Scriven, and G. Maxwell. Minneapolis: University of Minnesota Press.
- . 1962. Explanations, Predictions, and Laws. In *Minnesota Studies in the Philosophy of Science*, Volume III, eds. H. Feigl and G. Maxwell. Minneapolis: University of Minnesota Press.

“EXPLANATORY UNIFICATION AND CAUSAL STRUCTURE”

- . 1963. The Temporal Asymmetry between Explanations and Predictions. In *Philosophy of Science. The Delaware Seminar*, Volume I, ed. B. Baumrin. New York: John Wiley.