

8

Explanatory Knowledge and Metaphysical Dependence

I

There is a famous remark Aristotle made about knowledge: “Men do not think they know a thing unless they have grasped the ‘why’ of it” (*Physics* II, ch. 3; see also *Metaphysics* V, ch. 2). This remark is often quoted by writers on scientific explanation to underline the importance of explanation to scientific knowledge, and why, as philosophers, we should concern ourselves with understanding what explanation is—that is, to show that “analyzing” scientific explanation, or building a “model” of explanation, is a reputable philosophical enterprise. It is not uncommon to invoke Aristotle not only to stress the centrality of explanation to science but also to motivate a distinction between *explanatory knowledge* and *descriptive knowledge*—*knowing why* from *knowing that*.¹ We are told, for example, that to know that the spectral lines of light received from distant celestial bodies show a shift toward red is one thing, but quite another to know why these red shifts occur. The former, it is suggested, is merely knowing a fact, but the latter involves something deeper and more valuable, an understanding of why the fact obtains. The mark of a theoretical science, it is often claimed, is precisely that it aspires to go beyond “phenomenological descriptions”

My thanks to Enrique Villanueva, Marcelo Sabates, and Margarita Santana, each of whom presented an instructive set of comments on this paper at the 1993 SOFIA conference in Tenerife, Spain, and also to others at the conference who made useful suggestions and comments—in particular, Richard Foley and Peter Railton. Many thanks also to Paul Humphreys and Wesley Salmon for helpful and clarifying comments.

¹ For example, Wesley Salmon in his *Scientific Explanation and the Causal Structure of the World* (Princeton, NJ: Princeton University Press, 1984), p. 4.

of observed regularities and provide us with explanatory insights, an understanding of why things are the way they are.

From this one would expect theorists of explanation to go into detailed discussions of what explanatory knowledge consists in, and how it differs from knowledge that is merely descriptive; you would, in short, expect a theory of explanation to be an *epistemological theory*, a theory of explanatory knowledge, which presumably is a special kind of knowledge. And you would expect such a theory to be developed as part of a general theory of knowledge. But, as everyone knows, this is not what happens: once the writers get past the obligatory paragraphs about the importance of explanation and understanding, they quickly launch into highly technical constructions heavily laden with an array of terms like “law,” “derivation,” “causality,” “probability,” “simplicity,” and their forbidding technical kins, and we never see a serious discussion of just what these concepts have to do with understanding something or making something intelligible. The actual theories of explanation that we have accumulated to date, such as Hempel’s covering-law theory, the causal theories of Salmon, Lewis, and Humphreys, the pragmatic theories of Bromberger, Achinstein, and Van Fraassen²—don’t look much like theories of understanding or accounts of a type of knowledge. Although explanatory understanding is sometimes mentioned, as we shall see below, in the writings of some writers who take the simplification-unification approach to explanation, it seems quickly to be lost sight of when serious theory construction begins, and terms like “understanding,” “intelligibility,” and “explanatory knowledge” seldom make an appearance once the initial stage-setting is over.

The distortion introduced by this trend into the debate over the nature of explanation is reflected in the fact that the standard rule of the game, when the adequacy of some proposed “model of explanation” is at issue, used to be, and still largely is, the method of counterexamples: you test the model against your intuitive judgments about examples and nonexamples of explanation.³

² See Salmon, *Scientific Explanation and the Causal Structure of the World*; David Lewis, “Causal Explanation,” in Lewis, *Philosophical Papers II* (Oxford: Oxford University Press, 1986); Paul Humphreys, *The Chances of Explanation* (Princeton: Princeton University Press, 1989); Sylvain Bromberger, “An Approach to Explanation,” in R.S. Butler, ed., *Analytic Philosophy*, Second Series (Oxford: Basil Blackwell, 1962); Peter Achinstein, *The Nature of Explanation* (Oxford: Oxford University Press, 1983); Bas Van Fraassen, *Scientific Image* (Oxford: Oxford University Press, 1980).

³ As Peter Railton observes in a related context; see his “Explanation and Metaphysical Controversy,” in Philip Kitcher and Wesley C. Salmon, eds., *Scientific Explanation*, Minnesota Studies in the Philosophy of Science, vol. 13, (Minneapolis: University of Minnesota Press, 1989).

The models are seldom evaluated by explicitly considering whether they give a satisfactory account of understanding, and the stock examples and counterexamples that have played a key role in the rise and fall of models of explanation have not been seriously analyzed on principled grounds—exactly why they provide, or fail to provide, genuine explanatory understanding or knowledge. To be sure, our preanalytic and untutored judgments (“intuitions”) about them must reflect certain inexplicit and unarticulated general principles we hold that interrelate an amorphous group of concepts including understanding something, answering why-questions, knowing a cause of something, dispelling puzzles, and so on. But the central task of a theory of explanation must be to bring into the open these tacit principles, formulate them in precise and systematic statements, and provide them with a rationale in the context of a general theory of knowledge. We see in retrospect that somehow the concepts that ought to have been at center stage in the theory of explanation had been shunted aside. Think of a theory of justice that never mentions justice, or a theory of number that never mentions numbers. The anomaly would be exactly the same. In a nutshell, then, the trouble is that we have had *models*, but not *theories*, of explanation.

But philosophy of science, in which we normally place theories of explanation, is not the only field to blame; theory of knowledge is equally at fault. The traditional theory of knowledge has focused very narrowly on propositional knowledge, and its chief project has been the analysis of *knowing that p*, for a single, stand-alone proposition *p*. The mainstream analytic epistemology has paid little attention to the concept of understanding, or how bits and pieces of propositional knowledge can come to constitute an intelligible epistemic structure. The concept of explanation or of understanding surely does not undergo a radical, discontinuous transformation when we move from ordinary knowledge to scientific knowledge, any more than the concept of knowledge does. Explanations in the developed sciences may be more precise, more systematic and unified, calling on resources richer and more varied than what is available outside them, but surely the root concept of understanding must be much the same. It is only that understanding can be realized in a variety of ways. There is every reason then why explanation, that is, understanding, should be among the central concerns of general epistemology.

The point of going over all this is to remind ourselves that explanation is an epistemic affair, and that explanations are among our important epistemic

achievements. The idea of explaining something is inseparable from the idea of making it intelligible; to seek an explanation of something is to seek to understand it, to render it intelligible. These are simple conceptual points, and I take them to be untendentious and uncontroversial.

II

If we begin with these ideas firmly in mind, we can see that the following must be recognized as the central question of any theory of explanation:

The Epistemological Question: What is it that we know—that is, what exactly is our epistemic gain—when we have an explanation of p ?

Typically, when we seek an explanation of why p we already know, or at least believe that we know, that p is the case. We want an explanation of p presumably because we are in an epistemically imperfect state vis-à-vis the fact that p , and there is something we need to know, or at any rate something epistemic we must acquire, if we are to ameliorate our epistemic situation. It may be that there is a proposition, q , about some particular matters of fact such that if we were to know that q , we would have an explanation of p ; that is, the deficit may be one of insufficient information about specific matters of fact. But an explanation may be more than the possession of pieces of propositional knowledge; it may involve, or require, the seeing of a pattern, a certain relationship or configuration, in disparate propositions. Let E be a set of propositions that makes up an explanation of p (we will follow the standard terminology and call it an “explanans” of p). The general epistemological question then is this: when we have E as an explanation of p , what is it that we know? What exactly is the epistemic gain we have achieved?

So far the issues have been put in general terms so that they may in principle arise for any explanation involving an explanandum represented by a proposition. These issues could, it seems, be generalized to apply to all kinds of explanation, including, e.g., explanations of concepts and meanings, the rules of a game, the significance of a social ritual, the function of a biological organ or social institution, etc.⁴ From this point on, though, we

⁴ This of course doesn't mean that there will be an informative uniform answer that covers all varieties of explanations.

will focus, more or less exclusively, on explanations of events, as has been customary in philosophical discussions of explanation. In such an explanation, the explanandum is an event, or the occurrence of an event if you prefer, and it is represented in an explanation by a statement to the effect that the event occurred. Let us use lower-case letters, “*e*,” “*f*,” etc. to stand for individual events, and the corresponding upper-case letters, “*E*,” “*F*,” etc. for statements affirming that these events occur or occurred.⁵

Suppose then that an explanation of why an event, *e*, occurred is being sought, and that another event, *g*, has been invoked to explain it. Let us see how the situation might be represented under the Hempelian covering-law models, in particular, the deductive-nomological (“D-N”) model. On this model, an explanation of event *e* in terms of event *g* is a D-N argument, with *E* (the statement to the effect that *e* occurred) as the conclusion and *G* (the statement to the effect that *g* occurred) as a premise. In addition, the explanans will include a general law, *L*, which, together with *G*, deductively yields *E*. Thus, the proffered explanation is a linguistic-propositional structure, something that belongs in our epistemic system, the total body of our knowledge at a given time. To evaluate whether a proposed explanation is a genuine, correct explanation, we would, on Hempel’s account, need to determine various things: that *L* is a law, that *E* is logically derivable from (or entailed by) *L* and *G* but not from either alone, and that each statement in the explanans is true. Notice that apart from the determination of the truth of the explanatory premises, the evaluation of an explanation is a purely *internal* affair, in that the objects involved are statements and they are examined in respect of their forms and mutual logical relationships. In particular, there is no need to ascertain whether any special relationship obtains between the two events, the event to be explained and the event being invoked as its explanation. For Hempel, for *g* to explain *e* is nothing more, or less, than for *G* to nomologically imply *E*; the explanatory relation between *g* and *e* collapses into a logical relation between their descriptions. The basic relation that generates an explanatory relation is a logico-linguistic one that connects descriptions of events, and the job of formulating an explanation consists, it seems, in merely re-arranging appro-

⁵ Alternatively we could formulate our discussion in terms of “facts,” rather than events, without affecting any of the essential points.

prate items in the body of propositions that constitute our total knowledge at the time. In explaining something, then, all action takes place *within* the epistemic system, on the “subjective” side of the divide between knowledge and the reality known, or between representation and the world represented.

The foregoing, as I take it, is the official—that is, Hempel’s—story about covering-law explanations. There is, however, a very different story that can be told of what’s going on in a D-N explanation: one could say that the very point of a D-N derivation of *E* from *G* is to demonstrate that *g* is a *cause* of *e*, and that a D-N argument is explanatory only because, and to the extent that, it succeeds in depicting a causal relation between *g* and *e*. Ultimately it is this objective causal relation, something that holds *outside* our epistemic system, that grounds the fact that the D-N argument with *G* and *L* as premises and *E* as conclusion is an explanation.⁶ This brings us to the second fundamental question for theories of explanation:

The Metaphysical Question: When *G* is an explanans for *E*, in virtue of what relation between *g* and *e*, the events represented by *G* and *E* respectively, is *G* an explanans for *E*? What is the objective relation connecting events, *g* and *e*, that grounds the explanatory relation between their descriptions, *G* and *E*?

One possible response is to deny that any such objective relation exists, or needs to exist, to “ground” the explanatory relation between their descriptions. This, I believe, is the official stance of the covering-law theorist: there is no further fact about an explanation beyond the fact that it is a D-N argument showing that the event to be explained is nomologically expectable on the basis of the antecedent circumstances as described in the explanans. The scientific understanding produced by the explanation consists precisely in the grasping of the nomic expectability of the event to be explained. Consider the following interestingly revealing remarks by Hempel, in one of the few places where he explicitly mentions the issue of understanding in connection with his models of explanation:

⁶ Hempel’s position on causal explanations is complex and ambivalent. He would grant that many D-N explanations of individual events are causal explanations in the sense that the initial conditions represented by the singular premises are a cause of the event represented by the explanatory conclusion. However, he denies that all singular explanations are causal, and would deny that a given D-N argument constitutes an explanation in virtue of the fact that it correctly depicts a causal relation between the explanandum and the events invoked in the explanans.

Thus a D-N explanation answers the question “*Why* did the explanandum phenomenon occur?” by showing that the phenomenon resulted from certain particular circumstances, specified in C_1, C_2, \dots, C_k , in accordance with the laws L_1, L_2, \dots, L_r . By pointing this out, the argument shows that, given the particular circumstances and the laws in question the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred.⁷

Notice Hempel’s use of the causative “resulted” in his first sentence; this shows, we may speculate, that Hempel was intuitively pulled in the direction of giving a causal answer to the Metaphysical Question: a D-N explanation explains by showing that the event to be explained “resulted from,” that is, caused by, the events invoked in the explanans. But he immediately pulls back, and, in the sentence that follows, gives an entirely unrelated answer, the “official” position of the D-N theory: a D-N explanation explains by showing that the event being explained was to be expected given the events and laws cited in the explanans. As I take it, this represents Hempel’s real position; for him, nomic expectability is of the essence of scientific understanding.

On a view like Hempel’s, then, explanation is essentially an activity *internal* to an epistemic corpus: whether or not something is an explanation—a good, “true” or “correct” explanation—depends on factors internal to a body of knowledge, not what goes on in the world—except, of course, for the truth of the statements comprising the explanans. We may call an approach of this kind “explanatory internalism.”

If, on the other hand, you are disposed to look for a nontrivial answer to the Metaphysical Question, attempting to find a relationship between the explanandum event and the event invoked as its explanation, you are taking the stance of “explanatory realism” (or “explanatory externalism”), as we may call it. A realist about explanation believes that some objective relation between the events underlies, or grounds, the explanatory relation between their descriptions. That is, statement G constitutes a correct explanation of statement E in virtue of the fact that a certain relationship obtains between events g and e . What could such a relationship be? One strong traditional favorite of course is the causal relation: G is a correct explanans of E because

⁷ Carl G. Hempel, “Aspects of Scientific Explanation,” in his *Aspects of Scientific Explanation* (New York: The Free Press, 1965), p. 337 (the italics are Hempel’s).

g is a cause of e . On this view, what warrants the use of G in framing an explanation of event e , and thus an answer to the question “Why did e occur?,” is the fact that the event represented by G , namely g , is a cause of e . The fact that G is an explanans of E , therefore, depends on something outside one’s epistemic system; just as the claim to know that p requires an “objective correlate,” namely the fact that p , the claim that G is an explanans for E requires an objective correlate, namely the fact that g is a cause of e .⁸

But is the causal relation the only objective relation capable of grounding the explanans relation? Is it the only relation that is explanatory? David Lewis has defended an affirmative answer, at least for explanations of individual events.⁹ According to him, all explanations of individual events are causal explanations, each of them telling something about “the causal history” of the event being explained. Is Lewis right? We will touch on this question later, but for now Lewis’s position serves to illustrate the realist approach to explanation. For Lewis, every explanatory claim is a causal claim, and what makes an explanation of e in terms of g a correct explanation is the fact that an appropriate causal relation obtains between e and g , and the proffered explanation is informative as an explanation to the extent that it is informative about the causal history of e .

It bears repeating that a proponent of the D-N model need not be an explanatory internalist, although Hempel pretty clearly was an internalist. It is clear that the D-N model is exactly what an explanatory realist who believed in causation as the “explanation-grounding relation” (as we might call it) would embrace, if he also believed in the nomic-subsumptive conception of causation, the essentially Humean view that events are connected as cause and effect in virtue of being subsumed by an appropriate law. For such a person, the D-N model is, more directly, a model of what it

⁸ In *Scientific Explanation and the Causal Structure of the World* Salmon distinguishes among “three basic conceptions” of scientific explanation, the “epistemic,” the “modal,” and the “ontic” conceptions. The epistemic conception, under which he includes Hempel’s covering-law theory, appears to be a form of what I am calling “explanatory internalism” (but not exhaustive of it); both the modal and the ontic conceptions appear to be forms of explanatory realism, the former grounding the explanatory relation in the relation of necessitation among events and the latter grounding it in the causal relation. The term “ontic,” therefore, seems a bit of a misnomer: the modal approach is as “ontic” as the ontic approach. And it would seem that the term “epistemic,” as used by Salmon, is also misleading. I think it should be granted on all hands that explanation is epistemic (what else could it be?); the proper question to raise is whether there is something “ontic,” out there in the world, that grounds or underwrites (“corresponds to,” if you prefer) the explanatory relation.

⁹ In “Causal Explanation.”

is for events to be subsumed under a law,¹⁰ and a D-N argument is a perspicuous way of exhibiting events in a causal-explanatory relation.

One might argue that if you embrace the kind of causal-explanatory realism just sketched, there is little reason to accept the D-N model. For if the causal relation is what at bottom generates explanatory understanding, why should we restrict ourselves, one might ask, to a single way of representing that relation in an explanation? Why, that is, should we limit ourselves to D-N arguments to exhibit the causal relation and not avail ourselves of other ways of providing information about the causal facts involving the explanandum? (Notice that a parallel question can be raised in regard to any explanatory realist who believes in a “model” of explanation.) It would seem that the best course for the causal realist is to follow Lewis, who has said that “to explain an event is to provide some information about its causal history.”¹¹ I believe that the D-N theorist has a plausible line for an answer: she could say that a D-N argument is a particularly perspicuous way of representing the causal information involved. A successful D-N argument, in virtue of the requirement of nomological derivation, exhibits those features of the supposed cause that are efficacious or relevant in the production of the explanandum event, and this makes it explanatorily informative in a way in which a mere tagging of the cause, by a proper name or a description based on causally irrelevant features of the situation, is not. Consider: “Why was the entire crop of the island wiped out?” “It was caused by the event reported on p. 12 of the last *Sunday Times*”.

Although this is not an implausible response, it is not conclusive; granted that appropriate D-N arguments are particularly perspicuous as a vehicle for conveying causal information, there is no reason to exclude all other ways of representing such information. Understanding and explanatory toriness are matters of degree, and there seems little reason to insist that all explanations take a single logical form, or a small set of such forms. The D-N model makes explanation an all-or-nothing affair, and it must be granted that this does not sit well with explanatory realism. The same comment applies to most other “models” of explanation.

¹⁰ During the heyday of logical positivism causation apparently came under the positivist strictures against metaphysics, and there was little overt discussion of issues related to causation. It is possible to think of the Hempelian D-N account of explanation as a covert causal theory of explanation dressed up in a positivistically correct internalist guise.

¹¹ “Causal Explanation,” p. 217.

III

According to Salmon, another causal explanatory realist, to explain an event is to exhibit its place in the causal structure of the world.¹² As explanatory realists, therefore, Lewis and Salmon have a simple answer to the Epistemological Question. What it is that we know when we have an explanation is a causal fact; explanatory knowledge is causal knowledge. In general, then, the explanatory realist has a ready answer to the Epistemological Question: if *R* is identified as the explanation-grounding relation (as we might call it), what we know when we have an explanation is the fact that *R* obtains between appropriate events. Thus, explanatory knowledge is just additional bits of propositional knowledge: on causal explanatory realism, for example, to have an explanation of *e* in terms of *g* is to know, or accept, the proposition that *g* is a cause of *e*.

But this still leaves our initial epistemological issues about explanation unresolved; we still need an epistemology of understanding, for a theory of explanation, we argued, must be a theory of understanding. Take the causal theorist like Lewis and Salmon; we still need him to answer the following question: what does causal knowledge have to with understanding? That is, why, and in what way, does the knowledge that some event, *g*, caused *e* produce, or enhance, our understanding of *e*, or help make *e* intelligible? In general, then, the explanatory realist has the task of explaining why and how knowledge that the grounding relation *R* is instantiated in a given situation promotes our understanding of that situation.

Some explanatory realists seem to reject, or at any rate try to avoid, this as a responsibility of a theory of explanation. For example, Paul Humphreys thinks that considerations of understanding will lead to “a relativization of explanations to an individual,” since the epistemic situation differs from one individual to the next; he proposes to “set aside that whole issue of what constitutes or promotes understanding.”¹³ Salmon does raise the issue of explanation and understanding, writing:

¹² See *Scientific Explanation and the Causal Structure of the World*, pp. 19–20.

¹³ *The Chances of Explanation*, p. 127.

Perhaps the most important fruit of modern science is the understanding it provides of the world in which we live, and of the phenomena that transpire within it. Such understanding results from our ability to fashion scientific explanations.¹⁴

We can agree with what Salmon is saying here. We prize explanations, especially those in science, because understanding, as Salmon puts it, “results from our ability to fashion explanations.” That is almost tautological; what isn’t tautological, and in fact far from obvious, is why and how understanding results from our ability to fashion *X*, where *X* is what a given theory of explanation says explanations are. That is, it is far from obvious just how understanding springs from our ability to construct D-N arguments, or to discover and formulate causal judgments of the sort that Salmon would consider explanatory. And the section entitled “Explanation and Understanding,” from which the foregoing quotation has been drawn, contains nothing further on understanding, and the present issue is never faced, at least not directly.¹⁵

The explanatory realist must exercise special caution in dealing with this problem. For, on explanatory realism, explanatory knowledge turns out to be just additional bits of propositional knowledge. For the causal theorist like Salmon and Lewis, an explanation of an event is merely a causal proposition, to the effect that a certain event caused the explanandum event. To “have an explanation,” on this view, is to have such a proposition of that kind in one’s epistemic corpus. This puts the distinction between “explanatory knowledge” and “descriptive knowledge” in potential jeopardy; the distinction as it is standardly drawn faces an imminent collapse. Why isn’t the knowledge that the breaking of the dam caused the town to be flooded any less descriptive than the knowledge that the bridge collapsed at midnight. What the explanatory realist must say, it seems, is that although all knowledge is descriptive, some descriptive knowledge has a special explanatory role in virtue of the particular content it has. His job then is to explain why the particular type of descriptive content (e.g., causal facts) he has identified as explanatory serves to enhance understanding. On the other hand, explanatory internalism leaves open the possibility of giving an

¹⁴ *Scientific Explanation and the Causal Structure of the World*, p. 259.

¹⁵ Salmon’s criticism of “psychologism” about explanation earlier in the book, and his rejection of “the epistemic model” of explanation, can be read as a repudiation of the concept of understanding as an element in a theory of explanation, although this might not have been his real intention.

account of a type of knowledge that does not consist in propositional knowledge, something that supervenes on, but distinct from, such knowledge. This is an intriguing possibility, and in some ways an exciting one, to which we will recur below.

IV

Michael Friedman begins his influential paper, "Explanation and Scientific Understanding,"¹⁶ with some insightful remarks on the importance of understanding in a theory of explanation. He makes much the same point we have made at the outset, when he demands that "a theory of scientific explanation tell us what it is about the explanation relation that produces understanding."¹⁷ And Friedman has no trouble showing that traditional conceptions of explanation, including Hempel's D-N model, fail to provide an account of the relation between explanation and understanding.

Friedman himself sees the key to understanding this relation in the idea of *unification*: scientific explanation produces understanding by effecting unification in our epistemic system. He writes:

Once again, we have reduced a multiplicity of unexplained, independent phenomena to one. I claim that this is the crucial property of scientific theories we are looking for; this is the essence of scientific explanation—science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than one with more.¹⁸

For example, the explanatory power of Newton's mechanics consists in its unifying power over such diverse, and apparently unrelated, laws as Kepler's laws, Galileo's law of falling bodies, the law of simple pendulum, and countless others. By deriving these laws from Newton's system, we show that they are not independent phenomena, and that they could be brought

¹⁶ *Journal of Philosophy* 71 (1974): 5–19. Reprinted in Joseph C. Pitt, ed., *Theories of Explanation* (New York and Oxford: Oxford University Press, 1988). All references to this article are to the version in the Pitt volume.

¹⁷ "Explanation and Scientific Understanding." p.189.

¹⁸ *Ibid.* p.195.

under a single set of regularities. In this way, a significant reduction in the number of independent assumptions is achieved, and the result is a more unified, simpler system that makes the world a more intelligible, more comprehensible place.

Philip Kitcher, finding faults with Friedman's account of unification, offers another account of explanatory unification.¹⁹ In contrast to Friedman, who is interested in the reduction of the number of independent laws, Kitcher focuses on the idea of reducing the number of independent argument patterns. What Kitcher finds significant in such theories as Newton's mechanics and Darwin's evolutionary biology, both universally hailed for their great explanatory power, is the fact that they employ relatively few argument patterns to derive, and explain, large numbers of phenomena. Darwin's theory, Kitcher notes, does not attempt to give a detailed derivation of how certain particular characteristics of species have emerged; rather, what's impressive about it is that it produces a derivation schema that can be applied repeatedly to particular characteristics of all species to explain how they have emerged. The schema uses the principle of natural selection as the basic premise in conjunction with other specific premises about ancestral forms and the nature of their environment, and laws of variation and inheritance (largely unknown to Darwin). Kitcher's development of this idea is rather involved, and its details will not concern us here. What is noteworthy is that Kitcher fully endorses Friedman's demand that a theory of explanation must give an account of how explanation is conducive to the enhancement of understanding. He is also in agreement with Friedman in holding that unification is the key to understanding. They differ only in their analysis of unification—or about the kind of unification that makes a difference to understanding.

It is useful to view Friedman and Kitcher as explanatory internalists.²⁰ They appear to follow Hempel in taking explanations to be closely tied to law-based derivations. Neither tells us explicitly what explanations, in the sense of "products," are, but it is clear enough that, like Hempel, they seem to take explanations as involving nomological derivations. The crucial point of interest for us here is this: what makes these derivations explanatory, on

¹⁹ In "Explanatory Unification," *Philosophy of Science* 48 (1981): 507–531. Reprinted in Pitt, *Theories of Explanation*.

²⁰ Salmon would probably agree; he thinks that they adopt the epistemic conception of explanation, which roughly corresponds to what I am calling explanatory internalism.

Friedman's and Kitcher's accounts, is their relationship to other items within our epistemic system, not some objective facts about external events or phenomena. On Kitcher's account, for example, what makes a given D-N argument explanatory is the fact that it is a member of a class of arguments (a "systematization," as Kitcher calls it, of our belief system) which best unifies our belief system. And the measure of degrees of unification depends solely on factors internal to the epistemic system, such as the number of argument patterns required to generate the given class of arguments, the "stringency" of the patterns, etc., not on any objective relations holding for events or phenomena involved in the putative explanations. For Friedman, whether or not a given law explains another²¹ is determined crucially by the unifying power of the explaining law, and the concepts in terms of which the latter is explained are exclusively logical ones (equivalence, implication, etc.) and evidential ones ("independent acceptability"). In general, such virtues as simplicity and unity are primarily properties of theories and belief systems; they are epistemic virtues that concern properties of our representations of the world, but not of the world itself.

Let us now consider how these internalist theories answer the Epistemological Question about explanation: What is it that we know when we have an explanation? It is clear that, on either Kitcher's or Friedman's view, there is no particular proposition that we come to know when we gain explanatory knowledge. Explaining is not, for them, a matter of discovering, or imparting, more propositional knowledge; explanatory activity consists in constructing derivations whose structure and steps are logically or epistemically related in certain specified ways to the rest of the belief system. To put it somewhat crudely, explanation is a matter of the shape and organization of one's belief system, not of its content. Both Kitcher's and Friedman's accounts make explanation a holistic affair: whether or not a given derivation is an explanation cannot be determined locally, just by looking at the derivation; it depends on facts about the whole belief system.²² This means that the epistemic gain represented by an explanation cannot be measured or represented precisely. The fact that a particular derivation is an explanation relative to a belief system says something about the structure

²¹ Friedman is concerned only with explanations of laws, and therefore doesn't quite fit our scheme of classification; however, this does not obscure the internalist character of his account.

²² It seems that any unification approach to explanation will be holistic, although this isn't true of all internalist theories (compare, e.g., Hempel's covering-law theory).

and organization of the system and the place and role of the derivation within it, and this makes it impossible to say anything precise and useful about the epistemic gain to be associated with individual explanations.

This holistic picture is appealing in some ways. We are apt to think that explanatory understanding cannot just be a matter of having more information, more propositional knowledge; somehow, it seems like something that overlays, or supervenes on, but is distinct from, knowledge of particular propositions. On the other hand, the holistic picture doesn't fit certain other salient aspects of our explanatory practice. It may be that, as Friedman says, "a world with fewer independent phenomena is . . . more comprehensible than one with more," and that, as Kitcher holds, a better unified theory is more explanatory. On such a holistic conception, though, it is not even clear that the idea of explaining some individual item, whether this is an event or a law, can have a determinate meaning. In particular, causal explanations of individual events and actions seem to make no sense within such a picture. It seems obvious, however, that we construct and verify such explanations locally, and derive explanatory understanding from them. To ascertain a derivation as an explanation, Kitcher asks us to consider our whole belief system and whether or not the derivation is a member of the class of derivations that can be generated by a set of argument patterns which collectively have a certain unifying property. This seems wildly unrealistic: it isn't something we do or need to do; it probably isn't something that any of us *can* do!

But what does the unification approach tell us about explanation and understanding? Surprisingly little, I think. Kitcher complains about the Hempelian account by asking, "Why should it be that exactly those derivations which employ laws advance our understanding?" Good question! He then goes on to say, "As Friedman points out, we can easily connect the notion of unification with that of understanding."²³ But how is the connection supposed to go? Just how does unification manage to advance understanding? Kitcher himself says nothing further; so what does Friedman have to say? Against Hempel, who, as we saw, takes nomic expectability as the source of explanatory understanding, Friedman points out that this concept and that of understanding are "quite distinct notions,"²⁴ and that Hempel

²³ "Explanatory Unification," p. 168 (in the Pitt volume).

²⁴ "Explanation and Scientific Understanding," p. 190.

fails to provide an intelligible linkage between them. Friedman is of course right; but then is it so clear that the concept of unification, whether taken preanalytically or in either of the senses explicated by Friedman and Kitcher, is any closer to understanding? The preanalytic notion of unification does not seem to have anything significant to do with the idea of understanding. To be sure, a better unified, or simpler, theory is easier for us to “understand,” in the senses of learning or making use of it.²⁵ Given our finite intelligence, we do better with simpler concepts, fewer axioms and rules, and so on. But not always: as anyone who has studied axiomatic systems of logic knows, the simplest systems, simplest in terms of the number of primitives, axioms, and rules of inference, severely tax our ability to understand. We generally do better with middlingly complex systems with intuitive concepts and rules, and some redundancies built into them. In any case, this notion of understanding, roughly in the sense of learnability (for humans), does not seem to be what is relevant to the explanatory understanding of events and facts.

Kitcher’s concept of unification is built essentially from logical concepts, but Friedman’s makes use of epistemic concepts as well. The principal epistemic concept in Friedman is that of “independent acceptability,” which seems to belong to the same general genre of evidential concepts as that of “expectability,” which Friedman finds “quite distinct” from understanding. And it is not obvious just at what stage Friedman’s analysis of unification is supposed to make contact with understanding. If unification is to be intelligibly connected with understanding, the connection must be consciously forged, and to do that one would have to say a lot more about understanding. It would be unrealistic to expect that somehow understanding will magically materialize out of simplicity and unification, concepts that appear to be—and, on Kitcher’s and Friedman’s accounts, are—largely formal and logical.

V

In this final section, I want to sketch a realist approach to explanation which has a certain affinity with the simplification–unification approach. My

²⁵ When Friedman writes “a world with fewer independent phenomena is . . . more comprehensible than one with more” (ibid. p.195), he sounds as though he has in mind something like this sense of “understanding.”

approach will show that the unification approach is not intrinsically internalist, and that, on the contrary, it can be plausibly motivated in a realist setting; in fact, I claim that explanatory realism can make better sense of the idea that explanations unify and simplify. How unification or simplification is to be connected to explanatory understanding is a difficult question, and I will have nothing useful to add to what has already been said by Friedman, Kitcher, and others. My main object is to describe a broadly realist approach that exploits the idea that explanations explain by unifying and simplifying.

The key concept I will make use of is that of *dependence*, and my claim will be that dependence relations of various kinds serve as objective correlates of explanations. Dependence, as I will use the notion here, is a relation between individual events and states; however, it can also relate facts, properties, regularities between events, and even entities. We speak of the “causal dependence” of one event or state on another; that is one type of dependence, obviously one of central importance. Another dependence relation, orthogonal to causal dependence and equally central to our scheme of things, is *mereological dependence* (or “mereological supervenience,” as it has been called): the properties of a whole, or the fact that a whole instantiates a certain property, may depend on the properties and relations had by its parts. Perhaps, even the *existence* of a whole, say a table, depends on the existence of its parts. We can think of the principle of mereological supervenience as the metaphysical principle underwriting the research strategy of microreduction and the method of microreductive explanation, in the way in which the principle of causal determination provides a metaphysical basis for the research strategy of causal investigation and causal explanation. Just as the latter rationalizes our attempt to look for the diachronic, temporally antecedent determinants of phenomena, the former urges us to search for their synchronic micro-determinants. Causal dependence and mereological dependence may well be the two most basic, and important, relations of dependence that can serve as explanation-grounding relations.

Are there other kinds of dependence relations that can ground explanations? Although this is not the place to try to be exhaustive, there are some cases to consider. Socrates expires in the prison, and Xanthippe becomes a widow. The widowing of Xanthippe depends on the death of Socrates, and

this is perhaps still another kind of dependence.²⁶ There is a widespread belief in mind-body supervenience: the mentality of a creature depends on its physical nature.²⁷ Belief in evaluative supervenience, the idea that evaluative and normative properties supervene on factual, or nonevaluative properties seems even more widely accepted.²⁸ These and other dependence relations may generate explanations—explanations of mental phenomena in terms of their underlying neurobiology, and of evaluative facts (e.g., why some act is morally right, why some work of art is beautiful) in terms of the nonevaluative facts on which they supervene.

We think of the world as a system with structure, not a mere agglomeration of unconnected items, and much of the structure we seek comes from the pervasive presence of dependence relations. Dependence is asymmetric and transitive, and can generate relational structures of dependent events, states, and properties. The ontological contribution of dependence relations lies exactly in this fact: they reduce the number of independent events, states, facts, and properties we need to recognize. And that is precisely the unifying and simplifying power of dependence relations. Unity and structure go hand in hand; dependence enhances unity by generating structure.

These points apply to lawful regularities as well as individual events and states; in particular, we think of macrophysical regularities as arising out of microphysical regularities and hence explainable by them. The principle of micro-macro determination—that is, the doctrine of mereological supervenience—is applicable to laws and regularities as well as individual states and events. Here too, underlying dependency relations support explanations.

My main proposal, then, is this: *explanations track dependence relations*. The relation that “grounds” the relation between an explanans, *G*, and its explanatory conclusion, *E*, is that of dependence; namely, *G* is an explanans of *E* just in case *e*, the event being explained, depends on *g*, the event invoked as explaining it.

²⁶ See my “Noncausal Connections,” in *Supervenience and Mind* (Cambridge: Cambridge University Press, 1993).

²⁷ There is the possibility of construing mind-body supervenience as a form of mereological supervenience. See “Postscripts on mental causation” in my *Supervenience and Mind*.

²⁸ Moral realists will reject formulations of supervenience in terms of “properties;” however, there are formulations that should be acceptable to them, e.g., in terms of predicates or judgments.

On this proposal, therefore, the simplifying effect of an explanation is seen both in our belief system and in the world: by showing an event to be dependent on another, the explanation reduces the number of independent events in the world, and also the number of independent assumptions we need to accept about the world. This approach differs importantly from the internalist approaches of Kitcher and Friedman, in two crucial respects: simplicity and unity are features of the events and facts of the world as well of our beliefs and propositions, and they are, in the first instance, *local* features of the propositions and events involved in individual explanations, rather than *holistic, global* features of the whole belief system or world. But the account does not ignore the holistic aspect; far from ignoring it, it gives an account of how scientific understanding at the global level arises out of local explanations. For the dependency structures generated by individual explanations are cumulative, and contribute to the overall unity and simplicity of the system. In consequence, our account allows explanatory understanding both at the local and the global level, and relate them in a natural way, whereas on Friedman's account, scientific understanding is exclusively a global feature of a whole theoretical system, and there is no obvious way of making sense of scientific understanding contributed by individual explanations.²⁹

Within the general scheme of realism, explanatory realism is appealing and perhaps even compelling.³⁰ I am in agreement with Salmon when he says, "we must surely require that there be some sort of *objective* relationship between the explanatory facts and the fact-to-be-explained."³¹ The present approach does justice to this realist requirement. Moreover, it shows that the idea of unification and simplification can be combined with explanatory realism in a natural and plausible way, and opens the possibility of an account of understanding within a realist scheme. Producing a usable account of understanding is the hard part; one scarcely knows where to begin. The difficulty does not seem to lie on the side of unity and simplicity; I think we have a serviceably clear understanding of these notions. What we

²⁹ Friedman, "Explanation and Scientific Understanding," p. 197.

³⁰ David-Hillel Ruben's *Explaining Explanation* (London and New York: Routledge, 1990), chap. 7, contains a perceptive and enlightening discussion of the realist perspective on explanation, to which the present paper is indebted. I first discussed explanatory realism in "Explanatory Realism, Causal Realism, and Explanatory Exclusion," *Midwest Studies in Philosophy* 12 (1988): 225–240. Essay 7 in this volume.

³¹ *Scientific Explanation and the Causal Structure of the World*, p. 13.

lack is an understanding of understanding that is clear and rich enough for useful theorizing. To build a bridge from unity and simplicity to understanding and intelligibility, we need an epistemology of understanding, something that has by and large been neglected by contemporary analytic epistemology and philosophy of science.