

# Scientific Explanation

Michael Strevens

For the *Macmillan Encyclopedia of Philosophy*, second edition

The three cardinal aims of science are prediction, control, and explanation; but the greatest of these is explanation. Also the most inscrutable: prediction aims at truth, and control at happiness, and insofar as we have some independent grasp of these notions, we can evaluate science's strategies of prediction and control from the outside. Explanation, by contrast, aims at scientific understanding, a good intrinsic to science and therefore something that it seems we can only look to science itself to explicate.

Philosophers have wondered whether science might be better off abandoning the pursuit of explanation. Duhem (1954), among others, argued that explanatory knowledge would have to be a kind of knowledge so exalted as to be forever beyond the reach of ordinary scientific inquiry: it would have to be knowledge of the essential natures of things, something that neo-Kantians, empiricists, and practical men of science could all agree was neither possible nor perhaps even desirable.

Everything changed when Hempel formulated his DN account of explanation. In accordance with the observation above, that our only clue to the nature of explanatory knowledge is science's own explanatory practice, Hempel proposed simply to describe what kind of things scientists tendered when they claimed to have an explanation, without asking whether such things were capable of providing "true understanding". Since Hempel, the philosophy of scientific explanation has proceeded in this humble vein,

seeming more like a sociology of scientific practice than an inquiry into a set of transcendent norms. In keeping with its mission as a branch of philosophy, however, the study of explanation pursues a particular kind of sociological knowledge: it is concerned almost exclusively with the ideal at which scientists take themselves to be aiming, and barely at all with the steps and missteps taken on the way to realizing the ideal.

As Hempel saw it, scientific explanation was of a piece with prediction, requiring the same resources and giving a similar kind of satisfaction. No doubt this modest view of the explanatory enterprise played a part in making the study of explanation acceptable in the climate of post-war empiricism. The story of explanation in recent decades, however, is an expansionist one. Over the years, philosophers of explanation have gradually required more resources for, and made grander claims for the significance of, explanation's role in science. (For a comprehensive overview of the philosophy of explanation from 1948 to 1988, with a full bibliography, see Salmon (1990).)

**The DN Account** Hempel's deductive-nomological, or DN account (Hempel and Oppenheim 1948), is intended to capture the form of any deterministic scientific explanation of an individual event, such as the expansion of a particular metal bar when heated, the extinction of the dinosaurs, or the outbreak of the American Civil War.

According to Hempel, such an explanation is always a deductive derivation of the occurrence of the event to be explained from a set of true propositions including at least one statement of a scientific law. (The event to be explained is called the *explanandum*; the set of explaining statements is sometimes called the *explanans*.) In other words, a deterministic event explanation is always a sound, law-involving, deductive argument with the conclusion that the explanandum event occurred.

Intuitively, the premises of a DN explanation spell out the relevant initial, background, and other boundary conditions, together with the laws governing the behavior of the system in which the explanandum occurred. Hempel

cites the following argument, for example, as a typical DN explanation of the event of a thermometer's mercury expanding when placed in hot water:

The (cool) sample of mercury was placed in hot water, heating it,  
Mercury expands when heated, thus

---

The sample of mercury expanded.

Because the law or laws that must be cited in a DN explanation typically “cover” the pattern of behavior of which the explanandum is an instance, the DN account is sometimes referred to as the *covering law account* of explanation.

You will see that the DN account is not intended to give the form of probabilistic event explanations; Hempel offers a separate account of probabilistic explanation, to be discussed below. The explanation of phenomena other than events is, by contrast, apparently amenable to the DN approach. Hempel suggests that a scientific law can be explained, for example, much like an event, by deducing it from premises including at least one other law. However, he finds himself unable to make good on this proposal, for reasons connected to the relevance problem discussed in the next section.

Many scientific explanations of events and other phenomena undoubtedly have the form proposed by the DN account: they are logical derivations from laws and other information. Hempel does not entirely satisfy himself, however, with answering questions of form. Taking one step beyond sociological humility, he advances a thesis as to why deductive, law-involving arguments should confer understanding:

[A DN explanation] 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 (Hempel 1965a, 337).

Scientific understanding, then, takes the form of retrospective expectation: you might say (loosely) that, whereas prediction is concerned with what we should expect in the future, explanation is concerned with what we should have expected in the past. Explanation is, then, put on a par with prediction, and so made safe for empiricist philosophy of science. Hempel even goes so far as to say that the difference between explanation and prediction is merely pragmatic (Hempel and Oppenheim 1948, 249), though the DN account does not in itself entail such a thesis.

**Objections to the DN Account** Three kinds of objections to the DN account have been especially important for the subsequent development of the philosophy of explanation.

The first kind of objection, developed by Kyburg, Salmon, and others, points to the DN theory's inability to account for judgments of explanatory relevance. The paradigm is the following argument, which satisfies all of the DN account's criteria for a good explanation of the event of a particular teaspoon of salt's dissolving:

The teaspoon of salt was hexed (meaning that certain hand gestures were made over the salt),

The salt was placed in water,

All hexed salt dissolves when placed in water, thus

---

The salt dissolved.

The explanation appears to attribute the salt's dissolving in part to its being hexed, when in fact the hexing is irrelevant.

There are various responses to the counter-example that aim to preserve as much of the DN account as possible, for example, holding that the generalization about hexed salt is not a true law, or imposing the requirement that a DN explanation use the most general law available.

Salmon's much less conservative reaction is to conclude that Hempel is wrong to think of explanation in terms of expectability, therefore of explanations as kinds of argument. The relation between the factors cited in an

explanation and the explanandum itself, Salmon holds, is not epistemic, but *ontic*; it should be a physical relevance relation—a relation of statistical relevance, he first proposes (Salmon 1970), or a relation of causal relevance, as he later comes to believe (Salmon 1984). The faulty explanation of the salt's dissolving is to be discarded, argues Salmon, not because of some formal or logical defect, but because it cites an event, the hexing of the salt, that fails to bear the appropriate relevance relation to the explanandum.

Hempel himself declines (early in his career, at least) to give a DN account of the explanation of laws because of a related problem. Kepler's laws may be derived from a single "law" that is simply the conjunction of Kepler's laws and Boyle's law. Such a derivation is clearly no explanation of Kepler's laws, writes Hempel, yet it satisfies the DN account's requirements: the premises are true, the argument valid and law-involving (Hempel and Oppenheim 1948, 273n33).

The second important objection to the DN account is perhaps also the most famous: it shows, most philosophers would agree, that the DN account pays insufficient attention to the explanatory role of causal relations.

The height of a flagpole can be cited, along with the position of the sun and the law that light travels in straight lines, to explain the length of the flagpole's shadow. The DN account is well able to make sense of this explanation: it can be cast in the form of a sound, law-involving argument. But now take this same argument and switch the premise stating the height of the flagpole with the premise stating the length of the shadow. You now have a sound, law-involving argument for the height of the flagpole that cites, among other things, the length of the shadow—thus, according to the DN account, you have an explanation of the height of the flagpole that cites, as in explainer, the length of the shadow. This consequence of the DN account—that the height of a flagpole can be explained by the length of its shadow—seems obviously wrong, and it is wrong, it seems, because a cause cannot be explained by its own effects.

A further famous example strongly suggests that effects can only be explained by their causes, together with the laws and background conditions in virtue of which they are causes. Suppose that the arrival of a certain kind of weather front is always followed by a storm, and that a certain reading on a barometer is a sure sign that such a front has arrived. Then a barometer reading of this sort is always followed by a storm. The storm cannot be explained, however, by citing the barometer reading and the fact that such readings are always followed by storms, though these two facts together satisfy the requirements of the DN account. A constant, robust correlation is not, it appears, enough for explanation. What is needed, as Salmon eventually concludes, is a causal relation.

Hempel for a long time resisted the suggestion that facts about causation played any special role in explanation (see, for example, Hempel (1965a), §2.2). Over the years, however, due in part to the development of sophisticated empiricist accounts of causation, this has become a minority view.

The third class of objections to the DN account focuses on the account's requirements that every explanation cite a law, and that (except in probabilistic explanation) the law or laws be strong enough to entail, given appropriate boundary conditions, the explanandum.

One way to develop the objection is to point to everyday explanations that cite the cause of an event as its explanation, without mentioning any covering law, as when you cite a patch of ice on the road as the cause of a motorcycle accident.

More important for the study of explanation in science are varieties of explanation in which there is no prospect, and no need for, either the entailment or the probabilification of the explanandum. Perhaps the best example of all is Darwinian explanation, in which a trait  $T$  of some species is explained by pointing to the way in which  $T$  enhanced, directly or indirectly, the reproductive prospects of its possessor. Attempting to fit Darwinian explanation into the DN framework creates a host of problems, among which

the most intractable is perhaps the following (Scriven 1959): for every trait that evolved because it benefited its possessors in some way, there are many other, equally valuable traits that did not evolve, perhaps because the right mutation did not occur, perhaps for more systematic reasons (for example, the trait's evolution would have required a dramatic reconfiguration of the species' developmental pathways). To have a DN explanation of *T*, you would have to produce a deductive argument entailing that *T*, and none of the alternatives, evolved. You would have to be in a position, in other words, to show that *T had to evolve*. Not only does this seem close to impossible; more importantly, it seems unnecessary for understanding the appearance of *T*. You can understand the course of evolution without retrospectively predicting its every twist and turn.

Hempel is aware of the problem with Darwinian explanation. His response is to argue that there is no such thing: faced with a choice between the DN account and Darwinian explanation, we should opt for the former, and consider Darwinian stories to be at best partial explanations of traits (Hempel 1965c). He advocates a similar deflationary treatment of functionalist explanation in sociology and of historical explanations that are not entailments.

**The IS Account** Hempel's account of the probabilistic explanation of events, the inductive-statistical or IS account (Hempel 1965a, §3), in many ways parallels the DN account of deterministic event explanation. Like a DN explanation, an IS explanation is a law-involving argument giving good reason to expect that the explanandum event occurred. However, whereas a DN explanation is a deductive argument entailing the explanandum, an IS explanation is an inductive argument conferring high probability on the explanandum.

Hempel's example is the explanation of John Jones' swift recovery from a strep infection. The probability of a swift recovery without the administration of penicillin, Hempel supposes, is 0.1, while the probability with

penicillin is 0.9. Citing Jones's infection, his treatment with penicillin, and the resulting high probability of recovery, then, confers a high probability on Jones's swift recovery; in the circumstances you would expect him to recover swiftly. This inductive argument is sufficient, in Hempel's view, to explain the swift recovery.

Inductive soundness imposes one additional requirement that has no parallel in deductive logic. Suppose you know that Jones's strain of strep is resistant to penicillin. An inductive argument is said to be sound only if all relevant background knowledge is taken into account; consequently, an inductive argument for Jones's swift recovery must cite the infection's penicillin resistance. But once the new premise is added, the argument will no longer confer a high probability on its conclusion. This is what is wanted: there ought to be no inductive argument for swift recovery—you ought not to expect swift recovery—when the strep is known to be resistant.

Hempel imposes a similar requirement on IS explanations, which he calls the requirement of maximal specificity (see Hempel (1965a), §3.4 for the details). In virtue of this requirement, it is not possible to explain Jones's swift recovery by citing treatment with penicillin when the infection is known to be penicillin resistant.

As with the DN account of explanation, a number of objections to the IS account have exerted a strong influence on the subsequent development of the philosophical study of explanation. Versions of both the relevance and the causal objections apply to the IS account as well as to the DN account. I will briefly describe two other important criticisms.

First is the complaint that it is too much to ask that explanations confer high probability on their explananda. In many ways, this is the analogue of the third objection to the DN account above; in the same paper that Scriven expresses doubts about the existence of a DN treatment of Darwinian explanation, he describes the following example, best conceived of, I think, as an objection to the IS account. The probability that John Jones contracts



paresis, a form of tertiary syphilis that attacks the central nervous system, given that he has untreated secondary syphilis, is very low. But only syphilitics contract paresis. It seems reasonable to cite untreated syphilis, then, as explaining John Jones's paresis, though the explanation confers only a low probability on the explanandum.

The proponent of the IS account is committed to rejecting such attempts at explanation, as Hempel does, arguing that in such cases we have only a partial explanation of why the patient contracted syphilis. This is perhaps one of the most convincing of Hempel's defenses, but the paresis example is nevertheless widely regarded as posing a serious problem for the expectability approach to explanation.

A second objection to the IS account focuses on the requirement of maximal specificity. The requirement insists that all relevant background knowledge must be included in a probabilistic event explanation, but it does not require that relevant but unknown information be taken into account. In particular, if John Jones's infection is penicillin resistant, but this fact is not known to the explainer, then the IS account deems the explainer's appeal to the administration of penicillin as a perfectly good explanation of Jones's swift recovery.

As Coffa (1974) argues, this is surely not correct. If the infection is resistant to penicillin, then the administration of penicillin cannot explain the recovery, regardless of what the explainer does and does not know. The requirement of maximal specificity makes probabilistic explanation relative to the explainer's epistemic situation, then, in a way that it very much appears not to be. This objection hits right at the heart of the expectability conception of explanation, suggesting that explanation is not an epistemic matter in the least.

A third objection that is applicable to many accounts of probabilistic explanation will be raised in the discussion of the statistical relevance account below.

**The Statistical Relevance Account** In response to the DN account's relevance problem, Salmon suggests that the factors cited in an explanation must stand in a relation of statistical relevance to the explanandum. He does not intend this as a friendly amendment to Hempel's account, but rather as a radical reconceptualization of the nature of explanation: the function of an explanation, Salmon argues, is not to show that the explanandum was to be expected, but to describe factors—ideally, all the factors—statistically relevant to the occurrence of the explanandum (Salmon 1970).

From the beginning, statistical relevance is presented as an objective relation, that is, a relation holding independently of the explainer's background knowledge or other context. (Coffa's critique of the IS account, discussed above, discourages relativistic backsliding.) Salmon thus requires an account of probability that is both objective and also broad enough to encompass any possible explanandum.

For breadth, he settles on frequentism, the view that the probability of an event type is equal to the frequency with which it occurs in a reference class of outcomes. For objectivity, he works hard to develop what he calls a homogeneity constraint on the reference classes that can be used as bases for explanatory probabilities. Such a constraint, he believes, is strong enough to determine a single, observer-independent probability distribution over any set of outcomes of interest. Salmon (1984) summarizes the theory of homogeneity; for further information, see the discussion of the reference class problem in the separate entry on probability and chance.

Statistical relevance is a comparative concept: to say that a factor *A* is statistically relevant to the occurrence of an event *E* is to say that the probability of *E* (or for the frequentist, of events of the same type as *E*) in the presence of *A* is greater than the probability of *E* in the absence of *A*. Thus the determination of a relevance relation requires not only a reference class—a class of outcomes all of which occurred in the presence of *A*—but a contrast class, a class of outcomes all of which occurred in the absence of *A*. The contrast

class is not normally homogeneous. Thus for Salmon, the contrast probability must be a weighted sum of different homogeneous probabilities, each corresponding to a different way that *A* might have been absent, and giving the probability of *E* when *A* is absent in that way.

Perhaps inevitably, if not inescapably, Salmon arrives at the view that a full SR explanation is a complete table of relevance, describing not only factors that are present and statistically relevant to the explanandum, but also factors that are absent but would have been statistically relevant if they had been present. He further adds to the table all the alternatives to the explanandum *E* with respect to which there existed homogeneous probabilities, and a list of all the factors that would have been relevant to these alternatives, if they occurred. Consequently, the information proffered in an SR explanation of an event *E* not only explains the actual occurrence of *E*, but would also explain any occurrence of an event of the same type, even if different relevant factors were present, as well as the occurrence of any alternative to *E*.

As something of a corollary to this view, Salmon holds that negatively relevant factors—factors that lower the probability of the explanandum—are as explanatory as positively relevant factors, and that all factors should be mentioned regardless of their degree of relevance. Salmon's not discriminating among these factors is perhaps best understood as follows. Seeing that a factor is statistically relevant to the explanandum is an explanatory end in itself. That the factor made a particular kind of change—positive or negative, large or small—to the total probability of the explanandum would be important only if appreciating the value of the total probability were also an explanatory end. But it is not: knowing which relevance relations hold is all that matters.

I will consider several objections to the SR account. First, for all Salmon's justifications, an SR account seems to contain too much information. To explain *E* when *A* was absent, why is it necessary to know that, had *A* been

present, it would have been relevant? Why is it further necessary to know what would have been relevant to the occurrence of some alternative to a type *E* event that did not in fact occur? This information does not appear to be directly relevant to the explanatory task at hand, that of explaining *E* itself.

Second, the SR account seems vulnerable to the causal objection to the DN account: it seems to hold that *A* is explanatorily relevant to *E* whenever *A* is correlated with *E*, when in fact it is necessary that *A* be a cause of *E*. The barometer reading is statistically relevant to the storm in the example described above, but it does not thereby explain the storm.

Salmon is well aware of this problem, and proposes that only certain kinds of statistical relevance relations are explanatory, namely, those that survive a *screening off* test. (See discussion of probabilistic theories of causation?) A factor *A* that is correlated with *E* is screened off from *E* by another factor *B* if, conditional on *B*, *A* makes no difference to the probability of *E* (just as, for example, conditional on the presence of the front, the barometer reading makes no difference to the probability of the storm), but conditional on *A*, *B* does make a difference to the probability of *E*. When there is some *B* that screens off *A* from *E*, Salmon says that *A* is not genuinely statistically relevant to *E*. And *A*'s relevance will indeed disappear in a relevance table that also cites *B*. Note that Salmon's treatment does not make an explicit appeal to causal facts. Whether all problems concerning the role of causation in explanation can be solved in this way is unclear.

A third objection dogs all the probabilistic accounts of explanation to be considered in this entry. Suppose that I strap a small but unreliable bomb to one wheel of your car. The probability that the bomb detonates is 50%, in which case your tire goes flat. The trigger fails, but you drive over a nail and your tire does go flat. The bomb has increased the probability of the flat, but it plays no role in its explanation. (Does the presence of the nail screen off the presence of the bomb? No, if it is assumed that the nail's effect is, like

that of the bomb, probabilistic.) Sometimes statistically relevant factors are explanatorily irrelevant.

Finally, it is not easy to see how the SR account might be generalized to give an account of the explanation of phenomena other than events.

**The Unification Account** Friedman (1974) suggests that, while the logical empiricists' "official" account of explanation is the expectability account, they have an unofficial account, too, on which to explain a phenomenon is to see it as an instance of a broad pattern of similar phenomena. Hempel himself occasionally writes in this vein:

The understanding [explanation] conveys lies . . . in the insight that the explanandum fits into, or can be subsumed under, a system of uniformities represented by empirical laws or theoretical principles (Hempel 1965a, 488).

Friedman formulates what he calls a unification account of explanation, a particularly global version of this conception of explanation as pattern subsumption, on which a phenomenon is explained by the system of subsuming laws that best unifies all the phenomena there are. Kitcher (1981, 1989) amends and extends Friedman's account in various ways.

The unifying power of a theory is proportional, on both Friedman's and Kitcher's accounts, not only to the number of phenomena that can be subsumed under the theory, but also to the simplicity of the theory. (Kitcher imposes some additional desiderata.) The theory that best unifies all the phenomena, then, might be said to yield the most for the fewest: the most derivable phenomena for the fewest number of basic principles. It is characteristic of the unificationist position to insist that only the absolutely most unifying theory has full explanatory power, but this view does not in itself preclude the possibility of partial explanation by more weakly unifying theories.

Why be a unificationist? Friedman suggests that the virtue of the most unifying theory is that it reduces to a minimum the number of “fundamental incomprehensibilities”, that is, unexplained explainers. Perhaps a more common justification for unificationism is that suggested by Hempel: to understand something is to fit it into a wider pattern. Add that, the wider the pattern, the more powerful the explanation, and you are well on the way to unificationism.

Many of the virtues of the unification account stem from the great versatility of the pattern subsumption relation. A subsuming pattern need not be exceptionless, so not only probabilistic explanation, but also other forms of non-deductive explanation fit the unification mold. Darwinian explanation, for example, can be seen as accounting for a trait by seeing it as part of a widespread pattern of adaptedness in the biological world—though Kitcher, for one, resists this view of evolutionary explanation, and indeed, argues that all explanations can be formulated as deductive arguments (Kitcher 1989, §5). More inclusively, Kitcher has argued that unificationism supplies an effective account of mathematical, as well as scientific, explanation. For some further claimed advantages of the unification over the causal approach, see Kitcher (1989), §3.

Unificationism promises to give a powerful and subtle account of explanatory relevance. For example, an explanation of a teaspoon of salt’s dissolving that cites the law “all hexed salt dissolves in water” is rejected as insufficiently unifying, because the law is both more complex and covers fewer phenomena than the law “all salt dissolves in water”. More interesting, the unificationist can give an account of why many of the low-level details of the implementation of biological, psychological, economic, and social mechanisms seem to be irrelevant to understanding those mechanisms’ behavior; the details, however, have yet to be worked out (Kitcher 1984).

Two important classes of objections stand in the way of the unification approach to explanation. First is the familiar question concerning the role

of causation in explanation. Can the unification account explain why explanation so often, perhaps always, seems to follow the direction of causation? You might think not: the explanation of a flagpole's height in terms of the length of its shadow seems to cite just as unifying a pattern as the explanation of shadow length in terms of pole height—the very same pattern, in fact.

Kitcher takes up the challenge, arguing that the unification account reproduces the asymmetries in explanation usually put down to something causal (Kitcher 1981). On his view, a unifying pattern is an argument pattern. Since arguments have a direction, the pattern in which the pole height explains the shadow length is distinguished from the pattern in which the length explains the height. The unifying power of each must, therefore, be assessed separately. To solve the problem, the correct comparison is not between the unifying power of these two argument patterns, but between the unifying power of the pattern that wrongly explains pole height in terms of shadow length, and that of the pattern we usually cite to explain the height of a flagpole.

Kitcher calls this latter argument pattern an *origin and development* pattern, and claims that it is instantiated by, and so subsumes, every account we give of the properties of a thing that describes its origin and development, as when, for example, we tell the story of the construction and erection of the flagpole. The pattern is enormously general, then, and so easily wins the right to explain the height of the flagpole. Having argued, in effect, that unificationist explanation tends to proceed in the direction of causation, Kitcher then makes the dramatic claim that it is the order of explanation that determines the order of causation: our causal beliefs depend on and reflect our explanatory practice.

The second objection to explanatory unificationism is that it makes explanation an overly global matter. How one phenomenon is to be explained depends, according to the unificationist, on what best unifies all the other

phenomena, therefore on what the other phenomena are. To many writers, it seems that finding an explanation does not require, even in principle, knowledge extending to all corners of the universe. A more moderate or local unificationism is possible, of course, but another natural place to look for locality is in the causal approach to explanation.

**The Causal Approach** In 1965 Hempel could regard the idea that there is something causal to explanation over and above the exceptionless regularities cited by a DN explanation as lacking a “precise construal” (Hempel 1965a, 353). Since that time, philosophers have come to see claims about causal relations as having a rich empirical content that goes far beyond mere regularities and their instantiation (see Spirtes et al. (2000), though the tradition began well before Hempel made his remark, with Reichenbach (1956)). Even metaphysical empiricists, then, can agree that there is a distinctive causal approach to explanation. Thanks to the development of sophisticated but wholly empiricist accounts of causation (again beginning with Reichenbach), they can go further and in good conscience endorse the causal approach.

Strong arguments suggest that the causal approach is correct. The first and most persuasive is the equation of causal and explanatory direction suggested by the flagpole/shadow and barometer/storm examples. The second is the observation that a requirement of causal relevance between explainers and the explained will deal with the problem of the hexed salt and similar cases. The third is the fact that you can give a cause for a phenomenon without being able to predict it. In those counterexamples to the DN and IS accounts where grounds insufficient for prediction nevertheless seem to be sufficient for explanation—the explanation of paresis by syphilis, and of a trait’s evolution by its conferring a certain benefit—the force of the explanation might well be thought to lie in the aptness of the cited cause. The causal approach is now dominant in the philosophy of explanation.

The most important divide within the causal approach concerns the na-



ture of the causal relation called on to do the explanatory work. Salmon (1984) invokes a notion of causation close to fundamental physics, and declares the explanation of an event to consist of the sum total of causal influences on the explanandum in this fundamental level sense.

Such an account, however, appears to count far too many events as explanatorily relevant. As Salmon concedes, though a baseball causally influences the window that it shatters, and so rightly counts as a part of the explanation of the shattering, so do the shouts of the ball-players, which cause the window to vibrate even as it is struck by the ball. The shouts too, then, will be counted on Salmon's approach as a part of the explanation of the shattering. But they are surely (except perhaps in some very unusual cases) irrelevant.

A popular response to this worry begins with the observation that, while it is correct to say that the ball caused the window to shatter, it is not correct to say that the shouts caused the window to shatter. Such locutions suggest that there is another kind of causal relation, distinct from Salmon's fundamental physical relation, that holds between the ball and the shattering but not between the shouts and the shattering.

How can it be that Salmon's relation holds between the shouts and the shattering but the new causal relation does not? One response is that Salmon's relation is based on a faulty theory of causation, but this is not the answer normally given. Rather, the new causal relation is understood as relating events at all levels, whereas Salmon's causality relates events only at the lowest level.

The high level event of the shattering is the event that would have occurred no matter what the physical details of the shattering, that is, no matter which shards of glass flew where. The low level event is the event individuated by all of the shattering's physical details; this event only occurred, then, because the window shattered in exactly the way that it did. (Some writers call high level events *states of affairs* or *facts*, and hold that events

proper are always low level.)

When we ask for an explanation of the shattering, we are normally asking for an explanation of the fact that the window shattered, not the fact that it shattered in exactly the way it did. Thus we ask for the causes of the high level event, not the low level event. Even though the low- and high-level events are coextensive in space and time, it seems that there are causes of the former that are not among the causes of the latter, namely, the events that determine, given that the shattering occurred, exactly how it occurred. These detail-determining events, because they are not causes of the explanandum, the shattering, do not explain it. (For more on the potential for different causal relations between low and high level events, see Bennett (1988).)

The idea, in short, is that there are many different levels of explananda, corresponding to different levels of eventhood, and different causal relations at all these different levels. Salmon's fundamental physical causation, then, is only one among many different levels of causation. Add this conception of causation as a multilevel relation to the causal approach to explanation, and you get a theory on which the explainers of an event depend on the level of the event. (This level dependence of the explanation is also characteristic of the DN, IS, and SR accounts.)

The best known multilevel theory of causation is the counterfactual account. If the shouting had not happened, the high level shattering event would still have occurred, but because it would have happened in a different way, the low level shattering event would not have occurred. Thus, the high level shattering does not, whereas the low level shattering does, counterfactually depend on the shouting. On a counterfactual approach to causation, this implies that the shouting is a cause of the low level shattering but not the high level shattering, and so, taking this multilevel relation as the explanatory causal relation, that the shouting does not explain the high level shattering, even though—as its causation of the low level shattering shows—it is connected causally to the shattering in Salmon's sense. For this approach

to explanation, but based on a more sophisticated counterfactual account of causation, see Lewis (1986); for a different though related multilevel approach, see Woodward (2003).

An alternative to the multilevel approach is a two-factor approach to causal explanation, on which all explainers of an event must causally influence that event at the fundamental physical level, as prescribed by Salmon, but on which they must pass in addition a further test for explanatory relevance. Salmon himself suggests, late in his career, that the further test might be one of statistical relevance: only the causal influences that change the probability of an event explain the event (Salmon 1997). Strevens (2004) suggests a different two-factor approach.

An advantage of the two-factor approach is the relatively modest demands it makes of the metaphysics of causation, transferring as it does much of the burden of determining explainers to the “further test” for relevance. What, then, to say about claims apparently stating the existence of high level causal relations, such as “The ball’s hitting the window, but not the players’ shouting, caused the window to shatter”? Strevens (2004) suggests that locutions of this form are in fact causal-explanatory claims, asserting the explanatory relevance of certain causal influences (compare Kitcher’s theory of causation mentioned above).

Despite the popularity of the causal approach, it is relatively undeveloped. Little has been written about the causal explanation of laws, for example; it is usually said that they are explained by describing their “underlying mechanisms”, but not every law explicitly concerns causes and effects. Equally, not every event explanation appears to involve the delineation of causes. For examples of both kinds of worry, see Kitcher (1989), §3.

Work on the causal approach to probabilistic event explanation is more advanced. Two main currents can be distinguished in the literature. The first springs from the idea that probabilities themselves have the character of dispositions and are able to cause the events to which they are attached.

The probability of one half that a tossed coin lands heads, for example, is interpreted as a “statistical disposition” that causes the coin (in most cases) to land heads about one half of the time (Fetzer 1981).

The second current flows from the idea that other events or states of affairs can cause events by making a difference to the probabilities of those events. This view is compatible with the dispositional view of probabilistic causality, but it is compatible also with its rejection. Humphreys (1989) writes that “chance is literally nothing” (p. 65), and so certainly cannot cause anything itself, but that events nevertheless cause other events in an indeterministic world by making a difference to their probabilities. Because probability itself is impotent, Humphreys holds that the *kind* of difference a cause makes to the probability of its effect is irrelevant. It does not matter whether the change in probability is positive or negative, large or small (compare with the SR account). Whatever the change, the factor responsible for the change is a cause, and so ought to be cited in an explanation of the effect.

Railton (1978) offers an account of probabilistic explanation that makes room for both conceptions of the relation between probability and causation. On what Railton calls his DNP account, an event is explained by deriving its exact probability from the appropriate initial conditions, background conditions, and laws. Formally, a DNP explanation resembles, as its name suggests, a DN explanation, except that it is the probability of the explanandum, not the explanandum itself, that is deduced. In contrast to Hempel’s account of probabilistic event explanation, the DNP account does not require a high probability for the explanandum, and because it asks for an accurate derivation of the exact probability, it requires, like the SR account, that an explanation cite all factors probabilistically relevant to the explanandum, whether known or unknown, and (though Railton does not give a criterion for relevance) no irrelevant factors. Perhaps most important of all, the DNP account is, unlike Hempel’s various accounts, open to a causal interpretation: the factors that make a difference to the probability, and even

the probability itself, can be considered causes of the explanandum, and the explanation successful precisely because it specifies these causes.

An important lacuna in causal accounts of probabilistic explanation is a detailed treatment of probabilistic explanation in sciences such as statistical mechanics and evolutionary biology, where there is some possibility at least that the underlying processes producing the usual explananda are approximately deterministic. The consensus is to regard such explanations as not genuinely probabilistic; Railton (1981) suggests that they can be reinterpreted as reporting on the robustness of the underlying processes with respect to the event to be explained, that is, the processes' tendency to produce the same kind of outcome given a variety of initial and background conditions.

**Other Issues** Let me conclude with a brief sketch of some issues concerning scientific explanation not mentioned above. First is the question of pragmatics in explanation. Most writers hold that pragmatics affects the explanatory enterprise in only one, relatively minor, way: when an explanation is transmitted from one person to another, the act is subject to the usual pragmatics of communication. This position on pragmatics dovetails with the majority view that the explanatory facts are not essentially communicative: explanations exist independently of anyone's intention to explain anything to anyone else.

Both van Fraassen (1980), chap. 5 and Achinstein (1983) dissent from this majority, holding that there is no explanation without communication, and finding in the pragmatics of communication an account of many facets of explanatory practice. However, this literature has yet to answer the question why science treats explanations as preexisting facts to be discovered, rather than as entities created in the act of communication.

Second, it is an open question whether there is a single standard for evaluating scientific explanations that has remained constant since the beginning of modern science, let alone for the entire history of human explana-

tion. The accounts of explanation in this entry assume, of course, a positive answer, but most work on explanation lacks a substantial historical dimension.

A third issue is idealization in explanation: while almost every account of explanation surveyed above requires that explanations contain no false representations of reality, the practice of using idealized models in scientific explanation is widespread. These models deliberately misrepresent the nature of the systems they describe; the ideal gas model, for example, represents gas molecules as having zero volume, but despite this distortion of the facts, it is considered to explain certain behaviors of real gases. Some writers regard idealization as a temporary or practical measure, out of place in a perfected science. Strevens (2004) suggests that on both the unificationist and a certain causal approach to explanation, idealizations can be seen as serving a genuine and enduring explanatory role.

## References

- Achinstein, P. (1983). *The Nature of Explanation*. Oxford University Press, Oxford.
- Bennett, J. (1988). *Events and Their Names*. Hackett, Indianapolis, IN.
- Coffa, A. (1974). Hempel's ambiguity. *Synthese* 28:141–163.
- Duhem, P. (1954). *The Aim and Structure of Physical Theory*. Translated by P. P. Wiener. Princeton University Press, Princeton, NJ.
- Fetzer, J. (1981). *Scientific Knowledge: Causation, Explanation, and Corroboration*. D. Reidel, Dordrecht.
- van Fraassen, B. (1980). *The Scientific Image*. Oxford University Press, Oxford.
- Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy* 71:5–19.
- Hempel, C. (1965a). Aspects of scientific explanation. In Hempel (1965b), chap. 12, pp. 331–496.
- . (1965b). *Aspects of Scientific Explanation*. Free Press, New York.
- . (1965c). The logic of functional analysis. In Hempel (1965b), chap. 11, pp. 297–330. Revised version of a paper originally published in L. Gross (ed.), *Symposium on Sociological Theory*, Harper & Row, New York, 1959.
- Hempel, C. and P. Oppenheim. (1948). Studies in the logic of explanation. *Philosophy of Science* 15:135–175. Reprinted in Hempel (1965b), chap. 10.
- Humphreys, P. (1989). *The Chances of Explanation*. Princeton University Press, Princeton, NJ.

- Kitcher, P. (1981). Explanatory unification. *Philosophy of Science* 48:507–531.
- . (1984). 1953 and all that: A tale of two sciences. *Philosophical Review* 93:335–373.
- . (1989). Explanatory unification and the causal structure of the world. In P. Kitcher and W. Salmon (eds.), *Scientific Explanation*, volume 13 of *Minnesota Studies in the Philosophy of Science*, pp. 410–505. University of Minnesota Press, Minneapolis.
- Lewis, D. (1986). Causal explanation. In *Philosophical Papers*, volume 2, pp. 214–240. Oxford University Press, Oxford.
- Railton, P. (1978). A deductive-nomological model of probabilistic explanation. *Philosophy of Science* 45:206–226.
- . (1981). Probability, explanation, and information. *Synthese* 48:233–256.
- Reichenbach, H. (1956). *The Direction of Time*. University of California Press, Berkeley, CA.
- Salmon, W. (1970). Statistical explanation. Reprinted in *Statistical Explanation and Statistical Relevance*, pp. 29–87. University of Pittsburgh Press, Pittsburgh.
- . (1984). *Explanation and the Causal Structure of the World*. Princeton University Press, Princeton, NJ.
- . (1990). *Four Decades of Scientific Explanation*. University of Minnesota Press, Minneapolis.
- . (1997). Causality and explanation: A reply to two critiques. *Philosophy of Science* 64:461–477.



- Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science* 30:477–482.
- Spirtes, P., C. Glymour, and R. Scheines. (2000). *Causation, Prediction, and Search*. Second edition. MIT Press, Cambridge, MA.
- Strevens, M. (2004). The causal and unification accounts of explanation unified – causally. *Noûs* 38:154–176.
- Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford.