Scientific Explanation[*]

Bradford Skow

1 Identifying the Subject

The title of this chapter is misleading. "Scientific explanation" is the traditional name for a topic philosophers of science are supposed to have something to say about. But it is a bad name for that topic. For one thing, theories of scientific explanation don't have an activity that is exclusive to scientists as their subject matter. Non-scientists do it too, all the time; they just have less specialized knowledge to use, and direct their attention to less complicated phenomena. This suggests that we drop the adjective "scientific" and call the topic "explanation." But the word "explanation" all by itself is also a bad name for the topic. What philosophers have called theories of explanation are meant to say something about what is happening in scenarios like these:

(1) A physicist explains why the planetary orbits are stable.

(2) A father explains to his older child why his younger child is crying.

There is the word "explains," so what is the problem? The problem is that what philosophers have called theories of explanation are not intended to have anything to say about what goes on when

(3) A biologist explains what the theory of evolution says,

or when

(4) A policeman explains where the train station is.

[*]To appear in *The Oxford Handbook of the Philosophy of Science.*

The difference is clear: (1) and (2) are cases where someone explains *why* something is the case, while in (3) and (4) someone explains *what* or *where*.

So maybe a better name for what philosophers of science are after is a "theory of explaining why." But I have my doubts even here. Explaining why P is a speech act, and when someone successfully performs this speech act they, among other things, convey to their audience the answer to the question why P. But it seems that one can answer the question why P without explaining why P. It seems, for example, that I could tell you why the planetary orbits are stable, without explaining to you why the planetary orbits are stable. So there are two components to explaining why: (i) answering the relevant why-question, and (ii) doing this by performing the speech-act of explaining. How important is the second part to the philosophy of science? Does it matter to the philosophy of science what makes explaining different from other speech-acts?

Some theories of explanation do give the speech-act of explaining an important role to play, but in this survey I will focus just on the first component of explaining why. A theory that addresses that component is a theory of answers to why-questions.[1]

Why-questions are important, in everyday life and in the sciences. The natural curiosity that we are all born with, and that most of us retain throughout our lives, is in part a curiosity about why things happen. And scientists do not just happen to go around answering why-questions; answering why-questions is one of the aims

[1] Van Fraassen (1980) identifies theories of explanation with theories of answers to why-questions. Most other philosophers deny that their theory is meant to cover all and only answers to why-questions. Hempel (1965) tried to distinguish between "explanation-seeking" why-questions and "evidence-seeking" why-questions, and claimed that his theory was only about the first kind of why-question. I doubt that there is any distinction of this sort to be made, but lack space to develop those doubts here. Achinstein (1983) gave the speech-act of explaining an important role in his theory. Other philosophers think a theory of explanation should say something about explaining how, and so should say something about how-questions, not just why-questions: see (Hempel 1965: section 5.2), which cites (Dray 1957); see also (Cross 1991). But I doubt that there is a principle by which we can broaden a theory of explanation to cover how-questions and still exclude who-, what-, when-, and where-questions.

of science. It was a great scientific advance when physicists finally figured out why hydrogen has the spectrum it does. And the fact that they needed to use quantum mechanics to answer this question told in favor of that theory (which at the time was still new).[2]

So what does it take for a body of information to be an answer to the question why P? One very common form answers take is "P because Q." This is not the only form answers to why-questions take; "The chicken crossed the road to get to the other side" is an answer to "Why did the chicken cross the road?" But I lack the space here to go into the relationship between because-answers and to-answers, and anyway a great deal of science aims at because-answers, not to-answers.[3]

Having focused again on a narrower target, the question becomes: what have philosophers of science had to say about what it takes for it to be the case that P because Q? We want to see how the theories fill in the schema

(5) Necessarily, P because Q iff...[4]

Let us start at the beginning.

2 The DN Model

The "fountainhead" of almost all contemporary thinking about explanation is Carl Hempel's Deductive-Nomological (DN) "model" of explanation.[5] Put into the form (5), the theory says

[2]This is controversial; one form of scientific anti-realism says that science aims to systematically describe what happens, not to say anything about why things happen.

[3]Biology and psychology are major exceptions. We often explain people's behavior in terms of their purposes, and at least seem to explain the behavior of "lower" forms of life in terms of the functions that behavior serves. I regret that I cannot address the controversies around these kinds of explanation.

[4]Depending on one's metaphilosophical views, one might want something stronger than a necessarily true biconditional. I won't worry about those distinctions here. I will also suppress the "necessarily" in what follows.

[5]First presented in 1948 a paper joint authored with Paul Oppenheim (Hempel and Oppenheim 1948). The most elaborate development and defense is in (Hempel 1965). Wesley Salmon called it the "fountainhead" (1989: 12).

(6) P because Q iff (i) it is true that Q, (ii) the fact that Q entails the fact that P, (iii) the fact that Q has at least one law of nature as a conjunct, and (iv) the fact that Q would not entail the fact that P if the laws were removed.

A more traditional way to state the DN model is to say that "explanations are arguments" of a certain kind. The conjunction of the argument's premises expresses the fact that Q; the argument's conclusion expresses the fact that P; clauses (i) and (ii) correspond to the claim that the argument is sound; clause (iii) corresponds to the claim that at least one of the argument's premises must be (or express) a law of nature.

There are plenty of examples that Hempel's theory seems to get exactly right. Suppose you drop a rock from 1 meter above the ground, and it hits the ground at a speed of .45 meters per second. Why does it hit with that speed? The following seems like a good way to answer. Use Newton's second law and the law of universal gravitation to deduce the speed at impact from the height of the fall. Here the facts cited in the answer include laws of nature (Newton's), and the fact being explained is deduced from them, just as Hempel's theory requires. And this is not an isolated example. Hang out with the right sort of physicists, or applied mathematicians, and you will spend all day solving differential equations. It looks like the point of doing this is often to answer why-questions. Why is such and such system in such and such state at time T? Solve the differential equation describing its behavior over time; plug in its state at the initial time; you have thereby deduced its state at T from, among other things, a law (the equation). You have also thereby, it certainly seems, explained why it is in that state.

Hempel's DN model was not meant to apply to "statistical explanations." Still, even in its intended domain it was shown to be false in the 60s. Here are two of the many well-known counterexamples:

- **The ink bottle.** Jones knocks an ink bottle with his knee. It topples over and spills ink all over the rug. Then the rug is stained because Jones knocked the ink bottle. But "Jones knocked the ink bottle" does not contain a law of nature.[6]

[6] This example is originally due to Scriven (1959: 456). See also (Scriven 1962).

- **The flagpole.** A flagpole is 10 meters high, and it casts a shadow 10 meters long. The angle the line passing through the sun, the top of the flagpole, and the top of the shadow makes with the earth is 45 degrees. The the fact that the shadow is 10 meters long, and that the sun is in that position, together with the laws of geometric optics, entails that the flagpole is 10 meters high. But it is false that the flagpole is 10 meters high because this fact obtains. It is not 10 meters high because it casts a 10 meter long shadow.[7]

A diagnosis of where Hempel went wrong is not hard to find. In **The ink bottle** we find out why the rug is stained by learning what caused the stain, even though no laws are mentioned.[8] In **The flagpole** we see the height of the flagpole deduced from a law without any mention of any the causes of its having that height, but do not learn why the flagpole is that high. The obvious moral is that an answer to a why-question needs to say something about causes, but does not need to mention any laws.

3 Causal Theories of Explanation

The moral needs to be qualified and refined if it is to have any chance of being true.

First, we are clearly here focusing only on why-questions that ask why some given event occurred. Only events can have causes, so it only makes sense to impose a causal requirement on answers to why-questions about events. This limits the scope of the theory; it makes sense to ask why Galileo's law of free fall is true, or why every polynomial over the complex numbers has a root, even though these are not questions about events.

Second, "an answer to the question why E occurred must say something about the causes of E" is too vague. A philosophical theory needs to say something much more precise and detailed.

---

[7]This example is originally due to Bromberger. He never published it, though he did discuss similar examples in published work (see Salmon 1989: 72).

[8]In response to Scriven's example, Hempel held that laws were still in some sense implicit in the explanation (Hempel 1965). Woodward dismantles this reply in chapter 4 of his (2003).

There have been several attempts to do so. A naive causal theory of explanation might just say that to answer the question why E occurred it is necessary and sufficient to describe at least one cause of E. Philosophers have articulated more detailed, and nuanced, causal theories of explanation. Wesley Salmon's theory (1984) and David Lewis's theory (1986) are two prominent examples. Lewis's can be stated with fewer technical notions, so let us take a brief look at it.[9]

It is actually not easy to put Lewis's theory in the form (5), so let me put it in the terms he preferred. First let us distinguish between a complete answer to a question, and a partial answer. This distinction should be familiar. Suppose I ask you who is taking your class. You have twenty students, but you just say that Jones is enrolled. You have given me a partial answer. If you had said the names of all your students, and told me that those were all the students, you would have given me a complete answer.

Lewis's theory has two parts; one says what it takes for something to be a complete answer to a why-question, the other, what it takes to be a partial answer:

(7) A proposition P is the complete answer to the question why event E occurred iff P is a maximally specific proposition about E's causal history. (It says what E's causes are, what the causes of its causes are, and so on.)

(8) A proposition P is a partial answer to the question why event E occurred iff P is about E's causal history (it need not be maximally specific).

A window breaks and we want to know why. On Lewis's view, "Because someone threw a rock at it" is only a partial answer. For while the proposition that someone threw a rock is about the causal history of the breaking, it is not maximally specific. It does not tell us who threw the rock, or about any of the other causes of the breaking (for example, the causes of the throw). "Because Adam threw a rock at it" is also a partial answer, but is more complete. It pins down more facts about the causal history than the first answer does.

---

[9]There are also a lot more objections to Salmon's than to Lewis's, in part because Salmon's is tied to one particular (and particularly implausible) theory of causation. See (Hitchcock 1995) for some persuasive objections to Salmon's theory. (David Lewis of course also had a theory of causation — a few in fact — but one need not accept his views about causation to accept his theory of explanation.)

Obviously if (7) is right then complete answers to why-questions are impossible to write down. None of us has ever heard one. We trade only in partial answers.

Is some causal theory of explanation — Lewis's, Salmon's, or some other — correct? A lot of alleged counterexamples turn on controversial claims about causation, and apply only to some causal theories of explanation. A survey of all of them is too big a project for this article; here is just one example. I hit a vase, and it cracks. It cracked, in part, because it was fragile. But fragility is a disposition, and some philosophers hold that dispositions cannot cause anything. If that is right then the naive causal theory of explanation is in trouble. But this kind of example may not refute the naive theory; there are theories of causation on which dispositions can be causes.[10] And anyway, even if the naive theory is in trouble, Lewis's theory, for example, may not be. One could make the case that the fact that the vase is fragile tells us something about the causal history of the breaking. (Certainly the set of "possible causal histories" of a non-fragile vase is different from the set of possible causal histories of a fragile vase. It "takes more" to break a non-fragile vase.)

Lewis's theory is relatively undemanding; the class of facts that are about the causal history of some event E will typically be quite large, much larger than the class of facts that describe some particular cause of E. Still, some explanations appear to fail to meet even Lewis's undemanding criteria. Elliott Sober (1983) drew attention to "equilibrium explanations." R. A. Fisher (1931) explained why the current sex ratio among humans is one-to-one. In broad outline his answer was that if the sex ratio had ever departed from one-to-one, individuals who overproduced the minority sex would have higher fitness, and so the sex ratio would move back toward one-to-one. Elliott Sober argued that equilibrium explanations are non-causal. Here is Sober's summary:

> a causal explanation...would presumably describe some earlier state of the population and the evolutionary forces that moved the population to its present [sex ratio] configuration...Where causal explanation shows how the event to be explained was in fact produced, equilibrium expla-

[10] On the question of whether the fragility explanation is causal see Lange (2013: 494), who gets the example from Jackson and Pettit (1992). For a discussion of whether dispositions can be causes see (Fara and Choi 2012).

> nation shows how the event would have occurred regardless of which of a variety of causal scenarios actually transpired. (202)

If Sober had had Lewis's theory of explanation in mind, he might have continued: someone X could know every detail of every cause of the current sex ratio without knowing anything about what the sex ratio would have been if those causes had not occurred. And, the argument goes, if X were to learn more about what the sex ratio would have been, if those causes had not occurred, then after learning that information X would know more about why that ratio is 1:1. If all this is right then it looks like Lewis's theory is false.[11]

Another important class of examples worth mentioning are examples of "mathematical explanations of physical phenomena." Why are the periods of the life-cycles of cicadas prime numbers? Members of a cicada species emerges once every N years to mate and then die. If a predator species also has a periodic life-cycle, and its life-cycle in years is some number that divides N, then predators of those species will be around whenever the cicadas emerge. Predation will then reduce the size of the next cicada generation. It is therefore advantageous for the cicada to have a life-cycle period that, when measured in years, has as few divisors as possible. Periods that, when measured in years, are prime numbers, are therefore the best candidates. It looks here like the fact that prime numbers have the smallest number of divisors helps answer a why-question. Some doubt that this fact could count as causal-explanatory information.[12]

I think the jury is still out on whether a causal theory of explanation is defen-

[11]Few now accept that Fisher's explanation is a "non-causal explanation." Strevens, for example, argues in detail that equilibrium explanations count as causal on his theory (2008: 267-72).

There is a gap in the argument I presented that, to my knowledge, others have not exploited. From (i) X learned P, and (ii) after learning P (and nothing else) X knew more about why Q, it does not follow that P is part of the answer to the question why Q. It could just be that learning P put X in a position to see that something else she already knew is part of the answer to the question why Q.

[12]This example was introduced into the literature by Baker (2005). Lange (2013) discusses a host of examples that he calls "distinctively mathematical," and defends the claim that they are non-causal explanations.

sible.[13] But if no such theory is true, what is the alternative? Some opponents of causal theories see some other common factor in the alleged counterexamples we have surveyed. I'll say something about one attempt to isolate that factor — unificationist approaches to explanation — below. Another conclusion would be to think that the class of answers to why-questions exhibits a great deal of diversity and little unity. If that's right then the job of the philosopher reduces to that of cataloguing the different varieties of answer and putting them on display, as one might produce an illustrated list of the different species of parrot that inhabit some tropical island.

4 Probabilities in Explanation

Alongside his DN model of explanation, Hempel developed what he called the "inductive-statistical" model of explanation. What kinds of answers to why-questions is this model meant to apply to?

Suppose Jones has strep throat, and his doctor gives him penicillin. The penicillin works, and Jones recovers within a week. Why did he recover? It certainly looks like the fact that he took (an appropriate dose of) penicillin should be part of the answer. But taking penicillin does not guarantee recovery from strep in a week; only, say, 90% of those who take it recover that fast.[14] So this cannot be a DN explanation citing the "law" that everyone who suffers from strep and takes penicillin recovers in a week. For there is no such law. Hempel's IS model is meant to apply here instead:

- P because Q iff (i) it is true that Q, (ii) the probability of P given Q is high, and (iii) the fact that Q has at least one law as a conjunct.

The IS model, like the DN model, requires there to be a law, but relaxes the requirement that the fact that Q entail the fact that P, substituting a high probability requirement. If we assume Hempel's "explanations are arguments" point of view, then IS explanations are good inductive arguments with law-statements among their premises, where a good inductive argument is one in which the conclusion has high

[13] I defended a causal theory similar to Lewis's in (Skow 2014), and propose and defend a new causal theory in (Skow MS-a).

[14] I made this number up.

probability given the premises. In the penicillin example, Hempel was willing to count as a law the fact that the probability of recovery from strep in one week when given penicillin is 90%.

The IS model fails. Even if taking penicillin raised the probability of recovery within a week from, say, 10% to 20%, it would still be true that Jones recovered in a week because he took penicillin. In the wake of the IS model's failure several other theories have tried to state just what facts about probabilities must, or may, appear in an explanation. Wesley Salmon's "Statistical Relevance" theory of explanation was an important successor (Salmon et al 1971). It was meant to apply to why-questions of the form "Why is this x, which is A, also B?" ("Why is Jones, who had strep, a person who recovered from strep in one week?") Roughly speaking, Salmon required an answer to this question to (i) cite a factor C that is "statistically relevant" to being B — the probability that x is B, given that x is A and C, must be different from the probability that x is B, given that it is A; and also (ii) state just what these probabilities are.[15]

Is there any need to have a separate theory of statistical explanations? The naive causal theory of explanation does not need to be amended or augmented to deal with the penicillin example. True, penicillin doesn't guarantee recovery, it just makes recovery more likely. Nevertheless, Jones's taking penicillin still caused his recovery, and so the naive theory will say that the proposition that Jones took penicillin answers the question why he recovered. No statistics need to appear in the answer.

One might reply that if (i) one answers the question why some event E occurred by citing a cause of E, and (ii) there is an interesting statistical relationship between E-type events and C-type events,[16] then one's answer could be improved by citing that statistical relationship. The suggestion is that "Because he took penicillin, and penicillin raises the chance of recovery in a week to 90%" is a better answer to the question why Jones recovered than "Because he took penicillin" is.

[15] I have omitted many details of the theory.

[16] That Suzy threw a rock at the window did not guarantee that the window would break. But there is no interesting statistical relationship between rock-throwings and window-breakings.

If this is right then causal theories of explanation do need to say something special about statistical explanations. But one might doubt that the longer explanation is a better answer. Does the fact that penicillin raised the chance of recovery really add anything to the fact that the penicillin caused the recovery? The statistics might be helpful if we do not yet know whether penicillin caused the recovery; but once we do know this, the facts about the statistics can seem irrelevant. One might in fact hold that, if some causal theory of explanation is true, then no true answer to a why-question about an events mentions probabilities or chances.[17]

One complication here is the possibility that causation can be analyzed in terms of probabilities. Perhaps for C to be a cause of E is for the probability of E given C to be related in the right way to other probabilities. For example, one might hold that for C to be a cause of E is for the probability of E given C to be higher than the unconditional probability of E. If this view is right then maybe the fact that C is a cause of E is the very same fact as some fact about probabilities. Then one cannot say that an answer to a why-question about an event should cite facts about causes but not facts about probabilities.

In fact, if to be a cause of E is to raise the probability of E, then the rough version of Salmon's Statistical Relevance theory I have been working with is (almost) equivalent to the naive causal theory of explanation.[18] It seems to me that insofar as the Statistical Relevance theory (in its more complete version) gets things right, it is because it is using statistical relationships as a stand-in, or surrogate, for causal relationships. The idea that causation can be analyzed in terms of probabilities, however, faces a lot of challenges.[19]

There is a need to have a separate theory of statistical explanations if there are

[17]Compare (Humphreys 1989: 109) and (Kitcher 1989: 422). The restriction to why-questions about events is important. Even if answers to why-questions about events should not cite statistics, it might still be that answers to other kinds of why-questions should.

[18]Salmon allows conditions that lower the probability of E to explain E, even though they do not count as causes on the current proposal.

[19]See (Hitchcock 2010) for an overview. (Cartwright 1979) and chapter 4 of (Spirtes, Glymour, and Scheines 2001) contain arguments that statistical relationships between events underdetermine causal relationships.

"non-causal" statistical explanations. On some interpretations, quantum mechanics says that radioactive decay is a genuinely indeterministic process. When an atom decays by, say, emitting a beta particle, the laws of quantum mechanics, plus the exact condition of the atom and its environment just before the decay, did not determine that it would decay. Now suppose we ask why the atom decayed. Peter Railton held that we can answer this question by deducing the probability of decay from the laws of quantum mechanics and the prior condition of the atom and its environment (Railon 1978). This probability doesn't seem to have anything to do with the causes of the decay (it is not clear that the decay has any causes). So if Railton is right that facts about the probability help explain why the atom decayed, then we have a "non-causal" role for probabilities in explanations.[20] Probabilities also play an important role in the answers to why-questions that statistical mechanics offer. Some have held that the information about probabilities in these explanations is also not causal-explanatory information.[21]

5 Theories of Causal Explanation

Coming up with a theory that applies generally to all why-questions, a filling-in of (5)

(5) P because Q iff...

that covers all cases, is a daunting project. We have already seen some philosophers narrow their sights and just focus on questions that ask why some event occurred. Some philosophers narrow their sights even more. They say that they aim to state and defend a theory of causal explanation of events. Instead of saying what it takes for something to be an explanation "full stop," or even an explanation of an event, the project is that of saying what it takes for something to be a *causal* explanation

[20] This is a big "if"; Kitcher, for example, denies that quantum mechanics explains why the atom decayed (Kitcher 1989: 451). I should say that Railton himself thought explanations like this were causal.

[21] Chapter 10 of (Strevens 2008) attempts to locate a non-causal explanatory role for probabilities. Statistical mechanics is a popular hunting-ground for those looking for non-causal explanations. For some examples see (Batterman 1992) and (Reutlinger forthcoming).

of an event. James Woodward's book *Making Things Happen* (2003) and Michael Strevens' book *Depth* (2008) are two important recent works of this kind.

Each book builds its theory of causal explanation on a theory of causation. Woodward starts with a counterfactual theory of causation. His basic insight is that causation is a kind of conditional counterfactual dependence: C is a cause of E if and only if, holding fixed the occurrence of certain facts, had C not occurred, E would not have occurred. Woodward has a lot to say about just how this insight is to be developed and understood: a lot to say about what facts may be held fixed, and a lot to say about how to understand the way he is using counterfactual "had C not occurred, E would not have occurred." Since this is an article on explanation, not on causation, I must pass over these details.[22]

What does Woodward say it takes to causally explain why some event occurred? In one place he summarizes his theory like this: "[causal] explanation is a matter of exhibiting systematic patterns of counterfactual dependence" (191). So on Woodward's view citing a cause of E is a way to causally explain why E occurred because, he claims, citing a cause of E conveys information about what E counterfactually depends on. But if Woodward is right, then even after I have told you that C caused E, I can give you still more causal-explanatory information, *even if I do not tell you anything about E's other (token) causes*. For I might tell you more about the way in which E depends on C. Suppose that there is a pot of water on my stove, over a lit burner, and that the water is at a steady 110 degrees Farenheit. Suppose that it is the knob's being turned half-way that caused the water to be at that temperature. Then I can give some causal-explanatory information about the temperature of the water by citing this cause, and saying that the knob is turned

[22]Woodward's theory of causation makes use of "structural equations" for modeling causal facts. Two important technical works in this tradition are (Pearl 2009) and (Spirtes, Glymour, and Scheines 2001). I have described Woodward's theory of token causation between events; but more central to his theory is his theory of type-level causation between variables (a variable is a "generalized event"; events occur, or fail to occur, in each possible world, while variables take on one of many possible values in each possible world). Yablo (2002) is an important defender of a conditional dependence theory of token causation that does not rely on structural equations.

half-way. On Woodward's way of thinking about causation I have, in effect, told you that there is *some* other position of the knob P such that, if the knob had been at P (instead of at the half-way mark), then the water would have been at a different temperature. Obviously I could provide you with more of this kind of information. I could be more specific about the way the temperature depends on the knob's position. For example I could say that if the knob were turned all the way, the water would be boiling. Woodward holds that this information is also causal-explanatory information about the water's temperature. In general, information about how the event being explained is situated in a range of counterfactual alternatives is causal-explanatory information. Bare claims about the token causes of the event are one way to provide this kind of information, but not the only way.

Let me now turn to Strevens' theory. Strevens thinks that there are two kinds of causation. On the one hand, there is the low-level, microphysical production of precisely-specified events, like that event consisting in *these* atoms being in exactly *these* positions at this time. On the other hand, there is high-level causation between imprecisely-specified events, as when Suzy's throw causes the window to break. These events are imprecise because there are a huge number of microphysical ways Suzy could have thrown, and a huge number of ways in which the window could break.

Strevens analyzes high-level causation in terms of low-level causation; he says relatively little about what low-level causation is. Here, in brief outline, is how the analysis goes. He agrees with the common thought that the causes of an event "make a difference" to whether that event occurred. But he does not understand difference-making in counterfactual terms. Instead he starts with the notion of a set of facts "causally entailing" another fact. Set S causally entails F if and only if, first, S entails F, and second, this logical relation "mirrors" the low-level, microphysical production of F (or of the "maximal precisification of F") by the members of S. Suppose we have such an S and F. Strevens then subjects the set S to a process of "abstraction." Roughly speaking, this process weakens the members of S as much as possible, consistent with S still causally entailing F. So we might start with the fact that I threw a rock with a mass of 50 grams at that window, the fact that it had an initial velocity of 10 m/s, the fact that there was nothing around to interfere with the rock's flight, the fact that the window was made of normal glass and was 5mm thick, and the fact that certain laws of mechanics are true. This set causally entails that the window breaks. But none of the facts mentioned are difference-makers; they are "too specific." After Strevens' process of abstraction does its work on this set, the fact that the rock had a mass of 50 grams and was thrown at 10 m/s will have been replaced by the (weaker) fact that it was thrown with an initial momentum in a certain range. (It will be the range containing exactly those initial momenta that causally entail that a window of that sort will break under these conditions; the momentum .5 kg m/s will fall in this range.) *This* fact, about the initial momentum, is *not* "too specific" to be a difference-maker; it is just specific enough, and so counts as a cause of the breaking.[23]

So we start with a set of facts S that causally entails the occurrence of some event E; Strevens' abstract procedure returns a set of weaker facts S* that (i) also causally entails that E occurs, but now also (ii) contains facts that correspond to causes of E.[24] What then is Strevens' theory of causal explanation? He says that the deduction of the occurrence of E from S* is a "standalone" causal explanation of the occurrence of E. Other kinds of causal explanations are to be understood in terms of standalone explanations; they are the basic building blocks of his theory.

6 Unificationist Theories of Explanation

The original goal was to say what it takes for it to be the case that P because Q. Causal theories of explanation lower their sights, and focus only on explanations of events. Theories of causal explanation lower their sights further, and focus only on one particular kind of explanation of events. Unificationist theories of explanation

[23]Stephen Yablo was a pioneer of the idea that the causes of E are the events that are (i) "causally sufficient" for the occurrence of E that are also (ii) "not too specific" (Yablo 1992). The idea that the causes of E are facts that play an essential role in entailing the occurrence of E was a key component of J. L. Mackie's theory of causation (Mackie 1974). Officially Strevens' process is a process of "optimization," which has a process of abstraction as a part. I do not have space to go into the other components of the process.

[24]For example, the fact that this window broke at T corresponds to the event that consists in the breaking of the window.

aim back up at the original target. Their slogan is: to explain why is to unify.

So what is unification, and how might the notion of unification figure in a theory of answers to why-questions? Michael Friedman and Philip Kitcher are the founding fathers of the unificationist approach to explanation (see Friedman 1974, Kitcher 1981, and Kitcher 1989). For reasons of space I will just discuss Kitcher's theory. Central to Kitcher's theory is the notion of a set of argument-forms unifying a set of propositions. He illustrates this notion with a example:

> The unifying power of Newton's work consisted in its demonstration that one pattern of argument could be used again and again in the derivation of a wide range of accepted sentences. (1981: 514)

You want to prove that apples take so-and-so many seconds to fall from a tree of such-and-such a height? Newton showed that you could prove this by setting up a differential equation governing the apple's motion, using his three laws and the law of universal gravitation, and then solving that equation. You want to prove some facts about the moon's orbit? You can prove them using the same method. You want to prove that the tide is higher when the moon is overhead? Same form of argument again. In Kitcher's view, that's what it was for Newton's theory to unify all of these phenomena.

We need a more precise statement of Kitcher's notion of unification. The things to which Kitcher's notion of unification apply are sets of argument-forms. Different sets of argument-forms might unify a body of fact to different degrees. Explanation is going to go with being a "best" unifier, so we need to know: given a set of propositions K, what is it for one set of argument-forms to unify K better than another? Only sets of argument-forms that "generate K from one of its subsets" are even in the running. (These are sets of argument-forms for which there is some subset P of K — the "premise-set" — such that, for each proposition X in K, there is a valid argument for X from premises in P that instantiates one of the forms.) Among these, Kitcher says, sets of argument-forms that use smaller premise-sets unify better — other things being equal; smaller sets of argument-forms unify better — other things being equal; sets of argument-forms each of whose members have a relatively small set of instances unify better — other things being equal.

These criteria pull against each other, and it is not obvious how to arrive at an all-things-considered verdict. But let us pass over these details. Kitcher's theory of explanation makes use of this notion of unification in the following way:

- An argument is an answer to the question why P[25] iff it is a sound argument with P as its conclusion, and it instantiates one of the argument-forms that belongs to the set of argument-forms that best unifies the set of all "Humean" truths.[26] (To save writing, let us say that an argument-form that appears in the set of argument-forms that best unifies is a "good" argument-form.)

Kitcher takes from Hempel the idea that explanation is deduction. But for Hempel explanation is local: if you can deduce P from some laws then you have explained why P is true, no matter what the answers to other why-questions look like. For Kitcher explanation is global, or holistic: roughly speaking, a deduction of P from some laws gets to be an answer to the question why P only if arguments of the same form can be used to deduce many other truths.[27] Kitcher thinks the global nature of his theory is a virtue. The standard counterexamples to the DN model, he claims, involve a deduction of some proposition from some laws where the deduction does not instantiate one of the good argument-forms. We can deduce the height of the flagpole from the length of its shadow and the laws of optics. But, he argues, the set of argument-forms that best unifies the set of all truths will not include an argument-form for deducing phenomena from their effects; for the set of argument-forms that best unifies the phenomena will certainly need an argument-forms for deducing phenomena from their causes, and there do not seem to be any phenomena that can be deduced from the effects but not their causes.

[25] Kitcher's theory is slightly easier to digest if put in this form, rather than form (5).

[26] Among the non-Humean truths are those of the form "P because Q." See section 8.3 of (Kitcher 1989) for a more complete characterization of the Humean truths, as well as Kitcher's reasons for excluding them. It is worth noting that Kitcher's restriction means that on his view questions of the form "Why is it the case that P because Q?" do not have answers. This is a flaw, since "higher-order" explanations do exist. My statement of Kitcher's view is simplified in a few ways that will not matter for my discussion.

[27] For the reason why this is rough see the "natural worry" below.

Some have worried that Kitcher's theory, even if it says the right thing about the flagpole example, says it for the wrong reasons. If flagpoles were in the habit of springing up uncaused, then maybe we could unify the phenomena better if we had an argument-form for deducing events from their effects. Even in a possible world like that, though, it does not seem that the flagpole is 10 meters high because it casts a shadow 10 meters long.[28]

Kitcher had an unusual view about causation. He held that *what it is* for C to be a cause of E is for C to appear in the correct answer to the question why E occurred (for a more precise statement of his view see (Kitcher 1989: 497); see also (Kitcher 1986: 229)). So Kitcher would say that the world we are imagining, in which the length of a flagpole's shadow explains why the flagpole is as tall as it is, is thereby also a world in which the length of the shadow is a cause of the flagpole's height. (For what it's worth, I do not find this defense plausible.[29])

Here's another natural worry about Kitcher's theory: doesn't it entail that the world is necessary unified? That would be bad, because surely it is possible for the world not to be unified. But to become an objection this worry needs more development. For Kitcher's theory appears to allow for possible worlds in which the set of argument-forms that best unifies the Humean truths contains a huge number of argument-forms, each of which is instantiated only a few times. That is, it allows for worlds in which the set of argument-forms that best unifies the Humean truths unifies them only to a tiny degree.[30]

The specific details of Kitcher's theory aside, there is a general feature of uni-

[28]For objections to Kitcher like this see (Woodward 2003: 361) and (Paul and Hall 2013: 91-2).

[29]Generically, empiricists hold that both causation and explanation must be analyzed in terms of more epistemologically unproblematic notions. Kitcher's analyses of explanation in terms of unification, and of causation in terms of explanation, are empiricist in spirit. Perhaps a committed empiricist will find his view more plausible than I do. Kitcher's view is not, however, the only empiricist option; there are empiricist theories of causation that do not analyze it in terms of explanation.

[30]Section 8.2 of (Kitcher 1989) is headed by the question "What If the World Isn't Unified?" but it ends up addressing instead the objection that, on his view, a cause of some event E might not explain E. Kitcher presents the views about causation that I just mentioned in response to this objection.

ficationist theories that bothers me. Unification, it seems to me, should be analyzed in terms of explanation, not the other way around. It may be that there are many different kinds of unification,[31] but still, one central way for a theory to unify, I think, is for it to provide answers to many different why-questions. This conception of unification cannot appear in a theory of answers to why-questions, for its appearance there would make the theory circular.

If there is an objection to Kitcher's theory here, though, it is a frustratingly subtle one. Kitcher said that Newton's theory unified a lot of phenomena by allowing for deductions of all those phenomena, deductions that all instantiated the same argument-form. I want to say that I disagree; what it was for Newton's theory to unify all of those phenomena was, instead, for his theory to figure in the answer to the question why X occurred, for each of those phenomena X. But just what do Kitcher and I disagree about here? Kitcher will say that, since his theory of explanation is correct, and since the argument-form Newton's theory uses is a good one, the deductions instantiating it *do* answer the relevant why-questions about those phenomena. So my objection here is not that Kitcher cannot say that Newton's theory answers a lot of why-questions. Nor is my objection that Kitcher fails to provide correct necessary and sufficient conditions for unification.[32] Instead, my objection is to his "order of analysis." Kitcher says that Newton's theory unified *in virtue of the fact that* it made it possible to deduce many phenomena using a good argument form. I say that even if Newton's theory did make this possible, the theory did not unify those phenomena in virtue of making this possible. It unified in virtue of the fact that it answered many why-questions. Kitcher may be right about which theories unify, but I think he is wrong about what unification is. And again, if I am right about what unification is, then the notion of unification cannot appear in the correct theory of answers to why-questions.

[31] That there are many kinds of unification is a theme of (Morrison 2000).

[32] That is, my objection *here* is not that Kitcher fails to provide correct necessary and sufficient conditions. I do happen to think that his theory fails to do this. (For I think that his theory of explanation is wrong, and I think that his conditions on unification are correct only if his theory of explanation is right.) But I do not have space here to develop this objection in a way that does not make it look like it begs the question.

7 Radical Context Sensitivity?

In *The Scientific Hypothesis* Bas Van Fraassen defended a view about explanation that entails that there can be no interesting filling-in of schema (5). It is worth discussing here because, if true, it shows that every theory I have mentioned goes wrong right from the beginning.

Van Fraassen's theory makes use of the notion of context-sensitivity. The pronoun "I" is a paradigm case of a context-sensitive word, so it can serve as an illustration. There is no one fixed thing that the word "I" refers to. When I use it, it refers to Bradford Skow; when my oldest brother uses it, it refers to Erik Skow; you get the idea. Clearly the rule is: on any occasion when "I" is spoken, it refers to the person speaking.[33] Now the fact that a certain person is speaking is fact about the context in which "I" is being used. The context also includes facts about who is listening, where and when the conversation takes place, and also the beliefs and presuppositions of the parties to the conversation.[34] So, more generally, what "I" refers to depends on the context. That is what it is to say that "I" is context-sensitive. "Zebra," by contrast, is not context-sensitive. No matter who is speaking, or where, or what the parties to the conversation believe, "zebra" applies to all and only the zebras.

Another example of context-sensitivity, one that may be more helpful for what is to follow, is the word "tall." Suppose Jones is six feet tall. Then in a context in which we have been talking about the heights of various philosophers, "Jones is tall" is true. Jones has the property that "tall" expresses in that context. But in a context in which we have been discussing basketball players, "tall" expresses a different property, and "Jones is tall" is false.[35]

Van Fraassen claimed that why-questions are context-sensitive. To explain his view it will help to temporarily make a distinction. There is a familiar distinction between a declarative sentence and a proposition. Propositions are the "semantic

[33] Ignoring exceptional cases, as when I am reading out loud from your diary.

[34] There is a lot of debate in the philosophy of language about just what contexts are. See (Stalnaker 2014).

[35] This is an over-simplified model of the way in which "tall" is context-sensitive, but the simplifications do not matter here.

values" of declarative sentences: a declarative sentence, like "Jones is human," expresses a proposition. But there are other kinds of sentences besides declarative ones. Our interest is in interrogative sentences. Sometimes "question" is used as a synonym for "interrogative sentence," but it helps to distinguish between them. Let us say that questions are the semantic values of interrogative sentences. Then questions stand to interrogatives as propositions stand to declaratives.

Van Fraassen's thesis, then, is that why-interrogatives are context-sensitive. They express different why-questions in different contexts.

He claims that they are context-sensitive in two ways. First, he holds that why-questions are contrastive. That is, an interrogative of the form

- Why is it the case that P?

in a given context, expresses the question why it is the case that P rather than A1, or A2,..., or An. What goes in for "A1," "A2," and so on depends on the context. The proposition that A1, the proposition that A2, and so on, are alternatives to the proposition that P: each is incompatible with P and with all the others.

If this is right, if why-interrogatives are context-sensitive in this way, then which sentence expresses a true answer to a why-interrogative must be context-sensitive in a similar way. And there is good evidence that both of these claims are right. Van Fraassen asks us to think about Adam. I might ask a question using the sentence, "Why did Adam eat the apple?" And I might ask this in a context in which the salient alternative course of action was for Adam to eat nothing. Then, it seems, in this context I asked the question why Adam ate the apple rather than nothing at all. In another context the most salient alternative might have been for Adam to eat the pear instead. Then a use of "Why did Adam eat the apple?" serves to ask the question why Adam ate the apple rather than the pear. In the first context, the answer to "Why did Adam eat the apple?" is "Because he was hungry." This answer is not correct, however, in the second context. For in the second context "Because he was hungry" expresses the proposition that Adam ate the apple rather than the pear because he was hungry. And that is false. (The apple wouldn't do any more to alleviate Adam's hunger than the pear, and Adam knew this.)

This kind of context-sensitivity in "because" statements is no real threat to the project of finding an interesting way to complete (5). We just need to make a small change to the project. We just need to make the "hidden contrasts" explicit in our schema. The goal becomes to fill in the right-hand side of

(9) P rather than A1, or A2, or ... or An, because Q iff ...

Many philosophers have thought that causation is also a contrastive relation: a fully explicit claim about causation looks something like: "C rather than C* caused E rather than E*." Adam's being hungry, rather than satiated, caused him to eat the apple, rather than nothing at all. If this is the right way to think about causation, then those who want to develop a causal theory of explanation will not be bothered by the move from (5) to (9).[36]

But Van Fraassen thought that the context-sensitivity of why-interrogatives, and therefore of because-sentences, went beyond the role of context in supplying alternatives to the "P" in "why P?". He held that context also supplied a "relation of relevance." In brief, and ignoring from now on the fact that why-interrogatives and because-statements are contrastive, he thought the following about the semantics of because-sentences:

(10) "P because Q" is true in context C iff (i) it is true that P, and it is true that Q, and (ii) the fact that Q is "relevant-in-C" to the fact that P.

We can think of claim (10) this way: just as we all think that "tall" expresses different properties in different contexts, Van Fraassen thought that "because" expresses different relations (between facts) in different contexts.

How radical a claim (10) is depends on how many "candidate relevance relations" you think there are. One could accept (10) and still make the modest claim that there are only, say, two candidate relevance relations. In any context, one or the other of them is the semantic value of "because." But Van Fraassen was a radical. He appeared to believe

[36]For arguments in favor of taking causation to be contrastive see (Hitchcock 1996a) and (Schaffer 2005). On contrasts in explanation see, for example, (Hitchcock 1996b) and (Lipton 2004). The causal claim I wrote down is "contrastive on both sides" while in schema (9) there are only contrasts on the left-hand side. Die-hard contrastivists will say that (9) should be contrastive on both sides. I don't have space to go into these details here.

(11) For any two propositions there is a candidate relevance relation that the first bears to the second.

If (11) is true then there is not much to say about the nature of explanation beyond (10). Those who think that only causes explain were right — about one relevance relation; those who think that F can explain G even if F does not cause G are also right — about a different relevance relation.

But (11) is hard to believe. Consider "The flagpole is 10 meters high because it casts a shadow 10 meters long," or "The economy is weak because Saturn is in retrograde motion." These sound false. But if (11) is right, then there is a conversational context we can get into in which these sentences express truths. (Importantly, this is so even if we restrict our attention to contexts in which we still believe that the flagpole's shadow does not cause its height, and in which Saturn's motion does not have any causal influence on the economy.) In his book Van Fraassen tried to produce a context in which "The flagpole is 10 meters high because it casts a shadow 10 meters long" expresses a truth, but he did not convince many people.[37] I suspect that the most that Van Fraassen establishes is a modest form of context-sensitivity, one that is no threat to traditional theories of explanation.

It is worth mentioning that one can believe that there are many explanatory relevance relations, and that context determines which one is operative for a given use of "because," while still having a unified theory of explanation. One could say that every relevance relation is a restriction of some one single ur-relevance relation. A causal theory of explanation, for example, could say that the ur-relevance relation is causation. Some contexts select a relevance relation that only some of E's causes bear to E. Different causes will be relevant in different contexts. This kind of context-sensitivity is not radical and is no threat to the project of completing (5).

8 Other Topics

There are many topics I have not been able to cover, so I will end by just throwing out some references about three of them. First: we explain not just physical

[37] The most influential response to Van Fraassen's views on explanation is (Kitcher and Salmon 1987).

phenomena but also mathematical facts. Since mathematical facts cannot be caused it is unclear whether the work that has been done on causal explanation sheds any light on explanation in mathematics. (Steiner 1978) can be read as an attempt to find a surrogate for causal explanation in mathematics. A recent, and much deeper, investigation of explanation in mathematics is (Lange 2014).

Second: a distinctively metaphysical kind of explanation — "grounding," or "in virtue of" explanation — is currently the focus of a lot of attention in metaphysics. Some of that work is relevant to the philosophy of science also, since some scientific explanations are of this kind. One example is an explanation of the fact that it is 72 degrees Farenheit in this room that cites the motions of the room's air molecules. (Rosen 2010) and (Fine 2012) are standard references on grounding and explanation.

Third, and finally: some explanations are better than others. There are two obvious ways in which one may be better: an explanation may be better by being more complete, and an explanation may be better by being more relevant in context. (If what I want are the psychological factors that explain your behavior, then an explanation citing those factors is better in this sense.) Are there any other ways for one explanation to be better than another? We seem to have a word for one: depth. One explanation may be deeper than another, and being deeper seems like a way of being better. If this is right, then a complete theory of explanation should include a theory of depth. What does it take for one explanation to be deeper than another? Weslake (2010) surveys some answers and proposes his own.

References

Achinstein, Peter (1983). *The Nature of Explanation.* Oxford University Press.

Baker, Alan (2005). "Are There Genuine Mathematical Explanations of Physical Phenomena?" *Mind* 114: 223-38.

Batterman, Robert (1992). "Explanatory Instability." *Nous* 26: 325-48.

Cartwright, Nancy (1979). "Causal Laws and Effective Strategies." *Nous* 13: 419-437.

Cross, Charles (1991). "Explanation and the Theory of Questions." *Erkenntnis* 34:

237-60.

Dray, W. (1957). *Laws and Explanation in History.* Oxford University Press.

Fara, Michael and Sungho Choi (2012). “Dispositions.” *The Stanford Encyclopedia of Philosophy.*

Fine, Kit (2012). “Guide to Ground.” *Metaphysical Grounding: Understanding the Structure of Reality*, Fabrice Correia and Benjamin Schnieder, eds. Cambridge University Press: 37-80.

Fisher, R. (1931). *The Genetical Theory of Natural Selection.* Dover.

Friedman, Michael (1974). “Explanation and Scientific Understanding.” *The Journal of Philosophy* 71: 5-19.

Hempel, Carl (1965). “Aspects of Scientific Explanation.” In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* The Free Press, 331-496.

Hempel, Carl and Paul Oppenheim (1948). “Studies in the Logic of Explanation.” *Philosophy of Science* 15: 135-75.

Hitchcock, Chris (1995). “Salmon on Explanatory Relevance.” *Philosophy of Science* 62: 304-20.

— (1996a). “Farewell to Binary Causation.” *Canadian Journal of Philosophy* 26: 267-82.

— (1996b). “The Role of Contrast in Causal and Explanatory Claims.” *Synthese* 107: 395-419.

— (2010). Probabilistic Causation. *The Stanford Encyclopedia of Philosophy.* `http://plato.stanford.edu/entries/causation-probabilistic/`

Humphreys, Paul (1989). *The Chances of Explanation.* Princeton University Press.

Jackson, Frank and Philip Pettit (1992). “In Defense of Explanatory Ecumenism.” *Economics and Philosophy* 8: 1-21.

Kitcher, Philip (1981). “Explanatory Unification.” *Philosophy of Science* 48: 507-31.

— (1986). “Projecting the Order of Nature.” In R. E. Butts (ed.), *Kant’s Philosophy of Science.* D. Reidel Publishing: 201-35.

— (1989). “Explanatory Unification and the Causal Structure of the World.” In Kitcher and Salmon (eds.), *Scientific Explanation.* Volume 13 of *Minnesota*

Studies in the Philosophy of Science. University of Minnesota Press: 410-505.

Kitcher, Philip and Wesley Salmon (1987). "Van Fraassen on Explanation." *The Journal of Philosophy* 84: 315-30.

Lange, Marc (2013). "What Makes a Scientific Explanation Distinctively Mathematical?" *British Journal for the Philosophy of Science* 64: 485-511.

— (2014). "Aspects of Mathematical Explanation: Symmetry, Unity, and Salience." *The Philosophical Review* 123: 485-531.

Lewis, David (1986). "Causal Explanation." In *Philosophical Papers, Volume II*. Oxford University Press.

Lipton, Peter (2004). *Inference to the Best Explanation.* Routledge. 2nd edition.

Mackie, John L. (1974). *The Cement of the Universe.* Oxford University Press.

Morrison, Margaret (2000). *Unifying Scientific Theories.* Cambridge University Press.

Paul, L. A. and Ned Hall (2013). *Causation.* Oxford University Press.

Pearl, Judea (2009). *Causality*. 2nd edition. Cambridge University Press.

Railton, Peter (1978). "A Deductive-Nomological Model of Probabilistic Explanation." *Philosophy of Science* 45: 206-26.

Reutlinger, Alexander (forthcoming). "Why Is There Universal Macro-Behavior? Renormalization Group Explanation As Non-causal Explanation." *Philosophy of Science.*

Rosen, Gideon (2010). "Metaphysical Dependence: Grounding and Reduction." In *Modality: Metaphysics, Logic, and Epistemology*, Bob Hale and Aviv Hoffmann, eds. Oxford University Press: 109-36.

Salmon, Wesley (1984). *Scientific Explanation and the Causal Structure of the World.* Princeton University Press.

— (1989). *Four Decades of Scientific Explanation.* University of Pittsburgh press.

Salmon, Wesley, Richard Jeffrey, and James Greeno (1971). *Statistical Explanation and Statistical Relevance.* University of Pittsburgh Press.

Schaffer, Jonathan (2005). "Contrastive Causation." *The Philosophical Review* 114: 297-328.

Scriven, Michael (1959). "Truisms as the Grounds for Historical Explanation." In P. Gardiner (ed.), *Theories of History.* The Free Press, 443-75.

— (1962). "Explanations, Predictions, and Laws." In *Minnesota Studies in the Philosophy of Science III*, Herbert Feigl and Grover Maxwell, eds. University of Minnesota Press: 170-230.

Skow, Bradford (2014). "Are There Non-Causal Explanations (of Particular Events)?" *British Journal for the Philosophy of Science* 65: 445-467.

— (MS-a). "Explanations, Reasons Why, and Causes."

Spirtes P., C. Glymour, and R. Scheines (2001). *Causation, Prediction, and Search.* 2nd edition. MIT Press.

Sober, Elliott (1983). "Equilibrium Explanation." *Philosophical Studies* 43: 201-10.

Stalnaker, Robert (2014). *Context.* Oxford University Press.

Steiner, Mark (1978). "Mathematical Explanation." *Philosophical Studies* 34: 135-51.

Strevens, Michael (2008). *Depth.* Harvard University Press.

Van Fraassen, Bas C. (1980). *The Scientific Image.* Oxford University Press.

Weslake, Brad (2010). "Explanatory Depth." *Philosophy of Science* 77: 273-94.

Woodward, James (2003). *Making Things Happen.* Oxford University Press.

Yablo, Stephen (1992). "Mental Causation." *The Philosophical Review* 101: 245-80.

— (2002). "De Facto Dependence." *The Journal of Philosophy* 99: 130-48.