

Causal Patterns and Adequate Explanations

Angela Potochnik

Abstract

Causal accounts of scientific explanation are currently broadly accepted (though not universally so). My first task in this paper is to show that, even for a causal approach to explanation, significant features of explanatory practice are not determined by settling how causal facts bear on the phenomenon to be explained. I then develop a broadly causal approach to explanation that accounts for the additional features that I argue an explanation should have. This approach to explanation makes sense of several aspects of actual explanatory practice, including the widespread use of equilibrium explanations, the formulation of distinct explanations for a single event, and the tight relationship between explanations of events and explanations of causal regularities.

1 Three Puzzles for Causal Explanation

Recent decades have witnessed the ascendancy of causal accounts of scientific explanation. This is due in part to the broad acceptance of an “explanatory asymmetry” whereby causes can explain their effects, but effects cannot explain their causes (Bromberger, 1966). Famously, a flagpole’s height and the position of the sun can explain the length of the flagpole’s shadow, but the shadow’s length and the sun’s position ordinarily cannot explain the flagpole’s height. Woodward’s (2003) manipulationist approach to causal explanation is held in high regard by many, and has contributed to the dominance of (a version of) the causal

account of explanation. Strevens' (2008) more recent kairetic account is another notable example of an approach to causal explanation. Finally, a causal approach to explanation accords well with current interest in mechanistic explanation (e.g. Machamer et al., 2000, Craver, 2006).

I suspect that causal explanation is at least a prominent subspecies of scientific explanation, if not the only game in town. But I will not defend that view in this paper. I simply assume that illuminating causal relationships is a key task of (many) explanations. My aim is to show that, even for a causal approach to explanation, significant features of explanatory practice are not determined by settling how the causal facts bear on the phenomenon to be explained. Providing causal information cannot be the only goal nor the only requirement of scientific explanation; other considerations determine important features of explanations. I illustrate this with three puzzles. Versions of these puzzles are already addressed in parts of the literature on scientific explanation, but in my view their full significance is not widely appreciated. Each draws attention to a way in which explanations are underdetermined by the causal facts and, accordingly, indicates a feature that a causal account of explanation must possess.

Explanations That Neglect Causes

Why is the cup of coffee on my desk 70 degrees Fahrenheit? —because that is the ambient temperature of my office, and the coffee has been sitting there for hours.

Equilibrium explanations like this one indicate the structural features of a scenario that determine equilibrium or stasis conditions and the features that allowed the system to reach equilibrium conditions. This is a widely employed style of scientific explanation. Examples include explanations of the outcomes of thermodynamic processes that appeal to the ideal gas law; explanations of organisms' traits that appeal to the selective advantage

conferred by the trait; and explanations in biology and the social sciences that appeal to game theory. Any causal account of explanation that aspires to generality should be able to accommodate this broadly employed style of explanation, but some consider equilibrium explanations to be non-causal (Sober, 1983) or not straightforwardly causal (Rice, 2013). Others, including Woodward (2003), do judge equilibrium explanations to be causal. Yet even when equilibrium explanations are assimilated to a causal approach to explanation, principled reasons should be provided for the distinctive appeal of equilibrium explanations over alternative explanations that give more detailed causal information.

Different Explanations Without Different Causes

Why did the vervet monkey run up the tree? —because another vervet made an alarm call.

Why didn't it run under the brush instead? —because a snake was present, not an eagle.

Why did it move at all, when it stayed still last time? —the vervet who raised the alarm was an adult; last time, it was a young vervet whose call couldn't yet be trusted.

All of these are requests for explanations of a single explanandum, the proposition that a vervet monkey scaled a tree (on a particular, if imagined, occasion). Nor is the causal process in question: in this scenario, an adult vervet saw a snake, made the “snake” alarm call; the focal vervet heard this call, correctly identified its source and content, and proceeded to scale a tree. Yet the explanations provided all cite different causal factors. It is common to account for this type of variation by appealing to the role of contrast classes: an event is explained by different causal factors, depending upon what is taken to be the alternative, non-occurrent event. For present purposes, the point is that some additional consideration—

such as which contrast is focal—is at work, shaping the content of the causal explanation. Something needs to be said about this influence on which explanation is appropriate.

Causes That Aren't Satisfying Explanations

Why did Laura start to smoke cigarettes? —Laura has a variant at the *BDNF* genetic locus that affects one's response to social stress and can lead to smoking initiation.

According to recent findings, certain variants at this *BDNF* genetic locus do exert a significant effect on one's propensity to smoke (Amos et al., 2010). However, though our imagined Laura has one such variant, this is hardly a successful explanation of why she started to smoke cigarettes. Too much is missing: perhaps that, as a teenage girl, she is part of a demographic at particular risk of smoking initiation. Perhaps that she has a troubled home life, poor school attendance, high risk-taking behavior. Perhaps that she faced a particular source of stress immediately before smoking her first cigarette. Perhaps simply that Laura lives in twenty-first century United States, a society where social stress leads some to smoke. So, although this variant at the *BDNF* locus is a causal influence on Laura's initiation of smoking, it does not by itself successfully explain her smoking. To generalize: some true causal information may not be sufficient to explain an event. Yet it is impossible to cite *all* causal information, since causal processes stretch indefinitely far back into time and are often quite complex at any given stage. Accordingly, an account is needed to settle which causal factors are required for an adequate explanation and which can be omitted.

I do not claim that these are unrecognized issues nor that existing accounts of explanation cannot possibly accommodate them. Instead, I call these puzzles because they are left unresolved by adopting a causal approach to explanation and bringing an account of causation to bear. These three puzzles demonstrate that, even for a causal approach to

explanation, important features of explanatory practice are not determined by settling the causal facts. Any account of explanation must additionally address what makes some causes explanatorily significant and others explanatorily insignificant. In the remainder of this paper, I argue that each of these puzzles highlights a central feature of how science uses causes to explain, and I use those insights to develop a broadly causal approach to explanation. I employ insights from several other accounts of explanation, most notably those developed by van Fraassen (1980), Woodward (2003), and Strevens (2008). Yet the observation at the heart of this paper—that crucial features of explanatory practice are not settled by causal analysis—motivates an approach to explanation that differs from these and other accounts.

Each of the sections below addresses one of the three puzzles I have introduced. In Section 2 I motivate the idea that the best explanation of an event will showcase the causal pattern responsible for that event, viz., a causal dependence and the scope of that dependence. I thus endorse what I call broadly causal explanations, which may omit some causal factors or even avoid referencing any actual causal processes. This radically increases the number of potential causal explanations for a given event. Accordingly, in Section 3 I suggest a means for selecting which broadly causal explanation is appropriate to the circumstances of explanation. I argue that this is determined by which causal elements are focal to the research program in which an explanation is formulated. Allowing explanations to be shaped by researchers' interests in this way introduces the possibility of idiosyncratic explanations that say more about what researchers want to be true than about causal significance. Thus, in Section 4 I develop a way to keep explanatory role proportionate to causal role. This is accomplished with a requirement of explanatory adequacy that any causal pattern must satisfy in order to serve as an explanation.

2 Causal Patterns

Equilibrium explanations are recognized as a potential hurdle for a causal account of explanation. Sober (1983) distinguishes equilibrium explanation from causal explanation, and he defends the idea that an equilibrium explanation can generate more understanding of an event, even though it says less about the event's causes than a causal explanation would. Woodward (2003) criticizes the causal mechanical model of explanation, viz. Salmon's (1984) view, for its inability to accommodate explanations that do not "trace causal processes," of which equilibrium explanations are one type. There are two issues here. The first is whether equilibrium explanations can (and should) be assimilated to a causal approach to explanation. The second is whether a causal approach can account for the distinctive appeal of explaining with equilibria.

Let us turn first to the question of whether equilibrium explanations should be considered to be a type of causal explanation. Though equilibrium explanation is incompatible with some conceptions of causal explanation, it can be accommodated by others. Woodward (2003) claims this as an advantage of his manipulationist account of causal explanation—he intends his view to apply to equilibrium explanations. Kuorikoski (2007) employs Woodward's manipulationist account to more fully defend equilibrium explanations as a type of causal explanation, on the grounds that they capture causally relevant properties of the system. Such explanations do not detail the causal process that led to the phenomenon to be explained, but they still depict causal relationships:

The explanatory dependencies captured by these models are not causal dependencies between successive events, initial conditions and the end state, but constitutive dependencies between properties of a system and their causal basis (Kuorikoski, 2007, p.156).

This suggests a broadened conception of causal explanation that accommodates explaining

with equilibria. In Kuorikoski's view, explanatory information may include not just information about a causal process leading to the phenomenon to be explained—"causal dependencies between successive events"—but also the relationship between structural features of the system and the phenomenon to be explained. Such information does not describe a causal process, but it qualifies as a causal dependency on Woodward's manipulationist account of causation. An intervention on the cited structural properties would shift the equilibrium value, and thus change the phenomenon to be explained, whereas the relationship between the structural properties and equilibrium value is invariant across changes to initial values. Put in terms of the example of the coffee on my desk, a change to the ambient temperature would change the drink's equilibrium temperature, but the relationship between room temperature and equilibrium drink temperature does not vary, no matter whether my beverage is hot coffee or cold tea.

Notice that this way of assimilating equilibrium explanation to a causal account of explanation has additional implications. The broadened conception of causal-explanatory information sanctions explanations that do not depict the causal process leading to the event to be explained, so long as they depict causally relevant properties. Equilibrium explanations are but one class of such broadly causal explanations; another is explaining a phenomenon by citing a causal regularity of which it is an instance. Here is an example adapted from (Strevens, 2008). One may explain why table salt (sodium chloride) dissolves easily in water by pointing out that sodium is an alkali metal, and that all alkali metals have loosely bound outer electrons, which makes them prone to ionization. This explanation appeals to a causal regularity embodied by a class of entities and points out that the type of entity of interest is a member of that class. Such regularity explanations have commonalities with alternative approaches to explanation, such as the unification account (Kitcher, 1981).

Many causal regularities are not exceptionless. Accordingly, causal generalizations that have exceptions further expand the possibilities for broadly causal explanation. For example,

explaining the pressure of a gas in a rigid container at room temperature by appealing to the ideal gas law is an equilibrium explanation—and thus explains by depicting a causal dependency. This is so even though the ideal gas “law” holds only approximately, and breaks down in some conditions. Woodward (2003) fully endorses the explanatory value of causal generalizations that are limited in these ways. He argues that one of the primary advantages of his account of explanation is that it still counts as explanatory those causal generalizations that are limited in their scope and that hold only approximately.¹

This demonstrates how a range of generalizations about causal dependencies can be deemed to be broadly causal explanations. It thus shows how equilibrium explanations and causal regularity explanations—even those that are approximate and have exceptions—can be assimilated to a causal approach to explanation. Woodward’s manipulationist account, with its central concept of invariance, offers one analysis of how such generalizations convey causal-explanatory information. I am not wedded to the specifics of Woodward’s account, but it does identify a key feature of causal explanation. Explanations may not trace actual causal processes, e.g., understood in terms of Salmon’s (1984) mark transmission or Dowe’s (2000) conserved quantities. Instead, broadly causal explanations portray *causal patterns*, including patterns that are approximate and partial. This idea is further supported by the discussion that follows.

Let us now turn to the second question I raised at the beginning of this section, namely, the question of what accounts for the distinctive appeal of broadly causal explanations, including equilibrium explanations and causal regularity explanations. At least sometimes, these broadly causal explanations are better than explanations that trace the actual causal processes that led to the phenomenon to be explained. Consider my initial example of

¹This move to “broadly causal” explanations is different from the view that explanations cite causal laws primarily because the former allows causal regularity explanations that are not exceptionless. The present view also sanctions explanations that do not qualify as mechanistic explanations; an obvious example is equilibrium explanations.

an equilibrium explanation: explaining why the cup of coffee on my desk is 70 degrees Fahrenheit with the information that my office is 70 degrees and that the drink has been there for hours. An alternative explanation might depict features of the actual causal process by, e.g., specifying that the eight ounce coffee was 170 degrees Fahrenheit when placed in the 70 degree room four hours ago; how the processes of evaporation, conduction, convection and radiation combined for a certain rate of heat transfer; and how that rate and the relative masses and heat capacities of the substances involved resulted in the beverage cooling to 70 degrees well within four hours. I submit that the equilibrium explanation for the beverage's temperature is at least sometimes a better explanation than the causal process explanation. Another example is explaining why table salt dissolves easily in water by referencing a general causal regularity. According to Strevens (2008),

It is fine to explain sodium's reactivity by pointing to the fact that all sodium atoms have loosely bound outer electrons. But it is even better to point to the fact that sodium is one of the alkali metals, and that all the alkali metals have loosely bound outer electrons (245).

What can be said, from the perspective of a causal approach to explanation, to justify why such explanations are at least sometimes better than explanations that trace actual causal processes?

The answer lies, I think, in the ability of such explanations to depict causal patterns. Depictions of causal patterns have two features that together constitute their power to explain: (1) they feature one or more of the properties of a system upon which the phenomenon to be explained depends, and (2) they communicate information about the *scope* of that dependence. Equilibrium explanations provide information about the scope of a causal dependence in virtue of their setup: equilibrium models typically invoke the minimum assumptions required in order to generate the domain of attraction. Thus the ideal gas law omits reference to the type of gas molecules and intermolecular attraction; evolutionary game

theory models simply assume that a trait is heritable without indicating the exact nature of genetic transmission; and the equilibrium explanation of my abandoned coffee's temperature did not indicate its original temperature. Employing minimal assumptions makes these models more tractable, and it facilitates their application in any circumstances when the domain of attraction obtains. It also thereby indicates the full range of circumstances in which the domain of attraction obtains, viz., gives information about the scope of the causal dependence in question. Explanations that reference a broad causal regularity, e.g., the loosely bound outer electrons of not just sodium but all alkali metals, similarly indicate the full range of circumstances in which a causal dependence obtains. Causal pattern explanations thus not only depict causal dependencies, they also delimit the scope of the dependence.²

The second feature, information about the scope of the causal dependence, can be lacking from causal process explanations. Tracing the causal process that led to a phenomenon gives information about causal properties upon which the phenomenon depends, but it may not indicate anything about the circumstances in which the dependence holds. The more detailed the representation of a causal process, the less information is communicated about how the details of the process may vary without altering the causal dependency in question. Let me be clear. I do not claim that tracing specific causal processes can never be explanatory, nor do I claim that such explanations never indicate the scope of dependence. As an illustration, consider that some phenomena depend essentially on myriad factors, and a slight change to any one of these factors significantly alters the phenomenon. In such scenarios, tracing

²I have here characterized depictions of causal patterns, but one might also ask what a causal pattern is (rather than what it takes to depict one). Causal patterns are *patterns* in the sense that they are relationships that generally obtain in a range of circumstances of some breadth, be it wide or narrow. This is why a depiction of a causal pattern must give information about the scope of that pattern. Causal patterns are *causal* in the sense that the pattern is ultimately due to causal connections. They are only broadly causal: the pattern may not consist in causal connections between successive events. Finally, notice I said that causal patterns *generally* obtain in a range of circumstances. This is because causal patterns can exist despite exceptions, even within the pattern's range. I discuss the consequences of patterns allowing exceptions below.

the specific causal process may be the only explanation possible, and that explanation accurately indicates the (very narrow) scope of the dependence. In contrast, when explaining phenomena for which an equilibrium explanation or causal regularity explanation is available, a mark in favor of this approach to explanation—a reason to prefer such an explanation—is its ability to communicate the scope of the causal dependence in question.

Information about the scope of the causal dependence contributes to an explanation in a number of ways. It provides information about the circumstances in which phenomena of that type can be expected, such as the information that a beverage (in an open cup) will be room temperature anytime it has set undisturbed in the room for more than a few hours. It relates the phenomenon in question to similar phenomena, such as today’s hot coffee and yesterday’s cold tea both becoming room temperature. And it contributes to our store of information about general causal regularities, such as the factors upon which heat transfer depends. None of these is accomplished by causal information alone; the scope of causal dependence is also crucial information.

The explanatory value of information about what I am calling the scope of causal dependence has been posited by a number of philosophers. Arguments for the value of high-level explanations appeal to these explanations’ applicability in a range of circumstances that have certain high-level properties in common (e.g. Putnam, 1975; Garfinkel, 1981; Jackson and Pettit, 1992). Strevens (2008) urges that causal factors that do not truly make a difference to a phenomenon be omitted from its explanation; to this end, he argues that causal explanations should be as general as possible. Woodward’s (2003) account of explanation is particularly valuable here as well: the concept of an intervention is intended to identify causal dependencies, and the concept of invariance is intended to identify the scope of those dependencies.

Assimilating equilibrium explanations to a causal approach to explanation thus leads to the recognition of an important feature of causal explanation. The value of equilibrium

explanations and causal regularity explanations over causal process explanations stems from the former's ability to portray a causal pattern. The explanatory value of causal patterns consists in their ability to highlight a relationship of causal dependence, along with the scope of that dependence. Causal process explanations, in contrast, may fail to provide information about the scope of the causal dependence, in which case their value as explanations is diminished.

Another feature of this idea about the explanatory value of causal patterns deserves discussion. Recall from above that many causal regularities are not exceptionless, so explanatory causal generalizations may hold only approximately. This creates the possibility of a successful causal pattern explanation that accurately indicates the scope of dependence but nonetheless omits significant causal influences. Consider the ideal gas law. That generalization does not reference molecular size or intermolecular forces, but these are both significant causal variables. If molecules are too large or intermolecular forces too strong, the regularity captured by the ideal gas law ceases to obtain. Because an ideal gas law explanation of, say, a gas's pressure does not specify these two properties of the gas, it omits depicting at least two significant causal influences. To generalize from this example, there may be significant causal influences on a phenomenon that are not referenced in an explanation of that phenomenon. It follows that demonstrating that a factor is causally relevant does not prove its *explanatory* relevance.

This implication conflicts with accounts that aim to establish explanatory relevance solely on the basis of establishing causal relevance, including Woodward's and Strevens'. Showing that a feature of a system qualifies as causal (or difference-making) on Woodward's or Strevens' account is not sufficient to show that it belongs in an explanation. In my view, an explanation sometimes can be improved by omitting mention of some significant causal influence(s). I thus begin to depart from Woodward and Strevens at this point, though I have employed insights from both their accounts in accommodating and motivating equilibrium

explanations and other causal pattern explanations. This departure will become evident in the next section, where I further explore the disconnect between causal relevance and explanatory relevance.

3 The Principled Neglect of Causes

If Section 2 is right, then the class of potential causal explanations is greatly expanded. Included are not only what one might term *causal process explanations*—explanations that trace parts of the causal process that led to the event in question—but also *causal pattern explanations*—explanations that depict the relationship between structural features of the system and the event in question, neglecting mention of any actual causal steps.³ Moreover, causal pattern explanations may omit significant causal or structural features, even at the price of allowing some exceptions to the pattern they depict. This plethora of potential causal explanations should bring to mind the second puzzle from Section 1: a case in which one explanandum warrants different explanations in different circumstances, even when the causal facts are not in dispute. In the simple example I presented, the explanation for a vervet monkey scaling a tree was, depending on the question posed, (a) that another vervet made an alarm call; (b) that there was a snake present, not an eagle; and (c) that the vervet raising the alarm was an adult, not a juvenile. Some consideration other than the causal facts helps to determine which of these explanations is appropriate.

My vervet monkey example is akin to examples used to motivate the view that explanation is contrastive.⁴ Garfinkel (1981) claims that “explanation always takes place

³This distinction between causal process explanations and causal patterns explanations is similar to Dretske’s (2004) distinction between triggering and structuring causes, but there is an important difference. Dretske’s triggering and structuring causes each may be part of the causal process that led to the event to be explained, whereas causal pattern explanations may cite features of the scenario that are not parts of the causal process. For instance, the ideal gas law expresses a structural feature shared by a wide variety of causal processes, including those with reverse causal dependencies, such as pressure causing a change in temperature and temperature causing a change in pressure.

⁴See Dretske (1972) for an early treatment of contrastive statements.

relative to a background of alternatives,” and that what he terms “contrast spaces” are needed to represent that relativity (p.25). On this view, the intended contrast with a counterfactual state of affairs, whether explicit or implicit, determines which causal facts belong in an explanation. This accounts for the different vervet monkey explanations by appealing to the different salient alternatives, that is, the different contrast classes. The explanations I listed distinguish the vervet scaling a tree from, respectively, (a) doing nothing; (b) running under the brush; and (c) noting an alarm call but not responding.

Van Fraassen (1980) also stresses the significance of contrastive explanation, but he argues that contrast classes are only one way that contextual factors influence what information belongs in an explanation. I think this must be right; more goes into determining which causal information is called for in an explanation than merely contrast classes. One way to establish this is simply with an example where the causal facts *and* the contrast class are held fixed, and yet the best explanation seems to vary. Consider an explanation for why a tablespoon of salt dissolved in a glass of water, rather than settling at the bottom (the contrast class). The explanation advocated in Section 2 (following Strevens, 2008) is the causal regularity explanation that the sodium in the salt, like all alkali metals, has loosely bound outer electrons and is therefore highly reactive. However, this is a non-explanation for a chemist who well knows the properties of sodium, but thought she had added enough salt to create a saturated sodium chloride solution. In that context, the best explanation will point to the fact that the water had previously been sitting in the sun, so had warmed to significantly above room temperature, and the saturation point has thus increased. In this simple example, the two explanations direct attention to different causal dependencies, even though the explanandum, causal facts, and contrast class are all the same.

A range of philosophers have acknowledged and addressed such contextual influences on explanatory practice. However, these influences are generally relegated to the category “pragmatics of explanation,” and then either dismissed as beyond the scope of a core account

of explanation, or else treated as secondary, add-on features. Woodward (2003) takes the second approach, devoting only seven pages late in his book to the role of pragmatics. Setting aside contextual influences is often yoked to an ontological view of explanation, viz., that explanations are observer-independent facts, out there in the world. Strevens' (2008) account of explanation is a prime example.⁵ In my view, dismissing contextual influences or postponing consideration of them until after a core account of explanation is developed is not the right approach to take.

Almost all events result from a plethora of causal dependencies that can be represented in a multitude of ways. Depicting the full gamut in a single explanation is impossible. For one thing, causal processes stretch indefinitely far back in time. A more interesting problem is that a choice must be made about how to represent each causal dependency, and employing one representation can preclude others. For instance, an explanation must either represent a certain causal dependency in terms of the structural features that determine the equilibrium point or in terms of a dynamic causal process. The former representation shows how certain structural features result in (or make very likely) a certain outcome, whereas the latter traces the dynamics of the actual causal trajectory. The two representations are fully distinct and accomplish very different tasks, and yet they represent the same segment of causal history. To include both representations in an explanation would amount to characterizing the same segment of causal history twice; this is arguably providing two explanations instead of one. This is made more obvious by the fact that each representation is accomplished via a different methodological approach, so they cannot be neatly combined. I say more on this topic below.

Because it is impossible to generate an explanation untinged by representational choices, questions of representation cannot be ignored or deferred. But how explanations should represent the causal dependencies at play is influenced by context; consider again the

⁵See Potochnik (2011) for concerns about Strevens' account of explanation that are related to the point developed here.

example immediately above. In the common explanation of salt dissolving in water, a (possibly unstated) assumption is that the water is not already saturated; this is how the first explanation above represents the water's level of saturation. In contrast, the second explanation focuses directly on the saturation level, representing it more precisely via the claim that the solution is just below the saturation point because of its higher than expected temperature. Thus, an analysis of the role of contextual influences is a crucial part of any account of causal explanation, for the determination of how the causal structure should be represented—and which features to represent—requires consideration of context.

Happily, current philosophy of science has the resources needed to accommodate this feature of explanation. The place to look is recent treatments of strategic abstraction and idealization for purposes of explanation. A number of philosophers have espoused versions of the view that explanations can be improved by the incorporation of abstractions and idealizations, despite the resultant decrease in representational accuracy; see especially (Batterman, 2002, 2009; Weisberg, 2007; Strevens, 2008; Bokulich, 2011). Weisberg (2007) terms this minimalist idealization, which he defines as “the practice of constructing and studying theoretical models that include only the core causal factors which give rise to a phenomenon.” This accords with the point established in the previous section, namely that a causal explanation should depict a causal pattern—viz., a causal dependency along with the scope of the dependence. As Batterman (2009) points out,

The world is constantly changing in myriads of ways; yet despite this, we see the same patterns over and over again in different situations. Idealizing is a means for focusing on exactly those features that are constitutive of the regularity—those features that we see repeated at different times and in different places (p.430).

The incorporation of idealizations into explanations is justified by the isolation of a causal pattern—Weisberg's “core causal factors.”

What, then, distinguishes the parts of a causal process that constitute a pattern from the parts that are patternless change, to be idealized away? The scope of causal dependence is certainly relevant; a causal dependence must be sufficiently pattern-like in order to qualify as a causal regularity, after all. But recall from the end of the previous section that many causal regularities hold only approximately, and thus may omit significant causal influences. This makes it possible for one causal process to embody multiple, distinct causal patterns. I postulate that our world is one in which this scenario is commonplace: causal processes are often in accord with multiple patterns, each admitting of exceptions. For this reason, what is pattern-like, and thus explanatory, is distinguished from non-central causal factors in part by the explainer's agenda or, one might say, the research program. A scientific research program usually involves (at least) a choice of focal phenomena; hypotheses about the types of causal factors at work or similarities with other phenomena; and a methodology, viz. a type of mathematical model, simulation, manner of investigation, etc. The research program in which an explanation is formulated accounts for the contextual influence identified above. An explanation of an event or phenomenon always occurs in the context of some research agenda. This narrows the scope of investigation to a segment of causal history, brings certain types of causal relationships to the fore, and involves certain methods of investigation and representation.

My examples up to this point—explaining a vervet monkey running up a tree and explaining table salt dissolving in water—are too simple to really show off how different research programs result in different explanations. Let us turn to another example, one truer to the sort of explanations actually generated in scientific research. As I mentioned above, one species of equilibrium explanation is provided by evolutionary game theory. Consider the game theory explanation for why Harris sparrows within a single flock vary in color between pale and dark (instead of being a single shade—the contrast class). Coloration is unrelated to physical strength, survival, or reproductive success, but it is related to status:

dark birds almost always displace pale birds from food sources (Rohwer and Ewald, 1981). Maynard Smith (1982) explains the variation in color with the Hawk-Dove game theory model. Because the cost of an injury through conflict is greater than the value of the food source to a sparrow, this polymorphism is an equilibrium point. Dark, aggressive birds risk injury from other dark birds but have greater access to resources; pale, submissive birds have fewer resources but avoid injury. In a mixed population (at equilibrium) each is an equally successful strategy, and coloration is a badge that helps the birds divide resources while avoiding unnecessary injury.

Now consider the same explanandum and contrast class—why Harris sparrows vary in color (within a flock) instead of being a single shade—from the perspective of a different research program. If sparrow color is polyphenic, that is, if the same genotypes produce both light and dark birds, a biologist researching phenotypic plasticity may investigate how genetic factors and environmental variation together explain the variation in color. This biologist would seek one or more genetic factors influencing the trait of feather-color in a way that is mediated by environmental conditions, such as the preponderance of aggressive or submissive birds in the flock. This would focus on a different segment of the causal history than the game theory explanation.⁶ The game theory explanation accounts very differently for the segment of causal history upon which the phenotypic plasticity explanation focuses: it simply assumes that feather color is heritable, abstracting away from all features of the process. Recall from above that one must choose how to represent a causal dependency, and that this results in different explanations suited to different research programs. In the present example, one cannot simply insert a phenotypic plasticity explanation into the game theory explanation; the modeling technique does not exist. One could, of course, give *both* explanations. But

⁶Some would attempt to characterize the difference as a difference in explanandum by claiming that natural selection (represented in the game theory model) explains why the sparrows *evolved* this trait and genetics and environmental conditions explain why the sparrows *develop* the trait. But this is simply a way to direct attention to one part or another of the causal history of the same event, namely that the sparrows have evolved to develop the trait in question. See also (Potochnik, 2010).

the game theory explanation is enlightening to those who wonder about whether and how natural selection led to this variation in color, while the phenotypic plasticity explanation is enlightening to those who wonder about whether and how the environment mediates genetic influence to produce this variation in color. In each case, the other explanation is entirely beside the point.

Explanations that incorporate abstractions and idealizations are supposed to showcase what has been termed variously the core causal factors (Weisberg, 2007), a causal pattern (Batterman, 2009), the true difference-makers (Strevens, 2008), etc. Yet many accounts of explanation have gone wrong or are lacking, due to a failure to appreciate the essential role of a research agenda in deciding *which* causal factors (or, equivalently, pattern) should be showcased in this way. The influence of the research agenda is so crucial to determining the content of explanations that accounting for it cannot be ignored or postponed until after a “core” account of explanation is provided. Abstractions and idealizations stand in for causal information, including information about significant causal influences, in order to enable the representation of the focal causal factors as a modular, repeatable causal dependence. The game theory explanation for variation in Harris sparrow color benefits from ignoring the basis of heritability, since this showcases the similarities between that trait and other instances of the Hawk-Dove game, regardless of significant causal differences among them.

To sum up, I postulate that we live in a world of complex causal processes occurring in myriad permutations, but that nonetheless allow the discovery of patterns. A consequence of this complexity is that many causal patterns are riddled with exceptions. Abstractions and idealizations are introduced into explanations in order to showcase patterns and regularities, even at the expense of neglecting other significant causal influences. Yet an explanation that neglects significant causal influences cannot possibly be the *single* best causal explanation of an event. Instead, any explanation is relative to the background research agenda, that is, to what sort of causal regularities are of interest. Thus, there is always a third factor influencing

explanations, beyond the explanandum and the causal facts: the research program in which the explanation is sought.⁷ Some of the research program's influence may be conceived of in terms of contrast classes, i.e., what alternative to the explanandum is explored, but this does not exhaust the influence. The role of research programs in the formulation of explanations is, ultimately, what enables the principled neglect of some causes from causal explanations.

4 Attaining an Adequate Explanation

In the previous section I argued that any explanation is shaped in part by the research program in which it is formulated, and that perfectly good causal explanations can and should neglect some significant causal influences. Yet it seems equally clear that providing any old causal information does not always generate an acceptable causal explanation. Recall the third puzzle from above. That an imagined character Laura has a particular gene that makes social stress more likely to induce smoking cigarettes is, by assumption, a true causal fact. However, this information does not seem to explain why Laura began smoking, at least not by itself. Too much is missing; this does not yet account for the difference between Laura and nonsmokers (or a nonsmoking Laura). A causal explanation may fail if it does not give the "right" causal information. The task, then, is to determine what additional criteria must be met in order for causal information to qualify as an adequate explanation.

The idea that explanations should capture causal patterns suggests one possible approach: perhaps explanations should feature causal factors that have the most influence, while ignoring those with a minor influence. This idea is suggested by Garfinkel (1981), for one, who argues that explanations should feature causes that critically affect the outcome

⁷Above I said that a research program is often associated with choice of focal phenomena, hypotheses about types of causes or similarities with other phenomena, and a methodology to be employed. Research programs are also influenced by chance factors such as what equipment happens to be available and what techniques and target systems the researcher happens to be familiar with (Rob Skipper, in conversation). Such accidental circumstances thus also influence the explanations that are generated.

and omit overly specific details. In a similar spirit, Strevens (2008) suggests that minor causal influences, such as other planets' contribution to Mars' exact orbit, may not have any explanatory role to play. But this approach faces a problem: sometimes, even oftentimes, minor causal influences are center stage in the quest for scientific understanding, while significant causal factors are relegated to the role of background conditions. The smoking example used here is inspired by recent research into genetic influences on smoking behavior, namely (Amos et al., 2010). The researchers identify distinct genetic influences on three stages of smoking behavior: smoking initiation, current smoking, and smoking cessation. Yet they also emphasize that these genetic factors are only one part of the story; neurobiological and environmental influences are also significant. Even though the availability of tobacco and social norms are highly influential causal variables (and acknowledged as such by Amos et al.), the relatively minor effects of each genetic factor are the researchers' focus.

Strevens (2008) suggests a distinct requirement that he thinks causal information should meet in order to count as an explanation. He claims that, for any determined event, an explanation must "causally entail" the explanandum. This means that, to be a full-fledged explanation (of a determined event), a model must entail the explanandum, and the entailment must proceed in a way that accurately represents the causal process that in fact led to the event. This requirement imports some of the spirit of the classical deductive-nomological approach to explanation into a causal account. However, even for determined events, the requirement of causal entailment is too strong. Many of the most enlightening causal patterns—the ideal gas law, predator-prey relationships, evolutionary arms races, and *X* and *Y* chromosomes' role in sex determination, to name a few—have exceptions. Successful deterministic explanations can allow for exceptions and, moreover, nondeterministic explanations can explain determined events. To require that any causal explanation entail the explanandum would eliminate many of science's most successful explanations.

It seems, then, that relying on the strength of causal relationships to establish the minimal requirements for explanation cannot do the job properly, nor can stipulating that an explanation must entail the outcome to be explained. Yet both approaches are headed in a promising direction, and their failures are instructive. Each is an interpretation of the sensible requirement that an explanation should in fact account for, or demonstrate what is responsible for, the occurrence of the event to be explained. The question is how best to formulate that requirement.

As a first step, let us weaken Strevens' requirement of causal entailment in a straightforward way that allows causal patterns that have exceptions to nonetheless qualify as explanatory:

$$Pr(E|C_{expl}) \approx Pr(E|C) \tag{1}$$

Here E is the event to be explained, C_{expl} is the causal information comprising the explanation, and C is the full causal history of E . In other words, an adequate explanation of an event must provide causal information that accounts for the approximate probability of the occurrence of that event, given all events in its causal history. It follows that, for deterministic phenomena, an explanation should provide information that collectively entails or renders highly probable the event to be explained. But this formulation of the requirement serves equally well for non-determined events. For events that are truly probabilistic, an explanation should provide information that renders the occurrence of the event roughly as probable as it in fact was, given the full causal history.⁸

A few possible concerns with this basic approach can be averted by recalling the goal of this requirement. The aim is to establish a benchmark that must be met by any adequate

⁸These must be objective probabilities, but otherwise, I expect it is irrelevant which probability interpretation is employed. Notice that the requirement that the probabilities in question be approximately equal is vague. This is fitting, since it seems that the threshold of explanatory adequacy is also vague. Moreover, I expect the degree to which $Pr(E|C_{expl})$ must approximate $Pr(E|C)$ may vary somewhat in different circumstances of explanation. The degree of approximation required often allows for true causal influences of modest effect, and possibly even moderate effect, to be neglected.

explanation, viz. the minimal standard some causal information must meet in order to qualify as an explanation. This requirement, which I will call the requirement for explanatory adequacy, is thus not intended as a sufficient condition for successful explanations, but only a necessary condition to be met by any otherwise satisfactory explanation. Nor is the present requirement expected to establish the true causal factors. Attempts to establish causal or explanatory relationships on the basis of probabilistic relationships have been fraught with difficulty, but my requirement for explanatory adequacy is not designed for either of these purposes. The overall project of this paper presumes an account of causation, or a method for ascertaining causal facts, which I in no way provide. And then, it is the research interests, together with consideration of causal patterns, that determine which causal facts are explanatory.⁹ This condition of explanatory adequacy is simply an additional benchmark to be met by any adequate explanation. Finally, notice also that this requirement does not limit the inclusion of minor causal influences; it simply requires the additional inclusion of major causal influences. These points will be further discussed below after my suggested criterion for explanatory adequacy is finalized.

In its current formulation, the requirement for explanatory adequacy is not strong enough. To see why, consider how to explain an event that corresponds to a widespread causal pattern but with an unusual wrinkle. When I walk into the kitchen and flip the switch, the light goes on. Consider, though, the time I walk in, flip the switch, and nothing happens because the breaker is tripped. I then walk down to the basement, reset the breaker, and the light goes on. One candidate explanation for why the kitchen light went on is that I flipped the switch. This is indeed an essential part of the causal process leading to the light going on, and the explanation meets the requirement (1). Yet it feels as if the explanation is lacking:

⁹Along with my commitment to causal explanation, this distinguishes my condition of explanatory adequacy from Hempel's requirement of deductive or statistical relevance. The condition of explanatory adequacy I propose in this section is designed for a different job. Additional differences will emerge as the discussion proceeds, including especially that the satisfaction of my proposed requirement does not require explicitly *citing* all significant factors.

this time, the flipping of the switch does not in itself explain the light going on. There is more that must be said. Formulating the requirement for explanatory adequacy as follows disqualifies such explanations:

$$Pr(E|C_{expl}) \approx Pr(E|C_{expl} \& C_k) \text{ for all } C_k \quad (2)$$

The new term, C_k , stands for any subset of the events in E 's causal history. This ensures that the explanation does not neglect to take into account—somehow or other—any causal factors of monumental importance to the event to be explained. For the light switch example, the requirement is not met for the C_k that includes the breaker tripping but not me subsequently resetting it. (2) is the definitive version of the requirement for explanatory adequacy, which I dub $[EA]$. $[EA]$ entails the previous formulation (1), for if an explanation renders an event approximately as likely as it is when any combination of causal factors are also taken into account, then this also holds true when *all* causal factors are taken into account.

The requirement $[EA]$ is somewhat similar to the first proposal for explanatory adequacy considered in this section, the popular idea that explanations should feature any causal factors that have a great deal of influence over the outcome. The similarity is that, in order to satisfy $[EA]$, an explanation must take into account all causal factors with a significant influence on the probability of the event's occurrence. Yet two differences from the earlier proposal deserve emphasis. First, as I mentioned above, $[EA]$ does not require that minor causal influences be omitted from an explanation. Sometimes minor causal influences are center stage in the quest for understanding; the illustration used above is recent research into a variety of subtle genetic influences on smoking behavior. When such minor influences are the research focus, they may well be an important part of the explanatory causal pattern. Second, unlike the first proposal considered in this section, the satisfaction of $[EA]$ merely requires that significant causal factors be *somehow* taken into account by an adequate

explanation. This need not involve explicitly citing, or representing, those factors. Significant causal factors that are not central to the research program can still be set aside as advocated in §3. Their influence simply must be acknowledged. Accounting for significant factors without focusing upon them is made possible with the use of strategic abstractions and idealizations. Simplifying assumptions can be used to bring an explanation into alignment with the totality of causal facts, so that $[EA]$ is satisfied.

Here is an illustration of how causal factors might be taken into account without being explicitly cited or represented. Consider again the simple equilibrium explanation for my coffee's temperature of 70 degrees Fahrenheit. As discussed in §2, equilibrium models must include the assumptions needed to generate the domain of attraction, that is, the range of conditions that would lead to the equilibrium value. For the coffee explanation, one of those assumptions is that the coffee is in a container that is not well insulated. The explanation cites the facts that my office is 70 degrees and that the coffee has been there for hours, and assumes that the coffee container is not well insulated, thereby accounting for the causal factor of my coffee being in an open ceramic cup. Even clearer examples can be drawn from equilibrium models actually employed in the sciences. For instance, many evolutionary game theory models assume that the strategies under investigation propagate to the degree of their success. This assumption is an idealization, for it is, strictly speaking, false of genetically influenced traits. Instead, strategies in one generation and the next share a common cause, namely genes that both influence the strategy and propagate themselves. The assumption of strategies propagating to the degree of their success is a way of taking into account—without accurately representing—the role of genetic influence on the strategy and on genetic transmission.¹⁰

The requirement $[EA]$ is much weaker than it appears because of the variety of ways in

¹⁰No matter how complex and epistatic the genetic influences are, $[EA]$ is satisfied by the simple assumption of heritability. So long as the genetics in fact result in strong enough heritability for the assumption to be borne out, it is irrelevant to $[EA]$ that genetic interaction *could* have interfered.

which an explanation can take into account significant causal factors. They may be explicitly cited or represented, or if they are inessential to the research program at hand, they may instead be implicitly acknowledged via assumptions, including simplifying assumptions that abstract or idealize. Significant causal factors may also be implicitly taken into account via common background knowledge. If you and I both know that I reset the tripped breaker, I may explain to you that the light turned on by citing the fact that I had flipped the switch earlier (before addressing the breaker). This is an explanation in virtue of our shared background knowledge about the reset breaker; that shared background knowledge is required in order for the explanation to satisfy $[EA]$.

Now we can diagnose the failure of the example explanation. The variant Laura has at the *BDNF* genetic locus is by itself not enough to explain her smoking initiation. The causal role of that variant is highly sensitive to circumstances like social norms and tobacco availability, and so the explanation fails to satisfy $[EA]$. Surely it is this sensitivity to other factors that accounts for why the research inspiring my example begins by noting the range of non-genetic influences on smoking behavior before focusing on the genetic influences they discovered (Amos et al., 2010). The requirement $[EA]$ thus places a limit on the neglect of causal influences in the pursuit of a causal-pattern explanation. A causal explanation must account for all of the significant causal influences on an event, even those that are not central to the patterns scientists are presently seeking to understand.

5 Conclusion

The three puzzles with which I began show that providing causal information simpliciter cannot be the only goal nor the only requirement of a scientific explanation. I have suggested a way of accommodating each of these puzzles via a broadly causal approach to explanation. The approach I advocate draws from the accounts of Woodward (2003), van Fraassen (1980),

Strevens (2008), and others. In §2 I argued that the goal of a causal explanation is to display a causal pattern. This results in the possibility and desirability of equilibrium and causal regularity explanations, which do not trace causal processes. This solves the first puzzle: equilibrium explanations are broadly causal explanations, valuable for their ability to highlight the scope of the causal patterns they represent. In §3, I argued that what causal pattern is explanatory depends on the focus of the research program in which the explanation is sought. This leads to the neglect of many significant causal factors, in the interest of portraying the focal causal pattern. This solves the second puzzle: one event, resulting from a single causal process, may nonetheless be explained in multiple distinct ways in the context of different research programs. Finally, in §4 I argued that causal information must meet the requirement $[EA]$ in order to qualify as an adequate explanation. Explanations' neglect of causal factors in order to represent causal patterns of interest is thus reined in. This solves the third puzzle: sometimes citing true causal facts is not sufficient to generate an explanation—too much may be missing.

These elements combine to form the skeleton of a broadly causal account of explanation. The best explanation of an event E will showcase the causal pattern responsible for E , viz., a causal dependence and the scope of that dependence. More accurately, the best explanation will showcase the causal pattern, involving the causal elements upon which the current research program focuses, that is responsible for E . The satisfactoriness of an explanation thus is indelibly linked to the research program in which it is formulated. It does not follow from this that any old just-so story can become an explanation. First, this account of explanation assumes that a successful analysis of causation (which I in no way provide) is used to establish the relevant causal facts. Only true causal influences can be part of a causal explanation. Second, an explanatory causal pattern must, as I said, *be responsible* for E . This means that an explanation must satisfy the requirement of explanatory adequacy, $[EA]$.

The features of causal explanation addressed in this paper are not merely secondary questions, peripheral to an account of the central features of explanation. Instead, they are crucial to determining what features of the causal process leading to an event should be included in an explanation of that event and how those features should be represented. These observations yield insight into how it is that science actually uses causes to explain. Appreciating the role of causal patterns helps make sense of how giving causal information can generate understanding. It also shows that any causal approach to explanation must be broad enough to accommodate, and even motivate, equilibrium explanations and other causal regularity explanations. The role of causal patterns and the related idea that any explanation is relative to particular research interests resonate with actual scientific practice. Equilibrium explanations, for example, are in widespread use, and scientific explanations are regularly declared completed even though they omit much causal history. Finally, the idea that explanations feature causal patterns of interest in the context of a particular research program connects event explanations to the explanation of causal regularities and laws in a natural way. Both types of explanation occur in tandem. The search for an understanding of causal regularities inspires and informs the search for explanations of individual events. The event explanations generated by a successful research program in turn inspire and inform explanations of causal regularities, surely one of the paramount aims of science.

Acknowledgments

This paper was drafted during a Research Fellowship at the Michelle R. Clayman Institute for Gender Research at Stanford University, and it was completed with the help of a Charles Phelps Taft Summer Research Fellowship. The ideas have benefitted from the help of Michael Friedman, Peter Godfrey-Smith, Helen Longino, Elliott Sober, and Michael Strevens. Additionally, Chris Haufe, Robert Skipper, Elliott Sober, James Woodward, and

the participants in a conference in honor of Elliott Sober’s 65th birthday, as well as several anonymous referees, have provided helpful feedback on earlier drafts of this paper.

References

- Amos, Christopher I, Margaret R Spitz, and Paul Cinciripini (2010), “Chipping away at the genetics of smoking behavior”, *Nature Genetics* 42: 366–368.
- Batterman, Robert W. (2002), *The Devil in the Details*, Oxford: Oxford University Press.
- (2009), “Idealization and modeling”, *Synthese* 169: 427–446.
- Bokulich, Alisa (2011), “How scientific models can explain”, *Synthese* 180: 33–45.
- Bromberger, Silvain (1966), “Why-Questions”, in R. Colodny, ed., *Mind and Cosmos*, Pittsburgh: University of Pittsburgh Press, 86–111.
- Craver, Carl F. (2006), “When mechanistic models explain”, *Synthese* 153: 355–376.
- Dowe, P. (2000), *Physical causation*, Cambridge University Press.
- Dretske, Fred (1972), “Contrastive Statements”, *The Philosophical Review* 81: 411–437.
- (2004), “Psychological vs. Biological Explanations of Behavior”, *Behavior and Philosophy* 32: 167–177.
- Garfinkel, Alan (1981), *Forms of Explanation: Rethinking the Questions in Social Theory*, New Haven: Yale 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: 507–531.
- Kuorikoski, Jaakko (2007), “Explaining with Equilibria”, in J. Persson and P. Ylikoski, eds., *Rethinking Explanation*, Springer, 149–162.
- Machamer, Peter, Lindley Darden, and Carl F. Craver (2000), “Thinking about Mechanisms”, *Philosophy of Science* 67: 1–25.

- Maynard Smith, John (1982), *Evolution and the Theory of Games*, Cambridge: Cambridge University Press.
- Potochnik, Angela (2010), “Levels of Explanation Reconceived”, *Philosophy of Science* 77: 59–72.
- (2011), “Explanation and Understanding: An Alternative to Strevens’ *Depth*”, *European Journal for the Philosophy of Science* 1: 29–38.
- Putnam, Hilary (1975), *Philosophy and our Mental Life*, Cambridge: Cambridge University Press, *Philosophical Papers*, vol. 2, chap. 14, 291–303.
- Rice, Collin (2013), “Moving Beyond Causes: Optimality Models and Scientific Explanation”, *Nouss* doi:10.1111/nous.12042.
- Rohwer, S., and P.W. Ewald (1981), “The cost of dominance and advantage of subordination in a badge signalling system”, *Evolution* 35: 441–454.
- Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.
- Sober, Elliott (1983), “Equilibrium Explanation”, *Philosophical Studies* 43: 201–210.
- Strevens, Michael (2008), *Depth: An Account of Scientific Explanation*, Cambridge: Harvard University Press.
- van Fraassen, Bas C. (1980), *The Scientific Image*, Oxford: Clarendon Press.
- Weisberg, Michael (2007), “Three Kinds of Idealization”, *The Journal of Philosophy* 104: 639–659.
- Woodward, James (2003), *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.