Explanation and Understanding: An Alternative to Strevens' *Depth*

Angela Potochnik

Abstract

Michael Strevens offers an account of causal explanation according to which explanatory practice is shaped by counterbalanced commitments to representing causal influence and abstracting away from overly specific details. In this paper, I challenge a key feature of that account. I argue that what Strevens calls explanatory frameworks figure prominently in explanatory practice because they actually improve explanations. This suggestion is simple but has far-reaching implications. It affects the status of explanations that cite multiply realizable properties; changes the explanatory role of causal factors with small effect; and undermines Strevens' titular explanatory virtue, depth. This results in greater coherence with explanatory practice and accords with the emphasis that Strevens places on explanatory patterns. Ultimately, my suggestion preserves a tight connection between explanation and the creation of understanding by taking into account explanations' role in communication.

1 Introduction

At their best, scientific explanations "light up our minds with a bolt of insight" (Strevens, 2009, 136). To understand a phenomenon is to have explained it. Strevens (2004, 2009)

provides an account of explanation meant to capture this role in understanding. He says, "I take scientific understanding to be that state produced, and only produced, by grasping a true explanation" (2009, 3). Yet, despite this connection between explanation and understanding, Strevens distances his account from the communicative roles of explanation. He instead gives precedence to the ontological sense of explanation: "an explanation [is] something out in the world, a set of facts to be discovered" (2009, 6). In particular, on Strevens' view, explanations are sets of causal facts, for his is a causal account of explanation. The communicative purposes of explanation are taken to be secondary; they are simply attempts to convey those sets of explanatory causal facts.

Strevens (2004, 2009) outlines a two-factor causal account of explanation, according to which explanations are shaped both by a relation of causal influence and a difference-making criterion that governs which causal influences belong in an explanation. Though he also tackles probabilistic explanation and the explanation of laws, in this paper I focus on the core of Strevens' account: explanations of events that (by assumption) result from deterministic causal processes. After outlining Strevens' approach to event explanation in Section 2, I argue in Section 3 that one feature of that account is unmotivated. Strevens' endorsement of what he calls deep standalone explanations is not sufficiently supported by other features of his account nor by independent argument, and it does not cohere well with explanatory practice. Furthermore, this feature of his account is responsible for a disparity between explanations that Strevens and others find intuitively appealing and the explanations that Strevens officially sanctions. For these reasons, I argue that a preference for deep standalone explanations should be rejected in favor of explanations that rely on what Strevens calls explanatory frameworks. In Section 4, I suggest that the difficulty stems from Strevens' view that scientific explanation is an ontological matter. Returning instead to the connection between explanation and the creation of understanding—that "bolt of insight" that lights up our minds—clarifies the value of explanations formulated with particular goals in mind, that

is, within explanatory frameworks. This also brings Strevens' account of event explanation more in line with his views about the explanatory role of patterns.

2 Strevens' Kairetic Account of Event Explanation

Strevens' two-factor account of explanation comprises a metaphysical dependence relation and an explanatory relevance relation. For the former, Strevens employs what he takes to be an ecumenical conception of causal influence. He points out that many accounts of causation—including the conserved quantity, counterfactual, and manipulationist accounts (e.g. Dowe, 2000; Lewis, 1973; Woodward, 2003)—posit a fundamental-level causal relation. Strevens singles out this fundamental-level relation as the metaphysical dependence relation upon which his account of explanation is based. By considering only fundamental-level causal influence, Strevens hopes to avoid the metaphysical issue of levels of causation. However, by singling out fundamental-level relations, he *is* taking a stand regarding the explanatory role of higher-level causation (if such a thing exists). Strevens says, "My ecumenism entails ignoring, for explanatory purposes, the wealth of high-level causal relations offered by multilevel accounts of causation" (2009, 33). On his view, all explanations are grounded in fundamental-level causal influence. Accordingly, all high-level causal claims should be reinterpreted as claims about high-level *explanations*.

The second factor of Strevens' account, explanatory relevance, is provided by the kairetic account of difference-making. This is Strevens' method for ascertaining which aspects of a causal process made a difference to the occurrence of an event. The difference-makers are found by applying an optimizing procedure to a causal model for the event. A causal model for an event entails that event in a way that accurately represents the (fundamental-level) causal processes that led to the event. The optimizing procedure alters that model according to two desiderata: the model is (a) made as abstract as possible while (b) preserving

causal contiguity among the fundamental-level causal processes that realize the model. The desideratum of causal contiguity is satisfied to the degree that the "realizers constitute a contiguous set in causal similarity space" (Strevens, 2009, 104). Since Strevens only credits fundamental-level causal processes with explanatory potential, causal contiguity can be understood as contiguity among the trajectories of the model's various realizations through the state space of fundamental physics.[1] A causal factor that appears in *any* optimized model for an event counts as a difference-maker for that event.

According to Strevens, science aims at what he calls standalone explanations. A standalone explanation of an event is a causal model that preserves fundamental-level causal contiguity and contains only difference-makers for the event. Any standalone explanation is "complete," i.e., sufficient for understanding a phenomenon (2009, 117). Yet every event has multiple standalone explanations. The simplest are "atomic causal models," which are the direct results of optimizing causal models, but standalone explanations can be extended to reference more of an event's difference-makers. Strevens endorses two forms of extension. *Elongation* traces the chain of difference-makers further back in time. Strevens thinks that maximal elongation is "a kind of ideal," yet "an ideal to which those interested only in full understanding need not take the trouble to aspire" (2009, 154). In contrast, *deepening*—showing how laws in the explanation are manifestations of lower-level processes—Strevens declares to be compulsory. Explanations may cite "causal covering laws," viz., high-level laws that simply summarize the microphysical underpinnings, thereby preserving causal contiguity. Yet explanations that cite high-level laws that violate the requirement of causal contiguity are inadequate. Explanations must be deep.

What, then, of high-level explanations that reference multiply realizable properties and functional specifications? These are incohesive at the fundamental level; they lack depth.

[1]Strictly speaking, this is dynamic contiguity, which Strevens claims is necessary but not always sufficient for causal contiguity. Nonetheless, it serves as his proxy for the requisite causal contiguity.

Strevens has a workaround on offer: the explanatory framework. Some explanations cite difference-making relations relative to a background state of affairs. That background serves as the *framework* for the explanation and is exempt from the determination of difference-makers. Consider a simple example. When explaining why someone survived, given that he was poisoned, the poisoning is simply a background state of affairs. The poisoning is not a difference-maker for survival, but neither is it a *non*-difference-maker: it is a background consideration that is held fixed and according to which the difference-makers are established. That the person imbibed an antidote would be a difference-maker relative to the explanatory framework that stipulates he was poisoned.

On Strevens' view, multiply realizable properties and functional specifications may appear in explanations only when supported by explanatory frameworks. One of Strevens' examples is the Lotka-Volterra equation, which represents predator-prey interactions. The Lotka-Volterra equation employs the functionally defined properties of predator and prey, so it can explain population dynamics only relative to a framework stipulating that one population preys on the other. Strevens deems such an explanation acceptable, but inferior to one that does not need a framework. Practical considerations motivate the use of multiply realizable properties and functional specifications, but deep standalone explanations require specifying the causally contiguous fundamental-level realizers. The Lotka-Volterra equation can explain population dynamics given the proper explanatory framework, but it is inferior to a standalone explanation that specifies the nature of predation.

The explanatory framework also enables Strevens to account for the distinction often made between causes and background conditions. Causal influences that an explanation does not focus upon are relegated to the explanatory framework, which allows those influences to be treated as mere background conditions according to which the difference-makers are ascertained. For instance, the presence of oxygen is a causal factor that is generally treated as a precondition for fire rather than a cause of any particular fire. Explanatory frameworks are

also how Strevens accounts for contrastive explanations and explanations that cite omissions and preventions, though I omit those topics for brevity. However, regardless of their practical utility, frameworked explanations are in Strevens' view always inferior to deep standalone explanations. The latter provide full understanding and are the aim of scientific explanation. Strevens acknowledges that "you may have to visit many other university departments, finishing of course with the physics department" (2009, 161) to gain the depth—the relation to fundamental-level causal laws—that they require.

3 Revisiting the Explanatory Framework

Many aspects of Strevens' account deserve further discussion, but I will focus on the notion of explanatory frameworks. Strevens employs explanatory frameworks to accommodate various features of explanatory practice that otherwise would not align with his account. These include explanations that appeal to functional specifications, reference multiply realizable high-level properties, or sideline causal factors that qualify as difference-makers. Yet Strevens considers frameworked explanations to be inferior to deep standalone explanations. Here I will show that Strevens' argument for privileging deep explanations is tenuous, and I will argue that his account would be strengthened by a full endorsement of frameworked explanations.

The value that Strevens places on deep standalone explanations stems entirely from their satisfaction of his cohesion constraint—the requirement that explanations only abstract away from contiguous sets of fundamental-level causal processes. Yet the justification for this cohesion constraint is far from unassailable. Strevens initially adopts the constraint because it eliminates disjunctive explanations, but he mentions in passing that there may be alternate solutions to the disjunction problem (2009, 102). Even if a cohesion constraint is adopted, there is the question of how cohesion should be understood. Strevens' sole reason for

defining cohesion as fundamental-level causal contiguity is that he credits only fundamental-level causal relations with explanatory relevance. But Strevens offers no argument against the existence or explanatory relevance of higher-level causal relationships. Moreover, he expresses doubts about the requirement of fundamental-level causal contiguity (108, 163). It seems Strevens' 'causal ecumenism' is actually rather doctrinaire in the degree to which it weds his account to the fundamental level.

One repercussion is the cohesion constraint's inability to accommodate explanations that cite multiply realizable properties or functional specifications. Strevens acknowledges the appeal of explanations that ignore lower-level mechanisms, but he returns to his worry about the problem of disjunctive explanations:

> I do not see how to formulate a cohesion constraint that is able both to disbar disjunctive models from explanation and yet also to allow black-boxing in a deep—that is, framework-independent—mechanism (2009, 162).[2]

His solution is to allow frameworked explanations to incorporate functional specifications and multiple realizable properties because of their practical utility, while endorsing deep standalone explanations as superior alternatives. Yet Strevens' decision to rely only on fundamental-level causal relations and his tentative use of causal contiguity to solve the disjunction problem are tenuous grounds for this commitment to depth, the titular feature of his account.

Strevens thinks the best hope for allowing black boxes into cohesive explanations is to individuate causal processes according to standards that vary with the level of explanation, but he disallows this because he does not see how to formulate such a standard (2009, 163). As I see it, one approach would be to allow higher-level causal relations to play an explanatory role, which would enable causal contiguity to be cashed out in a level-relative way. But

[2]For a discussion of the disjunction problem, see (Cover and Curd, 1998, 786) and (Strevens, 2009, §3.45).

Strevens almost certainly would be dissatisfied with this suggestion.[3] A less-invasive approach would simply be to fully sanction frameworked explanations. The framework notion could still be employed to distinguish beneficial black-boxing from problematically disjunctive explanations. The distinction would rest on which causal influences were treated as part of the explanation and which were relegated to the framework, just as how Strevens uses explanatory frameworks to distinguish causes from background conditions.[4] This is the suggestion that I defend in the present paper.

Let us reevaluate the relative merits of deep standalone explanations and frameworked explanations. I agree with Strevens that "the most important source of evidence concerning our explanatory practice is the sum total of the explanations regarded as scientifically adequate" (2009, 37). On Strevens' view, deep standalone explanations cite higher-level properties if and only if they represent causally contiguous fundamental-level causal processes, whereas frameworked explanations freely appeal to multiply realizable properties, substitute functional specifications for lower-level causal dynamics, and/or willfully emphasize some factors while relegating others to the role of background conditions. If deep explanations are superior, we should expect the prevalence of scientific explanations that eschew these practices in favor of causal contiguity.

I see no evidence of this tendency. Multiply realizable properties and functional specifications are overwhelmingly common in science, and though effort is sometimes put toward understanding their range and means of application, they are seldom eliminated. Instead, as Strevens notes (160), black-boxing occurs without any regard for contiguity among the fundamental-level realizers. For example, any number of properties central to evolutionary explanations exemplify this: heritability, fitness, sexual reproduction, gene, male, aggressive, cooperative, camouflage... the list continues ad infinitum. Such properties

[3]As discussed above, Strevens believes all high-level causal claims are in fact claims about high-level explanations, explanations that succeed in virtue of fundamental-level causal relations.

[4]This is similar to the approach to the disjunction problem that I suggest in (Potochnik, 2010b).

are employed whenever expedient, taking into account only feasibility and the parts of the causal process that are of immediate interest. The explanations that appeal to such properties all qualify as frameworked on Strevens' account.

Indeed, most of Strevens' own examples are frameworked explanations. Explanations of predator-prey dynamics that employ the Lotka-Volterra equation do not qualify as deep, for the Lotka-Volterra equation itself is multiply realizable. An explanation of Rasputin's death that cites the conspirators' decision to murder him is, according to Strevens, right to black-box the means of assassination, for these are "effectively irrelevant" (Strevens, 2009, 169)—though this also results in a frameworked explanation.

Without a strong argument for the superiority of deep standalone explanations, and without any indication that explanatory practice conforms to that norm, there is no reason to privilege explanatory depth over frameworked explanations. Furthermore, the endorsement of frameworked explanations would resolve some difficulties facing Strevens' account of explanation. Most immediately, this would allow explanations—the *best* explanations—to employ multiply realizable properties and functional specifications. Many have argued that the strength of high-level explanations is their ability to group phenomena that behave similarly at a high-level of description, despite fundamental-level dissimilarities (Fodor, 1974; Putnam, 1975; Garfinkel, 1981; Kitcher, 1984; Jackson and Pettit, 1992; Sober, 1999). Multiply realizable properties are responsible for that strength. Strevens himself notes the explanatory value of multiply realizable properties; his only worry is that he cannot see how to allow those explanations without relying on frameworks (2009, 162). Granting frameworked explanations the status of full-fledged explanations sidesteps that problem.

Another difficulty facing Strevens' account that would be resolved by fully endorsing frameworked explanations regards tradeoffs. Explanations that cite multiply realizable properties trade off some cohesion for greater generality. Strevens cautiously endorses a different type of tradeoff: explanations that sacrifice a small amount of *accuracy* for

greater generality. His example is an explanation of Mars' exact orbit that leaves out the gravitational effect of the other planets. That gravitational effect is technically a difference-maker, but it has such slight effect that Strevens suggests omitting it for the sake of a more general explanation. This move accommodates the intuition that an explanation can neglect small causal factors when this allows "a more abstract scheme of description that better captures the high-level difference-making structure" (2009, 146-7).

Here Strevens seems to bite a bullet. In order to sanction this move, he breaks the connection between difference-making and explanatory relevance that is at the core of his account. He says, "the difference-makers dropped in the course of an accuracy tradeoff, then, literally make no contribution to our understanding of the explanandum" (2009, 146). Beyond the difficulties this creates for Strevens' account, it is also counterintuitive. It may be consistent with both intuition and explanatory practice for explanations to neglect some factors in order to depict high-level difference-makers. The problem is the idea that this *always* should be done. If small difference-makers "literally make no contribution to our understanding" of an event, then explanations err if they reference those factors. Yet factors of small effect sometimes can be explanatorily important. Surely there are situations in which the other planets' gravitational effects belong in the explanation of Mars' exact orbit.

Endorsing frameworked explanations would offer an alternate solution. Strevens shows how causal factors that are relegated to the explanatory framework are sometimes black-boxed. I suggest that they also can be ignored in other ways, such as being replaced by a general ceteris paribus clause or simply neglected entirely. Relegating those factors to the framework enables an explanation to neglect them in order to portray higher-level influences. However, causal factors are frameworked only according to the interests of the explanation-seekers. For instance, if one asks why Mars has its exact orbit, given that the sun has a particular gravitational influence, this re-frameworking puts the planets' gravitational influence at center stage.

Trading off accuracy for generality across the board is as counterintuitive as is devaluing multiply realizable explanations across the board. Various tradeoffs of accuracy and of cohesion to accomplish high-level explanations are found in actual scientific practice, but the degree to which each is employed varies. Whereas biologists will oftentimes sacrifice cohesion to employ the Lotka-Volterra equation's account of predator-prey dynamics in a community, occasionally they may seek the details of how those dynamics are realized over some period of time in that particular community. And whereas many accounts of Mars' orbit will sacrifice a degree of accuracy by omitting the gravitational influence of other planets, other accounts will instead highlight those very gravitational perturbations.[5] Strevens' innovation of the explanatory framework gracefully accommodates this feature of explanatory practice. The only obstacle is the insistence that frameworked explanations are lacking—inferior to deep standalone explanations. I have argued that there are no clear grounds for that claim.

4 The Role of Patterns in Explanatory Practice

I suspect that an early move in Strevens' approach to explanation is responsible for his commitment to deep standalone explanations. Recall that Strevens focuses on the ontological sense of explanation, according to which an explanation is a set of facts out in the world. He seems to assume that there is just one such set of facts for each event: "what explains a given phenomenon is a set of causal facts... communicative acts that we call explanations are attempts to convey some part of this explanatory causal information" (Strevens, 2009, 6). So Strevens searches for the single, best explanation of a given event. That search results in the type of explanation that he fully endorses: deep standalone explanations, which are purportedly sufficient for full understanding. Frameworked explanations are only partial explanations, influenced by immediate interests. They are thus poor candidates for good

[5]See (Potochnik, 2010a) for an extended case study of this variation for evolutionary explanations.

explanations in the ontological sense.

Returning to the connection between explanation and the creation of understanding leads to a different endpoint. Strevens notes how explanations should "light up our minds with a bolt of insight" (2009, 136). Accounting for this is aided by considering explanations' role in communication.[6] Explanations that yield understanding are those explanations that are actually formulated, that are found to be useful, that provide bolts of insight. In a world where multiple realization and complex causal interactions abound, those explanations often are improved by the neglect of many causal factors. This is why the ideal of exhaustive explanations that cite all difference-makers strikes even Strevens as exhausting (154), and it is also why explanatory frameworks are so valuable. Frameworks enable some causal factors to be sidelined to the end of understanding the influence of other factors.

Further support is provided by some of Strevens' other views. Strevens amends his account of event explanation when he discusses explanations of laws. There he suggests that explanations of events are improved by citing a pattern of *entanglement*—roughly, a general pattern of the co-occurance of properties. The details are complex, so I will suppress them here and simply provide an example. To explain why a sodium sample reacted when it came in contact with water, one should specify that sodium is an alkali metal; that alkali metals have loosely bound outer electrons; and that this loose binding enabled contact with water to remove electrons from the sodium atoms. The pattern of entanglement is the connection between being an alkali metal and having loosely bound outer electrons. The property of being sodium is also entangled with the property of having loosely bound outer electrons, but Strevens thinks that the pattern holding of all alkali metals is more explanatory because it is a more general entanglement. This idea fits with Strevens' view that explanatory practice is shaped "so as to direct our attention to the world's causal patterns" (2009, 264).

[6]Communicative uses are generally deemed inessential to the nature of scientific explanations; the views of Bromberger (1966) and Achinstein (1983) are exceptions.

This addition to Strevens' account further clarifies the value of frameworked explanations. Consider once more the example of explaining predator-prey dynamics with the Lotka-Volterra equation. This explanation relies on a framework, since it employs the functionally defined properties of predator and prey. According to Strevens' un-amended account, full understanding requires eliminating the explanatory framework, viz., supplying the fundamental-level causal details that realize the predator-prey relationship in a particular community. And since the causal details of the predator-prey relationship differ from community to community, "there is no single explanation of stability across the ecosystems in question" (2009, 159). The amended account is markedly different from this. It grants explanatory value to general patterns via the notion of entanglement, and those patterns may not be cohesive in the sense of fundamental-level causal contiguity.[7] Taking frameworked explanations as full-fledged explanations allows the range of applicability of the Lotka-Volterra equation to emerge as a startling revelation of pattern. There is a bolt of insight when one grasps that, *regardless of implementation*, the Lotka-Volterra equation captures a broad ecological pattern. Despite myriad differences among communities, their predator-prey relationships are all captured by a single mathematical relationship.

The same revelation of pattern occurs when a trait is explained as the outcome of natural selection pressures, while the trait's inheritance is only functionally specified (Rosales, 2005; Potochnik, 2009) , and when the more-or-less independent assortment of genes is explained as the result of chromosomal alignment during meiosis, while the molecular details are ignored (Kitcher, 1984). Though Strevens deems these frameworked explanations inferior to deep standalone explanations, they are valuable for their ability to direct attention to a broad causal pattern in the world—an explanatory goal endorsed by Strevens. Frameworked explanations are thus well-positioned to create understanding in virtue of, not despite, their

[7]Strevens does require that patterns of entanglement be cohesive, but in a different sense. A pattern of entanglement is cohesive "to the degree that there is a single reason for the entanglement" (2009, 255).

neglect of some difference-makers.

Acknowledgements

Michael Strevens has given me a good deal of helpful feedback on these ideas and the paper itself. Alistair Isaac, Teru Miyake, Joel Velasco, and two anonymous referees also provided helpful comments on earlier drafts of the paper.

References

Achinstein, P. [1983]: *The Nature of Explanation*, Oxford, UK: Oxford University Press.

Bromberger, S. [1966]: 'Why-questions', in R. Colodny (ed.), *Mind and Cosmos*, Pittsburgh: University of Pittsburgh Press, pp. 86–111.

Cover, J. A. and Curd, M. (eds.) [1998]: *Philosophy of Science: The Central Issues*, New York: W.W. Norton and Company.

Dowe, P. [2000]: *Physical causation*, Cambridge University Press.

Fodor, J. [1974]: 'Special sciences: The disunity of science as a working hypothesis', *Synthese*, **28**, pp. 97–115.

Garfinkel, A. [1981]: *Forms of Explanation: Rethinking the Questions in Social Theory*, New Haven: Yale University Press.

Jackson, F. and Pettit, P. [1992]: 'In defense of explanatory ecumenism', *Economics and Philosophy*, **8**, pp. 1–21.

Kitcher, P. [1984]: '1953 and all that: A tale of two sciences', *Philosophical Review*, **93**, pp. 335–373.

Lewis, D. [1973]: 'Causation', *Journal of Philosophy*, **70**, pp. 556–567.

Potochnik, A. [2009]: 'Optimality modeling in a suboptimal world', *Biology and Philosophy*, **24**(2), pp. 183–197.

Potochnik, A. [2010a]: 'Explanatory independence and epistemic interdependence: A case study of the optimality approach', *The British Journal for the Philosophy of Science*, **61**(1), pp. 213–233.

Potochnik, A. [2010b]: 'Levels of explanation reconceived', *Philosophy of Science*, **77**(1), pp. 59–72.

Putnam, H. [1975]: *Philosophy and our Mental Life*, volume 2 of *Philosophical Papers*, chapter 14, Cambridge, UK: Cambridge University Press, pp. 291–303.

Rosales, A. [2005]: 'John Maynard Smith and the natural philosophy of adaptation', *Biology and Philosophy*, **20**(5), pp. 1027–1040.

Sober, E. [1999]: 'The multiple realizability argument against reduction', *Philosophy of Science*, **66**, pp. 542–564.

Strevens, M. [2004]: 'The causal and unification accounts of explanation unified—causally', *Nous*, **38**, pp. 154–179.

Strevens, M. [2009]: *Depth: An Account of Scientific Explanation*, Cambridge, MA: Harvard University Press.

Woodward, J. [2003]: *Making Things Happen: A Theory of Causal Explanation*, Oxford, UK: Oxford University Press.