1 March 2012

Causal relations and explanatory strategies in physics

Andrew Wayne
Draft – please do not cite

Word count: 4834 words

Abstract

Many philosophers now regard causal approaches to explanation as highly promising, even in physics. This is due in large part to James Woodward's influential argument that a wide range of explanations (including explanations in physics) are causal, based on his interventionist approach to causation. This article focuses on explanations, widespread in physics, involving highly idealized models. These explanations are not causal, yet they do not fall under any of the types of non-causal explanation Woodward describes. I argue that causal explanation is simply not as widespread or important in physics as Woodward and others maintain.

1. Introduction

Many philosophers now regard causal approaches to explanation as highly promising, even in physics. In part this is because the major alternative, deductivist approaches to explanation, have fallen on hard times (Hempel 1965; Kitcher 1989). Problems of explanatory irrelevance and explanatory asymmetry (recall hexing spells and flagpoles) have motivated many to pay more attention to the role of causation in explanation. Preeminent among recent work on causal explanation is James Woodward's influential argument that a wide range of explanations, including explanations in physics, are causal explanations, based on his interventionist approach to causation (Woodward 2003; Woodward 2007). After reviewing Woodward's approach (Section 2), this paper argues that causal relations are insufficient for explanation because they do not account for the key feature of explanatory integration in physics (Section 3). Further, causal relations are unnecessary for explanations, widespread in physics, involving highly idealized models. These explanations are not causal, yet they do not fall under any of the types of non-causal explanation Woodward describes (Section 4). This constitutes a significant limitation on the scope of causal explanation in

physics that neither Woodward nor any other proponent of causal explanation has recognized. Causal explanation is simply not as widespread or important in physics as Woodward and others—such as Wesley Salmon, Phil Dowe and Michael Strevens—maintain (Salmon 1984; Dowe 2000; Strevens 2008).

2. Woodward on causal explanation

For Woodward, causal relations are captured in counterfactual claims about what would happen to an effect Y if an intervention on another variable (or set of variables) *X* were to occur. Causal explanations in turn appeal to these "interventionist" counterfactual dependencies. Woodward is clear that his account of causation is non-reductive, in the sense that it does not aim to give an account of causation exclusively in non-causal terms. Explanation is also non-reductive, for Woodward. He allows that not all causal explanations need be in terms of fundamental physics, and indeed that fundamental physics is an area in which explanations seem to be predominantly non-causal. He emphasizes that macro causal claims can often be more explanatory than causal claims about their micro realizers, and that these macro causal claims can be explanatory while offering only an approximate description of the relevant features of the target physical system.

Consider an explanandum consisting of the statement that some variable *Y* takes a particular value. For Woodward,

(1) [A] successful [causal] explanation will involve a generalization G [in the explanans] and explanans variable(s) X such that G correctly describes how the value of Y would change under interventions that produce a range of different values of X in different background circumstances (2003, 203).

What makes the causal generalization G explanatory is that it can answer a relevant range of "what-if-things-had-been-different" questions, and it does this by supporting the correct counterfactuals about what would happen under scientifically relevant interventions on the explanans variable X. To do this, G must be invariant (roughly, describe the same sort of dependence of Y on the X) under the relevant range of interventions and in a range of relevant background conditions. Unlike deductivist approaches, successful explanations are not just nomologically sufficient, that is, they cannot just subsume the explanandum under a regularity and thereby show it is to be expected given the truth of the statements in the explanans. Rather, they must also describe relevant dependency relations—they must show how this explanandum would change if the intervention or background conditions were to change. Explanation locates the explanandum within a space of relevant alternative possible explananda.

We have seen that on Woodward's account, causal explanation requires counterfactuals describing possible interventions and possible covariation in changes in the values of

variables, and a notion of scientifically relevant possibility guiding the selection of interventions, dependencies and alternative possible explananda. The other key component of his account, of course, is an account of causal relations, including the cause-effect relation between variable X in the explanans and Y in the explanandum. For Woodward, if some intervention on X produces a change in the value of Y, then X is a token direct cause of Y. Roughly speaking,

(2)   An *intervention I* is a hypothetical experimental manipulation on X such that,
       (i) *I* causes X,
       (ii) *I* changes the value of X in such a way that the value of X does not depend on the values of any other variables that cause X, and
       (iii) *I* changes the value of X in such a way that if any change occurs in the value of Y, it occurs only as the result of the change in X and not from some other source.

(See Woodward 2003, 98-107 for a more detailed account.) Woodward's notion of intervention is not limited to what humans can actually do with physical systems. Rather, it is defined in terms of possible or hypothetical manipulations of values of variables within a model.

Woodward rightly emphasizes that only some changes in the explanans and only some contrasts between the explanandum and its alternatives are of causal and hence explanatory relevance. As he puts it, "It is also true that if a large meteor had struck my office just as I was typing these words, I would not have typed them, but again, we are reluctant to accept the failure of the meteor to strike as part of the explanation for my writing what I did" (2003, 226). The problem here is not that causal omissions can never figure in genuine explanations—Woodward is clear that sometimes they can—but rather that in this context a meteor intervention is not what Woodward dubs a "serious possibility." Scientists approach empirical phenomena with a large stock of shared beliefs about which of the interventions or dependency relations are potentially causally and explanatorily relevant, and which alternatives to the explanandum are relevant as well. Woodward is clear that what counts as a causal factor is relative to a particular choice of variables and also to a particular range of values of these variables (Woodward 2003, 55-56). Different models—in Woodward's terms, different sets of structural equations, variables and directed graphs—result in a different set of causes and hence a different explanation.

So far explanation, causation and intervention have been defined in terms of statements about variables, values and dependency relations within a model. But not every transformation or modification one can perform on a model corresponds to a hypothetical manipulation on the physical system itself (in Woodward's sense), and only those that do so correspond can underwrite causal claims. Causation requires that the values and dependency relations of variables in the model represent physical features of the target system. As Woodward puts it, successful causal explanation requires that the statements (about

counterfactuals, dependency relations, values of variables, causal relations and so on) in the explanandum and in the explanans be true or approximately true of the target system (2003, 203). Without the truth or approximate truth of the explanandum, it fails to be an explanation of any physical phenomenon at all. Without the truth or approximate truth of the explanans, the statements about the model simply cannot describe any real causal relations in the target system.

For example, the period of a pendulum may be approximately derived and explained in terms of its length, in a fixed gravitational field, by appealing to counterfactual claims about the behaviour of an idealized pendulum model satisfying Galileo's pendulum law. The law states that the period of a pendulum is proportional the square root of its length:

(3) $T \alpha \sqrt{l}$

The relevant counterfactual claim is: if the length l were increased to l^*, in a fixed gravitational field, then the period T of the model pendulum would have increased to T^*, in accordance with (3). However, the model does *not* support an explanation of the length of the pendulum in terms of its period, because the relevant interventionist counterfactual is false of the model: it is false that if the period were increased to T^*—for instance by moving the pendulum to a weaker gravitational field—the length of the pendulum would have changed. Woodward uses this example to illustrate how his causal model of explanation solves the problem of explanatory asymmetry that bedevils deductivist approaches (2003, 197). For our purposes, the important point is that the interventionist counterfactual doing the explanatory work (and described in the explanans) is true of the model and is also approximately true of the target system. For Woodward, the fact that the dependency relations in the model approximate "what the real dependency relations in the world actually are" is fundamental to his account of causal explanation (Woodward 2003, 202).

3. Causal relations are insufficient for explanation

I contend that a consequence of Woodward's account is that causal relations are insufficient for explanation in physics, and in two steps. First, some causal derivations fail to be explanatory. They may satisfy (1) and (2) above, and they may have significant predictive or heuristic value, but they do not explain. Second, where a causal derivation is explanatory, it is never merely by virtue of satisfying (1) and (2); rather, explanation requires that the causal story be integrated with a global model of broad scope and explanatory power.

According to Woodward, what makes the causal generalization G in (1) explanatory is that it answers "what-if-things-had-been-different" questions, and it does this by supporting the correct counterfactuals about what would happen under interventions. Consider Woodward's example of the explanation of the period of a pendulum, but this time prior to Galileo's theoretical advances. Taking liberties with the actual historical order of events,

imagine (counterfactually) that Galileo had conducted his years of painstaking experimental observations of pendulums first, in advance of any other work on his new science of mechanics. Had he arrived at his pendulum law (3) and his idealized pendulum model this way, we would be inclined to say that his argument deriving the period of a pendulum is not explanatory. The pendulum model on its own supports a relevant and approximately correct set of counterfactual claims about interventions on a physical pendulum. Nonetheless, it would be merely a phenomenological or data model, as contemporary physicists would put it. It fits a given set of data well, and it may describe the correct dependency relations in an isolated model, but fails to connect with other, more global models. These sorts of models may have predictive and heuristic power, but they do not underwrite explanations in physics.

Unfortunately, Woodward's account yields the result that many phenomenological models do come out as explanatory, and this cannot be right. Woodward posits a base threshold of explanatoriness, above which stands a continuum running from less deep or good explanations to deeper and better ones (2003, 368). The worry is that (1) and (2) set the threshold very low indeed: generalizations that are invariant under any intervention at all exceed the threshold because they answer a "what-if-things-had-been-different" question (2003, 369). So Woodward would certainly view the counterfactual Galileo's standalone pendulum model as underwriting a bona fide explanation of the period of the pendulum. But we have good reason to maintain that it does not, nor do the plethora of other phenomenological models in physics that capture some of the dependency relations in their target physical systems.

As a matter of historical fact, the pendulum law is significant for Galileo precisely because it is a key step in his route to the fundamental laws of his new science of mechanics. Galileo measured the elapsed time of an object's vertical fall over a distance equal to the length of the pendulum, for various pendulum lengths (Drake 1989, xxvii). He obtained a constant ratio of free-fall times to time for the pendulum to swing to vertical. With the pendulum law and that ratio, Galileo could calculate the times for other distances of free-fall and then, removing pendulums entirely from the calculation, write down his famous law of motion: that all objects fall at the same rate, regardless of their composition or mass, and that objects starting at rest accelerate uniformly as they fall, i.e. their speed is proportional to the square of the elapsed time of fall. He found the law fit well his previous measurements of descents along inclined planes.

This suggests that the idealized model pendulum gets its explanatory power by its integration into Galileo's new science of mechanics. In this case, it is integration of a particularly simple sort: Galileo took his pendulum law to follow from his more general law of free fall, and the idealized model pendulum is simply a special case of a more general model covering falling objects in general. Newton's subsequent achievement was greatly to increase this integration by explaining the motions of bodies in terms of the forces acting on

them and providing a unified framework for all gravitational systems. The important point for our purposes is that it is not sufficient that the idealized pendulum model approximate the correct dependency relations in a physical pendulum for it to be explanatory.

Woodward does say that successful causal explanation must include *relevant* dependency relations and answer a *relevant* range of "what-if-things-had-been-different" questions, and that scientists share an understanding of which interventions and which dependency relations are *explanatorily relevant*. Woodward seems to recognize that merely describing local causal relations is not sufficient for explanation, while perhaps not fully appreciating the consequences for the role of causation in explanation. The challenge is not to rule out an explanatory role for the absence of falling meteors. Rather, the challenge is to underwrite the explanatory role of dependency relations in the local pendulum model. And this can be done only in the context of a wider integration with a global model in physics—here Galilean (or even better Newtonian) mechanics.

The point is not just that some causal derivations satisfying (1) and (2) fail to be explanatory, as in the contrary-to-historical-fact Galilean account of the pendulum. It is also that no causal derivation is explanatory merely by virtue of satisfying (1) and (2). This is because what makes the dependency relations described in the explanans relevant (*i.e.*, explanatorily relevant) is the integration of the local model described in the explanans with a global model of broad scope and explanatory power. Without such integration, the local model will generally fail to be explanatory, no matter how accurately it represents causal relations in the target physical system. And as we shall now see, with such integration the local model will generally be explanatory—even if it fails to represents any causal relations in the target physical system.

4. Causal relations are unnecessary for explanation

Woodward allows that not all explanations in physics need be causal and notes that fundamental physics is an area in which explanations seem to be predominantly non-causal. What Woodward has in mind, in these and other sorts of physics explanations he calls non-causal, are cases in which the notion of an intervention on a physical system is incoherent or inapplicable. This includes global applications of fundamental physics to the whole universe or to large portions of it, where the notion of a local intervention is inapplicable (2007, 91); explanations that appeal to alternative situations not plausibly characterized as an intervention, e.g., altering the dimensionality of space-time (2003, 220); and situations that lack the invariance or stability properties needed to define an intervention on the system (2007, 77). These sorts of cases, however, are merely the tip of a very large iceberg of non-causal explanation in physics.

The issue is that, aside from explanations in textbooks (from which Woodward's examples seem to be drawn), much of the explanatory practice in physics does not fit

Woodward's characterization. These are cases in which the idealized models that underwrite putative explanations are largely non-representative of target physical systems. So while they approximately model the explanandum behaviour, they do not approximate aspects of the physical system described in the explanans. Moreover, these models are not corrigible, in the sense that they cannot be refined in a theoretically justified, non-ad hoc way to bring them in closer agreement with the target system. The point is that these are cases of explanation in which physicists view the scientifically relevant claims about interventions and systematic patterns of dependency relations that figure in a potential explanans to be statements about a highly idealized model, statements that are not even approximately true of the target system containing the phenomenon to be explained. If the explanatory practice of contemporary physics is taken seriously, there are highly idealized models of significant explanatory value.

Valuable work has been done by philosophers of physics on the possible explanatory roles of highly idealized models (Rueger 2001; Batterman 2002; Bokulich 2008; Batterman 2010; Bokulich 2011). Alisa Bokulich, for instance, has argued that "fictional models" can be explanatory if they meet certain conditions. Bokulich focuses on semi-classical models, which mix classical and quantum features. These models are known not to represent successfully the physical system because, for example, they include quantum particles following definite classical trajectories. The earliest and most well-known of these models is Niels Bohr's model of the hydrogen atom. As Bokulich puts it, "I want to defend the view that despite being a fiction, Bohr's model of the atom does in fact explain the spectrum of hydrogen" (Bokulich 2011, 42). Robert Batterman is interested in how highly idealized models explain the universality of structural features, such as the common characteristic shape of droplets at breakup when water drops fall from a dripping faucet.

> We can explain and understand (for large scales) why a given drop shape at breakup occurs and why it is to be expected. The answer depends essentially upon an appeal to the existence of a genuine singularity developing in the equations of motion in a finite time. It is because of this singularity that there is a decoupling of the breakup behaviour (characterized by the scaling solution) from the larger length scales such as those of the faucet diameter. Without a singularity, there is no scaling or similarity solution. Thus, the virtue of the hydrodynamic singularity is that it allows for the explanation of such universal behaviour. The very break-down of the continuum equations enables us to provide an explanation of universality (Batterman 2009, 442-443).

Asymptotic analyses that systematically abstract away from micro details enable idealized models to explain underlying structural or universal features. Batterman calls these "asymptotic explanations" (Batterman 2002, Ch. 4).

One option for Woodward and other proponents of causal explanation is simply to reject any role for highly idealized models in explanation. These are putative explanations that fail to meet Woodward's requirement for causal explanation, nor do they fall under his class of

non-causal explanations in physics. These models are simply highly inaccurate representations of the physical world. One could argue that highly idealized asymptotic and semi-classical models have great heuristic and predictive value, but do not underwrite explanations. They can play no part in underwriting the true causal premises needed in an acceptable explanation. In my view, this kind of wholesale rejection of any role for highly idealized models in explanation would be a mistake. A closer look reveals a more nuanced and complex set of considerations.

In the case of the Bohr model and other semi-classical models, there is no consensus among physicists that these models are explanatory, and rightly so. Clearly, their explanatory merits need to be examined on a case-by-case basis. At the very least, we have good reason to be skeptical that the Bohr model of the atom has any explanatory value, especially in light of the quite impressive explanations of the hydrogen spectrum given in terms of relativistic quantum theory.

The situation with respect to asymptotic models is somewhat different. On the one hand, a case can be made that at least one of these models may be eliminated (in principle at least) in scientific explanation (Redhead 2004; Belot 2005). On the other hand, these sorts of models are used widely and are regarded as underwriting among the best explanations on offer in physics today. In addition to analyzing the use of asymptotic models to explain drop formation in hydrodynamics, Batterman has explored the use of asymptotic models to explain critical phenomena in thermodynamics and to explain the rainbow in catastrophe optics (Batterman 2002). Similar sorts of highly idealized, asymptotic models are accepted as explanatory in many areas of physics beyond those that are the focus of Batterman (and his critics). For instance, these sorts of models are taken to underwrite explanations of a wide variety of non-linear dynamical systems, from a damped, driven oscillator model of the human heart to gravitational waves ([self-reference omitted]).

The gravitational waves case is particularly interesting. Physicists take themselves to have explained gravitational waves using Einstein's General Theory of Relativity (GTR). However, even in the simplest models of binary systems that produce gravitational waves, the Einstein Field Equations (the equations of GTR) cannot be solved directly. The reason is that these are a set of coupled, nonlinear equations governing the relation between the distribution of matter and energy in the universe and the curvature of space-time (of which gravitational waves are one feature). An attempt to solve the Einstein Field Equations directly by applying regular perturbation methods results in divergences (infinities) in values for the properties of gravitational waves observable from earth. So physics takes what is by now a familiar strategy: replace the intractable original problem with a tractable one, called the post-Newtonian approximation, that makes essential use of singular perturbation theory and asymptotic models. The empirical results are predictions and explanations of gravitational wave phenomena. These phenomena have not been observed (at the time of

writing), but a handful of large gravitational-wave detectors should soon reach sensitivities high enough for direct detection of gravitational waves (Pitkin, Reid et al. 2011; [self-reference omitted]).

We have good reason to accept, at least provisionally, explanations in physics based on highly idealized models. However, I am not claiming to have presented a conclusive argument for doing so. Obviously, much work remains to be done. Further analysis of the details of Bokulich's and Batterman's examples is needed, and vastly more cases of putative explanation via highly idealized models in physics need to be examined in detail. The question that needs to be asked of each case is: does explanation of a phenomenon ineliminably require appeal to a highly idealized model in this case? Nor am I claiming that "model explanation" or "asymptotic explanation" are adequate normative accounts of explanation in physics that can underwrite this sort of explanatory practice. Rather, I am claiming that philosophers have good reasons to take seriously the fact that the explanatory practice of physics includes a large class of explanations based on highly idealized models, explanations that are clearly not causal on Woodward's (nor any other plausible) account. I should also note that rejecting these sorts of cases wholesale as explanatory failures has as a consequence that physicists are massively mistaken about the explanatory merits of their theories and about the scope of their understanding of the natural world. This runs counter to Woodward's own project of offering an account of explanation that has normative and descriptive elements in reflective equilibrium, an account "significantly constrained by prior usage, practice and paradigmatic examples" (2003, 8).

The best option is to accept these sorts of cases as explanatory and recognize that the explanations fall outside the scope of *causal* explanation in physics. We have seen how Woodward allows that explanations in physics may be noncausal where the notion of an intervention is incoherent or inapplicable. Explanations appealing to highly idealized models constitute a new way in which the notion of an intervention is inapplicable. In these explanations, the correct counterfactual dependencies between I, X and Y may well obtain such that Woodward's conditions (2)(ii) and (2)(iii) are satisfied. In other words, these cases fit very well Woodward's central idea that explanations include statements of counterfactual dependencies describing the results of a hypothetical manipulation of variables in a model. However, the explanation is not causal because (i) is surely false: I does not cause X, because the dependency relations in the model do not correspond to or represent—even in an approximative way—physical dependency relations in the target system. Choosing this option is to acknowledge that there is a distinct, large and important class of non-causal explanations that have not been recognized by Woodward, nor, I suggest, by other proponents of causal explanation in physics.

5. Conclusion

Recall that for Woodward, the notion of an intervention plays the crucial roles of underpinning both the *truth* and *explanatory relevance* of generalization G in the explanans of a successful causal explanation (1). In the context of physics, I have argued, "intervention" is simply not the right concept to play these roles. Even in cases where the notion of an intervention is coherent and applicable, it is not sufficient to meet the threshold of genuine explanatoriness in physics. As we have seen, what makes the dependency relations described in the explanans explanatorily relevant is the integration of the local model described in the explanans with a global model of broad scope and explanatory power. In other cases the notion of intervention is wholly unnecessary to underpin the truth of G, because G can be made true by facts about dependency relations in a model. These dependency relations are clearly not causal, because they are features of an idealized model that do not accurately represent corresponding features of the physical world.

Among the many virtues of Woodward's account of explanation are that it is explicitly model-based and that it makes explanation trace systematic patterns of dependencies rather than simply describing nomologically sufficient conditions. However, the argument given above that much successful explanation in physics involves highly idealized models counters Woodward's claim that many (non-fundamental) explanations in physics are causal. I suggest that the argument against Woodward's causal account tells equally strongly against other prominent defences of causal explanation in physics (e.g., Salmon 1984; Dowe 2000; Strevens 2008). There is good reason to believe that outside of textbook presentations, causal explanation is not as widespread in physics as its proponents have claimed. This point likely generalizes to other areas of science in which complex non-linear dynamical systems are modeled, such as biology and chemistry. These areas seem to have the same sorts of non-reductive explanations appealing to highly idealized, partially non-representative models. If this is right, causal concepts are not as useful in scientific explanation as many philosophers currently believe, and certainly causal theories of explanation are not as successful as the current consensus holds. Perhaps deductivist approaches to explanation merit renewed interest.

References

Batterman, Robert W. 2002. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford: Oxford University Press.

---. 2009. "Idealization and Modeling." *Synthese* **169**: 427-446.

---. 2010. "On the Explanatory Role of Mathematics in Empirical Science." *British Journal for the Philosophy of Science* **6**(1): 1-25.

Bokulich, Alisa. 2008. *Reexamining the Relationship between Classical and Quantum Mechanics: Beyond Reductionism and Pluralism*. Cambridge: Cambridge University Press.

---. 2011. "How Scientific Models Can Explain." *Synthese* **180**: 33-45.

Dowe, Philip. 2000. *Physical Causation*. Cambridge: Cambridge University Press.

Drake, Stillman. 1989. "Introduction." *Two New Sciences, Including Centers of Gravity and Force of Percussion*. Stillman Drake. Toronto: Wall & Thompson**:** i-xxxv.

Hempel, Carl Gustav. 1965. *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science*. New York: Free Press.

Kitcher, Philip. 1989. "Explanatory Unification and the Causal Structure of the World." *Minnesota Studies in the Philosophy of Science, volume XIII*. Philip Kitcher and Wesley C. Salmon. Minneapolis: University of Minnesota Press**:** 410-506.

Pitkin, Matthew, Stuart Reid, et al. (2011) "Gravitational Wave Detection by Interferometry (Ground and Space)." Living Revies of Relativity **14**.

Redhead, Michael. 2004. "Asymptotic Reasoning." *Studies in History and Philosophy of Modern Physics, vol. 35, pt. B, no. 3, pp*: 527-530.

Rueger, Alexander. 2001. "Explanations at Multiple Levels." *Minds and Machines* **11**(4): 503-520.

Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton, N.J.: Princeton University Press.

Strevens, Michael. 2008. *Depth : An Account of Scientific Explanation*. Cambridge, Mass.: Harvard University Press.

Woodward, James. 2003. *Making Things Happen : A Theory of Causal Explanation*. Oxford: Oxford University Press.

---. 2007. "Causation with a Human Face." *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Huw Price and Richard Corry. Oxford: Clarendon Press**:** 66-105.