Causal Theories of Explanation and the Challenge of Explanatory Disagreement

Lina Jansson*†

When evaluating the success of causal theories of explanation the focus has typically been on the legitimacy of causal relations and on putative examples of explanations that we cannot capture in causal terms. Here I motivate the existence of a third kind of problem: the difficulty of accounting for explanatory disputes. Moreover, I argue that this problem remains even if the first two are settled and that it threatens to undercut one of the central motivations for causal accounts of explanation, namely, the causal account of the directionality of scientific explanation.

1. Introduction. Woodward (2003) and Strevens (2008) offer two of the most prominent accounts of scientific explanation to appear in recent years. They share the feature that they are both, at basis, causal accounts of explanation. Further, Woodward and Strevens both develop these accounts independently of any specific metaphysical account of causation and, indeed, stay largely neutral on what such an account might look like.

Causal accounts such as these are very attractive for two reasons. First, the flexibility in the causal relations that they use allows them to account for a great range of cases of scientific explanation. For example, Woodward (2003, 6) is willing to count even equilibrium explanations as causal. Second, by invoking merely general features of causal relations, they solve many

Received October 2013; revised January 2014.

*To contact the author, please write to: Nanyang Technological University, Philosophy Group, 14 Nanyang Drive, Singapore 637332; e-mail: idajansson@ntu.edu.sg.

† I owe a great deal to many graduate students, faculty members, and visiting speakers at the University of Michigan for helpful comments, suggestions, and criticisms. A particular thanks to Gordon Belot, Laura Ruetsche, and Larry Sklar. Thank you also to the participants in the Michigan graduate student candidacy seminar, the 12th Annual Pitt-CMU Graduate Student Philosophy Conference, Larryfest, and the National University of Singapore seminar series. Finally, thank you to several anonymous referees for their comments. Part of this work was supported by a start-up grant from Nanyang Technological University.

Philosophy of Science, 81 (July 2014) pp. 332–348. 0031-8248/2014/8103-0008$10.00

of the counterexamples that plagued the deductive-nomological account. For example, many cases of explanatory asymmetry seem to track causal asymmetries, many cases of explanatory irrelevance seem to track causal irrelevance, and many cases of explanatory preemption seem to track causal preemption.

Kitcher (1981) nicely summarizes why we want an account of scientific explanation and what we hope that such an account can do for us. "Firstly, we would like to understand and to evaluate the popular claim that the natural sciences do not merely pile up unrelated items of knowledge of more or less practical significance, but that they increase our understanding of the world. . . . Secondly, an account of explanation ought to enable us to comprehend and to arbitrate disputes in past and present science" (508). In this article I present a problem for sophisticated causal accounts that stems from a failure to allow us to comprehend (and perhaps adjudicate) disputes in past and present science.

2. A Note on Methodology. The idea that a philosophical account of explanation should allow us to arbitrate scientific disputes might, at first glance, seem rather too ambitious. I think that philosophy has a role to play here, but the methodology that I will adopt accords philosophical accounts of explanation a moderate role. I see such accounts as an explication of the normative commitments that we already employ (although perhaps not explicitly) in our scientific practices, rather than as imposing a norm external to that practice. This is also the project I take the accounts that I discuss in this article to be concerned with.[1] Of course, if we are explicating an implicit norm then we do not have to accurately capture every instance of what has been thought to be an explanation. However, this methodology does commit us to being able to capture central, widely known, and well-understood cases of scientific explanation.

Further, and crucially for this article, I take it that this project not only commits us to delivering the correct verdict in central and well-understood cases of scientific explanation but also gives us a pro tanto commitment to the first part of Kitcher's second criterion. We should be able to comprehend as well as let theory arbitrate disputes in past and present science. Insofar as these debates are central and well understood, our account should be able to allow us to understand how the features that are invoked in the debate are latching on to some aspect that is of importance for explanatory concerns. If the account that we have developed cannot do this even for the

1. Although Strevens (2008, sec. 1.5) calls his project descriptive, I think that this is largely a terminological difference. Strevens takes his account to be a description (rather than an evaluation) of our practices, including what determines what we call explanatory.

central and well-known cases, then we have good reason to think that the account is not a successful explication of the norms implicit in the practice at all but part of a more radically revisionary project. In the following sections, I assume this methodological standpoint and present a challenge for the dominant causal accounts of explanation that arises from a failure to allow us to comprehend central debates over the explanatory standing of certain theories.

3. Challenges to Exclusively Causal Accounts of Explanation. Although there are several versions of causal accounts of explanation available in the literature, I will focus the discussion on those developed by Woodward (2003) and Strevens (2008). They are good representatives for causal accounts in general since they provide two of the most sophisticated developments available and yet take strikingly different approaches.

Strevens's account starts with a metaphysical assumption that there are relations of causal influence that his theory takes as an unanalyzed starting point. On the basis of these, we understand our explanatory practices as aiming to show what made a difference to the causal production of the phenomenon in question and how it did so. Strevens's kairetic account is constructed to give a careful way of selecting the explanatorily relevant causal influences—the difference-making ones—from the basis of all the causal influences for a given phenomenon. Although the notion of causal influences is liberally construed, it is nonetheless a substantive component in Strevens's account since it allows him to address the directionality of explanation in a causal way. Whether or not one takes the relation of explanation to be antisymmetrical—so that if one fact, event, and so on, plays a role in explaining another, it is not the case that the other plays a role in explaining the one—there are still many cases to illustrate that the relation of explanation is not symmetrical.[2] Many such cases are by now familiar; for example, while it seems as if the height of a flagpole can explain the length of its shadow, it does not seem as if the length of the shadow can explain the height of the flagpole.[3] Similarly, while the presence of a distant mass can play a role in explaining the motion of some body, the motion of that body does not explain (although it could play a role in predicting) the mass of the distant body.

Causal accounts of explanation identify the direction of explanation with the direction of causation and thereby identify the explanatory asymmetry with a causal asymmetry. This is the crucial constraint that Strevens places on the theories of causal influence that are compatible with

2. In order to keep the terminology already in use, I will write about explanation being asymmetrical (in the ordinary but not the mathematical sense) to capture this.

3. This case was put forward (with slightly different details) by Bromberger (1966).

his account. His model of explanation demands of causal relations only "just what is needed, and no more, to provide a causal resolution of the explanatory asymmetries" (Strevens 2008, 35).

Woodward's account of causal explanation starts from a completely different point of view by taking inspiration from manipulability theories of causation. Explanation is now a matter of "exhibiting systematic patterns of counterfactual dependence" (Woodward 2003, 191). The counterfactuals that are relevant are those that show how the target system would behave under certain manipulations (or rather under, to use Woodward's technical notion, *interventions*). The idea is, roughly, that explanatory relationships show how the variable being explained would change under manipulations of the variables doing the explaining.

Although it is not apparent at first glance, Woodward's solution to the asymmetry problem also relies crucially on the asymmetry of causal relations (as will become clearer in sec. 3.2). In Woodward's account we see causal notions represented by directed graphs where a directed edge represents a particular type of causal relation. "A *directed graph* is an ordered pair $\langle \mathbf{V}, \mathbf{E} \rangle$ where $\mathbf{V}$ is a set of vertices that serve as the variables representing the relata of the causal relation and $\mathbf{E}$ a set of directed edges connecting these vertices. A directed edge from vertex or variable X to vertex or variable Y means that *X directly causes Y*" (Woodward 2003, 42). The notion of a direct cause is spelled out in terms of interventions, which themselves, as we will see below, are understood by making use of causal notions captured by the directed graphs. "A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set $\mathbf{V}$ is that there be a possible intervention on X that will change Y. . . when all other variables in $\mathbf{V}$ besides X and Y are held fixed at some value by interventions" (55).

Although it is tempting to think of Woodward's interventionist account as involving a conceptual reduction, in order for this circularity to not be vicious, it is crucial that it does not. While we need to make use of causal reasoning to figure out whether X causes Y, we do not, in general, need to assume that we already know what the causal relationship between X and Y is. Here too, in the end, causal relations provide an unreduced foundation for the account. With these quick summaries of the relevant aspects of the two accounts on the table, I turn to the case that will occupy the rest of this article.

3.1. Newton's Theory of Universal Gravitation. On any reasonable measure Newton's theory of universal gravitation now stands as a great explanatory achievement. The theory of universal gravitational attraction was extremely successful, showing a wide range of phenomena to be of the same type and predicting the behavior of different types of systems from a few laws. However, the theory was not accepted as clearly explanatory when it was first proposed. The seeming appeal to action at a distance was thought to be troubling and raised questions as to whether the theory really could have identified the physical causes of the motions predicted. This debate over the explanatory status of the theory, famously captured in the Leibniz-Clarke correspondence, is a very good candidate for a debate that we would like our theories of explanation to both adjudicate and help us understand.

The difficulty facing causal accounts of explanation is that they force us to rule out certain prima facie plausible ways of understanding this debate.[4] If we accept the methodological restrictions of section 2 together with the claim that this a central and well-researched case of disagreement over scientific explanation, then this debate is a test for our theories of scientific explanation. For reasons that will become clear in section 4, it is also an interesting case to focus on since it blocks certain common argumentative moves that we are tempted to take when accepting a causal account of explanation.

Put simply, the problem of understanding this debate from a causal explanatory perspective stems from the reluctance, on both sides, to take there to be a straightforward causal explanation given by the theory. Objections that were brought against the theory by, for example, Leibniz partly follow the pattern expected by theories of causal explanation. Leibniz rejects the explanatory status of the theory on the grounds that it does not provide a causal account. Moreover, this rejection is based on the failure of a causal interpretation of the theory to adhere to a mechanistic conception of causation by invoking unmediated action at a distance. These objections to action at a distance were known to Newton, and he was not insensitive to them. Whether or not he privately accepted the restriction that causal explanations should be local, in several places of his writings we find him struggling with the problem of separating the search for causes from the search for laws, and he stresses both that he does not have a causal account and the priority of a law-based one. "Mathematics requires an investigation of those quantities of forces and their proportions that follow from any conditions that may be supposed. Then, coming down to physics, these proportions must be compared with the phenomena, so that it may be found out which conditions [or laws] of forces apply to each kind of attracting bodies. And then, finally, it will be possible to argue more securely concerning the physical species, physical causes, and physical proportions of these forces" (Newton 1999, *Principia*, Bk. I, scholium to proposition 69, 588–89).

4. I make a much more detailed historical argument for the interpretative views that I merely take to be plausible in this article in Jansson (2013).

This looks like it should be good news for causal accounts, since Newton seems to accept the view that Strevens requires, namely, that there are relations of causal influence at the foundational level. Yet Newton both clearly rejected the suggestion put forward by Leibniz that his theory was an instrumental and nonexplanatory one and in several places stresses the causal agnosticism of his account (see Leibniz 1961/2008, 258; Newton 1961/2008, 287). "Hitherto we have explained the phaenomena of the heavens and of our sea by the power of gravity, but have not yet assigned the cause of this power. . . . I have not been able to discover the cause of those properties of gravity from phaenomena, and I frame no hypotheses" (Newton 1995, *Principia*, Bk. III, General Scholium, 442).

Taking this passage on face value, Newton seems to hold that the theory of universal gravitation can supply explanations from the laws that he has identified but that these explanations are compatible with a causal agnosticism.[5] Within the methodological framework of section 2, this puts pressure on our accounts of scientific explanation to make sense of this, at least seemingly conceptually coherent, position. However, this is harder than it might appear at first glance. In particular, since the explanatory status is tied to the causal one, in the face of agnosticism about the causal relations—in response to the worries about causal action at a distance—we are forced into an explanatory agnosticism too.

Both Woodward and Strevens consider the case of Newtonian gravity in some detail. In the quote below, Woodward gives what I take to be a common debunking story as to why worries about action at a distance were taken to be relevant. The debunking story is essentially one that notes that, while many causal interactions are spatiotemporally continuous processes, there is no good reason to take this to be an a priori constraint on causal relations. The suggestion is that this is the mistake that one can make, and that people did make, thereby mistakenly ending up taking action at a distance to be worrying for the explanatory status of Newtonian gravity since it constitutes a worry about the existence of a requisite causal relation.

> It is perfectly true that Newton himself regarded this feature [action at a distance] as unsatisfactory or at least as indicating an important incompleteness in his theory, but there seems no reason to deny that his theory describes a causal relationship between the two bodies, and this seems to have been the conclusion reached by most physicists a generation or two after Newton . . . *If* . . . a causal interaction involves transfer of energy-momentum in accord with a conservation law, that interaction will be mediated by spatiotemporally continuous processes that propagate at finite velocity. However, although many causal interactions involve energy-

5. Here I will only rely on the claim that this is a reasonable interpretation.

> momentum transfer from cause to effect, not all do. . . . Moreover, both Lorentz invariance and the conservation of energy-momentum are clearly empirical truths and not a priori constraints that follow just from the notion of causation. (Woodward 2003, 148)

While I think that it is correct that there are no a priori constraints on causal relations that rule out action at a distance, this diagnosis seems to overlook an important aspect of the case. In order to reasonably doubt whether one has a causal explanation of a given phenomenon, it is not required that it be conceptually incoherent for there to be causation of the kind required by the causal explanation under consideration. All that is required is that one has an empirically based theory of what causal influences in the actual world are like that does not allow for the kind of influences that the theory postulates. Moreover, to have held a theory of causal influences that rules out action at a distance seems perfectly responsible and not to necessarily involve any conceptual confusion. Hesse notes this in her discussion of Bacon's lists of the various phenomena for which he can find no mechanical explanation. "The phenomena which he is most ready to ascribe to action at a distance without any material medium are those which savour most of witchcraft, magic, astrology, and telepathy, and since these were the beliefs most discredited by the subsequent advance of physical science, the fact that action at a distance was discredited with them is not surprising" (Hesse 1961/1965, 95).

This way of viewing action-at-a-distance worries when it comes to causal explanation acknowledges that it is conceptually possible to have action at a distance. What is worrying about postulating it is that it does not conform to our theory of how causation works. Moreover, a causal confusion does not do justice to Newton's resistance to accept an instrumentalist interpretation of his theory. If a mistaken a priori commitment to all causal influences being local was what was driving Newton, we would, on causal accounts of explanation, expect him to also deny the explanatory status of the theory (on pain of conceptual confusion), but this is not what we see. Rather, he seems to be opposed to proponents of a specific causal theory of explanation while agreeing (or at least allowing) that he did not have a complete causal explanation. If we adopt a causal account of explanation, what otherwise seems like a sensible position to hold—namely, that the laws are explanatory even given agnosticism about the causal mechanism—becomes conceptually incoherent.

This case is particularly interesting because the considerations involved in it generalize. We are familiar with worries about the existence of appropriate causal relations that stem not directly from our concept of causation but rather from our empirically founded theories of the nature of causal influence. After all, if we hold, for example, that no causal influences can propagate faster than the speed of light, we probably do not take this restriction to be part of the concept of causation but rather take it to be supported by the theory of special relativity. That is, here we recognize a distinction between our concept of causation and our theory of causal influences. Moreover, here too we find similar debates over the explanatory status of, for example, certain quantum predictions such as the correlations found between measurements on two particles prepared in an entangled state and then separated such that no signal traveling below superluminal speeds could connect them. All of these examples are cases in which, at least at first glance, we can have a greater confidence in the explanatory status of our theories than we have in them describing causal relationships.

If we take the above seriously as a sensible reading of a central historical case (with extensions to other contemporary cases), then it is hard to defend ruling this out as a conceptual impossibility by our account of explanation or causation. This cannot be a knockdown defeater of such accounts, but it does put pressure on these accounts to come up with a reasonable debunking story of how this kind of conceptual confusion arises, and, given the methodological view proposed in section 2, to simply deny the explanatory status of the laws is not an option. In the next section, I consider some suggestions and sharpen the problem by bringing out the difficulties faced when trying to accommodate this case within a causal explanatory framework.

3.2. A First Pass at Some Solutions. An immediate solution that suggests itself is that Strevens could argue that the explanation fits the criteria for being a particular "given-that" kind of explanation and that Woodward could argue that Newton would accept that his account is causal in the broader sense of the term employed by Woodward but not causal in the narrower sense of the term used in the action-at-a-distance debate. I address the second suggestion first. This suggestion has some initial plausibility since it seems reasonable to say that Newton would have agreed that, if you could change the mass of one of two gravitating bodies, then the motion of the other body would alter in predictable ways. This seems to capture the intuition behind Woodward's notion of causation, so perhaps these kinds of counterfactuals can, on their own, give us a broader notion of causation that Newton could embrace even in the face of agnosticism about the more specific causal processes and their nature.

However, this intuitive gloss cannot be borne out by Woodward's account. The reasons for this go back all the way to the motivations that pushed us toward abandoning the deductive-nomological account and toward causal accounts of explanation in the first place, namely, the prospect of using the directionality of the causal relation to capture the directionality of the explanatory relation. To achieve this, Woodward's account gives the relevant counterfactuals a causal underpinning. It is crucial that not all counterfactuals count for the purposes of causal explanation; rather, only the ones that have to do with what would happen under interventions of a certain kind count. However, the application of the notion of an intervention on *X* with respect to *Y* demands that we have settled questions about causal relations between the variables in the variable set to a large enough extent to be able to answer questions about the existence of directed paths. Here causal notions enter the picture.

> *I* is an intervention variable for *X* with respect to *Y* if and only if *I* meets the following conditions:
>
> (**IV**)
>
> **I1:** *I* causes *X*.
>
> **I2:** *I* acts as a switch for all the other variables that cause *X*. That is, certain values of *I* are such that when *I* attains those values, *X* ceases to depend on the values of other variables that cause *X* and instead depends only on the value taken by *I*.
>
> **I3:** Any directed path from *I* to *Y* goes through *X*. That is, *I* does not directly cause *Y* and is not a cause of any causes of *Y* that are distinct from *X* except, of course, for those causes of *Y*, if any, that are built into the *I*–*X*–*Y* connection itself: that is, except for (a) any causes of *Y* that are effects of *X* . . . and (b) any causes of *Y* that are between *I* and *X* and have no effect on *Y* independently of *X*.
>
> **I4:** *I* is (statistically) independent of any variable *Z* that causes *Y* and that is on a directed path that does not go through *X*.
>
> (Woodward 2003, 98)

The notion of a directed path is defined relative to a directed graph (where a directed edge represents a direct cause, as we saw above). If there is a directed edge between X_1 and X_2 and between X_2 and X_3, and so on, to X_n, then this is a directed path from X_1 to X_n.

Now we can try to evaluate the earlier intuitive claim that, if the mass of one of the bodies were to be changed, then the movement of the other body would change. In order to evaluate this claim, we need to adjudicate the claim as to whether there is an intervention on the mass with respect to, say, the acceleration that is such that any directed path from the intervention variable to the acceleration variable goes through the mass variable. At first it seems as if this is impossible to answer without more information about what the intervention is, and, as Woodward notes, it seems quite possible that there is no such intervention that is physically or nomologically possible. That would be bad news for a solution to the problem along these lines, since now we should be agnostic or even skeptical as to whether the explanatory counterfactuals hold. Woodward addresses this worry directly (although in a slightly different guise) and responds that the intervention can be a merely conceptually possible one in which we understand how to evaluate the relevant counterfactuals from the theory itself.

> There must be a way of disentangling—perhaps merely conceptually or analytically rather than in actuality—the effect on *E* of changing just *C* from the effects on *E* of changes in other potentially confounding variables, including direct effects from the intervention process itself. . . . Newtonian gravitational theory and mechanics themselves provide the needed basis. Although it may be true that any actual physical process that changes the position of the moon will also directly influence the tides, Newtonian theory and familiar rules about the composition of forces tell us how to subtract out any direct influence from such a process so that we can calculate just what the effect of say, doubling the moon's orbit (and no other changes) would be on the tides, even though it also may be true that there is no way of actually realizing this effect alone. In other words, Newtonian theory itself delivers a determinate answer to questions about what would happen to the tides under an intervention that doubles the moon's orbit, and this is enough for counterfactual claims about what would happen under such interventions to be legitimate and to allow us to assess their truth. (Woodward 2003, 131)

This is close to the intuitive idea that the laws themselves deliver the required answer that supports the counterfactual and hence the explanation. I think that this is exactly what we want to say for the intuitive counterfactual, but Woodward's account cannot accommodate this. Woodward is also careful not to say this. After all, we know that when it comes to the laws of the theory we get a symmetry of prediction that is not reflected in our judgments about explanation. This is the problem that causal accounts in general, and in Woodward's account the notion of an intervention in particular, are supposed to address. While we can take the theory to give us answers to these claims, on Woodward's account we have to have already given the theory a causal interpretation for it to do so.

In the face of the kind of causal agnosticism that Newton seems to accept, we cannot, therefore, run the same story of how the theory itself delivers the required answer to the question whether there is (conceptually) the right kind of intervention. Moreover, we cannot even claim that Newton, since he is committed to there being a causal process of some kind, is committed to there being an unknown causal structure that would support the existence of an intervention of the right kind. Without the causal information at hand, we simply cannot judge whether there is a conceptually possible intervention that fulfills criteria I2 and I3 above and whether there are interventions that will fix the appropriate variables at some values. At a deeper level, we cannot even, without more knowledge, judge whether the variable choice is such as to allow for the concept of an intervention to be successfully applied at all. The idea that the notion of an intervention can fail, even when intuitively plausible, in light of more information about the theory, is what underpins Woodward's rejection of the idea that an intervention on the spin of one particle with respect to the other in a quantum entangled state is available (Hausman and Woodward 1999, 566). Yet here, too, we might want to claim that once we have measured the spin of one particle this explains the spin state of the other.

Strevens also addresses the case of Newtonian gravity directly, and I think that his account goes further toward being able to accommodate the kind of view that I sketched in section 3.1.[6] He notes that we could try to take what he calls the post-Newtonian explanation of Kepler's laws to be simply the claim that "first, . . . all change in planetary motion, is caused by masses and their arrangement; second, . . . the dependence between the acceleration a of any particular body due to another body of mass M and distance r takes the approximate form $a = GM/r^2$; and third, . . . the accelerations due to different bodies compose through simple vector addition" (Strevens 2008, 327–28). For our purposes here, what is particularly interesting is the first and the second of these claims. If we want to allow for a causal agnosticism on Newton's part, we have to be careful in how we understand the claim that all change in planetary motion is caused by masses and their arrangements. Newton clearly thinks that the cause acts in a way proportional to mass, but it does not follow straightforwardly from this that the mass itself is the cause. Even more strongly, it does not follow that we have identified all that is causally relevant, or even essential, if we allow that mass is a causally salient variable. Finally, we have to be careful in how we understand the dependence between acceleration a and the body of mass M at a distance r. In order to get the asymmetry of dependence in a causal way, here we need to assume that M and r causally influence a but not vice versa.

If we take this to be the only causal information, then this way of thinking about the explanation lets too much in. If all we needed in order to have an explanation were to identify a causally salient variable and to then correctly describe some aspect of the formal relationship, we would get explanations rather cheaply. Strevens notices this when he says that such a post-Newtonian explanation would probably not be cohesive. "I suspect, however, that the post-Newtonian model is not cohesive. By specifying only the form of the causal relationship between mass and planetary acceleration, the model black-boxes the relationship. Different implementations of this inverse-square black box may be causally quite different from one another" (Strevens 2008, 328). To sharpen Strevens's point here, the

6. I am setting aside the fact that we now think that Newton's theory is not true. This is, however, part of the focus of Strevens's discussion.

formal constraint relating the salient cause to the phenomenon of interest could be actively misleading as to the causal production of the phenomenon while still remaining accurate. Instead, Strevens suggests that we can relegate black boxes to the explanatory framework (that is, very roughly, what the explanation is taking as a given) and thereby let them stand in for some specific mechanisms. "When a frameworked mechanism is black-boxed, the black box itself stands in for the mechanism, which is to say, it functions as a placeholder for a certain, determinate mechanism, either known or unknown. (In the case where the mechanism is unknown, it must of course be placed in the framework by a phrase or intention that picks it out indirectly, for example 'the mechanism that is actually responsible for such and such a behavior of such and such a component of the system.')" (Strevens 2008, 152). This gets closer to an account that could make sense of Newton's position. After all, he does think that there is a causal account to be given, but one that is unknown. Perhaps we can then black-box that causal mechanism as part of the explanatory framework and instead think of the explanations offered as why, given the actual mechanism, we find the movements of the planets to be such and such.[7]

The problem with this is that none of it, on its own, seems to get us to the stage that Strevens rightly notices is something that we need, namely, the causal interpretation of the formal constraint that allows us to give a causal account of the asymmetry problem. Strevens makes this explicit when he spells out the black-boxing as changing the question to be one of why "given an inverse-square dependence" (Strevens 2008, 329) we see the kind of behavior of the planets that we do. Here the dependence, and so the directionality, itself has been built into the framework. This leaves open the question how it is that we are able to place the dependence in the framework to start with. To make this seem like a justified move, we need to have some way of extracting the directionality from the formal relationship.

Even if we take Strevens to be giving a metaphysical account of the kind that Kim (1994) has in mind, this account is still beholden to our epistemic limitations. That is, it is not enough to claim that we take there to be a certain relation, we need to show how such a claim could be backed up (at least well enough to account for our level of explanatory confidence). Finally, under the assumption that we take this case seriously as one of casual agnosticism, our understanding of the asymmetry of the dependence cannot be relegated to our causal understanding of the theory. Here, then, is the rub. The advantage of basing the kairetic account on a primitive notion of causal influence was that it allowed us to make use of the causal solution to the asymmetry problem. Now we have a situation in which we have greater confidence in the laws than we do in our knowledge about the causal mechanisms. Rather

7. Here I use the term "mechanism" in Strevens's contemporary sense.

than using our causal knowledge to impose the required directionality on the laws, we seem to be using the laws to help us figure out features of the causal mechanism. We still need to find some way in which we can justify imposing the directionality on the formal relationship captured by the law, but we cannot directly appeal to our knowledge of the causal influences. In the order of the investigation that has been proposed, that would put the cart before the horse.

4. The Problem of Simple Ecumenism. So far I have argued that the causal accounts on the table cannot, at least as they stand, easily capture a seemingly sensible way of understanding a central case of a dispute over the explanatory standing of a theory in the history of science. However, both Woodward and Strevens stress that their accounts only cover causal explanations and allow that there can be other kinds of explanations available. Here I argue that this does not allow us to escape the problem. If they fail to capture this (and similar) case(s), it threatens to undercut one of the main motivations for causal accounts of explanation.

The question that is obliquely raised through the example of section 3 is whether it is possible to have explanations, in this case law-based ones, independently of causal knowledge.[8] That the intuitive pull toward accepting natural laws as one available foundation for explanations is strong is, of course, not a surprise. It is one of the features that made the deductive-nomological account attractive. One of the central motivations for abandoning such a view was the difficulty in accounting for the asymmetry of explanation and the ease by which causal accounts capture this. However, in order to succeed in diagnosing and remedying this problem, causal accounts have to rule out certain kinds of explanatory pluralism. In particular, allowing both laws and causal relationships to act as genuine and independent explanatory relations destroys the causal solution.[9] The causal diagnosis only works if we demand that only causal relations do explanatory work. Otherwise we have not yet accounted for why, for example, the application of the laws to derive the mass of a central body from the motion of an orbiting body does not explain why the central body has that mass, as opposed to merely not being a causal explanation.[10] This blocks the simple

8. In the rest of this section, I focus on law-based explanation since that is what I take to be at stake in the example in sec. 3.1. However, the problem is more general since it concerns any noncausal explanations.

9. A causal account could, of course, allow that laws can do explanatory work, but to keep the solution to the asymmetry problem, the ability of laws to do so has to be parasitic upon the (or a) causal explanation. That is, the laws are not independently explanatory.

10. This is a problem for any account that is prescriptive and that is permissive about the explanatory relations in this way. This means that is also a challenge for the large literature on mechanistic explanations, which typically relies on causal notions in this permissive way. See, e.g., Machamer, Darden, and Craver (2000, 2, 6), and for an ex-

move of merely allowing that there are noncausal explanations as a way of trying to understand the explanatory disagreement.

Causal accounts of explanation could try to delineate a domain where we have causal explanations rather than law-based ones. If we can do so in a way that is not ad hoc, we could allow laws to do explanatory work outside this domain without destroying the solution to the asymmetry problem within this domain. This is the strategy that I take Woodward to employ. Woodward (2003, sec. 5.9) is clear that he does not claim to have an account that covers all cases of explanation and explicitly discusses mathematical explanation and mathematical dependence as exceptions. Moreover, he provides a suggested delineation. "Roughly, any explanation that proceeds by showing how an outcome depends (where the dependence in question is not logical or conceptual) on other variables or factors counts as causal" (6). However, this delineation will not do the job of allowing us to understand the case of explanations from Newton's law of gravity.[11] After all, this is not a case in which the dependence in question seems to be mathematical, conceptual, or logical.

Strevens allows for there to be noncausal explanations, but he takes such explanations to demand an underlying asymmetric relationship. In particular, he mentions the possibility of mathematical dependence and Railton's example of the stopping point of stellar collapse.[12] However, he does not give an account of what this relation would look like.

> What relation holds between the law and the arrest, then, in virtue of which the one explains the other? Let me give a partial answer: the relation is, like causal influence, some kind of metaphysical dependence relation. I no more have an account of this relation than I have an account of the influence relation, but I suggest that it is the sort of relation that we say "makes things happen." For example, because this (asymmetric) dependence relation holds between the exclusion principle and the arrest, we are apt to say that the exclusion principle makes the arrest happen. (Strevens 2008, 178)

plicit reliance on Woodward's account in the mechanistic literature, see Craver (2006, 371–72). It is also a challenge for Bokulich (2008, 2011), which she takes on in Bokulich (2012), as well as for anyone developing the ideas of structural explanation found in Clifton (1998) or Hughes (1989/2003) into a prescriptive account. Hughes (1989/2003, 198) gives up on a prescriptive account of explanation, while Clifton (1998, 19–20) discusses broader frameworks. The point here is only that one needs some such broader framework in order to be ecumenical if one is to keep a prescriptive account of scientific explanation.

11. This delineation is also similar to the one suggested by Woodward and Hitchcock (2003, 6–7).

12. Discussed in Lewis (1986), but attributed to Peter Railton.

Even if we are happy to invoke a causal primitive in our theories, we should be concerned by finding that there are examples within the same domain that do not allow us to use the causal solution to the problem of asymmetry. Once we give up the causal solution that works by identifying the direction of explanation (at least in a non–ad hoc domain) with the direction of causality, we have simply reintroduced the problem. We now need a new solution to this problem that is noncausal.

The case that I have been pressing in the previous sections is particularly challenging since it puts pressure on the idea that a causal solution to the asymmetry could work even within a neatly delineated domain, such as all explanations of physical events. Here we have a domain where we would expect to find causal explanations and where Newton expected that such explanations were there to be found. After all, explanations of individual motion are particularly well suited for causal explanations, and yet we seem to be able to make sense of having a kind of explanation without knowledge of the causal production of the phenomenon. This means that this case applies also to the view that is put forward by Saatsi and Pexton (2013). They too notice the tension in Woodward's account but conclude that it can be alleviated by taking causal explanations to provide the solution to asymmetry problems in the case of particular event explanations but that the intuition of asymmetry is fragile or nonexistent when it comes to explanations of regularities (and so presumably we do not need causal information to solve the problem here). Whether or not we agree that the explanations of regularities have this feature, the case that I have been focusing on here is easily construed as a particular explanation as much as a regularity one. After all, this is a putative explanation of the particular motions of the particular bodies in our solar system (it is an explanation of "our heavens and our sea").

5. Conclusion. There are several ways to meet the challenge of this article. One way is to try to give a debunking account of the intuition that it makes sense to hold a position such as the one that I have ascribed to Newton. I cannot rule such an option out. However, I have tried to stress here that given that we accept a methodology in which we take our clues from the central kinds of explanations and the central kinds of explanatory debates that we have seen in the sciences, we incur a burden to at least give a debunking account that shows how the features that lead us to take the law-based explanations as explanatory roughly track some explanatorily relevant features. Finally, I hope to have convinced you of the difficulties that face the most straightforward strategies for accounting for this case that are available to causal accounts in general and to Woodward's and Strevens's accounts in particular.

Although causal explanations seem ubiquitous, we cannot escape the problem of explanatory asymmetry by merely appealing to causal asymmetry, even in these cases. For this solution to work, we need to assume that no other kind of explanation is available, and as long as such explanations cannot be ruled out in the domain in question, we cannot escape the work of giving a more general solution to this problem.

There are, of course, different ways of tackling this problem. One is to give up on the idea of a general, prescriptive account of explanation altogether and accept something like van Fraassen's (1980, 132–34) view that explanation is, through and through, a pragmatic phenomenon. On the other hand, we can keep the central intuition from Woodward's and Strevens's accounts, namely, that to explain is to track relations of dependence or influence, but take on the challenge from Lipton. "The existence of non-causal explanations show that a causal model of explanation cannot be complete. One reaction would be to attempt to expand the notion of causation to some broader notion of 'determination' that would encompass the non-causal cases (Ruben 1990: 230–3). This approach has merit, but it will be difficult to come up with such a notion that we understand even as well as causation" (Lipton 2004, 32).

Here I have argued that if we wish to maintain a prescriptive account with even a fairly moderate explanatory pluralism when it comes to the relations that can do explanatory work, then we have no easy way to avoid taking on something like this challenge. The trouble for causal accounts is far more serious than one of merely being incomplete. Finally, I take the way forward to be to examine the cases of explanatory asymmetry at hand and determine what information our judgments of asymmetry stem from. If the view stressed in this article is right, we should not expect it all to be causal.

REFERENCES

Bokulich, Alisa. 2008. *Reexamining the Quantum-Classical Relation: Beyond Reductionism and Pluralism*. Cambridge: Cambridge University Press.

———. 2011. "How Scientific Models Can Explain." *Synthese* 180 (1): 33–45.

———. 2012. "Distinguishing Explanatory from Nonexplanatory Fictions." *Philosophy of Science* 79:725–37.

Bromberger, Sylvain. 1966. "Why-Questions." In *Mind and Cosmos*, ed. R. Colodny, 86–111. University of Pittsburgh Series in the Philosophy of Science 3. Pittsburgh: University of Pittsburgh Press.

Clifton, Robert. 1998. "Scientific Explanation in Quantum Theory." PhilSci Archive. http://philsci-archive.pitt.edu/91/.

Craver, Carl. 2006. "When Mechanistic Models Explain." *Synthese* 153 (2): 355–76.

Hausman, Daniel, and James Woodward. 1999. "Independence, Invariance and the Causal Markov Condition." *British Journal for the Philosophy of Science* 50 (4): 521–83.

Hesse, Mary. 1961/1965. *Forces and Fields: The Concept of Action at a Distance in the History of Physics.* Repr. Totowa, NJ: Littlefield, Adams.

Hughes, R. I. G. 1989/2003. “Bell’s Theorem, Ideology, and Structural Explanation.” In *Philosophical Consequences of Quantum Theory: Reflections on Bell’s Theorem*, ed. J. Cushing and E. McMullin. Repr. Notre Dame, IN: University of Notre Dame Press.

Jansson, Lina. 2013. “Newton’s ‘satis est’: A New Explanatory Role for Laws.” *Studies in History and Philosophy of Science* 44:553–62.

Kim, Jaegwon. 1994. “Explanatory Knowledge and Metaphysical Dependence.” *Philosophical Issues* 50:51–69.

Kitcher, Philip. 1981. “Explanatory Unification.” *Philosophy of Science* 48 (4): 507–31.

Leibniz, Gottfried. 1961/2008. Letter to Isaac Newton, March 7/17, 1693. In *The Correspondence of Isaac Newton*, vol. 3, ed. H. W. Turnbull. Repr. Cambridge: Cambridge University Press.

Lewis, David. 1986. “Causal Explanation.” In *Philosophical Papers*, vol. 2, 214–40. Oxford: Oxford University Press.

Lipton, Peter. 2004. *Inference to the Best Explanation*. 2nd ed. New York: Routledge.

Machamer, Peter, Lindley Darden, and Carl Craver. 2000. “Thinking about Mechanisms.” *Philosophy of Science* 67:1–25.

Newton, Isaac. 1995. *The Principia*. Trans. Andrew Motte. Amherst, NY: Prometheus.

———. 1999. *The Principia: Mathematical Principles of Natural Philosophy; A New Translation*, trans. I. B. Cohen and A. Whitman. Berkeley: University of California Press.

———. 1961/2008. Letter to Gottfried Wilhelm Leibniz, March 7, 1693. In *The Correspondence of Isaac Newton*, vol. 3, ed. H. W. Turnbull. Repr. Cambridge: Cambridge University Press.

Ruben, David-Hillel. 1990. *Explaining Explanation*. New York: Routledge.

Saatsi, Juha, and Mark Pexton. 2013. “Reassessing Woodward’s Account of Explanation: Regularities, Counterfactuals, and Noncausal Explanations.” *Philosophy of Science* 80 (5): 613–24.

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

van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Clarendon.

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

Woodward, James, and Christoper Hitchcock. 2003. “Explanatory Generalizations.” Pt. 1, “A Counterfactual Account.” *Noûs* 37 (1): 1–24.