Synthese (2021) 198:3893–3918
https://doi.org/10.1007/s11229-019-02317-3

Asymmetry as a challenge to counterfactual accounts of non-causal explanation

Marc Lange[1]

Received: 17 April 2019 / Accepted: 4 July 2019 / Published online: 8 July 2019

Abstract
This paper examines some recent attempts that use counterfactuals to understand the asymmetry of non-causal scientific explanations. These attempts recognize that even when there is explanatory asymmetry, there may be symmetry in counterfactual dependence. Therefore, something more than mere counterfactual dependence is needed to account for explanatory asymmetry. Whether that further ingredient, even if applicable to causal explanation, can fit non-causal explanation is the challenge that explanatory asymmetry poses for counterfactual accounts of non-causal explanation. This paper argues that several recent accounts (Woodward, in: Reutlinger and Saatsi (eds) Explanation beyond causation: philosophical perspectives on non-causal explanations, Oxford University Press, Oxford, pp 117–140, 2018; Jansson and Saatsi in Br J Philos Sci, forthcoming; Jansson in J Philos 112:7–599, 2015; Saatsi and Pexton in Philos Sci 80: 613–624, 2013; French and Saatsi, in: Reutlinger and Saatsi (eds) Explanation beyond causation: philosophical perspectives on non-causal explanations, Oxford University Press, Oxford, pp 185–205, 2018) fail to meet this challenge. The paper then sketches a more positive proposal for dealing with explanatory asymmetry in non-causal explanations.

Keywords Explanation · Counterfactuals · Non-causal · Woodward · Mathematics · Relativity · Topology · Causation

1 Introduction

It has long been widely accepted that many scientific explanations are asymmetric: f helps to explain g but g does not help to explain f. A standard account of many of these familiar explanatory asymmetries is that f identifies g's causes whereas g does not identify f's causes (Salmon 1984, p. 95; Woodward 2003, pp. 155, 361). Recently,

✉ Marc Lange
mlange@email.unc.edu

[1] University of North Carolina at Chapel Hill, CB#3125, Chapel Hill, NC 27599-3125, USA

some philosophers have argued that various scientific explanations are non-causal.[1] It is difficult to see how the asymmetry of non-causal explanations could be accounted for by appealing to causal asymmetry. This paper critically examines some philosophical attempts to resolve this problem by generalizing the causal approach to explanatory asymmetry.

In particular, some philosophers have proposed that counterfactuals not only reflect the causal asymmetry that accounts for the asymmetry of causal explanations, but also can be used to account for the asymmetry of non-causal explanations. These philosophers recognize that even where there is explanatory asymmetry (*f* helps to explain *g* but *g* does not help to explain *f*), there is sometimes symmetry in counterfactual dependence: had *f* not obtained, *g* would not have obtained, *and* vice versa. Therefore, something other than mere counterfactual dependence is needed for counterfactuals to account for explanatory asymmetry. Whether that further ingredient (even if applicable to causal explanation) can be generalized to fit non-causal explanation is the challenge that explanatory asymmetry poses for counterfactual accounts of non-causal explanation.

In this paper, I will examine several recent attempts to address this challenge. I will argue that none of them is satisfactory. These attempts differ in the targets of the asymmetric non-causal scientific explanations for which they aim to account. In Sect. 2, I will examine proposals that aim to use counterfactuals to account for the asymmetry of non-causal explanations of facts concerning particular designated individuals (such as the bridge arrangement in a given city during a certain period, or the knot in Jones's shoelaces at a given moment, or Mother's attempt to divide her strawberries evenly among her children). I will refer to these as explanations of "singular facts" although the explanandum need not be a fact about just one event (that a given attempt to traverse a certain token arrangement of bridges fails); the explanandum may be a regularity (that no one ever successfully traverses a given token arrangement of bridges) or a modal fact (that the token arrangement cannot be traversed). In Sect. 3, by contrast, I will examine non-causal explanations that target natural laws (or the regularities they entail). These explanations concern types rather than tokens (e.g., any trefoil knot rather than the token knot in Jones's laces at a given moment).

Many of these proposals regarding non-causal explanation aim to generalize Woodward's (2003) interventionist account of causal explanation. Where there is explanatory asymmetry despite symmetry in counterfactual dependence, Woodward aims to use the notion of an "intervention" to carve out counterfactuals that mirror causal relations, hence are asymmetric, and so can generate explanatory asymmetry. But the notion of an intervention is a causal notion and so is not obviously applicable to non-causal explanation.

I begin (in Sect. 2.1) by examining Woodward's (2018) proposal for generating explanatory asymmetry in non-causal explanations of singular facts by generalizing from his use of interventionist counterfactuals to capture explanatory asymmetry in causal explanations of singular facts. In Sects. 2.2 and 2.3, I examine proposals from Jansson and Saatsi (forthcoming) and Jansson (2015), respectively. In Sect. 2.4, I conclude my look at non-causal explanations of singular facts by abstracting from

[1] See (for instance) Lange (2017) and Reutlinger and Saatsi (2018), and references therein.

some of the difficulties encountered by the above proposals. In Sect. 3, I turn to non-causal explanations of laws (or the regularities they entail). In Sect. 3.1, I examine the approach taken by Saatsi and Pexton (2013), which aims to avoid relying on interventions in extending Woodward's account of the causal explanation of regularities. In Sect. 3.2, I critique the strategy taken by French and Saatsi (2018), which aims to use interventions to account for the asymmetry in non-causal explanations of conservation laws by symmetry principles. I argue that all of these approaches encounter severe obstacles.

I take for granted that the explanations analyzed by these various philosophers are indeed non-causal and asymmetric. This paper (like Jansson 2015, p. 595) has no need to presuppose that *all* scientific explanations are asymmetric; the challenge presupposes only that the examples under discussion are asymmetric. Of course, one way to avoid having to resolve this challenge that explanatory asymmetry poses for non-causal explanations would be to deny that these non-causal explanations are asymmetric, as some philosophers have done.[2] This approach (along with any approach that accepts the asymmetry of non-causal explanations but avoids using some sort of counterfactual dependence to account for it) falls outside the scope of this paper. Furthermore, in considering the viability of counterfactual accounts of non-causal explanation, I will not examine any difficulties they encounter that do not concern explanatory asymmetry.

I do not regard the difficulties that I highlight for various counterfactual accounts as conclusively foreclosing the prospects for such accounts. Conclusive arguments are difficult to give. My aim is to advance the discussion by exposing some of the difficulties that counterfactual accounts face in responding to the challenge that explanatory asymmetry poses for non-causal explanations. In the final section, I will briefly say something more positive about the sort of response to this challenge that is (in my view) more likely to succeed.

2 Explanations of singular facts

2.1 Woodward's account

On Woodward's (2003) account, a causal explanation of event Y works by conveying that (and how) Y would have been different under a possible "intervention" (with respect to Y) on some event X that would have replaced X with some alternative. The relations sustaining these counterfactuals would still have held, had the intervention taken place, but are contingent. (In being counterfactually invariant yet contingent, these relations are like natural laws as traditionally understood.) Thus, a causal explanation of a singular fact works by conveying information about Y's systematic counterfactual dependence, that is, about the answers to certain what-if-things-had-been-different questions ("w-questions"). The explanation is causal because such a pattern of counterfactual dependence necessarily accompanies a causal relation. To

[2] This approach is taken by Reutlinger (2017, pp. 253): "I think there are good reasons to hold that *some* non-causal explanations are not asymmetric…I believe that *some* (for instance, Euler's explanation [in the bridge example discussed below in Sect. 2]) … lack such an asymmetry because the counterfactual dependence in question is symmetric."

accommodate non-causal explanations, Woodward (2003, p. 221; 2018) proposes that a given non-causal explanation shares many of the features of causal explanations, such as conveying answers to w-questions, but may differ in certain respects, such as in not involving interventions or in involving invariant relations that are (mathematically or conceptually) necessary rather than contingent.

To examine whether this proposal underwrites explanatory asymmetry in *non-causal* explanations of singular facts, we must first briefly review how the notion of an intervention enables this proposal to underwrite explanatory asymmetry in *causal* explanations of singular facts. An "intervention" I on X with respect to Y is an operation that would cause X to change and would result in a change to Y, if at all,

(i) not by I's causing Y to change as a cause of changing X

$$I \rightarrow Y \rightarrow X$$

(ii) not by I's causing Y to change by a separate causal pathway from the pathway by which I affects X

$$Y \leftarrow I \rightarrow X$$

(iii) not by I's being statistically correlated with changes to some of Y's causes where those changes are not (effects of) changes to X

$$X \leftarrow I - - - -(statistical\ correlation) - - - -Y's\ causes \rightarrow Y$$

but (iv) rather as an effect of changing X

$$I \rightarrow X \rightarrow Y$$

(Woodward 2003, p. 98). Thus, I would cause Y to change, if at all, exclusively by causing X to change. For example, suppose that X is Y's effect or (though not Y's effect or cause) shares a common cause with Y. Then an intervention on X with respect to Y would not bring about a change in Y. So to preclude Y's causal explanation from appealing to an effect of Y (that is, to capture the asymmetry of causal explanation), Woodward's account associates explanation not simply with how Y would have been different had X been different, but with how Y would have been different had X been changed *by an intervention on* X *with respect to* Y.

Some *non-causal* explanations of singular facts, according to Woodward, work like *causal* explanations of singular facts except that the invariant relations supporting the relevant counterfactuals are necessary rather than contingent. This approach, it seems to me, sometimes succeeds in capturing explanatory asymmetry when the explanandum and explanans concern particular events and so are the kinds of entities that could (at least in principle) be the targets of interventions. For instance, let's see how explanatory asymmetry is achieved when a pendulum's length L (partly) causally explains its period T (as discussed by Woodward 2003, pp. 197–198)—and how explanatory asymmetry is achieved in the same way when some city's bridges' possessing the "Euler feature" (that either zero or two nodes in their graph have an odd number of edges) non-causally (partly) explains why Jones succeeds in his attempt to traverse those bridges (by a continuous, landlocked path, crossing each bridge exactly once). (In a moment, I will turn to Königsberg's famous bridges, which cannot be traversed and lack the Euler feature.[3])

In the case of the pendulum, there is a possible intervention on *L* with respect to *T*—roughly (as I just said), a way to vary *L* that would change *T* (if at all) only as an effect of varying *L*. Such an intervention could consist in shortening the rod by which the bob is suspended. Under such an intervention, *T* would have been different. Because of this counterfactual dependence, *L* can (partly) causally explain *T.* Now let's see the explanatory asymmetry arise when we try to use *T* to causally explain *L.* There is a possible intervention on *T* with respect to *L*—roughly, a way to vary *T* that would change *L* (if at all) only as an effect of varying *T.* This intervention cannot consist in shortening the rod and thereby causing *T* to change, since that operation violates condition (i) above on such an intervention (because the operation causes *L* to change as a cause of changing *T*). Rather, such an intervention could consist in moving the pendulum to a location with a different gravitational acceleration. But under this intervention on *T*, *L* would remain unchanged. Thus the explanatory asymmetry is generated.

The same approach applies to the non-causal explanation of Jones's success in bridge-traversing. There is an intervention on *E* (possessing the Euler feature) with respect to *S* (Jones's success in traversing the bridges)—that is, roughly, a way to vary *E* that would change *S* (if at all) only as an effect of varying *E*. Such an intervention could consist in adding or removing certain bridges. Under such an intervention, *S* would have been different (Jones would have failed). Because of this counterfactual dependence, *E* can (partly) explain *S.* Now let's see the asymmetry arise when we try to use *S* to explain *E.* There is an intervention on *S* with respect to *E*—roughly, a way to vary *S* that would change *E* (if at all) only as an effect of varying *S.* This intervention cannot consist in adding or removing certain bridges and thereby causing *S* to change, since that operation violates condition (i) above on such an intervention (because the operation causes *E* to change as a cause of changing *S*). Rather, such an intervention could consist in posting a threatening personage on one of the bridges, making Jones too frightened to try to cross it. But under this intervention on *S*, *E* would remain unchanged. Thus the explanatory asymmetry is generated.

I presume the above to be roughly Woodward's view of this example. But although Woodward (2018, pp. 127–128) discusses a non-causal explanation concerning an arrangement of bridges, the explanandum he considers is not Jones's success (or failure) in attempting to traverse the bridges. Rather, his explanandum is the Königsberg bridges' non-travers*ability N*; Königsberg's arrangement of bridges lacks the Euler feature. I regard ~*E* in this case as non-causally explaining *N* and I take this explanation to be asymmetric: *N* does not explain ~*E*! Woodward says that despite the fact that the invariant relation is mathematically necessary rather than contingent, his account of causal explanation can easily be extended to this case: the notion of intervening

3 Among those who regard solutions to the "Königsberg bridge problem" as supplying non-causal explanations are Pincock 2007 and Reutlinger 2017, pp. 245–256 (see note 2).

on the bridge configuration (e.g., "by constructing additional bridges or removing some") "does not seem strained or unclear. This also fits naturally with an account of the example in terms of which it is explanatory in virtue of providing information to w-questions" (2018, p. 128).

Woodward (2018, p. 138) says that although *E* is the sort of thing that can be the target of an intervention, *N* (or ~ *N*) is not, since it concerns what is impossible (or possible) and so "is not the sort of thing that can be a causal effect or a target of causal explanation." On this view, posting a threatening personage on one of the bridges, thereby making Jones too frightened to try to cross it, is an intervention on *S* but not on ~ *N* since unlike *S*, *N* is a modal matter. (Indeed, posting such a person would not even change ~ *N* to *N*; it would not render the bridges untravers*able* in the relevant, topological sense. I will return to this point in Sect. 2.3.) Therefore, although there are interventions on *E* that cause *S* to change, there are no interventions on ~ *E* that *cause* *N* to change. To embrace the explanation of *N*, the account of non-causal explanation must regard the relevant counterfactuals as reflecting not how the explanandum would have been *caused* to be different by an intervention on the explanans, but merely how the explanandum would have *been* different under such an intervention. Presumably, Woodward regards this feature (along with the invariant relation's necessity) as helping to make the explanation non-causal.

Woodward (2018, p. 128) says that "the direction of the dependency relation seems unproblematic" in this case. Does Woodward's approach account for the explanatory asymmetry here? Of course, there are interventions on ~ *E* under which *N* would have been different whereas (according to Woodward) there are no interventions on *N* under which ~ *E* would have been different—simply because (Woodward says) *N*, as a modal matter, cannot be the target of an intervention. But even if we accept that *N* cannot be the target of an intervention, this cannot be the way that Woodward's account is supposed to generate explanatory asymmetry, since if it were, then in every non-causal explanation, the explanans would have to be eligible to be the target of an intervention. Woodward recognizes this condition to be too restrictive; to so restrict the explanans in a non-causal explanation would make the range of non-causal explanations too narrow to fit scientific practice. For this reason, Woodward loosens this restriction; he says (2018, p. 122) "one possible form of non-causal explanation answers w-questions … but does not do so by providing answers to questions about what happens under *interventions*." Woodward (2018, pp. 123–125) thereby leaves room for space's three-dimensionality to partly explain why stable planetary orbits are possible—even though space's dimensionality cannot be the target of an intervention. (I'll shortly consider Woodward's account of this explanation's asymmetry.)

But if the account allows explanations to derive their explanatory power by virtue of answering w-questions concerning what would have happened under changes that are *not* interventions, then such a w-question *can* concern what would have happened to ~ *E* had *N* been different. And any change to *N* would—as a matter of topological necessity—be accompanied by a change to ~ *E*. The counterfactual dependence is then symmetric and so no explanatory asymmetry is generated.[4]

[4] Woodward (2018, p. 128) says that the bridges' arrangement "has perfectly ordinary causes rooted in human decisions to construct one or another particular configuration. Because these decisions cause

Woodward may be proposing that although not every case of non-causal explanation involves an explanans that is eligible to be the target of an intervention, nevertheless in those cases where only one of the explanans and explanandum is eligible, that one is the explanans. In other words, Woodward may be proposing that the source of the explanatory asymmetry in some (but not all) cases of non-causal explanation is that the explanans is eligible to be the target of an intervention whereas the explanandum is not—and that the bridge example (with ~ *E* and *N*) is such a case.

This proposal applies to this bridge example only because *N* is a modal matter whereas ~ *E* is not. There are cases very similar to this bridge example (that is, where topology explains the impossibility of doing something) in which *both* the explanandum and the putative explanans are modal; this proposal cannot be applied to these cases. Yet it would be strange for whatever settles the order of explanatory priority in this bridge example not also to apply to these similar topological cases. Furthermore, in cases where both the explanandum and the putative explanans are modal, we could de-modalize either one, making *both* directions satisfy the above sufficient condition for explanatory asymmetry that Woodward may be proposing. Again, it would be strange for *both* directions to be explanatory.

Knot theory supplies examples of this kind. Consider a given token knot (not a type of knot, such as a trefoil knot, but the particular knot currently in—let's say—Jones's left shoelace). That it cannot be untied ("cannot" in the mathematical sense—that is, roughly speaking, permitting all operations except cutting the shoelace) has often been thought (e.g., by Kitcher 1989, p. 426; Lange 2017, p. 8) to have a non-causal explanation in the same way as does the fact that a given arrangement of bridges cannot be traversed. Just as a token bridge arrangement cannot be traversed if it lacks the "Euler feature", so a token knot cannot be untied if it is "tricolorable" (defined in Fig. 1).[5] Since tricolorability (i.e., being able to be "tricolored") is a modal matter (unlike ~ *E*), the proposed sufficient condition fails to settle the direction of explanatory priority between tricolorability and untie-ability, even though a knot's capacity to be untied seems much like a bridge arrangement's capacity to be traversed. Furthermore, we can de-modalize either of these modal properties of knots: that a knot has been tricolored entails that it cannot be untied and that a knot is untied entails that it is not tricolorable. Moreover, there is an intervention on a knot's being tricolored (involving clipping a

Footnote 4 continued
the configuration, it is clear that [*N*] is not somehow part of an explanation of the configuration." Of course, I grant that the bridges' arrangement (and hence ~ *E*, let's suppose) is causally explained by earlier human decisions. But I do not see how it follows, from the interventionist account's entailing that human decisions causally explain ~ *E*, that the account entails that "the direction [of explanation] must run from the configuration to the impossibility of traversing." Whether the account says that *N* non-causally explains ~ *E* depends on what the account says a non-causal explanation consists in; nothing in the account, as far as I can tell, ensures that if ~ *E* has a causal explanation, then nothing can satisfy the requirements for being a non-causal explanation of ~ *E*. (Furthermore, some philosophers (e.g., Salmon 1989, p. 183; Lange 2017, pp. 58–64) have maintained that a given fact can have both causal and (at least partly) non-causal explanations.) As we have seen, that *N* is unsuitable to be a target of intervention does not entail, on Woodward's account, that *N* cannot be the explanans in a non-causal explanation of a singular fact since (Woodward 2018, p. 122) "one possible form of non-causal explanation answers w-questions … but does not do so by providing answers to questions about what happens under *interventions*."

[5] I do not know whether in some cases a token knot's tricolorability in fact explains why it cannot be untied; see Sect. 2.4. But some knot invariants presumably do explain why some knots can(not) be untied, and I would not want to preclude tricolorability's doing so.

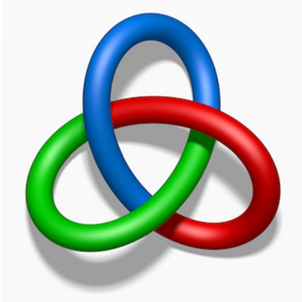


Fig. 1 A knot is "tricolorable" iff each "strand" (each continuous piece going from one undercrossing to the next) can be colored one of three colors, where at least two colors are used and at each crossing, the three strands are either all the same color or all different colors. In this figure, a trefoil knot is successfully "tricolored"

certain string segment in or out) that would change its untie-ability, and there is a similar intervention on a knot's being untied that would change its non-tricolorability. We have thereby placed *both* directions within the scope of the above proposed sufficient condition for establishing explanatory asymmetry, since each involves counterfactual dependence under an intervention on a putative explanans that is eligible to be the target of an intervention along with a putative explanandum that is modal and so ineligible. But it seems odd (though not strictly circular) for explanation to proceed in *both* directions—that is, for one knot's being tricolored to explain why it is untie-able while another knot's being successfully untied explains why it is non-tricolorable.

Presumably, some epicycles could be added to the above proposed sufficient condition in order to avoid these unattractive consequences. The challenge would be for these epicycles not to be ad hoc and somehow to exploit the concept of an intervention to pick out the sort of counterfactual dependence that tracks asymmetric explanatory dependence.

Let's now see whether explanatory asymmetry in the bridge case can instead be generated by the approach that Woodward uses to generate asymmetry in the case of space's three-dimensionality D non-causally explaining the possibility P of stable planetary orbits.[6] For I to qualify as an intervention on X relative to Y, I must (by condition (iii) above) be statistically uncorrelated with (i.e., must be "independent" of) changes to Y's causes where those changes are not (effects of) changes to X. Although neither D nor P can be the target of interventions, Woodward (2018, pp. 123–125) thinks that something like this "independence" condition must be satisfied in a non-causal explanation of one of these facts by the other. That is, such an explanation would convey information about how the explanandum would have been different, had the explanans been changed in a way where something like condition (iii) is satisfied—that is, where the explanans' change is statistically uncorrelated with changes to the explanandum's other explainers (as long as those changes are not explained by changes to the explanans). Woodward sees this condition as accounting for the explanatory asymmetry in the case of D and P. The pattern of entailments there is

[6] Woodward (2003, pp. 220–221) mentions this explanation, which is also discussed originally by Ehrenfest 1917 and more recently by Callender 2005. Although an explanation of the possibility of stable planetary orbits is not an explanation of a singular fact, I mention it in this section (devoted to proposals regarding explanations of singular facts) in order to examine whether Woodward's strategy in this case can be used to account for the asymmetry in the explanations of various singular facts (regarding token bridge arrangements.

symmetric: *P* follows logically from *D* plus auxiliary premises *A* (Newton's laws of motion and the form of the gravitational potential) and *D* follows logically from *P&A*. According to spacetime substantivalists, Woodward says, there is a way for *D* to have been different without *A* having been different (namely, by a change to space, where space—being a substance—is independent of the dynamics *A*), so this change satisfies condition (iii). Had *D* been changed in this way, *P* would have been different (by the same sort of deduction by which *D&A* entails *P*), underwriting *D*'s explaining *P*. Now let's see why (according to Woodward) substantivalists cannot run an explanation in the opposite direction, yielding the desired explanatory asymmetry. Admittedly (by the same sort of deduction by which *P&A* entails *D*), had *P* been different but *A* remained unchanged, *D* would have been different. But according to substantivalists, this counterfactual dependence does not underwrite *P*'s explaining *D*. Why not? Because (Woodward says) this counterfactual's antecedent ("had *P* been different but *A* remained unchanged") does not concern *P*'s change being brought about by something like an intervention, since condition (iii) is not satisfied: had *P* been different, *A* would *not* have been left unchanged. Rather, had *P* been different, space (being a substance and so independent of the dynamics) would have been unchanged and so *A* would have been different. Substantivalists (Woodward says) see explanatory asymmetry here because they believe that there is no change to *P* that satisfies condition (iii), whereas there is a change to *D* that satisfies condition (iii), and a change to *P* (or *D*) must satisfy condition (iii)—an "independence" condition—in order for what would have happened under that change to be the kind of counterfactual dependence associated with explanation.[7]

Therefore, the source of the explanatory asymmetry (according to substantivalism), on Woodward's account, is that had *D* been different, *A* would have been unchanged, whereas had *P* been different, *A* would *not* have been unchanged. Whatever the merits of this approach in the case of *D* and *P* (I will return to it in Sect. 4), it cannot account for the explanatory asymmetry between ~*E* and *N*. That is because in the latter case, there are no auxiliaries to (fail to) be independent; there is nothing to play the role of *A*. *N* follows from ~*E* alone, and vice versa. So far, then, Woodward's account fails to generate explanatory asymmetry between ~*E* and *N*.

2.2 Jansson's and Saatsi's account

Consider another putative example Woodward (2018, pp. 126–127) discusses of non-causal scientific explanations of singular facts: That 3 fails to divide 23 evenly (along, perhaps, with other such arithmetic facts), together with Mother's having 23 strawberries and 3 children, non-causally explains why Mother cannot divide her strawberries evenly among her children (without cutting any).[8] This explanation appears well suited to a counterfactual approach since there is a pattern of counterfactual dependence. For

[7] Woodward does not endorse *D*'s explaining *P* (or vice versa); he sees little empirical basis for maintaining that *A* would still have held, had *D* (or had *P*) been different. He is not prepared to endorse (or to deny) substantivalism. But this demurral makes no difference to his point, which is that his account of non-causal explanation correctly identifies what it would take for *D* to explain *P* or vice versa—e.g., correctly identifies where *substantivalists* see the explanatory asymmetry as coming from.

[8] Lange (2013, p. 488) introduced this example as a non-causal explanation.

instance, had Mother possessed one more strawberry, she would have succeeded in distributing her strawberries evenly without cutting any. But can a counterfactual approach account for the explanatory asymmetry here? Jansson and Saatsi (forthcoming—henceforth "J&S") propose that this explanatory asymmetry arises from another asymmetry. J&S hold that "it's true that had the system's (non-)divisibility-by-three been different, then the number of strawberries would have been different" (fn. 17). J&S say that this counterfactual does not underwrite a non-causal explanation because "fixing the system's (non-)divisibility-by-three does not fix the number of strawberries to any particular value. In contrast, the number of strawberries being twenty-three does fix the system's (non-)divisibility by three" (just before fn. 17). J&S base their account of explanatory asymmetry on the principle that "Fixing the explanans variable to its actual value should fix the explanandum variable to its actual value" (p. 17).[9]

However, it seems to me that whether this principle is satisfied in the given case depends on the particular variables involved. Suppose we take the variable's values not to be the precise numbers of strawberries and children. Instead let v = 0 [v = 1] represent that the number of strawberries divided by the number of children is [not] a whole number. Then v's actual value (1) *is* fixed by the fact that it is mathematically impossible for Mother to succeed in her strawberry-distribution task. Therefore, the principle cited by J&S does not preclude v = 1 from being explained by the impossibility of Mother's success. Nevertheless, it seems to me that this gets the explanatory direction backwards: v = 1 explains why Mother must fail, not vice versa.

I do not want to put much weight on this "tricky" change of variables. I intend it merely to highlight how J&S's approach yields the correct verdict regarding the strawberry example only because of what seems like a rather incidental feature of that example: that although the explanans would have been different, had the explanandum been different, there is no specific way that the explanans would have been different. Indeed, there are plenty of causal explanations of singular facts where had the cause not occurred, the effect would not have occurred, but there is no more specific *p* where *p* would have obtained instead of the effect (or the cause), had the cause not occurred. (Had Suzy not thrown the stone, the window would not have broken—but what would Suzy have been doing instead of throwing stones?) The same phenomenon occurs in *non*-causal explanations of singular facts. For instance, given Minkowski spacetime, that a particle is massless (i.e., has zero "rest mass") entails that and non-causally explains why it must have speed *c*. Nothing about its speed (other than that it is not *c*) follows from its having a given non-zero mass value. In this example, the asymmetry of non-causal explanation is not accompanied by any asymmetry in counterfactual dependence or entailment; each of these is symmetric. That a body must have speed *c* (together with the natural laws) entails that it is massless.

Furthermore, although the fact that Mother cannot succeed is not enough to entail the precise numbers of strawberries and children, the fact that Mother cannot succeed provides some information about the numbers of strawberries and children—and more precise information about how she must fail entails more precise information about the numbers of strawberries and children. For instance, the fact that after distributing

[9] Presumably, J&S would allow this principle to be loosened to permit statistical explanations of outcomes governed by indeterministic laws.

7 strawberries to each child, she must have 2 strawberries left over, one short of being able to distribute her strawberries evenly, ensures that she has 3 children and 23 strawberries.[10]

2.3 Jansson's account

Jansson (2015) offers another proposal for using counterfactuals to capture explanatory asymmetries in explanations of singular facts. She recognizes that simple counterfactual dependence can be symmetric even when explanation is asymmetric, and by avoiding any appeal to causal asymmetry or interventions, she aims to leave room for non-causal explanation (597). (She cites (587) space's three-dimensionality as helping explain the stability of planetary orbits.) Jansson is concerned with explanations that appeal to derivative laws, such as the explanation of a pendulum's period T that appeals to its length L and the pendulum law $T = 2\pi\sqrt{(L/g)}$. This law has "conditions of application" such as that the swings are small, there is no air resistance, and the string has been connected to the bob and set into motion. Jansson proposes (592) that L explains T rather than vice versa because of a counterfactual asymmetry involving the law's conditions of application: had those conditions failed, then L would have been no different but T would generally have been different. For example, had the string not been connected to the bob (or had there been air resistance), the string's length would have been no different but it would have had no disposition to swing with period T. This is not an interventionist counterfactual; this approach "does not tackle failures of explanatory symmetry by appeal to causal asymmetry" (593). So it seems well suited to covering non-causal explanations too.

It seems to me that Jansson is correct that this counterfactual asymmetry often parallels the explanatory asymmetry. However, the reason they often run in parallel is not, I think, that the counterfactual asymmetry is (as Jansson proposes) the source of the explanatory asymmetry. Rather, the reason that these two asymmetries often line up is that oftentimes the derivative law's conditions of application involve the absence of "disturbing factors" and had there been disturbing factors, the actual causal influences would still have been present (alongside the disturbing factors). Accordingly, the fact being explained would not still have held but the facts actually explaining it would still have held. (The law's conditions of application also often involve the physical system's having been assembled from its components. Had it not been assembled, its components would still have existed but the fact being explained would not have obtained.)

[10] Baron et al. (2017) argue for an account of counterfactuals that would permit the nontrivial truth of countermathematicals such as "Had 23 been divisible into a whole number by 3, then Mother would have been able to divide her 23 strawberries evenly among her 3 children without cutting any." They see such countermathematicals as underwriting "extra-mathematical" scientific explanations, thereby "extend[ing] the counterfactual theory of explanation to non-causal cases" (p. 1). But it seems to me that their approach to these counterfactuals inevitably also endorses "Had Mother been able to divide her 23 strawberries evenly among her 3 children without cutting any, then 23 would have been divisible into a whole number by 3." With that symmetry in counterfactual dependence, there would be explanatory symmetry on their account. They do not raise this issue; they (p. 28) defer to Woodward to defend using counterfactual dependence to understand explanatory dependence.

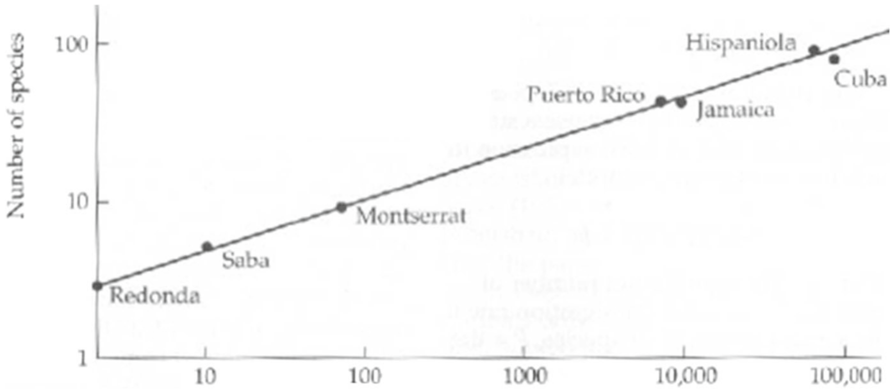


Fig. 2 The area (in square kilometers) of various islands in the West Indies is depicted on the *x*-axis. The number of amphibian and reptilian species on each island is depicted on the *y*-axis. (After MacArthur 1972, p. 104)

On my diagnosis, the explanatory asymmetry does not come from Jansson's counterfactual asymmetry, but they tend to line up when the possible "disturbing factors" tend to operate independently of the explanatory factors captured by the derivative law. This independence is less common in sciences giving complex multifactorial explanations citing many interconnected factors. Consequently, Jansson's approach tends to fail in those cases—even in cases of *causal* explanations appealing to derivative laws. Let's look at a plausible example from ecology where sometimes the actual causal influences would *not* still have been present if there had been disturbing factors.

By "one of ecology's few ironclad laws" (Pounds and Puschendorf 2004, p. 107), the number S of species in a given taxonomic group on an island in a given archipelago (e.g., land birds on Indonesian islands) is explained (see Fig. 2) by the island's area A and the "species-area relation" $S = cA^z$ (for non-zero constants c and z that are particular to the given taxon and archipelago). The law's conditions of application include that the islands have existed long enough for S to equilibrate, that life on Earth has not been wiped out by a cataclysm, that the archipelago's islands are isolated from one another by regions (typically, seawater) that the relevant species cannot inhabit, that the islands are all about equally distant from a source pool of potential immigrants (typically, a continent), and that the islands are each about equally diverse in their environmental conditions (e.g., temperature, elevation, rainfall). MacArthur and Wilson (1963, 1967) have given one influential proposal for explaining this law: a larger island tends to have larger available habitats for its species, so it can support larger populations of them, making chance extinctions less likely. Let's suppose as well that larger islands also present larger targets for stray creatures (Lomolino 1990). Therefore, under the law's conditions of application, larger islands have larger immigration rates and lower extinction rates, and so they tend to equilibrate at higher biodiversity. Nevertheless, a smaller island nearer the mainland may have greater biodiversity than a larger island much farther away. Likewise, a smaller island with much greater habitat heterogeneity may support greater biodiversity than a larger, much more homogeneous island. So the law can fail outside of its conditions of application.

In some respects, this example accords nicely with Jansson's proposal. Had the law's conditions of application not held because life on Earth was extinguished by a "nuclear winter", then the islands' areas would presumably have been no different but the numbers of species inhabiting them would obviously have been different. However, some possible "disturbing factors" are not independent of the islands' areas in that had the law's conditions of application failed in one of these ways, then the explainer *A* (not merely the explained *S*) would have been different. For instance, if the islands had differed greatly in that some of them contained a much more diverse range of elevations than others, then that might have been because a given volcano on one island had failed to collapse into the sea long ago. That island's current area *A* would then have been much larger than it actually is. Likewise, had the islands not been isolated by uninhabitable regions, that might have been because sea levels had been lower so that at low tide, there had been land bridges connecting some of the islands. But had sea levels been lower, some islands would have been larger; some of the submerged seabed surrounding them would instead have been above sea level at all times. Thus, there are many ways it can be that had the law's conditions of application failed, the explainer *A* would have been different. The explanatory factor *A* captured by the area law is not always independent of the presence of disturbing factors, and in those cases, Jansson's proposal will not yield the explanatory asymmetry.

Perhaps Jansson could address this problem by arguing that had the law's conditions of application failed in any of these ways, both *A* and *S* would have been different, and had the laws' conditions of application failed in other ways, then only *S* would have been different, but there are no ways for the law's conditions of application to fail under which *A* would have been different whereas *S* would not. I'm not sure that there are no such circumstances. In any event, my point in giving this case is not to give a one-off counterexample—to show that Jansson's approach sometimes fails to generate the explanatory asymmetry. Rather, my point is that this case suggest that even in those cases where Jansson's approach yields the correct answer (i.e., the actual explanatory asymmetry), it does so *not* because that explanatory asymmetry comes from a certain counterfactual asymmetry. Instead, even when the account yields the explanatory asymmetry, it does so only because the counterfactual asymmetry it identifies happens to run parallel to the explanatory asymmetry. In the "area law" case, that parallelism sometimes fails, and Jansson's account then fails to yield the explanatory asymmetry.

Although the species-area relation is often characterized as an ecological law, its nomic status remains controversial in island biogeography (see Kingsland 1995). Perhaps it is not a law at all; perhaps there are in fact no general laws of island biogeography. But this is not a problem for the use to which I am putting this example. I am offering it not as a simple counterexample to Jansson's account, but rather to show how it is possible for disturbing factors to interlock with the explanatory factors captured by a derivative law so that had certain disturbing factors been present, the explanatory factors might well have been different. The "area law" can illustrate how easily this can happen—especially in fields (like ecology) where complex multifactorial explanations are common—whether or not the "area law" turns out in fact to be an ecological law.

There is another serious obstacle that Jansson's account faces. In many non-causal explanations, especially those where the "law" is a mathematical fact, the "law" has no conditions of application, so it is difficult to see how Jansson's approach would apply. For instance, there are no limits to the conditions under which 3 fails to divide 23 evenly, but this "law" figures in the explanation of Mother's failure to divide her strawberries evenly.

Likewise, the "law" that a network lacking the Euler feature is non-traversable has no conditions of application but helps to explain why Jones failed in attempting to traverse Königsberg's bridges. Jansson might reply that the law's conditions of application include (for instance) that the bridges are the only way to get from one Königsberg island to another region of land—that there are no ferryboats, for instance.[11] After all (on this suggestion), had there been ferryboats, then it would have been possible to traverse all of the bridges, each exactly once, simply by crossing a bridge and then taking a ferryboat to the start of the next bridge. (And if there had been ferryboats, then the bridge arrangement would have been no different—exactly the asymmetry that Jansson emphasizes.) However, this suggestion fails to save Jansson's proposal: the absence of ferryboats is not a condition of the law's application because the explanandum is the impossibility of traversing (or the failure to traverse) each bridge exactly once *by a continuous, landlocked path* (etc.). To use ferryboats to "traverse" the bridges would be cheating. So even if there had been ferryboats, it would have been impossible to traverse the bridges *in the requisite way*. The law has no conditions of application to the relevant sort of bridge traversal.

Likewise, in Sect. 2.1 I mentioned that the posting of a threatening personage on one of the bridges does not make a bridge arrangement possessing the Euler feature non-travers*able* in the relevant sense. Of course, the posting of a threatening personage on a bridge in some traversable network could explain why Jones (who was frightened) failed in his attempt to traverse the bridges in that network. But the absence of threatening personages on the bridges is not a condition of application of the law that a network is traversable in the relevant sense iff it has the Euler feature.

That Jansson's approach fails to apply when the derivative law has no conditions of application could mean merely that Jansson's approach is not fully general; it must be supplemented with another approach for explanations where the derivative law has no conditions of application. Indeed, in Sect. 4 I will suggest a pluralist view myself—that in different non-causal explanations, different considerations establish the order of explanatory priority. But there is no reason independent of Jansson's account to have expected non-causal explanations involving laws that have conditions of application to work differently from non-causal explanations involving laws having no conditions limiting their application. If we found a promising account of the explanatory asymmetry of non-causal explanations involving laws having no conditions of application, we would naturally expect it to extend to explanations involving laws with conditions of application, rendering Jansson's account otiose.

[11] My thanks to a referee for suggesting that I consider such a possible reply on Jansson's behalf.

2.4 Conclusion

Let's step back from the specific proposals I have just discussed for using counterfactuals to derive explanatory asymmetry in non-causal explanations of singular facts. The overall strategy is to derive the asymmetry in *non-causal* explanations of singular facts by much the same means as the account uses to derive asymmetry in *causal* explanations of singular facts. The main obstacle to this strategy is that the account of causal explanation of singular facts can take for granted the asymmetries in causal arrows whereas an account of non-causal explanation, by contrast, cannot take for granted any asymmetries in non-causal arrows. That is, the account of non-causal explanation cannot begin by presupposing something asymmetric, such as (where, once again, ~E is lacking the Euler feature and N is non-traversability)

$$\text{Decisions by bridge designers and builders} \rightarrow \sim E \rightarrow N$$

If the first arrow is interpreted as causal, then it can safely be presupposed (since giving a philosophical account of causal relations and their asymmetry is an independent project from giving an account of non-causal explanation). But the second arrow cannot be causal. If the second arrow is to be understood as broadly logical (e.g., conceptual, mathematical) necessitation, then it can be presupposed. But it is $\leftarrow\rightarrow$, not merely $\rightarrow$, and so it cannot generate asymmetry. The same applies if the second arrow is interpreted as counterfactual dependence: it, too, is symmetric. Had ~E been different, then N would have been different, and vice versa.[12] If the second arrow is understood simply as non-causal explanation, then it is asymmetric ($\rightarrow$, not $\leftarrow\rightarrow$). But in an account of non-causal explanation, such an arrow cannot be presupposed; it must be accounted for. So we cannot account for the explanatory asymmetry here by claiming that there is a way to change ~E by an operation (e.g., building a bridge) that changes N (if at all) only as a non-causal result of ~E's change, but there is no way to change N except as a non-causal result of changing ~E. To make this claim about possible operations is simply to presume that the arrows of non-causal explanation point from ~E to N rather than from N to ~E.

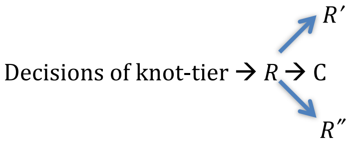
Let me make this point with one final example, returning to knot theory (from Sect. 2.1). Just as a bridge arrangement can be traversed only if it possesses the "Euler feature", a given knot can be untied only if it has a Laurent polynomial of 1. So it might seem that just as we can explain N by ~E, so we can explain C (that the knot cannot be untied) by R (that its Laurent polynomial is not 1):

[12] It might be suggested that the counterfactual-dependence arrow would be one-way if ~E grounded N. For instance, Socrates's existence grounds the existence of singleton {Socrates}. Had Socrates not existed, {Socrates} would not have existed. But had {Socrates} not existed (because nominalism about sets held), Socrates would still have existed. Could the asymmetry of non-causal scientific explanation of singular facts (where explanations supply information about counterfactual dependence) arise in this way?

Whether in the bridge example ~E (partially) grounds N depends on what "grounding" consists in—which is not a topic that I can address here. But whatever we may ultimately say about grounding, the counterfactuals in the strawberries example are not parallel to those in the Socrates case. Had Socrates not existed, {Socrates} would not have existed, but had nominalism about numbers held so that it is not the case that 3 fails to divide 23 evenly, Mother would still have been unable to divide her 23 strawberries evenly among her 3 children without cutting any.

Decisions of knot-tier in tying the knot $\rightarrow$ R $\rightarrow$ C

The first arrow is causal, but what about the second? It could be an arrow of entailment or of counterfactual dependence. But those arrows may run in both directions.[13] For instance, the arrow of counterfactual dependence is symmetric, since it can also run from C to R, as in this example from a mathematics article: "were [the trefoil knot] equivalent to the unknot [i.e., capable of being untied] it would … have 1 for its [Laurent] polynomial" (Millett and Lickorish 1988, p. 8). Furthermore, there are many other knot invariants besides the Laurent polynomial, such as tricolorability, the unknotting number, the bridge number, and the minimal crossing number. For each of these invariants, a knot can be untied only if the knot invariant has a certain value. So if the second arrow is entailment or counterfactual dependence, then each of these knot invariants (R', R'',…) has the same arrows as R. Which of them explains C? It may be that some of these knot invariants are more explanatorily fundamental than others (that is, some may be non-causally explained by others) and it may be that some explain C whereas others do not. For instance, perhaps the arrows of explanation are as follows:



But these distinctions among the knot invariants are not evident if the arrows represent entailment or counterfactual dependence; every one of the knot invariants will then have exactly the same arrows (from the knot-tier's decisions to it, and from it to C). The knot invariants may have different arrows if the arrows represent non-causal explanation.[14] But then an account of non-causal explanation must identify whatever is responsible for the order of explanatory precedence here; the account cannot simply build explanatory precedence into the acyclic graph and proceed from there.

3 Explanations of laws and the regularities they entail

3.1 Saatsi's and Pexton's account

Saatsi and Pexton (2013—hereafter "S&P") propose that explanations of lawlike regularities (whether the explanations are causal or non-causal[15]) work by supplying information about patterns of counterfactual dependence, thereby answering

[13] Some knot invariants give necessary and sufficient conditions for the knot's being capable of being untied. For instance, a knot on the surface of a torus is specified by a pair of coprime integers where it can be untied if and only if either integer is 1 or − 1.

[14] An analogy: Let C be that a given bishop currently on a black square (in a chess game) cannot ever occupy a given (white) square, let R be that bishops move only diagonally, and let R' be that "blackness" must be conserved in a bishop's moves. R' identifies a "bishop invariant" that entails C (and had R' not obtained, C would not have obtained). But R' does not explain C.

[15] S&P term some regularity explanations "causal" on p. 621, for example.

w-questions. In this respect, S&P say, explanations of lawlike regularities are like explanations of singular facts. However, in the case of explanations of lawlike regularities, the antecedents of the counterfactuals expressing these patterns of dependence do not posit interventions, unlike in explanations of singular facts. As an example, S&P sketch an allegedly non-causal explanation of the ¾ exponent in Kleiber's scaling law $B \propto M^{3/4}$ relating an organism's basal metabolic rate B to its mass M. The explanation consists of the law's derivation from the fact that organisms "employ fractal-like resource distributing networks" (620). The derivation acquires its explanatory power, S&P maintain, from its applicability to arbitrary dimension d, showing that the law's exponent is $d/(d+1)$ and thereby showing what the exponent would have been, had the organisms been of a different dimension. In this way, the explanation answers w-questions. Those questions do not concern what would have obtained under interventions since (S&P maintain) dimensionality is not capable of being the target of an intervention (as Woodward maintained regarding space's three-dimensionality explaining the stability of planetary orbits).

S&P acknowledge that interventions are "indispensable … in responding to a familiar puzzle about explanatory asymmetries in connection with explananda concerning singular states of affairs" (pp. 614–615). So without interventions, how does S&P's approach hope to account for asymmetry in explanations of these regularities?

S&P consider Woodward's (2003, pp. 187–193) example where from Coulomb's law (specifying a point charge's contribution to the electrostatic field), we can derive and thereby explain a regularity about the electrostatic field E at various distances r from a long, thin, straight wire with uniform charge density λ: that $E \propto \lambda/r$. According to Woodward, the source of this derivation's explanatory power is that the same sort of derivation answers w-questions about what the field would have been had the charge been distributed differently (e.g., nonuniformly on the wire, uniformly on a sphere). Although these counterfactuals all posit *interventions* on the charge distribution, S&P maintain that the derivation's explanatory power does not depend on this feature of the counterfactuals. Rather, S&P (616) say, all of the explanatory work is being done by the fact that the counterfactuals locate the explanandum within a range of alternative possibilities.

However, the positing of interventions and, more broadly, the direction of causal dependence seem to me indispensable to the account's capturing this explanation's asymmetry: that Coulomb's law explains the line-charge law, not vice versa. The derivation acquires its explanatory power partly by tracing the direction of causal dependence: that the charge elements in the wire cause the electric field rather than the reverse. If we ignore this causal element by considering counterfactual antecedents that do not posit interventions, then we open the door to an explanation that runs in the opposite direction. Of course, the line-charge law is not enough to entail Coulomb's law. However, suppose we ask why a point charge's contribution to a central electrostatic field is proportional to $1/r^2$ rather than to some other power of r. (This question presupposes that a point charge makes an r-dependent contribution to a central electrostatic field.) The $1/r^2$-dependence in Coulomb's law can then be derived from the line-charge law with its $1/r$-dependence, and the same sort of derivation can answer w-questions about what r's exponent in the point-charge law would have been had r's exponent in the line-charge law taken other values. This pattern of counterfactual dependence would have to be deemed explanatory if the counterfactuals relevant to explanation did not need to posit interventions. (These counterfactuals, which posit a different exponent to the line-charge law, are like the counterfactuals associated with the explanation of Kleiber's law, which posit a different dimensionality. None of them posits interventions, on Woodward's account.) An appeal to interventions blocks this direction of counterfactual dependence from counting as explanatory.

Thus, I disagree with S&P's view that the electrostatic derivation's explanatory power can be understood without reference to causation. (Indeed, Woodward regards this derivation of the line-charge law as a paradigmatically causal explanation.) What about a genuinely *non*-causal explanation of a regularity that follows from a law? Here S&P would seem to be on safer ground in setting interventions aside. However, I will now argue that even in the non-causal case, it is difficult to see how explanatory asymmetry can be generated.

Consider, for instance, special relativity's explanation of why the Lorentz transformations hold. The Lorentz transformations are deducible from various spacetime symmetries (such as the principle of relativity) together with the invariance of the spacetime interval (or the invariance of some finite speed). The same derivation, but with the temporal interval in place of the spacetime interval, yields the Galilean transformations in place of the Lorentz transformations. Thus, the derivation supplies answers to some w-questions, such as what the spacetime transformations would have been like, had the temporal interval been invariant instead of the spacetime interval.[16] Of course, the counterfactuals answering these w-questions do not posit interventions. S&P should find this congenial, since S&P hold that the power to explain regularities depends only on counterfactual dependence, not on the counterfactuals' tracing causal connections by positing interventions. Felline (2018) likewise sees length contraction as given a "structural explanation" by its counterfactual dependence on the spacetime interval's invariance.

However, such mere counterfactual dependence cannot suffice for such non-causal explanation. Instead of the spacetime interval's invariance, the relativity of simultaneity could be used (together with the same spacetime symmetries as before) to derive the Lorentz transformations. The same derivation, but with simultaneity's invariance in place of its relativity, yields the Galilean transformations in place of the Lorentz transformations. Thus, the derivation supplies answers to some w-questions, such as what the spacetime transformations would have been like, had simultaneity been invariant. (Again, these counterfactuals do not posit interventions.) So if counterfactual dependence sufficed for explanation here, then we would have to regard the Lorentz transformations as explained by the relativity of simultaneity (together with spacetime symmetries). But this is not the direction of explanation that is commonly accepted; rather, the Lorentz transformations and the relativity of simultaneity are thought to have a common explanation in the spacetime symmetries and the spacetime interval's invariance. By the same token, the relativistic formula for the addition of parallel

[16] For these derivations, see Lange 2017, pp. 96–112, 145–149. The symmetry principles sustaining these counterfactuals, such as the principle of relativity, are taken to be invariant under these counterfactual antecedents. Caution: we have two senses of "invariance" operating here! Invariance under counterfactual antecedents, which the principle of relativity possesses, is distinct from invariance across reference frames, which the speed c possesses.

velocities could be used (together with the same spacetime symmetries) to derive the Lorentz transformations. The same derivation, but with the classical velocity-addition formula in place of the relativistic formula, yields the Galilean transformations in place of the Lorentz transformations, again underwriting a counterfactual about what the transformations would then have been. But the velocity-addition law is not generally taken to explain why the Lorentz transformations hold. Rather, they have common explainers.

In short, the non-causal explanation of regularities cannot be identified simply with the existence of a pattern of counterfactual dependence because such a pattern can exist without corresponding explanations. S&P might reply by insisting that all counterfactual dependences are explanatory (even if explanatory circles would result). In fact, S&P (622) say that "intuitions about explanatory asymmetry are fragile or nonexistent for regularity explananda." But I am not sure that they truly believe this. For instance, they do not argue that an aspect of the line-charge law can explain something about Coulomb's law or that explanatory circles are harmless in the non-causal explanation of regularities.[17] For that matter, the scientific practice of explaining regularities is often very uniform in embracing certain directions of explanation rather than others. Famously, for example, spacetime symmetry principles are regarded as explaining conservation laws rather than the reverse. In fact, French and Saatsi (2018—hereafter "F&S") have recently proposed that this explanation "can be naturally captured in terms of a *counterfactual-dependence* account in the spirit of Woodward (2003), liberalized from its causal trappings" (p. 185). Let's examine F&S's proposal for capturing this explanatory asymmetry.

3.2 French's and Saatsi's account

Within a Lagrangian dynamical framework, various conservation laws (e.g., of energy, momentum, and angular momentum) are entailed by various symmetry principles (such as symmetry under arbitrary time-displacement, space-displacement, and rotation, respectively). But F&S (2018, p. 198) recognize that these derivations also run in reverse: the conservation laws (within a Lagrangian framework) entail the symmetries. (Noether's theorem underwrites both directions.) Therefore, although F&S (2018, p. 185) assert that "a symmetry fact … can contribute to provision of what-if-things-had-been-different information, showing how an explanandum *depends on* the symmetry," it seems that counterfactual dependence runs in both directions, allowing many more w-questions to be answered than correspond to explanatory relations. For instance, that a given physical system's Lagrangian is time-displacement symmetric (considering that this symmetry entails that the system's energy is conserved) tells us

[17] Admittedly, any test case (even one where there is overwhelming scientific agreement about the direction of explanation) might be questioned by an otherwise well-supported philosophical account of explanation. At a minimum, though, we would like an account of explanation to be able to explain away any places where scientific practice departs greatly from the account's verdicts—and, in a case where there is significant disagreement among scientists about the order of explanatory priority, we would like an account of explanation to illuminate the source of that disagreement. (For an example of a philosophical account of non-causal explanation identifying in one historical case the source of scientific disagreement about the direction of explanation, see Lange 2017, pp. 150–186).

that had the system existed at some other time with the same Lagrangian, it would still have conserved energy. But by the same token, that a given system conserves energy (considering that this conservation entails that its Lagrangian is time-displacement symmetric) tells us that had the system's kinetic energy initially been greater but its energy still conserved, then its Lagrangian would still have been time-displacement symmetric. F&S add another counterfactual dependence:

> [A]ssume that the closed system we are concerned with is the *whole* universe with its dynamical laws, represented via the Lagrangian, exhibiting certain symmetries. We can … answer counterfactual questions of the sort *'What if the universe were not symmetrical in this or that way?'* Answers to such what-if-things-had-been-different questions bring out the way in which particular conservation laws are counterfactually related to the symmetries at stake, even though it is not clear that counterfactuals regarding alternative symmetries can be interpreted in causal terms, with reference to possible manipulations or interventions. (2018, p. 200)

But F&S say nothing suggesting that this counterfactual dependence fails to run in reverse. Just as energy conservation would not have been a law if the universe's Lagrangian had not been time-displacement symmetric, so mustn't F&S say that the universe's Lagrangian would have failed to be time-displacement symmetric had energy conservation not been a law? Explanatory asymmetry seems to disappear once F&S allow explanations to arise from counterfactual dependence where the counterfactual antecedent does not posit an intervention.

However, F&S insist that in scientific practice, there is explanatory asymmetry here: symmetries explain conservation laws, not the reverse. To save this phenomenon, F&S aim to derive it from the notion of an intervention: not on the global symmetries of dynamical laws or on the global conservation laws (on which no intervention is possible), but rather on a feature of some particular physical system:

> The [system's] Lagrangian and its properties [such as its symmetries] reflect the relevant properties of the system being described: kinetic and potential energy functions, and whatever constraints there are to its dynamics.[18] When we consider changes to these features of the system, we consider changing, for example, the spatial distribution of mass or charge, or their quantity. These changes can have an effect on regularities manifested by the system as it evolves over time: different features of the system may become constants of motion … The point is that there is no way to alter these regularities concerning the system's behaviour – these constants of motion – directly as it were, without acting upon the features of the system that determine the system's behaviour. And it is the latter that feature in the Lagrangian, the symmetries of which thereby determine the constants of motion in a way that supports explanatory what-if-things-had-been-different counterfactuals. (2018, p. 199)

Any intervention on a particular system's momentum's remaining pointed in a certain direction (or on its Lagrangian's having a given symmetry) must change its momentum

[18] Here "constraint" is being used in the sense familiar from Lagrangian mechanics—e.g., that the rolling marble must remain in contact with the inclined plane, that the plane is rigid.

(or Lagrangian) by means of causing changes to "the features of the system that determine the system's behaviour" (such as the location or velocity of some system component or whether the system's bodies are electrically charged). That is, there is no possible intervention on the system's momentum having some value (or on its Lagrangian's having some symmetry) *with respect to* those underlying features (recalling Woodward's notion of "intervening on *X* with respect to *Y*" from Sect. 2.1). That those features are what figure in the Lagrangian is (according to F&S) the source of the explanatory priority of the system Lagrangian's symmetries over the system's conserved quantities:

> From the perspective of the counterfactual-dependence account, this explanatory priority [of symmetry over conservation] is underwritten by the fact that in a typical application of these results to a particular system (e.g., the solar system) there is a natural sense in which the conserved quantities *depend on* features of the system represented by the Lagrangian and its symmetries, but not the other way around. … Changing the potential energy function, either in its strength (by varying the amount of mass or charge at the centre), or in its spatial geometry by breaking the spherical symmetry in favour of some other symmetry, will have effects on the dynamical behaviour of bodies moving under the potential. These effects are reflected also in the regularities of the dynamics captured by the constants of motion. (2018, p. 199)

I will return to this argument, but before doing so, let me show where F&S think it leads. F&S take this explanatory asymmetry regarding a *particular system* to be the source of the explanatory asymmetry in the *general laws of nature*—that is, between symmetry principles and conservation laws:

> We think the reason that physicists often give explanatory priority to symmetries over conservation laws has to do with the fact that in analogous applications of Noether's theorem to particular *sub*systems of the universe, such as the central-force system examined above, the explanatory priority is transparent, partly due to the applicability of notions of manipulation and interventions. Explanatory reasoning about the relationship between conserved quantities and symmetries is naturally extended from such subsystems, involving e.g., central or harmonic forces, to symmetries of the laws covering the whole universe. Given the tight connection between conserved quantities and continuous symmetries in the Lagrangian framework – a connection which Noether's theorem captures in highly general terms – we naturally understand and explain conservation laws in terms of symmetries. This provides a non-causal explanation of particular conservation laws, capturing pervasive regularities of dynamical systems. (2018, pp. 200–201)

Thus, although (as we saw) F&S declare their aim to be the removal of "causal trappings" from the counterfactual-dependence account, their proposal surprisingly rests the explanatory asymmetry of this non-causal explanation (of global conservation laws by global spacetime symmetries) ultimately on the concept of an intervention. Some philosophers might think it inapt for an account of these non-causal explanations to appeal to a notion as thoroughly causal as "intervention". But F&S may deem it an insight that even these non-causal explanations are indebted to causal concepts for their direction of explanatory priority.

One problem with F&S's strategy, it seems to me, is that it begins by emphasizing that any intervention on a system Lagrangian's having a certain symmetry (or on a system's conserving a given quantity) must operate by manipulating one of the system's "features", such as a given system component's mass or charge. Since F&S's strategy is to begin with interventionist counterfactuals exclusively, those counterfactuals must specify how the Lagrangian's symmetry (or the conservation) would have been different under some intervention operating "directly" on those underlying features. As we just saw, F&S emphasize that "there is no way" to alter a given system Lagrangian's symmetry or the system's constant of motion "directly", i.e., other than by means of "acting upon the features of the system" such as the mass or charge of one of its components. Hence, if the counterfactuals associated with such explanations are interventionist counterfactuals, then the system's underlying features explain both its Lagrangian's having some symmetry and the system's conserving some quantity. The symmetry and conservation then have a common explainer, but neither is thereby given explanatory precedence over the other. Of course, if we relax the requirement that the counterfactuals associated with explanations be interventionist, then the system Lagrangian's having some symmetry is eligible to explain the system's conserving some quantity (since had that symmetry been absent, the system would not have conserved that quantity). But this counterfactual dependence goes in the other direction as well: had the system not conserved that quantity, its Lagrangian would not have possessed the corresponding symmetry. Again, no explanatory asymmetry arises.

In short: F&S maintain that for a particular system, its conserving various quantities can be altered only by acting on the components' masses or other such features. Its Lagrangian's symmetries can be altered only in this way as well. No interventionist asymmetry is thereby introduced between the Lagrangian's symmetries and the conservation—nor is any asymmetry introduced by the fact that the features of the system's components figure in the Lagrangian.[19]

This problem concerns F&S's account of how an individual system's having a symmetrical Lagrangian explains the system's conserving some quantity. But F&S's proposal also faces a problem in its next step: moving from individual systems to global symmetries and conservation laws. As we saw, F&S's approach aims to explain the explanatory asymmetry between spacetime symmetry principles and conservation laws by deriving that asymmetry from the fact that for an individual system, "[t]he

[19] Perhaps F&S think that the explanatory asymmetry arises from the way in which facts about the system's symmetries and conserved quantities can be deduced from facts about the system's manipulable properties. From the latter facts, we can derive the system's Lagrangian, then identify its symmetries, and from them infer that certain quantities are conserved. By contrast (F&S may be arguing), we cannot take the properties of (including the relations among) the system's components and infer directly that certain quantities are conserved, and from there infer the Lagrangian's symmetries. Rather, in Lagrangian mechanics, we can infer the conservation of certain quantities only by the intermediate step of first identifying the Lagrangian along with its symmetries.

To this strategy for saving explanatory asymmetry, I would reply that a different derivation would allow us to use the conserved quantities to arrive at the symmetries. We could begin by taking the components' properties and using Newtonian (rather than Lagrangian) mechanics to derive the system's time-evolution and thus that certain quantities are conserved. We could then use their conservation to deduce the Lagrangian's symmetries. Thus, explanatory asymmetry cannot be rescued in this way.

Lagrangian and its properties [such as its symmetry] reflect the relevant properties of the system being described [such as the distribution of mass or charge in the system]" (p. 199). However, by starting with individual systems (the features of which are subject to intervention) rather than the global principles (which are not subject to intervention), F&S's approach fails to reflect the fact that in modern physics, symmetry principles and conservation laws are understood to be modally much stronger than the features of individual systems and even the dynamical laws governing those systems. As Nobel laureate physicist Steven Weinberg (1992, p. 158) puts it, nowadays a symmetry principle is taken "as a fundamental fact … that stands on its own, independent of any detailed theory of nuclear forces." Whereas F&S focus on interventionist counterfactuals such as "Had this body possessed greater charge, the given system's Lagrangian would still have been time-displacement symmetric," the symmetry principle's modal status is better reflected in non-interventionist counterfactuals such as that time-displacement symmetry would still have obtained even if the dynamical laws had been different—for instance, even if there had been additional fundamental kinds of forces besides those specified by the various actual force laws.

Some philosophers [such as Morrison (1995) and Lange (2017)] and some physicists [such as Wigner (1972, p. 10) and Greene (2005, p. 225)] have tried to capture the role of spacetime symmetries in modern physics by characterizing these symmetry principles as "meta-laws". As laws governing first-order laws, meta-laws constrain what the first-order dynamical laws could have been and hence what individual Lagrangians there could have been.[20] By contrast, no interventionist counterfactual can capture the symmetry principles' status as meta-laws since no interventionist counterfactual posits additional kinds of fundamental forces or other changes to the fundamental dynamical laws. Thus, it would be very difficult for an approach like F&S's to reflect the symmetry principles' status that ultimately gives them explanatory priority over conservation laws.

The fact that symmetry principles are laws "which the [first-order] laws of nature have to obey" (Wigner 1985, p. 700), transcending the details of the first-order laws, was the revolutionary discovery about symmetries that motivated much of twentieth century fundamental physics. Nobel physics laureate David Gross (1996, p. 14256) contrasts the superseded view of symmetries and conservation laws as mere "consequences of the dynamical laws of nature" with the "great advance" of 20$^{\text{th}}$-century physics that "put[s] symmetry first [by] regard[ing] the symmetry principles as the primary feature of nature that constrains the allowable dynamical laws." By focusing on interventions on individual systems, F&S's approach cannot do justice to the framework that leads physicists to understand symmetry principles as explaining conservation laws.

4 Conclusion

I have argued that serious obstacles face various recent attempts to use counterfactuals to understand the asymmetry of non-causal scientific explanations. Although I doubt

[20] This is not the notion of "constraint" in note 18.

that counterfactual approaches will find a way around these problems, let me close by suggesting a way forward in understanding explanatory asymmetry in non-causal scientific explanations.[21]

I suggest that there is no fully general account of what makes some facts explanatorily prior to others in non-causal scientific explanations. Rather, the order of explanatory priority is fixed by different considerations in different non-causal explanations. For instance, in some non-causal explanations, a fact in the explanans derives its explanatory priority over the explanandum from its possessing greater modal strength than the explanandum has. For example (as I discussed in the previous subsection), a symmetry principle (such as the principle of relativity) is a meta-law governing the first-order laws analogously to the way that the first-order laws govern particular facts. For this reason, the principle of relativity can help to explain a first-order law such as the Lorentz transformations (as mentioned in Sect. 3.1 above), and likewise time-displacement symmetry can help to explain energy conservation (as mentioned in Sect. 3.2).

In other non-causal explanations, however, the explanans and explanandum do not differ in modal strength. Then the order of explanatory priority is fixed by other considerations. As we saw Woodward maintain (see Sect. 2.1 above), spacetime substantivalism takes space's three-dimensionality as helping to non-causally explain the possibility of stable planetary orbits. That is because substantivalism (I suggest) regards space as a kind of theater stage, its geometry constraining the individual actors. The container's features impose limits on its contents and thereby help to explain certain features of them.

In other non-causal explanations, still other considerations are responsible for the direction of explanation. The Lorentz transformations and the relativistic formula for the addition of parallel velocities concern frame-dependent quantities (such as spatial intervals, temporal intervals, and speeds). These quantities reflect not only reality, but also the reference frame from which events are being described. By contrast, the spacetime interval is invariant; it reflects reality alone, uncontaminated by the choice of reference frame.[22] Facts about reality help to explain facts about appearances, not the reverse. Accordingly, the spacetime interval's invariance is explanatorily prior to the Lorentz transformations and the relativistic velocity-addition law.

Of course, I have only sketched the approach I recommend. There is no way to guarantee in advance that it will not run into some of the same problems as the counterfactual approaches that I have examined. However, it does have one dimension of flexibility that they do not have: it is not limited to appealing to considerations originally developed for an interventionist account of causal explanations. For instance, when Woodward tries to account for the substantivalist's view that space's three-

[21] I cannot do more here than sketch the following positive views. I elaborate and defend them more fully in Lange (2017).

[22] These ideas are often expressed. For instance: "In physics, the frame-dependent quantities…are taken to be non-fundamental. … Frame-independent quantities, on the other hand, *do* correspond to fundamental, objective features of the world. The space–time interval *is* a fundamental, objective feature of the world, according to the theory of special relativity. … Reality is observer-independent. It does not depend on our arbitrary descriptions or conventions." (North 2009, pp. 63, 67) For many similar passages, see Lange (2017, pp. 141–145).

dimensionality (*D*) takes explanatory priority over the possibility of stable planetary orbits (*P*), Woodward must find some way to capture this explanatory priority by somehow extending the framework developed to handle interventions. As we saw in Sect. 2.1, he attempts to exploit the "independence" condition for interventions, which can be made to yield the right answer, but only thanks to the fortunate role played by Newton's laws and the gravitational potential in connecting *D* and *P*. Some cases have no auxiliary hypotheses playing this sort of role, and so this strategy cannot be applied to them. By contrast, as we have just seen, the approach I favor gives substantivalism a much more direct role in establishing explanatory priority.

The problem of explanatory asymmetry in *causal* explanation has received great attention. I predict that the same will ultimately be true regarding the problem of explanatory asymmetry in *non*-causal scientific explanation.[23]

References

Alan Pounds, J., & Puschendorf, R. (2004). Ecology: Clouded futures. *Nature, 427*(6970), 107–109.

Baron, S., Colyvan, M., & Ripley, D. (2017). How mathematics can make a difference. *Philosophers' Imprint, 17*(3), 1–19.

Callender, C. (2005). Answers in search of a question: "Proofs" of the tri-dimensionality of space. *Studies in History and Philosophy of Modern Physics, 36,* 113–136.

Ehrenfest, P. (1917). In what way does it become manifest in the fundamental laws of physics that space has three dimensions? *Proceedings of the Amsterdam Academy, 20,* 200–209.

Felline, L. (2018). Mechanisms meet structural explanation. *Synthese, 195,* 99–114.

French, S., & Saatsi, J. (2018). Symmetries and explanatory dependencies in physics. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations* (pp. 185–205). Oxford: Oxford University Press.

Greene, B. (2005). *The fabric of the cosmos*. New York: Vintage.

Gross, D. (1996). The role of symmetry in fundamental physics. *Proceedings of the National Academy of Sciences USA*, *93*, 14256–14259.

Jansson, L. (2015). Explanatory asymmetries: Laws of nature rehabilitated. *Journal of Philosophy, 112,* 577–599.

Jansson, L., & Saatsi, J. (forthcoming). Explanatory abstractions. *British Journal for the Philosophy of Science.*

Kingsland, S. (1995). *Modeling nature*. Chicago: University of Chicago Press.

Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In W. Salmon & P. Kitcher (Eds.), *Scientific explanation, minnesota studies in the philosophy of science* (Vol. 13, pp. 410–505). Minneapolis: University of Minnesota Press.

Lange, M. (2013). What makes a scientific explanation distinctively mathematical? *British Journal for the Philosophy of Science, 64,* 485–511.

Lange, M. (2017). *Because without cause*. New York: Oxford University Press.

Lomolino, M. V. (1990). The target area hypothesis: The influence of island area on immigration rates of non-volant mammals. *Oikos, 57,* 297–300.

MacArthur, R. H. (1972). *Geographic ecology*. Princeton: Princeton University Press.

MacArthur, R. H., & Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution, 17,* 373–387.

MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. Princeton: Princeton University Press.

Millett, K. C., & Lickorish, W. B. R. (1988). The new polynomial invariants of knots and links. *Mathematics Magazine, 61,* 3–25.

23 My thanks to Chris Dorst, Alex Reutlinger, and two referees for helpful comments on earlier drafts.

Morrison, M. (1995). The new aspect: Symmetries as meta-laws—structural metaphysics. In F. Weinert (Ed.), *Laws of nature: Essays on the philosophical, scientific, and historical dimensions* (pp. 157–188). Berlin: de Gruyter.

North, J. (2009). The 'structure' of physics: A case study. *Journal of Philosophy, 106,* 57–88.

Pincock, C. (2007). A role for mathematics in the physical sciences. *Nous, 41,* 253–275.

Reutlinger, A. (2017). Does the counterfactual theory of explanation apply to non-causal explanations in metaphysics? *European Journal for Philosophy of Science, 7,* 239–256.

Reutlinger, A., & Saatsi, J. (Eds.). (2018). *Explanation beyond causation*. Oxford: Oxford University Press.

Saatsi, J., & Pexton, M. (2013). Reassessing Woodward's account of explanation: Regularities, counterfactuals, and noncausal explanations. *Philosophy of Science, 80*(5), 613–624.

Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Salmon, W. (1989). Four decades of scientific explanation. In W. Salmon & P. Kitcher (Eds.), *Scientific explanation, minnesota studies in the philosophy of science* (Vol. 13, pp. 3–219). Minneapolis: University of Minnesota Press.

Weinberg, S. (1992). *Dreams of a final theory*. New York: Pantheon.

Wigner, E. (1972). Events, laws of nature, and invariance principles. In J. Schwinger (Ed.), *Nobel lectures: Physics 1963–1970* (pp. 6–19). Amsterdam: Elsevier.

Wigner, E. (1985). Events, Laws of Nature, and Invariance Principles. In A. Zuchichi (Ed.), *How far are we from the gauge forces* (pp. 699–708). New York: Plenum.

Woodward, J. (2003). *Making things happen*. New York: Oxford.

Woodward, J. (2018). Some varieties of non-causal explanation. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations* (pp. 117–140). Oxford: Oxford University Press.