Reassessing Woodward's account of explanation: regularities, counterfactuals, and non-causal explanations

Juha Saatsi and Mark Pexton

10th January 2013

Abstract

We reassess Woodward's counterfactual account of explanation in relation to regularity explananda. Woodward (2005) presents an account of causal explanation. We argue, by using an explanation of Kleiber's law to illustrate, that the account can cover also some non-causal explanations. This leads to a tension between the two key aspects of Woodward's account: the counterfactual aspect and the causal aspect. We explore this tension and make a case for jettisoning the causal aspect as constitutive of explanatory power in connection with regularity explananda.

1 Introduction

According to a currently popular view, explaining in science is a matter of providing *counterfactual information* about the world. (Woodward 2003; Woodward and Hitchcock 2003; Hitchcock and Woodward 2003; Ylikoski and Kuorikoski 2001) Explanatory power is a matter of providing information that answers *what-if-things-had-been-different questions* ('w-questions'), as Woodward puts it.

> [A]n explanation ought to be such that [it enables] us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways. (2003, 11)

For Woodward the explanatory information is typically also *causal* information; his is a counterfactual 'manipulationist' account of causal explanation.

> [One] ought to be able to associate with any successful explanation a hypothetical or counterfactual experiment that shows us that and how manipulation of the factors mentioned in the explanation would be a way of manipulating or altering the phenomenon explained. (ibid.)

Woodward's account of explanation, involving both counterfactual and causal information, is Janus-faced in this way, but largely harmonious due to Woodward's (non-reductive, circular) counterfactual 'analysis' of causation, allowing explanatory modal information to be often interpreted as causal information. The two 'faces' are not joined at the hip, however. With half-hearted reference to potential examples of non-causal explanations, Woodward happily welcomes the *possibility* that the counterfactual aspect of his account may come apart from its causal aspect, so that the counterfactual idea may be applicable to some genuinely non-causal explanations. (cf. 2003, §5.9)

Others have similarly welcomed the potential of Woodward's account to naturally subsume non-causal explanations. (e.g. Bokulich 2008, 2011) We also view as more fundamental the counterfactual aspect of the account. To motivate this we will present (§3) one concrete example of a non-causal explanation that can be analysed in counterfactual terms, in close parallel with one of Woodward's *exemplars* of causal explanation. This supports the idea that the counterfactual analysis of explanation should not be wedded to a causal manipulationist interpretation of explanatory modal information.

If we accept that some non-causal explanations can be subsumed under Woodward's counterfactual account, there's work to be done. The ensuing separation of the two aspects of the account leads to a subtle internal tension, the main topic of this paper. We can begin to sense the tension with the following question. If counterfactual information pure and simple—without the possibility of causal interpretation—can *sometimes* function as a source of explanatory power, then what indispensable *explanatory* role is there for the causal interpretation of counterfactual information in those cases in which such interpretation happens to be available? In other words: given that some of Woodward's causal explanations are relevantly similar—in terms of their overall explanatory profile—to some non-causal explanations in which counterfactual information suffices to explain, what further philosophical work (vis-à-vis the account of explanatory power) is done by the causal talk in connection with those explanations?

One might suggest that in the interest of offering a unified account of both causal and non-causal explanations the theory should eschew giving a central role to *contingent* features of explanatory counterfactual information, such as the 'manipulability' or otherwise of some variables. Accordingly, the suggestion would have it, causal explanations should not be seen as explanatory *by virtue of* being causal. Rather, the causal aspect should be viewed as a contingent feature independent of the analysis of explanatory power.

Would it be an option for Woodward to relinquish the causal aspect of his theory in the interest of such a unified account? Perhaps, but this comes with a price. For the causal aspect of Woodward's account plays an indispensable role in responding to a familiar puzzle about explanatory asymmetries in connection with explananda concerning singular states of affairs.[1] And retaining the causal aspect of the theory for singular states of affairs, whilst relinquishing it for regularity explanations, goes against providing a unified account of these latter two types of explanation, which is something Woodward explicitly aims at.

In sum, there is a tension in Woodward's account between (i) the aim of offering a unified account of causal explanations of singular states of affairs and causal explanations of regularities, on the one hand, and (ii) offering a unified account of causal and non-causal explanations of regularities, on the other hand. One or the other aim has to be given up, and there is work to be done to determine which one it should be. There is also work to be done to explain the pertinent difference between the two types of explanation that lie at the root of the ensuing, unavoidable disunity.

In the rest of the paper we explore further this tension and motivate giving up (i). Section 2 reviews the relevant features of Woodward's account and his key exemplar of causal explanation. Section 3 presents an outline of a non-causal, geometrical explanation of allometric scaling laws, that nicely fits (we submit) Woodward's broader framework. Section 4 examines the tension that arises from subsuming this non-causal explanation under Woodward's account. Finally, having taken a stance on how the tension should be interpreted, in section 5 we conclude by comparing clarifying our position and comparing it with Bokulich (2008, 2011).

[1]The causal aspect of Woodward's analysis also plays an indispensable role in specifying the truth-conditions for a large class of counterfactual statements. But this role in the 'semantics' of counterfactuals can be taken to be independent from Woodward's analysis of explanatory power in terms of the amount of counterfactual information provided.

2 Woodward's account

We now review the key aspects of Woodward's account (as presented in 2003, ch. 5). Woodward begins by contrasting an exemplar of a textbook physics explanation with a counter-example to the DN-model: a caricature 'explanation' of why a certain raven a is black.

> All ravens are black.
>
> a is a raven.
>
> a is black.

Let's agree that this deduction of 'a is black' is not explanatory of a's blackness. In contrast to this, Woodward presents as an exemplar of genuine explanation a physics textbook deduction that exhibits 'systematic patterns of counterfactual dependence.' The explanandum at stake is a *regularity*: the magnitude of the electric intensity at a perpendicular distance r from a very long fine wire with a positive, uniform charge distribution, as given by

$$E = \frac{1}{2\pi\epsilon_0}\frac{\lambda}{r} \tag{1}$$

where λ is the charge per unit length on the wire, and E is at right angle to the wire.

The explanation of this regularity is given by a derivation of equation (1) by using Coulomb's law to integrate over infinitesimal charge elements in the wire. This derivation is explanatory, but not simply by virtue of being a deduction (in a model) from Coulomb's law and the boundary conditions. Rather, says Woodward, the derivation (within its theoretical context) is explanatory by virtue of providing *counterfactual information*: it shows how the explanandum *would change* if the initial and boundary conditions used in the derivation *were changed* in various ways.

In other words: the derivation employes ingredients that suffice to answer a range of w-questions, bringing out the dependence of the regularity expressed in equation (1) on factors such as the shape of the wire, the specific charge distribution, etc.

> We can see from the [derivation] how certain factors (e.g., the geometry of the conductor, the distribution of charge along it, in some cases the distance from the conductor) all make a systematic difference to the intensity and direction of the field. (2003, 192)

It's worth also noting that Coulomb's 'law' need not be a universal truth for it to play a role in providing such explanatory information. It is enough that it is *invariant* with respect a range of potential alternative values in the relevant variables, such as the geometry of the wire, for example.[2]

We think that Woodward's exemplar effectively and convincingly captures the contrast between a genuinely explanatory derivation and a non-explanatory deduction. We also think that all the work here is done by the categorical difference in the counterfactual information provided. The deduction of the raven's blackness does not answer *any* appropriate w-questions about the explanandum; it does not in any way locate the latter within a range of alternative possibilities. By contrast, the explanation from electrostatics readily leads us to see the actual explanandum as one of a range of possible alternatives, corresponding to different boundary conditions (e.g. different shapes of the wire). The electrostatics' explanation is explanatory by virtue of encompassing this counterfactual information.

In presenting the exemplar and the contrast Woodward makes occasional, passing reference to causation. He notes, for example, that the counterfactual information at stake is 'also information that is relevant to the manipulation and control of the phenomena described by these explananda.' (2003, 191) He also states, without

[2]More precisely, Coulomb's law is in a technical sense *invariant-under-intervention* on these variables.

any explicit justification, that:

> [T]he demand that a successful explanation answer w-questions is tantamount to the requirement that explanations must provide information about the causes of their explananda, if "cause" is understood along the manipulationist lines... (2003: 194)

There is nothing in Woodward's discussion of the exemplar *per se* that justifies, as a general conclusion, this step from the demand for counterfactual information to a demand for causal information. (Call this step the 'manipulationist assumption.') To be sure, the counterfactual information in this exemplar *happens to be* such that it can be construed as causal information (when armed with Woodward's interventionist conception of causation).[3] But the requisite contrast to the non-explanatory raven case gets drawn purely in terms of the mere *existence* of counterfactual information, without any further qualification regarding the nature of that information. In as far as the exemplar is concerned, all we learn about explanation is that it hangs on the provision of counterfactual information.

Later in the chapter there is a justification to be found for Woodward's insistence on the causal interpretation of explanatory counterfactual information. This turns on the problem of explanatory asymmetries. The problem is well-known as a counterexample to the DN-model.[4] For example, the period T of a simple pendulum can be explained in terms of its length l and the gravitational field g, by appealing to the 'law':

$$T = 2\pi\sqrt{l/g} \tag{2}$$

[3]The shape of the electric wire, for example, can be construed as a cause of a specific electric intensity at a particular point outside the wire, so that manipulating the shape causes different electric intensities.

[4]Woodward also appeals to the classic problem of 'explanatory irrelevance' to motivate the causal aspect of his account. But the explanatory failure of cases such as the man-on-the-pill who fails to get pregnant has to do with the lack of counterfactual information simpliciter.

But we arguably cannot explain the length of the pendulum in terms T and g by appealing to a simple transform of (2):

$$l = \frac{T^2 g}{4\pi^2} \tag{3}$$

Yet, if we assume that the (mathematically equivalent) equations (2) and (3) are invariant under a range of relevant counterfactual suppositions regarding alternative values of the three variables, then (3) does provide us counterfactual information concerning states of affairs such as:

> If the period T were twice as long, then the length would have to be four times as long (for constant g).

If to explain is just to provide counterfactual information, then by using (3) we can explain the length in terms of the period. But surely the period does not explain the length! The lesson, says Woodward, is that not all counterfactual information is explanatory: in order to explain the counterfactual information has to track objective causal dependence. Since the length is not caused by the period—a notion Woodward makes precise in his interventionist framework—we cannot use equation (3) (read 'from right to left') to explain.

Woodward's response to the explanatory asymmetry problem has a great deal of intuitively pull. But at the same time it sits uncomfortably with the fact—the potential of which Woodward himself envisioned—that in *some* cases counterfactual information *can* be explanatory without tracking any causal dependence relations. We will return to this tension after offering an example of such a case.

3 An explanation of allometric scaling laws

'Allometric scaling laws' in biology refer to regularities that relate an organism's body mass M to some biological observable Y, via a simple proportionality: $Y = Y_0 \cdot M^b$, where Y_0 is a normalisation constant, and b is the scaling exponent. Among a vast array of scaling regularities there are some that biologists have found both impressive and puzzling. The most famous scaling regularity, *Kleiber's law*

$$B \propto M^{3/4} \tag{4}$$

relates basal metabolic rate to body mass. Arguably it has been observed to extend over 21 orders of magnitude. (West et al. 2000) In addition to this truly impressive range, the specific scaling exponent in Kleiber's law, $b = 3/4$, has been a subject of numerous studies. On simple geometrical grounds one might expect metabolic efficiency to conform to Euclidean scaling, suggesting for example that $b = 2/3$ as given by the surface-to-volume ratio. Furthermore, Kleiber's law is but one amongst several allometric scaling laws with '1/4' factor in the exponent.[5] Whence the ubiquitous 'quarter-power scaling'?

In order to answer this, mathematical models have been devised in theoretical systems biology to derive Kleiber's law. We will focus on a model by Brown, Enqvist, and West, not because it is uncontroversial or universally accepted, but because it provides a significant potential explanation that exhibits a close parallel to Woodward's exemplar. There are two components to the overall explanation provided by these theorists: a claim about the evolutionary history and a claim about the evolutionary optimality of some ubiquitous geometrical features of *life-supporting networks*. The mathematical model relates to the latter claim.

At the heart of the model are two basic sets of assumptions. First of all, all

[5] Here are some: for heart rate $b \approx -1/4$; for life span $b \approx 1/4$; for aorta diameter $b \approx 3/8$

life is supported by resource-distributing networks, such blood circulatory or plant vascular systems. Secondly, there are three simple constraints on these networks. (a) The supply of resources must be democratic, serving all of the organism. (b) The final branches of a network are assumed to be size-invariant units: e.g. capillaries in blood circulation. (c) The energy required for the distribution of resources through the network is minimised. From these extremely general assumptions the model derives the specific, ubiquitous geometrical characteristics of life-supporting networks and the $3/4$ scaling exponent in Kleiber's law.

In very rough outline, the derivation proceeds as follows. First, the metabolic rate is related to fluid flow through the network, as the former is proportional to oxygen consumption, which itself is proportional to the rate of fluid flow. Then, fluid flow (and resource distribution) in a network is characterised in terms of a small number of variables: length and radius of a branch segment at a given level branching, the branching ratio, etc. These variables quantify also the hierarchical branching structure, and are deduced in the model from the assumptions (a)–(c). From (a) and (b) it follows that both the branching ratio and the ratio of subsequent branch lengths are (on average) constant throughout the network: the geometry of the network is a self-similar and fractal-like. From (c) it follows that the branching is area-preserving: the sum of daughter branches' cross-sections is equal to that of their parent. Finally, from (c) it also follows that the total fluid volume is proportional to body mass, and from (b) that the total number of capillaries (the final network units) is proportional to a power b of body mass, where b is just the scaling exponent in Kleiber's law. From the ensuing equations b can be solved: $b = 3/4$. (For details see West et al., 2000)

The derivation takes place in a model incorporating idealisations comparable to those in Woodward's exemplar. But what is it about the derivation that renders it explanatory of the quarter-power scaling? Here it is natural to appeal to Woodward's idea that the explanatory power of the derivation lies in the counterfactual information it provides. Studying the derivation (within its theoretical context) enables us to answer a range of w-questions. For example, it is important that the derivation of the fractal-like geometry of the network, and the utilisation of this geometry in determining the scaling-exponent, are straightforwardly generalizable to a network in an arbitrary dimension d. That is, for d-dimensional organisms we have the scaling exponent $b = d/(d + 1)$. (West et al., 2000) The increase in the scaling exponent from a corresponding Euclidean scaling can also be related in general geometrical terms to the fractal-like nature of the network. (West et al., 1999) The model seems to explain by virtue of relating the actual $3/4$ scaling-exponent in this way to the fact that the organisms covered by Kleiber's law are three-dimensional, and employ fractal-like resource distributing networks. The quarter-power scaling is thus located within a range of alternative possibilities: had things been suitably different so that all organisms were effectively two dimensional (like the flat-worm), then the scaling exponent would be $2/3$, for example.

In this way the derivation shows how the scaling exponent counterfactually varies with the dimensionality of organisms.[6] But this explanatory modal information is not easily construed as causal dependence: the scaling *regularity* does not causally depend on the dimensionality of organisms, and it is not natural to think of dimensionality as a variable that could be intervened with to somehow manipulate the explanandum.[7] What we have is a geometrical, non-causal explanation that complements the complex causal story that explains the ubiquitous evolution of thus optimised networks.

In addition to answering w-questions, the derivation also incorporates an element that is invariant with respect to the explanans variables. Namely, the deriva-

[6]The derivation also displays the dependence of the scaling exponent on other contingent, non-geometric assumptions that feed into the derivation.

[7]The explanandum concerns a regularity that spans a massive range of different types of organisms, and there is no corresponding explanandum for an individual organism or even a single species.

tion turns on the invariant geometrical fact that for the types of systems in question the effective fractality of a network leads to an increase in the scaling exponent. (West et al., 1999)

All in all, this non-causal explanation closely matches Woodward's exemplar from electrostatics in its general explanatory profile, apart from the manipulability aspect.

4 A unified account of regularity explanations?

The close parallel between the two explanations strongly suggests that this particular non-causal explanation can be subsumed under Woodward's account. Hence, counterfactual information can sometimes be explanatory without being simultaneously causal information. But if this is correct, then we should wonder what contribution to explanatory power in Woodward's exemplar is due to the fact that counterfactual information in this case can be construed as causal information? Why is explanatory power sometimes down to tracking causal relations, sometimes not?

One can have a unified account of the two explanations only by dropping Woodward's insistence in taking the explanatoriness of his exemplar as somehow deriving from the causal character of the relevant counterfactual information. A unified account would take explanatory power in both cases to be a matter of answering a range of w-questions by providing objective counterfactual information, period. The difference in the character of that counterfactual information, its semantics, its connections to our capacity to manipulate the world, and so on, would not come into play in accounting for the *explanatoriness* of either explanation. All that matters for explanatory power, it would be said, is there being objective facts about the relevant counterfactual circumstances, and having epistemic access to these facts.

We find the aim of achieving of a unified account of the two regularity explanations attractive. There is logical room for this, since nothing in Woodward's account shows that manipulability *per se* actually plays an explanatory role in the exemplar, instead of being a merely contingent feature of the relevant modal information. What Woodward has shown, rather, is that manipulability plays an indispensable role *elsewhere* in his account, in responding to the explanatory asymmetry problem in connection with *singular states of affairs*, concerning, for example, the length of a *particular* pendulum. Unless one has an alternative solution to offer to the asymmetry problem, here the causal aspect really must be viewed as constitutive of explanatory counterfactual information. So, a unified account of causal and non-causal regularity explanations is possible, but it has a price: such an account cannot give a unified analysis of explanatory power that covers both explanations of regularities and singular states of affairs.

Such a unified account requires driving a wedge between explanations of regularities and singular states of affairs. How steep a cost is this? This depends on how the accounts of these two kinds of explanations can be made to cohere. We do not attempt to answer this complex question here. What we will now do, rather, is make a case for accepting that explanations of regularities may fundamentally differ from explanations of singular states of affairs in a way that motivates driving the wedge here.

It is noteworthy, first of all, that all the intuitions supporting the causal aspect of Woodward's account derive from cases concerning individual states of affairs. These intuitions can be accommodated by accepting that we cannot explain the length of a *particular* pendulum, for example, purely by citing modal facts regarding different possible values of its period and the surrounding gravitational field (since the latter do not cause the former), whilst maintaining that explaining *a regularity* concerning all pendulums is *just* a matter of supplying such modal information.

Secondly, intuitions about explanatory asymmetry are fragile or non-existent for regularity explananda. Take, for example, the regularity that for all simple pendulums of 1 second period, the length of each pendulum is directly proportional to the strength of the gravitational field g (as opposed to some other other power of g).[8]

$$l \propto g \quad (T = 1\mathrm{s}) \tag{5}$$

It seems perfectly sensible to explain this regularity by deducing it from equation (3). The derivation is rather simple, of course, but not non-explanatory like the deduction of raven a's blackness. The explanatory power of the derivation turns on counterfactual information that does not track causal dependency: equation (3) brings out, for example, the fact that had all the pendulums in the sample had some alternative constant period, l would have been similarly directly proportional to g. Thus, the derivation (within its theoretical context) locates the explanandum within a range of alternative possibilities. This explanation seems explanatorily analogous to Woodward's exemplar from electrostatics, apart from not satisfying the manipulationist assumption. And if this is not a bona fide explanation at all (against our intuitions), a case has to be made for this that does not turn on the manipulationist assumption (on pain of begging the question).

Thirdly, insisting on the causal character of explanatory counterfactual information has counterintuitive consequences in connection with regularity explananda, as illustrated by the following example. According to Woodward, the ideal gas law in conjunction with temperature T and pressure P can explain the volume of a

[8]For a more striking thought-experiment effect, assume that in our environment there is a lot of variation in g, but little variation in pendulum periods.

particular container of N moles of gas if and only if that container is *non-rigid* so as to causally respond to manipulation of P and T. (2003, 234) This seems right regarding explanations of singular states of affairs. But now consider a *regularity* exhibited by all containers of 1 mole of (ideal-enough) gas at 10 ˚C: their volume is inversely proportional to pressure:

$$V \propto P^{-1} \tag{6}$$

Can we explain this regularity on the basis of the ideal gas law? Adopting the manipulationist assumption implies that the answer is 'no', *just in case some of the containers in question happen to be rigid.* For if some of the containers are rigid, then we cannot construe the variable V as something that could be manipulated by causally intervening on P. But it is counterintuitive to claim that whether or not this regularity can be explained depends on the rigidity or otherwise of the individual containers. It is more natural to think that the ideal gas law explains (6) regardless, and it always does so simply by providing counterfactual information.

5 Conclusion

An advocate Woodward's counterfactual theory of explanation can well try to offering a unified account of explanation that subsumes both causal and non-causal regularity explanations. This is motivated by real examples of non-causal explanations from science that fit Woodward's central idea that explanatory power is a matter of providing counterfactual information.[9] The notion of causation as a constitutive feature of explanatoriness is well motivated *only* for explanations of singular states of affairs. This notion becomes unmotivated and even counterintuitive when extended to regularity explanations.

[9]There are various non-causal, typically geometrical explanations that could potentially be analysed in these terms.

Nothing we have said amounts to denying that explanation and causation are intimately connected, even for regularity explananda. For example, we don't deny that (type-level) causal claims are 'explanatory by virtue of providing counterfactual information' (Woodward 2003, 205). And it still makes sense to talk about a 'causal explanation' in connection with Woodward's exemplar, for example, to signify the important fact that the relevant counterfactual information can be causally construed. All this is compatible with denying that when some fact is explanatory of a regularity it is so *by virtue* of being causal. The lesson is that the causal dimension of the exemplar need not be *constitutive* of its explanatory power.

We are not alone in scoping out the potential of Woodward's account to capture various non-causal explanations. But we wish to stay much closer to Woodward's own line of thought than Bokulich (2008; 2011), for example, who claims to

> adopt Woodward's account of explanation [but to jettison a construal of] modal dependence narrowly in terms of the possible causal manipulations of the system. (2008, 226)

In Bokulich's account of 'structural (model) explanation' scientific models—and even fictional elements in those models—can explain if 'the counterfactual structure' of a model is 'isomorphic in the relevant respects to the counterfactual structure' of the system being explained, where the explanandum exhibits a 'pattern of dependence' on the elements of the model as a consequence of the structural features of the theory employed. (2009, 39-40)

In as far as we understand the proposal, we worry that Bokulich's construal of Woodward's account is too liberal in its conception of explanatory modal information. Even if it is not inconsistent with Woodward's ontological conception of explanation—viz. genuine explanatory power comes only from latching onto relevant objective features of the world—we are not convinced that capturing a *mere* *abstract structure* ('up to isomorphism') of counterfactual relations can be explanatory. Whilst we also happily jettison the causal aspect of Woodward's account (in connection with regularity explanations), we still hold onto the idea that explanatory power is a matter of providing information about 'first-order' modal relations (as opposed to mere structural features of such relations).

Finally, having identified a tension in Woodward's account and motivated a way of interpreting it, we want to be upfront about critical questions that remain. Why is explaining a singular state of affairs different from explaining a regularity in this way? What is the connection between a regularity explanation and an explanation of a corresponding singular state of affairs (if one exists)? There's more work to be done to resolve the tension.

References

Bokulich, Alisa. 2008. "Can Classical Structures Explain Quantum Phenomena?" *British Journal for the Philosophy of Science* 59: 217–235.

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

West, Geoffrey, James H. Brown, and Brian Enquist. 1999. "The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organism." *Science* 284: 1677–1679.

— 2000. "The Origin of Universal Scaling Laws in Biology." In *Scaling in Biology*, ed. James H. Brown, and Geoffrey West, 87–112. Oxford: Oxford University Press.

Woodward, James. 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, James, and Christopher Hitchcock. 2003. "Explanatory Generalizations, Part 1: A Counterfactual Account. *Noûs* 37: 1–24.

Hitchcock, Christopher, and James Woodward. 2003. "Explanatory Generalizations, Part 2: Plumbing Explanatory Depth." *Noûs* 37: 181–199.

Ylikoski, Petri, and Jaakko Kuorikoski. 2010. "Dissecting Explanatory Power." *Philosophical Studies* 148: 201–219.