**In *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanation,* ed. Reutlinger and Saatsi**

**Some Varieties of Non-Causal Explanation[[1]](#footnote-1)**

**James Woodward**

**HPS Pittsburgh**

1. **Introduction**

The topic of non-causal explanation is very much in vogue in contemporary philosophy of science, as evidenced both by this volume and many other recent books and papers. Here I explore some possible forms of non-causal scientific explanation. The strategy I follow is to begin with the interventionist account of causal explanation I have defended elsewhere (Woodward 2003) and then consider various ways in which the requirements in that account might be changed or loosened to cover various putative non-causal explanations. I proceed in this way for a variety of reasons. First, causal explanations are generally regarded as at least one paradigm of successful explanation, even if there is disagreement about how such explanations work and what sorts of features mark them off as causal. A general account of explanation that entailed that causal claims were never explanatory or that cast no light on why such claims are explanatory is, in my opinion, a non-starter. Moreover, although it is possible in principle that causal and non-causal explanations have no interesting features in common, the contrary assumption seems a more natural starting point and this also suggests beginning with causal explanations. Second, if one is going to talk about “non-causal” explanation, one needs a clear and well-motivated notion of causal explanation to contrast it with. Third, we have a fairly good grasp, in many respects of the notion of causation, and how this connects to other concepts and principles that figure in science. These include connections to probability, as expressed in, e.g., the principle of the common cause and the Causal Markov condition and, relatedly connections between causal independence and factorizability conditions, as described in Woodward (2015b). Also of central importance is the connection between causal claims and actual or hypothetical manipulations or interventions, as described in Woodward (2003). Within physics, notions of causal propagation and process, where applicable, are connected to (and expressed in terms of) other physical claims of various sorts – no signaling results in quantum field theory, prohibitions on space-like causal connections, and so on. To a considerable extent, we lack corresponding connections and constraints in connection with non-causal forms of explanation. This is not a good reason for neglecting the latter, but again suggests a strategy of using what we understand best as a point of departure.

Finally, another important point about the contrast between causal and non-causal explanations: It is tempting to suppose not just that these are different (which is a presupposition of any discussion of this topic) but that there are scientific theories that exclusively provide one rather than the other; in other words, that there are non-causal explanations that proceed independently of any variety of causal explanation (or independently of any sort of causal information) and perhaps conversely. It seems to me that the truth is often more complicated and nuanced; often plausible candidates for non-causal explanation rest on or make use of causal information of various sorts. Thus even if it is appropriate to think of these explanations as non-causal, they will often be intertwined with and dependent on causal information.

As an illustration, consider explanations that appeal to facts about the structure of networks in ecology, neurobiology, molecular biology and other disciplines, as described in Huneman (2010). In many cases such networks are represented by undirected graphs and (I agree) in some cases there is a prima-facie case for thinking of these as figuring in non-causal explanations. However, when we ask about the evidence which forms the basis for the construction of the networks or what the networks represent, it seems clear they rest on causal information. For example, an undirected network in ecology may represent predator/prey interactions (with the undirected character implying that it does not matter which nodes corresponds to predators and which to the prey.) Such interactions (on the basis of which the graph is constructed) are certainly causal. Similarly, a network model in neurobiology, again represented by an undirected graph, may be constructed on the basis of information about which neural regions causally influence others, so that the network is understood as not merely representing correlational or structural information, although it does not represent causal direction. I do not conclude from this the explanations provided by these models are all causal, but the examples illustrate the extent to which causal and non-causal information can be intertwined in explanatory contexts. This provides another reason for not neglecting causal explanation in our discussion of the non-causal variety.

Before turning to details two more preliminary remarks: First, my focus will be entirely on possible forms of explanation of empirically contingent claims about the natural world. It may well be, as a number of writers have claimed, that there are mathematical explanations of purely mathematical results – e.g., proofs of such results that are (mathematically) explanatory and which contrast in this respect with other (valid) proofs that are not mathematically explanatory, but I will not address this possibility in this paper.

 Second, the notion of explanation (as captured by the English word and its cognates in many other languages) has, pre-analytically, rather fuzzy boundaries, particularly when one moves beyond causal explanation. This vagueness encourages the use of what might be described as an “intuitionist” methodology in discussions of non-causal explanation; an example is presented and the reader is in effect asked whether this produces any sense of understanding – an “aha” feeling or something similar. It is not always easy to see what turns on the answer one gives to this question. I have found it difficult to entirely avoid this intuition-based manner of proceeding but in my view it should be treated with skepticism unless accompanied by an account of what is at stake (in terms of connections with the rest of scientific practice or goals of inquiry) in labeling something an explanation. In some cases, as with the explanations of irrelevance considered in Section 5, such connections seem obvious enough; in other cases (such as Mother and the strawberries – cf. Section 4) not so much.

**2. An Interventionist Account of Causation and Causal Explanation**

 **2.1 Interventions and Counterfactual Dependence**. According to Woodward (2003), causal claims must correctly describe patterns of counterfactual dependence between variables playing the role of causes and variables playing the role of effects. The relevant notion of counterfactual dependence is understood in terms of *interventions*: *C* causes *E* if and only if there is a possible intervention that changes *C* such that under that intervention, *E* would change. An intervention can be thought of as an idealized experimental manipulation which changes *C* “surgically” in such a way that any change in *E*, should it occur, will occur only “through” the change in *C* and not via some other route. For our purposes, we may think of a *causal explanation* as simply a structure that exhibits or traces such a pattern of dependence, perhaps with the additional qualification that the exhibition in question must satisfy some sort of non-triviality requirement.[[2]](#footnote-2) When an explanation satisfies this condition, Woodward (2003) described it as satisfying a what-if-things- had-been-different- requirement (w-requirement) in the sense that it identifies conditions in its *explanans* such that if those conditions had been different, the *explanandum*-phenomenon would have been different. (My label for this requirement now seems to me a bit misleading, for reasons given below.) When the variables cited in a candidate *explanans* meet this requirement there is an obvious sense in which they are “relevant to” or “make a difference to” the *explanandum-*phenomenon.

Although Woodward (2003) relied heavily on the idea that explanations work by conveying what-if-things-had-been different information, virtually nothing was said about such questions as how representationally “realistic” a theory or model must be to convey such information. This has led some readers (e.g., Batterman and Rice 2014) to interpret the w-requirement as a commitment to the idea that only theories that are realistic in the sense of mirroring or being isomorphic (or nearly so) to their target systems can be explanatory. I don’t see the interventionist view as committed to anything like this. Instead, what is crucial is (roughly) this: an explanatory model should be such that there is reasoning or inferences licensed by the model that tell one what would happen if interventions and other changes were to occur in the system whose behavior is being explained. This does not require that the model be isomorphic to the target system or even “similar” to it in any ordinary sense, except in the inference-licensing respect just described. To anticipate my discussion in Section 5, a minimal model (and inferences performed within such a model) can be used to explain the behavior of real systems via conformity to the w-requirement even if the minimal model is in many respects highly dissimilar (e.g. of different dimensionality) from the systems it explains. The justification for using the minimal model to explain in this way is precisely that one is able to show that various “what-if” results that hold in the minimal model will also hold for the target system.

 Turning now to a different subject, the interventionist account requires that for *C* to cause *E*, interventions on *C* must be “possible”. Woodward (2003) struggled, not particularly successfully, to characterize the relevant notion of possibility. I will not try to improve on what I said there but will assume that there are some clear cases in which we can recognize that interventions are not (in whatever respect is relevant to characterizing causation) possible. An intervention must involve a physical manipulation that changes the system intervened on and there are cases in which we cannot attach any clear sense to what this might involve. Examples discussed below include interventions that change the dimensionality of physical space and interventions that change a system into a system of a radically different kind – e.g. changing a gas into a ferromagnet. We do possess theories and analyses that purport to tell us how certain systems would behave if they had different spatial dimensions or were a ferromagnet rather than a gas but I assume that such claims should not be interpreted as having to do with the results of possible interventions, but rather must be understood in some other way.

**2.2. Invariance.** As described above, the characterization of causal explanation does not require that this explicitly cites a *generalization* connecting cause and effect. Nonetheless, in many, perhaps most scientific contexts, generalizations (laws, causal generalizations etc.), explicitly describing how the *explanandum-*phenomenon depends on conditions cited in the *explanans*, are naturally regarded as part of explanations that the various sciences provide. According to Woodward (2003), if these generalizations represent causal relations, they must satisfy *invariance* requirements: for example, at a minimum, such generalizations must be invariant in the sense that they will continue to hold under some range of interventions on factors cited in the *explanans.* Often, of course, we expect (and find) more in the way of invariance in successful explanations than the minimal condition described above: we are able to construct explanations employing generalizations which are invariant *both* under a *wide range* of interventions on the variables cited in the *explanans*, *and* under changes in other variables and conditions not explicitly cited in the *explanans* – what we may call *background conditions*. Note that, as characterized so far, invariance claims are understood simply as empirical claims about the stability of relationships under variations in the values of various sorts of variables, including variations due to interventions. We will consider below various possibilities for broadening the notion of invariance to include stability under other sorts of variations, including those that do not involve interventions but are rather conceptual or mathematical in character.

 **2. 3**. **Causal Relationships distinguished from Conceptual and Mathematical Relationships**. Woodward (2003) (tacitly and without explicit discussion) adopted the common philosophical view that causal (and causal explanatory) relationships contrast with relationships of dependence that hold for purely conceptual, logical, or mathematical reasons.[[3]](#footnote-3) To employ a standard illustration, Xantippe’s widowhood (W) “depends” in some sense on whether Socrates dies (S) but the dependence in question appears to be conceptual (or the result of a convention) rather than causal – one has the sense that (S) and (W) are not distinct in the right way for their relationship to qualify as causal. This is so even though there is an obvious sense in which it is true that by manipulating whether or not Socrates dies, one can alter whether Xantippe is a widow. Thus we should think of the interventionist characterization of causation and causal explanation described above as coming with the rider/restriction that the candidates for cause and effect should not stand in a conceptual or logico-mathematical relationship and that, in the case of causal explanation, the explanation should “work” by appealing to a relationship that does not hold for purely conceptual reasons. This contrast between conceptual/mathematical and causal relationships will figure importantly in my discussion below since some plausible candidates for non-causal explanations seem to involve relations between *explanans* and *explanandum* that are non-causal because “mathematical” or “conceptual”.

 **2.4**. **Interventionism is a Permissive Account of Causation**. The account of causation and causal explanation described above is broad and permissive – any (non-conceptual) relationship involving intervention-supporting counterfactual dependencies counts as causal, even if it lacks features that other accounts claim are necessary for causation. For example, there is no requirement that causal claims or explanations must provide explicit information about the transfer of energy and momentum or trace processes through time. Similarly, variables that are abstract, generic, multiply realizable, or “upper level” can figure in causal relationships as long as these variables are possible targets for intervention and figure in intervention-supporting relations of counterfactual dependence. The diagonal length of a square peg can figure in a causal explanation of its failure to fit in a circular hole of a certain diameter (with no reference to the composition of the peg or the forces between its component molecules being required) as long as it is true (as it presumably is) that there are possible interventions that would change the shape of the peg with the result that it fits into the hole.

Summarizing, the picture of causal explanation that emerges from these remarks has the following features: (i) causal explanations provide answer what-if-things-had-been-different questions by telling us how one variable *Y* will change under (ii) interventions on one or more others (*X1*…*Xn*). Such interventions must be “possible” in the sense that they correspond to conceptually possible or well-defined physical manipulations. As discussed below, explanations having the structure described in (i) and (ii) will also provide, indirectly, information about what factors do not make a difference to or are irrelevant to the explanandum, but in paradigmatic causal explanations, it is difference-making information that does the bulk of the explanatory work. Finally, (iii) when the relationship between *X1*,…, *Xn* and *Y* is causal, it will be invariant in the sense of continuing to hold (as an empirical matter and not for purely mathematical or conceptual reasons) under some range of interventions on *X1*,…, *Xn* and some range of changes in background conditions.

Relaxing or modifying (i)- (iii) either singly or in combination yields various possible candidates for forms of non-causal explanation, which will be explored in subsequent sections. For example, one possible form of non-causal explanation answers w-questions (thus retaining (i)), but does not do so by providing answers to questions about what happens under *interventions,* insteadsubstituting claims about what would happen under different sorts of changes in *X1*,…, *Xn* – e.g., changes that correspond to a purely mathematical or conceptual variations not having an interpretation in terms of a possible physical interventions, as in Bokulich (2011), Rice (2015), among others. Another possible form of non-causal explanation involves retaining (i) and (ii) but dropping requirement (iii), or perhaps retaining (i) but dropping both (ii) and (iii). Here one countenances “explanations” that answer w-questions, but do so by appealing to mathematical, non-empirical relationships. Yet another possibility is that there are forms of explanation that do not tell us anything about the conditions under which the explanandum-phenomenon would have been different, as suggested in Batterman and Rice (2014). (These include the explanations of irrelevance discussed in Section 5).

**3. Non-causal explanations not involving interventions**

Woodward (2003) briefly considered the following candidate for a non-causal explanation. It is possible to show that given assumptions about what the gravitational potential would be like in an n-dimensional space (in particular, that the potential is given by an n-dimensional generalization of Poisson’s equation), Newton’s laws of motion, and a certain conception of what the stability of planetary orbits consists in, it follows that no stable planetary orbit are possible for spaces of dimension n ≥ 4. Obviously orbits of any sort are impossible in a space for which n=1, and it can be argued that n=2 can be ruled out on other grounds, leaving n=3 as the only remaining possibility for stable orbits. Is this an explanation of why stable planetary orbits are possible (in our world)?

 Let’s assume that this derivation is sound.[[4]](#footnote-4) Presumably even if one countenances talk of what would happen under merely possible interventions, the idea of an intervention that would change the dimensionality of space takes us outside the bounds of useful or perhaps even intelligible application of the intervention concept: it is unhelpful, to say the least, to interpret the derivation described above as telling us what would happen to the stability of the planetary orbits under an interventions changing the value of n. Nonetheless one might still attempt to interpret the derivation as answering a w- question – it tells us how the possibility of stable orbits (or not) would change as the dimensionality of space changes. In other words, it might be claimed that the derivation satisfies some but not all of the requirements of the interventionist model of causal explanation – it exhibits a pattern of dependence of some kind (perhaps some non-interventionist form of counterfactual dependence) between the possibility of stable orbits and the dimensionality of space, even though this dependence does not have an interventionist interpretation. And since it seems uncontroversial that one of the core elements in many explanations is the exhibition of relationships showing how an *explanandum* depends on its associated *explanans*, one might, following a suggestion in Woodward (2003), take this to show that the derivation is explanatory.

 Moreover, if it is correct that causal explanations involve dependence relations that have an interventionist interpretation, one might take this to show that the derivation is a case of non-causal explanation – in other words, that one (plausible candidate for a) dividing line between causal and non-causal explanation is that at least some cases of the latter involve dependencies (suitable for answering w-questions) that do not have an interventionist interpretation.[[5]](#footnote-5) Put differently, the idea is that the dependence component in explanation and the interventionist component are separable, drop the latter and retain the former, and you have a non-causal explanation. Suggestions along broadly these lines have been made by a number of writers, including Bokulich (2011), Rice (2015), Saatsi and Pexton (2012), and Reutlinger (forthcoming). For example, Reutlinger argues that explanations of the universal behavior of many very different substances (including gases and ferromagnets) near their critical points in terms of the renormalization group (RG) exhibit the pattern above – the RG analysis shows that the critical point behavior “depends upon” such features of the systems as their dimensionality and the symmetry properties of their Hamiltonians, but the dimensionality of the systems and perhaps also the symmetry properties of their Hamiltonians are not features of these systems that are possible objects of intervention.[[6]](#footnote-6) In both the case of the stability of the solar system and the explanation of critical point behavior, the “manipulation” that goes on is mathematical or conceptual, rather than possibly physical – e.g., in the former case one imagines or constructs a model in which the dimensionality of the system is different and then calculates the consequences, in this way showing what difference the dimensionality makes. Similarly, in the RG framework, the investigation of the different fixed points of Hamiltonian flows that (arguably) reveal the dependence of critical phenomena on variables like spatial dimensionality does not describe physical transformations of the systems being analyzed, but rather transformations in a more abstract space.

Let us temporarily put aside issues about the structure of the RG explanation (and whether its structure is captured by the above remarks) and focus on the candidate explanation for the stability of the planetary orbits. There is an obvious problem with the analysis offered above. One role that the notion of an intervention plays is that it excludes forms of counterfactual dependence that do not seem explanatory. For example, as is well-known, there is a notion of counterfactual dependence (involving so-called backtracking counterfactuals) according to which the joint effects of a common cause counterfactually depend on one another but this dependence is not such that we can appeal to the occurrence of one of these effects to explain the other. In the case of ordinary causal explanation, requiring that the dependence have an interventionist interpretation arguably rules out these non–explanatory forms of counterfactual dependence. The question this raises is whether non-explanatory forms of counterfactual dependence can also be present in candidates for non-causal explanation (thus rendering them non-explanatory) and, if so, how we can recognize and exclude these if we don’t have the notion of an intervention to appeal to.

To sharpen this issue, let me add some information that I have so far suppressed: one may also run the derivation described above backwards, deriving the dimensionality of space from the claim that planetary orbits are stable and assumptions about the gravitational potential and the laws of motion. Indeed, the best-known derivations in the physics literature (such as those due to Ehrenfest (1917) and Buchel (1969)) take this second form. Moreover, they are explicitly presented as claims about explanation: that is, as claims that the stability of the planetary orbits *explains* the three dimensionality of space.[[7]](#footnote-7) The obvious question this raises is: which, if either, of these facts (dimensionality, stability) is correctly regarded as the *explanans* and which as the *explanandum*? Is it perhaps possible both for stability to explain dimensionality and conversely, so that non- causal explanation can be (sometimes) a symmetric notion? On what basis could one decide these questions?

As Callender (2005) notes, the claim that the stability of the orbits explains the three dimensionality of space is generally advocated by those with (or at least makes most sense within the context of the assumption of) a commitment to some form of relationalism about spacetime structure: if one is a relationist, it makes sense that facts about the structure of space should “depend” on facts about the possible motions of bodies and the character of the force laws governing those bodies. Conversely, if one is a substantivalist one will think of facts about the structure of space as independent of the motions of bodies in them, so that one will be inclined to think of the direction of explanation in this case as running from the former to the latter.

Without trying to resolve this dispute, let me note that independence assumptions (about what can vary independently of what else) of an apparently non-causal sort seem to play an important role in both purported explanations.[[8]](#footnote-8) In the case in which the dimensionality of space is claimed to explain the stability of the explanatory orbits, it is assumed that the form of the equation for the gravitational potential is independent of the dimensionality of space in the sense that an equation of the same general form would hold in higher dimensional spaces. Similarly, Newton’s laws of motion are assumed to be independent of the dimensionality of space – it is assumed that they also hold in spaces of different dimensions, with the suggestion being that in such a different dimensioned space (n ≠3), the orbits would not be stable. In the case in which the explanation is claimed to run from the (possible) stability of the orbits to the dimensionality of space, the apparent assumption is that the form of the gravitational potential and the laws of motion are independent of the stability of the orbits in the sense that the former would hold even if the planetary orbits were not possibly stable (in which case the apparent suggestion is that the dimensionality of space would be different). I confess that I find it hard to see what the empirical basis is for either of these sets of claims, although the first strikes me as somehow more natural. As I note below, in other cases of putative non-causal explanations (such as the Konigsberg bridge case), there seems to be a more secure basis for claims about explanatory direction.

**4. Non-causal Explanations Involving Mathematical Dependencies but with Manipulable Explanatory Factors**

 In the previous section we considered putative cases of non-causal explanation in which the *explanans* factors do not seem to be possible targets for interventions, but in which the relationship between the *explanans* and *explanandum* essentially involves assumptions that, however their status is understood, are not a priori mathematical truths. In the example involving the stability of the planetary orbits, the assumption that the gravitational potential for n dimensions takes the form of a generalization of Poisson’s equation is not a mathematical truth and similarly for the assumption that the Newtonian laws of motion hold in spaces of dimensionality different from 3. (It is hard to understand these assumptions except as empirical claims, even if it is unclear what empirical evidence might support them.)

I now want to consider some cases that have something like the opposite profile: at least some of the variables figuring in the candidate *explanans* are possible targets for manipulation (although one might not want to regard the manipulations as interventions in the technical sense, for reasons described in footnote 4) but the *connection* between these and the candidate *explanandum* seems (in some sense) purely mathematical. Marc Lange has described a simple example which arguably has this structure:

That Mother has three children and twenty-three strawberries, and that twenty-three cannot be divided evenly by three, explains why Mother failed when she tried a moment ago to distribute her strawberries evenly among her children without cutting any. (Lange, 2013)

Here a mathematical fact (that 23 cannot be divided evenly by three) is claimed to explain Mother’s failure on this particular occasion. (And, one might think, if this is so, this mathematical fact also explains why mother always fails on every occasion and why equal division is “impossible”.)

Without trying to decide immediately whether this is an “explanation”, let’s see how it might be fitted into the framework we are using. There is presumably no problem with the notion of manipulating the number of strawberries available to Mother. Perhaps all of the available strawberries must be drawn from a basket and we can add or remove strawberries from the basket. As we vary the number in the basket, we find, e.g., that adding one to the twenty-three makes even division possible, subtracting one makes it impossible and so on. The overall pattern that emerges is that even division among the children is possible when and only when the number of strawberries is evenly divisible by three. It is not a huge stretch to think that the observation that this pattern holds fits naturally (in this respect) into the w-question framework and that the observation thus isolates a factor on which the *explanandum* (whether even division is possible) “depends”. On these grounds one might think that an interesting similarity is present between this example and more paradigmatic cases of causal explanation and that this warrants regarding the example as providing a genuine explanation.[[9]](#footnote-9)

Of course there is also the obvious disanalogy mentioned earlier: given the particular facts in the example (number of strawberries and children) the connection between these and the candidate *explanandum* (whether equal division is possible) follows just as a matter of mathematics, without the need for any additional assumptions of a non-mathematical nature. Presumably this is why it does not seem correct to think of the relationship between the particular facts cited in the candidate *explanans* (and the failure to divide, or impossibility of dividing equally) as causal. Instead, as in the case of the relationship between Socrates’ death and Xantippe’s widowhood, it seems more natural to express the dependence between the possibility of equal division and the number of strawberries and children by means of locutions like “brings about by” that are appropriate for cases of non-causal dependence: by varying the number of strawberries or children one brings it about that Mother succeeds or fails at equal division.

Our reaction to this example may be colored by the fact that the mathematical fact to which it appeals is trivial and well-known; this may contribute to the sense that many may have that in this case citing the mathematical fact does not greatly enhance understanding, so that (at best) only in a very attenuated sense has an explanation been provided. However, there are other cases, such as the well-known Konigsberg bridge problem, which seem to have a similar structure where many will have more of a sense that an explanation has been furnished. Suppose we represent the configuration of bridges and landmasses in Konigsberg by means of an undirected graph in which bridges correspond to edges, and the land masses they connect to nodes or vertices. An *Eulerian path* through the graph is a path that traverses each edge exactly once. Euler proved that a necessary condition for a graph to contain an Eulerian path is that the graph be connected (there is a path between every pair of vertices) and that it contain either zero or two nodes of odd degree, where the degree of a node is the number edges connected to the node.[[10]](#footnote-10) This condition is also sufficient for a graph to contain an Eulerian path. The Konigsberg bridge configuration does not meet this condition – each of the four landmasses is connected to an odd number of bridges – and it follows that it contains no Eulerian path.

One might think of this demonstration in the following way: we have certain contingent facts – the connection pattern of the bridges and land masses of Konigsberg. Given these, one can derive via a mathematical argument that makes use of no additional empirical premises that it is impossible to cross each bridge exactly once. (That is, the connection between *explanans* and *explanandum* is entirely mathematical rather than empirical.) Moreover, the derivation makes use of information that can be used to answer a number of w-questions about the *explanandum* – as just one sort of possibility, the derivation tells us about alternative possible patterns of connectivity which would make it possible to traverse an Eulerian path among the bridges as well as about other patterns besides the actual one in which this would not be possible. In doing this the explanation also provides information about the many features of the situation that do not matter for (are irrelevant to) whether it is possible to traverse each bridge exactly once: it does not matter where one starts, what material the bridges are made of or even (as several writers note) what physical laws govern the bridges, as long as they provide stable connections. These assertions about the irrelevance of physical detail is bound up with our sense that Euler’s analysis isolates the abstract, graph-theoretical features of the situation that are relevant to whether it is possible to traverse an Eulerian path. Note, however, that this information about irrelevance figures in the analysis only against the background of information about what *is* relevant, which has to do with the connectivity of the graph.

 Note also that despite this mathematical connection between *explanans* and *explanandum*, the notion of changing or manipulating the bridge configuration – e.g., by constructing additional bridges or removing some – and tracing the results of this 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.

It is also worth noting that in this case, in contrast to the example involving the dimensionality of space in section 3, the direction of the dependency relation seems unproblematic. The configuration of the bridges has perfectly ordinary causes rooted in human decisions to construct one or another particular configuration. Because these decisions cause the configuration, it is clear that the impossibility of traversing an Eulerian path is not somehow part of an explanation of the configuration. Rather, if this is a case of explanation, the direction must run from the configuration to the impossibility of traversing, with the configuration instead having the causes described above. This shows one way in which the problem of distinguishing explanatory from non-explanatory patterns of dependence in connection with candidates for non-causal explanation might be addressed.

**5**.**The Role of Information about Irrelevance in Explanation**.

 As noted above, the w-question conception focuses on the role of factors that are explanatorily relevant to an *explanandum* – relevant in the sense that variations in those factors make a difference to whether the *explanandum* holds. A number of recent discussions have instead focused on what might be described as the role of irrelevance or independence in explanation – on information to the effect that some factors do *not* make a difference to some *explanandum*, with some writers seeming to suggest that some explanations work primarily or entirely by citing such independence information and that interventionist and other difference-making accounts of explanation cannot accommodate this fact (see, e. g. Batterman and Rice (2014), Gross (2015)). Indeed, it sometimes seems to be suggested that some *explananda* can be explained by citing *only* factors that are irrelevant to it, with difference-making factors playing no role at all. In this section I want to explore some issues raised by the role of information about irrelevance.

First we need to clarify what is meant by the notions of relevance and irrelevance. The interventionist understanding of relevance is that *X* is relevant to *Y* as long as *some* interventions that change the value of *Y* are associated with changes in *Y*; *X* is irrelevant to *Y* if there are no such interventions. Suppose that *X* and *Y* can take a range of different values and that *X* and *Y* are related by *F(X)*= *Y*. Assume *F* specifies that some changes in *X* are associated with changes in *Y* and that others are not – in other words, *F* is not a 1-1 function, although it does not map all values of *X* into the same value of *Y*. In such a case , *X* is relevant to *Y*, although of course we may also go on to describe more specifically which changes in the value of *X* are relevant to *Y* and which others are not. My understanding of the what-if-things-had been-different idea has always been that in such cases *F* provides w-information and is explanatory in virtue of describing the pattern of dependence of *Y* on *X* even though that pattern is such that some changes in *X* make no difference to the value of *Y.*[[11]](#footnote-11) We may also generalize the w-account to include (in) dependence information that is not understood in terms of interventions, as suggested above, in which case similar remarks (e.g. that the dependence need not be 1-1) apply.

 As suggested in passing above, information about independence and irrelevance is in many ways the flip side of the dependence or relevance information emphasized in the interventionist account, since the latter can often be “read” off from the former. To take the most obvious possibility, when (5.1) some variable *Y* is represented as dependent on others *X1,…,Xn*, this (at least often) implicitly conveys that other variables Z1,…,Zn, distinct from *X1,…,Xn* that are not explicitly mentioned in (5.1) are irrelevant to *Y*. When the gravitational inverse square law and Newton’s laws of motion are used to explain the trajectories of the planets, one does not have to explicitly add the information that the colors of the planets are irrelevant to their trajectories since the use of these laws conveys this information. In these respects, virtually all explanations convey information about irrelevance – this is not a distinctive feature of some specific subclass of explanations.

 In an examples like the one just described there is an obvious sense in which the dependence information seems to be doing the explanatory work, with the independence information following derivatively from the dependence information. One indication of this is the independence information by itself, apart from dependence information, does not seem explanatory: Presumably no one would be tempted to think that one could explain the motions of the planets just by citing information to the effect that factors such as color are irrelevant to that motion.

 In other cases, however, one has the sense that information about independence or irrelevance may be playing a more substantial explanatory role. Consider an *equilibrium explanation*, where one component of the explanation involves showing that the outcome being explained would result from a large number of possible initial states. As an illustration, suppose the final state of a gas (e.g. that it exerts a certain equilibrium pressure after being allowed to diffuse isothermally into a fixed volume) is explained by means of a demonstration that almost all initial states of the gas compatible with certain macroscopic thermodynamic constraints (e.g. the temperature of the gas and the volume of the container) will evolve to the same equilibrium outcome. Another illustration is provided by Fisher’s well-known explanation of sex allocation among offspring and the various generalizations of this due to Hamilton, Charnov and others, where the factors influencing equilibrium outcomes are shown to be independent of the details of specific episodes of fertilization.

 Such explanations are often claimed to be non-causal or not captured within a difference-making framework since they do not involve tracing the actual trajectory of the specific events leading to the explanandum-outcome. In a brief discussion, Woodward (2003) objected to the characterization of such explanations as non-causal: virtually always such explanations do invoke dependency or difference-making information (which can be understood in terms of interventionist counterfactuals) in addition to information to the effect that many initial states will lead to the same outcome. For example, in the case of the gas, the pressure will of course depend on the maintained temperature and the container volume – vary these and the temperature will vary. Since these dependency relations can be given an interventionist interpretation, one can interpret the explanations as providing causal explanations of why one particular equilibrium rather than another obtains.

While these observations still seem to me to be correct, I think it is also true that independence or irrelevance information seems to play a somewhat different role in these equilibrium explanations than it does in the explanation of planetary trajectories, where its role seems essentially trivial. Perhaps one aspect of the difference (not the only difference – see below) is this: a property like color plays no interesting role anywhere in mechanics or in most of the rest of physics and no one will be surprised by the observation that the influence of gravity on the planets is unaffected by their color. Indeed, the question: why does color not matter for planetary trajectories does not seem to arise in any natural way nor is it obvious what would serve as an answer to it. On the other hand, facts about the detailed trajectories of individual molecules are among the sorts of facts that physics pays attention to: they are relevant to what happens in many contexts and are *explananda* for many physical explanations. There thus seems to be a live question about why, to a very large extent, details about individual molecular trajectories don’t matter for the purposes of predicting or explaining thermodynamic variables. Replacing details about the individual trajectories of the 1023 molecules making up a sample of gas with a few thermodynamic variables involves replacing a huge number of degrees of freedom with a very small number which nonetheless are adequate for many predictive and explanatory purposes. It is natural to wonder why this “variable reduction” strategy works as well as it does and why it is that, given the values of the thermodynamic variables, further variations in the molecular trajectories almost always make no different to many of the outcomes specifiable in terms of thermodynamic variables.

Here we seem to be asking a different kind of question than the questions about the identification of difference-makers that characterize straightforward causal analysis; we are asking instead why variations in certain factors do *not* make a difference to various features of a system’s behavior, at least given the values of other factors. Put slightly differently, we are still interested in w-questions but now our focus is on the fact that if various factors had been different in various ways, the *explanandum* would *not* have been different and perhaps on understanding why this is the case.[[12]](#footnote-12) (I have so far not tried to provide any account of what such an explanation would look like – that will come later.)

Note, however, that these observations do *not* support the idea that one can explain why some outcome occurs by just citing factors that are irrelevant to it. In the example above and others discussed below, it seems more natural to regard the claims about irrelevance as *explananda* (or at least as claims that are in need of justification on the basis of other premises) rather than as part of an explanans (or premises that themselves do the explaining or justifying). That is, rather than citing the irrelevance of *V* to *E* in order to explain *E*, it looks as though what we are interested in explaining or understanding is why *V* is irrelevant to *E*. Explaining why *V* is irrelevant to *E* is different from citing the irrelevance of *V* to explain *E*. Moreover, independently of this point, in the examples we have been looking at, the irrelevance of certain factors to some outcome is conditional on the values of other factors that *are* identified as relevant, with the form of the explanatory claim being something like this: (5.1) Given the values of variables *X1,…,Xn* (which are relevant to outcome *E*) --e.g. temperature and volume-- variations in the values of additional variables *V1,…,Vn* (e.g., more detailed facts about individual molecular trajectories) are irrelevant to *E[[13]](#footnote-13)*. Thus insofar as the irrelevant variables or the information that they are irrelevant have explanatory import, they do so in the context of an explanation in which other variables *are* relevant.

 What might be involved in explaining that certain variables are irrelevant to others (or irrelevant to others conditional on the values of some third set of variables)? Although several writers, including Batterman and Rice (2014), defend the importance of such explanations and offer examples, I am not aware of any fully systematic treatment. Without attempting this, I speculate that one important consideration in many such cases is that there is an underlying dynamics which, even if it is not known in detail, supports the claims of irrelevance – what we want is insight into how the working of the dynamics makes for the irrelevance of certain variables. For example, in Fisher’s well-known treatment of sex allocation, it is not just that many fertilization episodes that differ in detail can be realizers of the creation of females or males.[[14]](#footnote-14) The equilibria in such analyses are (or are claimed to be) *stable* equlibria in the sense that perturbations that take populations away from equibrium allocations are soon returned to the equilibrium allocation because of the operation of natural selection – it being selectively disadvantageous to produce non-equilibrium sex ratios. In other words, there is a story to be told about the structure of the dynamics, basins of attraction, flows to fixed points etc. that gives us insight into why the details of individual episodes do not matter to the outcome. Similarly for the behavior of the gas. There is nothing similar to this in the case of explaining the irrelevance of colors to the trajectories of planets, which is why it is hard to see what non-trivial form such an explanation would take.

In the cases considered so far in this section the notion of irrelevance has an obvious interventionist interpretation. However, there are other cases, discussed below, in which we need to broaden the notions of relevance and irrelevance to include reference to variations or changes that do not have an interventionist interpretation or where is at least not obvious that such an interpretation is appropriate. These include cases in which it follows as a matter of mathematics that, given certain generic constraints, variations in values of other variables or variations in structural relationships make no difference to some outcome, but where the variations in question are not (or may not be) the sort of thing that can be produced by interventions.

A possible illustration is provided by the use of the method of arbitrary functions and similar arguments to explain the behavior of gambling devices. An obvious explanatory puzzle raised by such devices is to understand why they produce stable frequencies strictly between 0 and 1 despite being deterministic, and despite the fact the initial conditions characterizing any one device will vary from trial to trial (and of course also vary across devices) and that different devices are governed by different detailed dynamics. Moreover, these relative frequencies are also stable in the sense that they are unaffected by the manipulations available to macroscopic agents like croupiers. Very roughly, it can be shown that provided that the distribution of initial conditions on successive operations of such devices satisfies some generic constraints (e.g. one such constraint is that the distribution is absolutely continuous) and the dynamics of the devices also satisfy generic constraints, the devices will produce (in the limit) outcomes with well-defined probability distributions and stable relative frequencies – in many cases (when appropriate symmetries are satisfied) uniform distributions over those outcomes. It is natural to think of these sorts of analyses as providing explanations of the facts about irrelevance and independence described above.

In such cases it is not clear that all of the variations under which these devices can be shown to exhibit stable behavior have an interventionist interpretation. For example, the information that any one of a large range of different dynamics would have generated the same behavior seems to have to do with the consequences of variations within a mathematical space of possible dynamics rather than with variations that necessarily have an interventionist interpretation. Relatedly, it is arguable that those features of the system that the analysis reveals as relevant to the achievement of stable outcomes – the generic constraints on the initial conditions and on the dynamics – are not naturally regarded as “causes” of that stability in the interventionist sense of cause. For example, it is not obvious that the fact that the distribution of initial conditions satisfied by some device is absolutely continuous should count as a “cause ” of the device’s behavior. On the other hand, if we follow the line of thought in previous sections and extend the notion of information that answers w-questions to include cases in which the information in question does not have to do with interventionist counterfactuals but rather with what happens under variations of different sorts (in initial conditions, dynamics etc.) and where the answer may be that some outcome or relationship does *not* change under such variation (i.e. the variations are irrelevant) we can accommodate examples of this sort. That is, we can think of these as explanations of irrelevance where the irrelevance in question is irrelevance under variations of a certain sort but where the variations do not have an interventionist interpretation. In such cases, irrelevance is demonstrated mathematically by showing that the mathematical relationships between the variations and some phenomenon or relationship is such that the latter does not change under the former.

I conclude this section by briefly exploring some additional issues about irrelevance in the context of some recent claims made by Batterman and Rice (2014) about minimal models and their role in explanation. Abstractly speaking, we can think of a minimal model as a model which captures aspects of the common behavior of a class of systems (and of the behavior of more detailed models of such systems in this class). A minimal model serves as a kind of stand-in for all of the systems for which it is a minimal model –  for an appropriate class, results that can be shown to obtain for the minimal model must also hold for other models and systems within the delimited class, no matter what other features they possess. Thus one can make inferences (including “what if” inferences) and do calculations using the minimal model, knowing that the results “must” transfer to the other models and systems. Here the “must” is mathematical; one shows as a matter of mathematics that the minimal model has the stand-in or surrogative role just described with respect to the other models and systems in the universality class. Renormalization group analysis is one way of doing this – of justifying the use of a minimal model as a surrogate. In this respect, RGA delimits the “universality class” to which the minimal model belongs.

A striking example, discussed by Batterman and Rice, is provided by a brief paper by Goldenfeld and Kadanoff (1999) which describes the use of a minimal model for fluid flow (the lattice gas automaton or LGA). The model consists of point particles on a two-dimensional hexagonal lattice. Each particle interacts with its nearest neighbors in accord with a simple rule. When this rule is applied iteratively and coarse-grained averages are taken, a number of the macroscopic behaviors of fluids are reproduced. As Goldenfeld and Kadanoff explain, the equations governing macroscopic fluid behavior result from a few generic assumptions: these include locality (the particles making up the fluid are influenced only by their immediate neighbors), conservation (of particle number and momentum) and various symmetry conditions (isotropy and rotational invariance of the fluid). These features are also represented in the LGA and account for its success in reproducing actual fluid behavior, despite the fact that real fluids are not two dimensional, not lattices and so on.

Batterman and Rice make a number of claims about the use of minimal models in explanation. First, they seem to suggest in one passage that such models are explanatory because they provide information that various details are irrelevant to the behavior of the systems modeled.[[15]](#footnote-15)

 [The] models are explanatory because of a story about why a class of systems will all display the same large-scale behavior because the details that distinguish them are irrelevant. (Ibid.: 349)

 Elsewhere they write, in connection with the use of the renormalization group to explain critical point behavior:

The fact that the different fluids all possess these common features (having to do with behavior near their critical points) is also something that requires explanation. The explanation of this fact is provided by the renormalization group-like story that delimits the universality class by demonstrating that the details that genuinely distinguish the fluids from one another are irrelevant for the explanandum of interest. (Ibid.: 374)

Second, they claim that the features that characterize the minimal model are not causes of (and do not figure in any kind of causal explanation of) the fluid phenomena being explained:

We think it stretches the imagination to think of locality, conservation, and symmetry as causal factors that make a difference to the occurrence of certain patterns of fluid flow. (Ibid.: 360)

Although this may not be their intention, the first set of passages makes it sound as though they are claiming that the common behavior of the fluids can be explained just by citing factors that are irrelevant to that behavior or by the information that these factors are irrelevant. Let me suggest a friendly amendment: it would be perspicuous to distinguish the following questions: First, (a) why is it justifiable to use this particular model (LGA) as a minimal model for a whole class of systems? Second, (b) why do systems in this class exhibit the various common behaviors that they do? I agree with what I take to be Batterman and Rice’s view that the answer to (a) is provided by renormalization-type arguments or more generally by a mathematical demonstration of some kind that relates the models in this class to one another and shows that for some relevant class of behaviors, any model in the class will exhibit the same behavior as the minimal model. I also agree with Batterman and Rice that in answering this question one is providing a kind of explanation of (or at least insight into) why the details that distinguish the systems are irrelevant to their common behavior. But, to repeat an observation made earlier, the *explanandum* in this caseis a claim about irrelevance; this answer to (a) does not support the contention that irrelevance claims by themselves are enough to explain (b).

Instead, it seems to me that the explanation for why (b) holds is provided by the minimal model itself in conjunction with information along the lines of (a) supporting the use of the minimal model as an adequate surrogate for the various systems in the universality class. Of course the minimal model does not just consist in claims to the effect that various factors are irrelevant to the common behavior of the systems, (although its use certainly implies this) so we should not think of this explanation of (b) as consisting just in the citing of irrelevance information. Instead the minimal model also provides information about a common abstract structure shared by all of the systems in the universality class – structure that (as I see it) is relevant to the behavior of these systems. Here, as in previous cases, relevance and irrelevance information work together, with the irrelevance information telling us, roughly, why it is justifiable to use a certain minimal model and why various details that we might have expected to make a difference to systems in the universality class do not and the relevance information identifying the shared structure that does matter.

Regarding this shared structure several further questions arise. First, does the structure furnish a *causal* explanation of (b)? Here I agree with Batterman and Rice that the answer is “no”, or at least that it is “no” given an interventionist account of causation. The features characterizing the structure are just not the sort of things that are well-defined objects of intervention – one cannot in the relevant sense intervene to make the interactions governing the system local or non-local, to change the dimensionality of the system and so on. However, I would contend that we should not necessarily infer from this the minimal model does not cite difference-making factors at all or that these difference-making factors have no explanatory significance; instead it may be appropriate to think of the model as citing non-causal difference-making factors which have explanatory import in the manner that some of the putative explanations in Section 3 do. One reason for thinking that something like this must be the case is that the LGA and associated RG type analyses are not just used to provide insight into why various details distinguishing the systems are irrelevant to certain aspects of their behavior; they are also used to calculate (and presumably explain) various other more specific features of the systems in question – critical exponents, relations among critical exponents, deviations from behavior predicted by other (e.g., mean field) models and so on. These are not *explananda* that can be derived or explained just by citing information to the effect that various details are irrelevant or non-difference-makers; one also needs to identify which features are relevant to these behavior and it is hard to see how this could fail to involve difference-making information, albeit of a non-causal sort.

 I thus find Reutlinger’s recent suggestion (forthcoming) that explanations of the RG-sort under discussion work in part by citing what-if-things-had- been-different information of a non-causal sort plausible. I will add, however, that, for reasons described above, I do not think that this captures the whole story about the structure of such explanations; Batterman and Rice are correct that explanations of irrelevance also play a central role in such explanations.

**6. Conclusion**

In this paper I have tried to show how the interventionist account of causal explanation might be extended to capture various candidates for non-causal explanation. These include cases in which there is empirical dependence between explanans and explanandum which does not have an interventionist interpretation, and cases in which the relation between explanans and explanandum is conceptual or mathematical. Examples in which claims about the irrelevance of certain features to a system’s behavior are explained or justified are also acknowledged and discussed, but it is contended that difference-making considerations also play a role in such examples.

**References**

Batterman, R. and Rice, C. (2014), ‘Minimal Model Explanations’ *Philosophy of Science* 81: 349–76.

Bokulich, A. (2011), ‘How Scientific Models Can Explain’, *Synthese* 180: 33-45

Buchel, W. (1969), ‘Why is space three-dimensional?” Trans. Ira M. Freeman. *American Journal of Physics* 37: 1222–24.

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: 113–36

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.

Goldenfeld, N. and Kadanoff, L. (1999), ‘Simple Lessons from Complexity’, *Science* 284: 87–89.

Gross, F. (2015), ‘The Relevance of Irrelevance: Explanation in Systems Biology’, in Braillard and Malaterre (eds.) *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*, 175–98.

Huneman, P. (2010), ‘Topological Explanation and Robustness in Biological Systems’, *Synthese* 177: 213–45.

Lange, M. (2013), ‘What Makes a Scientific Explanation Distinctively Mathematical?’, *British Journal for the Philosophy of Science* 64: 485-511.

Jansson, L. and Saatsi, J. (forthcoming) “Explanatory Abstractions” *British Journal for the Philosophy of Science*.

Saatsi, J. and Pexton, M. (2012), ‘Reassessing Woodward’s account of explanation: regularities, counterfactuals, and non-causal explanations’, *Philosophy of Science* 80: 613–24.

Saatsi, J. (forthcoming), “On explanations from ‘geometry of motion”, *British Journal for the Philosophy of Science.*

Reutlinger, A. (forthcoming) “Is there a Monist Theory of Causal and Non-causal Explanation? The Counterfactual Theory of Scientific Explanation”, *Philosophy of Science.*

Rice, C. (2015), ‘Moving beyond causes: Optimality models and scientific explanation’, *Noûs,* 49: 589-615.

Woodward, J. (2003), *Making Things Happen* (New York: Oxford University Press).

Woodward, J. (2016b), ‘Causation in Science’, in P. Humphreys (ed.), *Oxford Handbook of Philosophy of Science*.

Woodward, J. (2016a), ‘The Problem of Variable Choice’, *Synthese* 193: 1047–72.

Woodward, J. (forthcoming) “Explanatory Autonomy: The Role of Proportionality, Stability and Conditional Irrelevance”

1. Many thanks to Bob Batterman, Alexander Reutlinger, Collin Rice and Juha Saatsi for helpful comments on earlier drafts. [↑](#footnote-ref-1)
2. Consider the claim that (2.1) the cause of *E* is the cause of *E*. If *E* has a cause (2.1) is true and some intervention on the cause of *E* will be associated with a change in *E*. Most, though, will regard (2.1) as no explanation of *E*, presumably because it is trivial and uninformative (other than implying that *E* has some cause.) [↑](#footnote-ref-2)
3. For additional discussion of some the subtleties surrounding this notion, see Woodward (2016a). [↑](#footnote-ref-3)
4. For discussion and some doubts about the soundness claim, see Callender (2005). [↑](#footnote-ref-4)
5. It is worth emphasizing that the candidate *explanandum* in this case is the *possibility* or not of stable orbits. A natural thought is that if stable orbits are possible, then whether or not some particular planetary orbit is stable is the sort of thing that might be explained causally, but that the possibility of stable orbits is not the sort of thing that can be a causal effect or a target of causal explanation. (The underlying idea would be that causal explanations have to do with what is actual or not, rather than what is possible or impossible.) I lack the space to explore this idea here. [↑](#footnote-ref-5)
6. These claims are not uncontroversial – they are rejected, for example, by Batterman and Rice (2014). [↑](#footnote-ref-6)
7. Buchel’s paper is entitled, “Why is Space Three-Dimensional?” [↑](#footnote-ref-7)
8. Independence assumptions also play an important role in judgments of *causal* direction – see Woodward (2016b). On this basis one might conjecture that if there is some general way of understanding such assumptions that is not specifically causal, this might be used in a unified theory of causal and non-causal explanation: roughly the idea would be that if *X* and *Y* are independent and *Z* is dependent on *X* and *Y*, then the direction of explanation runs from *X* and *Y* to *Z*, and this holds for non-causal forms of (in)dependence. [↑](#footnote-ref-8)
9. For a similar treatment of this example, see Jansson and Saatsi, forthcoming . [↑](#footnote-ref-9)
10. This is unmysterious when you think about it. Except for the starting and end point of the walk, to traverse an Eulerian path one must both enter each land mass via a bridge and exit via a different bridge. If each bridge is to be traversed exactly once, this requires that each such non-terminal land mass must have an even number of edges connected to it. At most two land masses can serve as starting and end points, with an odd number of edges connected to them. It is interesting to note (or so it seems to me) that it is a *proof* or argument along lines like this which does whatever explanatory work is present in the example rather than just the specification of the difference-making conditions itself. [↑](#footnote-ref-10)
11. I mention this because some writers (e.g. Gross (2015)) interpret me as holding the contrary view that when the relation between *X* and *Y* is not 1-1, this relationship is not explanatory (because it is not a dependence or difference-making relationship). Gross describes a biological example in which (put abstractly) some changes in the value of *X* are relevant to *Y* and many others are not; he claims that in this case the interventionist account cannot capture or take notice of the biological significance of this irrelevance information. My contrary view is that this is an ordinary dependence or difference-making relation and, according to interventionism, explanation can proceed by citing this relationship. [↑](#footnote-ref-11)
12. This is why I said earlier that my use of the phrase “w-information” in Woodward (2003) was a bit misleading or imprecise: I had in mind the specification of changes in factors in an explanans under which the explanandum would have been different but of course it may be true that under some changes in the explanans factors, the explanandum would not have been different. [↑](#footnote-ref-12)
13. For further discussion of this sort of conditional irrelevance (as I call it) see Woodward, forthcoming. [↑](#footnote-ref-13)
14. This is one (of several reasons) why thinking of such examples (just) in terms of multiple realizability misses important features. [↑](#footnote-ref-14)
15. Batterman informs me that this is not what the quoted passage was intended to express: his idea was rather that what justifies the use of the minimal model for explanatory purposes is the RG story about irrelevance of other actual details omitted by the minimal model. [↑](#footnote-ref-15)