Explanatory Abstractions

Lina Jansson and Juha Saatsi

August 2, 2016

Abstract

A number of philosophers have recently suggested that some abstract, plausibly non-causal and/or mathematical, explanations explain in a way that is radically different from the way causal explanation explain. Namely, while causal explanations explain by providing information about causal dependence, allegedly some abstract explanations explain in a way tied to the *independence* of the explanandum from the microdetails, or causal laws, for example. We oppose this recent trend to regard abstractions as explanatory in some *sui generis* way, and argue that a prominent account of causal explanation can be naturally extended to capture explanations that radically abstract away from microphysical and causal-nomogical details. To this end, we distinguish different senses in which an explanation can be more or less abstract, and analyse the connection between explanations' abstractness and their explanatory power. According to our analysis abstract explanations have much in common with counterfactual causal explanations.

1 Introduction

There is broad agreement that many explanations derive their explanatory power from information about *dependence*. Causal explanations involve causal dependence, and various kinds of non-causal explanations can arguably be similarly understood in terms of non-causal dependence of the explanandum on the explanans.[1] Indeed, perhaps (almost) all explanations involve information about what the explanandum (causally or non-causally) depends on? This hypothesis, if tenable, could be part of a unified account of explanations, according to which explanations—causal or otherwise—explain by virtue of providing information about dependences.

A consideration against this hypothesis comes from *explanatory abstractions*. Explanations vary in their abstractness: some explanations appeal to relatively abstract features of reality, whereas others turn on much less abstract explanans. We value explanatory abstraction, when appropriate: it filters out details that are irrelevant to the explanation in question. But abstraction does not directly concern what the explanandum depends on. Rather, abstraction is about *independence* of the explanandum from the irrelevant details. It is tempting to think, along with various recent authors, that this

[1] See e.g. Woodward [2003], Woodward and Hitchcock [2003a,b], Ylikoski and Kuorikoski [2010], Strevens [2008]. Ruben [1990] is an early proponent.

information about *independence* is where the explanatoriness of many abstract explanations lies. This mode of explaining, associated with what is irrelevant to the explanandum, can seem *sui generis*, different in kind from explanations in terms of dependences.

We will argue that explanatory abstractions do not point to a sui generis mode of explaining. There is no need to revise the idea that explanatory power is a matter of providing dependence information (and the more the better); no need to abandon the hope for a unified account of explanatory power. Here we stand against a clear trend that urges the opposite: that some abstract explanations do not fit an account of explanatory power that is focused on (non-)causal dependence.[2] More specifically, we aim to undercut prominent arguments supporting the trend we oppose, and to present and argue for one such unified account, extending the account of causal explanation developed by Woodward [2003] and Woodward and Hitchcock [2003a,b] to allow it to capture influential exemplars of abstract explanations behind the anti-unificationist trend.

Before we get to the nitty-gritty, let us illustrate explanatory abstraction and sketch some of the key points of the paper. The explanation of why Mother fails to divide her twenty-three strawberries equally among her three children (without cutting any strawberries) turns on the fact that twenty-three is not evenly divisible by three.[3] Citing the causal details or laws involved in Mother's attempt, or the physical constitution of the strawberries, do not suffice to explain this fact. The explanation is *independent* of such details; the explanatory mathematical facts hold irrespective of contingent causal laws and details. So where does the explanatory power associated with mathematics come from? Mathematical facts do not seem to capture dependence relations between numbers that would be analogous to contingent causal and nomological connections. So perhaps the explanatory contribution of mathematics here cannot be captured in terms of dependences? Perhaps mathematics' explanatory contribution is somehow sui generis, turning on mathematics' *independence* from the details of the physical constitution and the causal laws? We will be concerned with explicating and rebutting this intuition.

We see little reason to shift focus from dependence to independence in analysing abstract mathematical explanations of this kind. One easily overlooks a wealth of dependence information that we take for granted here. The explanation does not just tell us that Mother must fail as a matter of mathematical necessity. It also tells us what the failure depends on—the number of strawberries—and *how*. For example, if Mother had two fewer strawberries or one strawberry more, then she could have succeeded. This change relating *what-if-things-had-been-different* information is exactly the kind associated with explanatory dependence.[4] We will argue that this information is doing all the explaining, and that this is captured by a natural extension of the ideology of causal explanation according to which explanations function by providing what-if-things-had-been-different information. It has been broadly thought that this ideology is ill-fitting for capturing abstract, non-causal and/or mathematical explanations, since the relevant modal and counterfactual notions only apply to causal connections.[5] This, we will ar-

[2] See e.g Lange [2013], Pincock [2007], Pincock [2015] (when arguing that none of the existing dependence accounts can capture highly abstract explanations), Batterman and Rice [2014].

[3] Cf. Lange [2013], and Braine [1972].

[4] Most prominently in Woodward [2003].

[5] There are exceptions to this. For example, Bokulich [2008a,b, 2012], Rice [2015], and Reutlinger [forth-

gue, is mistaken. The reasons given for leaving behind the counterfactual dependence ideology associated with causal explanations in connection with abstract explanations are faulty. And so are the arguments for sui generis, independence-based accounts of abstract explanations.

The key to seeing the latter is to recognise, first of all, that practically all explanations have some degree of abstraction, and provide some dependence information. What we need is a way of gauging the explanatory importance of dependence information, relative to information about independence. We will argue that the explanatory importance of dependence information is revealed by varying it while keeping the degree of abstraction otherwise fixed. In the context of the strawberry case, consider removing the dependence information *altogether*, first of all, pretending that we can take away all dependence information based on background arithmetical knowledge, only retaining the information about the explanandum's independence from causal laws and features. If *all* we know is that as a matter of mathematical necessity Mother's attempt with twenty-three strawberries must fail, so that we are unable to answer any questions about possible changes that would render her success possible, then it is not clear that we have *any* explanation of Mother's failure.

Consider, now, adding a *little bit* of information about dependence, pretending that we only get to know that Mother could have succeeded with twenty-four strawberries. Assume that is all the information we have, on top of knowing that twenty-three is not divisible by three. Here we may have an explanation, but it looks like a very thin one. Yet, in terms of independence from particular causal laws, processes, and microphysical conditions, it is on a par with the general explanation that appeals to background arithmetical knowledge. If the original explanatory information explains by virtue of showing the explanandum to be independent from physical details and laws, then we would expect the modified explanation to explain to a commensurate degree. Yet, we do not see this. When we vary the amount of information about dependence, we also seem to vary the degree of explanatoriness, in a way that suggests that it has little to do with independence information.

To develop the line of thought sketched above we need to lay some conceptual groundwork. To this end we distinguish three different dimensions of abstraction (§2). Then we show how these distinctions apply to prominent recent conceptions of explanatory abstractions (§3). Our key claim is that not all 'dimensions' of abstraction are responsible for explanatoriness, even if good explanations are typically abstract along several dimensions. Against the backdrop of these conceptual distinctions we will then present an alternative counterfactual account of the explanatoriness of abstract explanations in the spirit (but not the letter) of the dependency oriented account developed by Woodward [2003] and Woodward and Hitchcock [2003a,b]. In particular, we focus on how to select the right what-if-things-had-been-different counterfactuals (§4). The virtues of this unified account are illustrated by revisiting a graph-theoretic explanation of Königsberg's bridges (§5) and further underlined in relation to other challenges faced by less unified views of explanation (§6).

---

coming]. We take ourselves to be in the same tradition as these accounts. However, we will differ from these accounts in our argument that we need to specify more precisely the *kind* of what-if-things-had-been-different information that counts as explanatory (which we do in §4).

2 Three dimensions of abstraction

As said, we will challenge the idea that explanatoriness turns on independence of the explanation from the particular physical laws, processes, or concrete physical structures. Yet, we think that abstract explanations typically do exhibit this kind of independence. The task of this section is to distinguish three conceptually different ways in which explanations can be abstract. They will all involve independence in one sense or another. However, we will later argue that only one kind of independence carries information about explanatory dependence, and, as a consequence, contributes to explanatory power.

Let us recall some simple examples of abstract, plausibly non-causal, mathematical explanations (what we call 'abstract explanations') in order to bring out intuitions about different senses of explanatory abstraction. The following examples have featured prominently in recent philosophical analyses: (1) Economical bees build hexagonal honeycombs because this is the most resource-efficient way to divide a Euclidean plane into regions of equal area with least total perimeter [Lyon and Colyvan, 2008]. (2) Plateau's laws hold for soap film geometry, because this geometry minimizes the system's energy by minimizing the surface area [Lyon, 2012, Pincock, 2015]. (3) Königsberg cannot be toured by crossing each and every bridge exactly once because of the bridges' relational configuration [Pincock, 2007]. (4) The strawberry case discussed above is another example.

We will focus on (3) and (4). Abstract explanations hinge on (relatively) abstract features of reality that are non-causal (or perhaps causal in a very broad sense of 'causal').[6] This abstractness is obviously partly a matter of relative independence of the explanans from the *actual physical structure* of the entities involved: strawberries, bridges, soap films, honeycombs, etc. This is the first intuitive aspect of explanatory abstraction that we wish to highlight. There is nothing special about the physical nature of strawberries, for example, that makes them indivisible in this way. The very same explanation essentially explains why twenty-three marbles (say) cannot be thus evenly divided. The strawberry explanation abstracts away from whatever concrete features need to be in place in order for there to be a set of twenty-three *individuals* that can be divided into mutually exclusive and collectively exhaustive subsets.

A second, more radical aspect of abstraction pertains to the fact that the explanations above are independent from the *actual laws of nature* that underlie the physical processes presupposed by the why-question: the processes of group division, bridge crossing, or comb building, for example. There is nothing special, as far as these explanations are concerned, about the actual physics that underlies the solidity of honeycomb walls, strawberries, or bridges. The same explanations would work just the same in more exotic, counter-legal possible worlds where the underlying physical laws are rather different (assuming the relevant why-questions still make sense). The explanation of Plateau's laws, for example, transcends the nomologically possible causes or grounds of the area-minimizing action of soap bubbles. This explanation is modally robust with respect to variation in the underlying physical laws, as long as the laws still give rise to area-minimizing action for some surfaces. To this extent the explanation abstracts away

[6] The issue of exactly what kinds of dependences count causal is convoluted and largely orthogonal to our interests in this paper. What we want to understand here is how various abstract explanations explain.

from the actual nomological facts pertaining to the systems in question.

So far we have identified a conceptual distinction between two senses of abstraction. There is also a third aspect of abstraction in the above explanations; one that is more easily missed. This has to do with the explanatory regularities involved, and the degree to which an explanatory generalisation is independent from the actual value of an explanans variable. For example, the fact that twenty-three is not evenly divisible by three is but an instance of the more general fact that no integer apart from whole multiples of three is evenly divisible by three. In as far as we judge that the explanatory work is really done by this more general arithmetical background knowledge, we can say that essentially the same explanation would answer the question of why Mother would have failed had she had twenty-two strawberries (or why she could have succeeded had she had twenty-one). The explanation turns on a regularity that is robust by virtue of allowing the same explanation to be given for a wide range of changes to the numbers of strawberries. The wider this range is, the more abstract the explanation, in this third sense of abstraction.

The above three intuitive aspects of explanatory abstraction all clearly have something in common. Namely, each involves a way in which an explanation can effectively be independent from some feature of the system in question. This independence is naturally thought of as follows. An explanation can cover a range of possible systems. Explanatory abstraction is a matter of independence of the explanation from the details of 'realization'. In the above cases of abstract explanations the extent of unnecessary details omitted is striking. Consider the Königsberg case, for instance: the explanation is independent of all the physical and geometrical features of the bridges and the crossing processes, and of the underlying nomological facts pertaining to the bridges and their possible crossings, only relying on a very high-level global structural feature associated with a wide-ranging explanatory regularity. (We will make this more precise later in the paper.)

The extent of such independence is most striking in the case of abstract explanations, but more or less any explanation actually exhibits the three dimensions of abstraction to some degree. Consider the mundane case of gravitational pendulums, for instance. Why is the period of a given (more or less) 'ideal' pendulum 2 seconds? Explaining this causally, in terms of its length l and gravitational acceleration g, relies on a law-like regularity that supports the same explanation in a range of possible cases that vary in these two parameters. (The third dimension of abstraction, above.) Furthermore, the explanation can apply just the same regardless of whether the underlying nomological facts are rooted in a classical Newtonian world with gravitational force acting at a distance, or whether the pendulum occupies a general relativistic world with gravity being a manifestation of curved spacetime. (The second dimension above.) Finally, the why-question presupposes that the pendulum cord is inextensible, but the explanation is independent from whatever microstructural facts underlie this property. (The first dimension above.) In light of the fact that more or less all explanations are abstract to some degree along all three dimensions, it is prima facie surprising that there would be two radically different sources of explanatory power in play: one for the cases that are commonly thought of as causal and a different one for cases of 'abstract explanation'. We will argue that the difference between this mundane causal explanation and 'abstract explanations' is one of degree, not of kind.

The notion of explanatory abstraction as a matter of independence is clearly a modal notion: it concerns the range of possible systems covered by the explanation. Different aspects of explanatory abstraction have to do with different dimensions of modal variation with respect to these possible systems. We can think of possible systems that vary from the actual explanandum only in the (nomologically) possible physical structures of the entities involved; possible systems that vary in the microphysical or dynamical laws they obey; possible systems that vary from the actual explanandum by varying a variable that explicitly features in the explanatory regularity employed. Although modal variation along these different dimensions is often interlinked and metaphysically intertwined in complex ways, the dimensions themselves are conceptually independent.[7] In the strawberry case, thinking about the explanatory regularity with respect to a system of twenty-three vs. twenty-five strawberries, one need not think about variation in the physical structures of these entities or the division processes. Similarly, in thinking about the explanation with respect to systems that vary in the physical structures of the entities to be divided, one need not think of variation in their number, or of counter-legal possibilities where the underlying physical laws are different. And vice versa.

Often abstract non-causal and mathematical explanations—like the ones above—exhibit a striking level of abstraction in *each* of the three senses. We believe this has confounded some recent commentators. If one thinks, prima facie, that abstraction in a particular sense is the source of explanatoriness, it is all too easy to find corroborating evidence when so many good explanations are highly abstract in that sense. Now that we have conceptually separated the different dimensions of abstraction, however, we can ask: If explanatory abstraction has something to do with explanatory power (as is commonly accepted), exactly which dimension(s) of abstraction can contribute to it?

In its full generality, we cannot hope to comprehensively answer this question in a single paper. There is, however, a narrower question that we can answer. It is commonly agreed that explanation (and the associated notion of explanatory power) has both worldly and pragmatic/communicative aspects. Here, we will largely set aside the communicative and pragmatic aspects. What is at stake in the debate at hand is whether abstract explanations require facts tied to independence to function as a source of explanatory power, instead of (or in addition to) facts about dependence. In particular, are facts related to independence a worldly underpinning of explanatory power for the paradigmatic cases in the literature (mentioned above)? Our first aim is to undermine prominent recent arguments that answer 'yes' to the last question.

Our second aim is, more positively, to show how these abstract explanations can be accommodated in more unified terms, in the spirit of the counterfactual account developed by Woodward and Hitchcock [2003a,b] (the W-H account). According to the W-H account, explanations provide a particular kind of information about dependence. This account associates explanatoriness essentially with our third dimension of abstraction. (Roughly, an explanatory generalisation that has a broader range under which we can vary the values of the variables provides more explanatory information about the dependence of the explanandum on these variables than a corresponding explanatory

[7]The dimensions are metaphysically intertwined because, for example, various possible changes in the physical laws are bound to imply changes in the physical structures.

generalisation with a more restricted range of invariance.) This is in sharp contrast to recent claims according to which (some) abstract explanations explain either (i) by virtue of abstracting away from all the concrete physical features of the system in question (our first dimension above), or (ii) by virtue of abstracting away from the underlying laws of nature (our second dimension). The W-H account is well liked in connection with various causal explanations (even amongst the advocates of alternative analyses of abstract explanations), but there has been resistance to extending this counterfactual account to abstract explanations that are (plausibly) non-causal or mathematical. We will argue that this resistance is entirely unnecessary. Issues concerning explanatory power and its relation to abstraction are tractable in very natural terms in the spirit of the W-H account also for abstract explanations, after the key notions of this account are appropriately clarified and refined.

We will next critically review some recent claims regarding explanatory abstraction, questioning the intuitions that support them. After that we will more systematically discuss the third dimension of abstraction, which has a better claim to be associated with the source of explanatory power than either dimension one or two.

3 Explanatory power and independence

Various philosophers have recently argued (in different ways) for the disunity of explanatoriness by pointing to the sui generis role of abstraction as a source of explanatory power. We will focus on two authors who particularly clearly associate explanatory power with either the independence from concrete physical details (dimension one above), or the independence from the particular laws (dimension two above).

Pincock [2007, 257] brought the Königsberg bridge example into philosophical prominence by identifying it as an 'abstract explanation' that 'appeals primarily to the formal relational features of a physical system.' Relational features that are 'formal' are clearly meant to stand apart from causal relations. Furthermore, regarding the explanation of the impossibility of touring the town by crossing each and every bridge exactly once, Pincock [2007, 259] accounts:

> [A]n explanation for this is that at least one vertex [in the formal graph structure instantiated by the bridges] has an odd valence. Whenever such a physical system has at least one bank or island with an odd number of bridges from it, there will be no path that crosses every bridge exactly once and that returns to the starting point. If the situation were slightly different [so that] the valence of the vertices were to be all even, then there would be a path of the desired kind.

This is all surely correct. The question is how to capture this explanation in philosophical terms. Pincock [2007, 260] intimates that the explanation critically turns on abstraction:

> The abstract explanation seems superior [to a microphysical explanation] because it gets at the root cause of why walking a certain path is impossible

> by focusing on the abstract structure of system. Even if the bridges were turned into gold, it would still have the structure of the same graph, and so the same abstract explanation would apply. By abstracting away from the microphysics, scientists can often give better explanations of the features of physical systems.

The notion of 'abstracting away from the microphysics' is naturally construed as being along the first dimension identified above: a matter of relative independence of the explanans from the physical structure of the entities involved. As Pincock [2011, 213] puts it: 'the explanatory power [in this case] is tied to the simple way in which the model abstracts from the irrelevant details of the target system.' The intuition is that the explanation comes from stripping away as irrelevant all the physical details pertaining to the bridges' make-up, length, location, angles, et cetera, thereby highlighting what is relevant, namely the formal 'mathematical structure found in the target system itself' Pincock [2011, 213].

We can see Pincock's intuition, but we are unable to see a good reason for viewing this as a sui generis 'abstract explanation'. The explanation is undeniably highly abstract and plausibly non-causal. But the notion that good explanations only provide relevant information, leaving out unnecessary details, is common ground with dependence accounts of explanation. According to dependence accounts one ought to provide suitable information about what the explanandum depends on, and one also ought not to claim (or imply) that the explanation depends on something that it does not. The leaving out of irrelevant detail is insufficient to motivate the view that 'abstract explanations' are *sui generis* in the way Pincock regards them.

How about the notion that the explanatorily relevant features of Königsberg are 'formal-cum-mathematical'? Does this motivate a departure from familiar counterfactual accounts of explanation? We do not think so. We do not regard the explanans formal or mathematical in any sense that makes the application of counterfactuals difficult. The explanation involves *applied* mathematics; it is not an intra-mathematical explanation. There is a clear sense in which the explanans is just concerned with how many bridges there are to/from each of the 'islands'. It is even unclear why the 'valence' of an island—there being an even or odd number of bridges to/from it—is not a high-level *physical* feature of it. And similarly for the yet more abstract feature of there being *at least one* odd 'island' in the whole system of many 'islands'. Furthermore, there seems to be a clear and straightforward sense in which the explanandum at stake *depends on* these, arguably high-level physical, features: the explanation tells us how the explanandum would change if there were a different number of bridges (as Pincock himself notes; see the first quote above). Whether this explanatory dependence is causal or not is neither here nor there for the prospects of sticking to the core idea of the W-H account of explanation: explanatoriness is ultimately a matter of telling us what the explanandum *depends on*.[8] When it comes to identifying the worldly source of explanatory power,

[8] After heavily emphasising the idiosyncratic features of abstract explanations Pincock [2015] suggests that it may be possible to provide a unified account in terms of dependence. However, Pincock's [2015] notion of dependence is explicitly not in the spirit of the W-H account. We will address Pincock's [2015] objection to the W-H account in section §4.

we have not yet been given a reason to think that in the Königsberg case explanatoriness derives (even partly) from the *independence* of the explanandum from some worldly features.

Let us now move on to consider Lange's [2013] analysis of the explanatory power of the cited abstract explanations (as 'distinctly mathematical').[9] Lange [2013, 486–488] argues that his analysis of these explanations reveals 'a fundamental difference' between causal explanation and abstract non-causal explanation, and its significance 'lies in what it reveals about the kinds of scientific explanations there are'. Abstract explanations, Lange argues, explain not by supplying information about the world's network of causal relations but by 'showing how the fact to be explained was inevitable to a stronger degree than could result from the causal powers [actually] bestowed by the possession of various properties.' This is a nod towards Wesley Salmon's 'modal conception' of scientific explanations, according to which such explanations 'do their jobs by showing that what did happen had to happen' [Salmon, 1985, 293]. As Lange [2013, 505] puts it:

> [They explain] not by describing the world's actual causal structure, but rather by showing how the explanandum arises from the framework that any possible causal structure must inhabit, where the 'possible' causal structures extend well beyond those that are logically consistent with all of the actual natural laws there happen to be.

According to Lange, abstract explanations may *happen* to also provide modal information that a counterfactual account (in the spirit of W-H account) would count as explanatory. But he explicitly rejects the idea that their explanatoriness in any way derives from such modal information, relating it instead to information about the *independence* of the explanandum from some contingent features of the system in question.[10] In contrast to Pincock, the critical sense of independence for Lange is in the direction of the second dimension of abstraction identified above: abstraction away from the actual laws underlying the features of the system presupposed in the context of the relevant why-question. In relation to the Königsberg case, for instance, Lange [2013, 505–506] writes:

> [The] explanation of the repeated failure to cross the Königsberg bridges shows that it cannot be done (where this impossibility is stronger than physical impossibility) [...] The explanans consists not only of various mathematically necessary facts, but also [...] of various contingent facts presupposed by the why question that the explanandum answers, such as that the arrangement of bridges and islands is fixed. The distinctively mathematical explanation shows it to be necessary (in a way that no particular force law is) that, under these contingent conditions, the bridges are not crossed.

[9] The term 'distinctly mathematical' is a term of art in Lange [2013]. It is aiming to capture many of the cases that Pincock [2015] would call 'highly abstract'.

[10] In relation to the strawberry example Lange argues that the counterfactuals pertaining to the number of Mother's strawberries provide *causal* information, noting e.g. that 'clearly manipulation of the numbers of strawberries or children would bring about corresponding changes in the outcome of Mother's attempt.' Yet he maintains that 'this explanation is non-causal because it does not work by describing the outcome's causes or, more broadly, the world's network of causal relations.' See Lange [2013, 493–496].

Lange's account thus emphasizes the way in which the actual physical laws do not come into play in explanations such as this. Undoubtedly, Lange is right to note that the sense of abstraction at play goes beyond the first dimension concerning the irrelevance of nomologically possible physical realizations of the contingent features presupposed in the context of the why-question. The same explanation would work, in the same way and to same extent, in a far removed possible world with alien properties and laws—as long as there is a system with 'traversible bridges' for which the why-question makes sense.

But, having said that, why should we regard *this* sense of abstraction as a source of explanatoriness in connection with abstract explanations? The answer to this question is surprisingly difficult to find in Lange [2013].[11] There is little to directly motivate this, beyond the commonplace idea that good explanations do not mention irrelevant details, which in this case include all causal laws. Yet again, the idea that good explanations only provide relevant information, leaving out unnecessary details, is common ground with any account according to which we ought to provide information about dependence, *and* we ought not to claim (or imply) that the explanandum depends on something that it does not. Thus, the independence from the actual laws seems insufficient in itself to motivate the view that abstract 'mathematical' explanations are *sui generis* in the way Lange regards them.

Both Lange and Pincock are contrasting their views to causal accounts of explanation. It is natural, therefore, to take them to provide a competing account of the worldly source of explanatory power. So far we have undermined broad motivations for thinking that abstract explanations require a source of explanatory power that is not centred on dependence. Our view is that two of the three dimensions of abstraction—even though undeniably exhibited by abstract explanations—do not actually provide a source of explanatoriness.[12] The problem is that these dimensions of abstraction do not provide information about explanatory dependence in the W-H sense. We do not, unlike Lange and Pincock, take it to be the case that explanations that are highly abstract in the first two dimensions require an account of explanation where the source of explanatory power is

[11]Lange's motivations are best understood in the broader context of his related views on other kinds of non-causal explanations (e.g. Lange [2015]), and on the modal metaphysics of laws of nature in general [Lange, 2009]. Like other philosophical views, Lange's account of abstract explanations can gain indirect support from being a coherent part of a 'bigger picture'. Here we do not wish to assess the pros and cons of this bigger picture; we will focus on Lange's association of explanatoriness with the second dimension of abstraction in connection with the paradigmatic examples of abstract explanations.

[12]This is *not* to say that they can never be relevant to *any* aspect of good explanation. To only provide relevant information is a shared commitment of any account of explanation (cf. Strevens [2008] on causal explanation). This goes some way towards explaining the intuition that dimension one and two provide a separate source of explanatory power. For example, if we assume that we have a mistaken belief that the island taken as a starting point is relevant to explaining the failure to complete a round-tour of Königsberg, then it is enlightening to find out that the starting island is irrelevant. However, a dependence account can easily capture this intuition. The failure to complete a round-tour of Königsberg does not depend on the starting island. A good explanation cannot cite misinformation about the dependences. Note that the intuition that this kind of information about independence is a source of explanatory power disappears once we are not focusing on a case of correcting a mistaken belief about dependences. We are not tempted to regard as explanatory the information that failure to complete the tour is independent of the fact that the bridges are made of stone and not wood. We never mistakenly believed that the failure to complete the tour depended on these features.

not in the W-H spirit.[13] In the rest of the paper we will discuss this specific notion of explanatory dependence and how it connects to the third dimension of abstraction.

But before we move on, let us address a natural worry. It may seem that dependence and independence are conceptually so closely connected that we cannot cleanly separate the two. In particular, any information about independence is also information about dependence: 'E is independent from A' means that 'E does *not* depend on A', and if we assume that an explanandum E depends on something, then 'E is independent from A' entails that 'E depends on something else than A'. It is thus the case, the worry goes, that information about independence along every dimension of abstraction can provide information about dependence.

In response, we simply note that it is *not* true that information about independence ipso facto provides information about dependence that is *explanatory*—that is, dependence information in the W-H sense. It is undeniably true that one provides what-if-things-had-been-different information, broadly speaking, by showing that even if the laws of nature were different, Mother could not divide twenty-three strawberries equally among three children. (What if things had been different with respect to laws of nature? Then Mother would have failed just the same!) But this is *not* the sort of what-if-things-had-been-different information that counts as explanatory in the counterfactual framework that we favour, according to which only what we call *change relating* counterfactual information counts as explanatory.[14] To understand Mother's failure, we need to be able to say under what conditions Mother could have succeeded. We now move on to discuss this framework.

4 Abstraction in a counterfactual framework

The starting idea of the W-H account is that explanation 'is a matter of exhibiting systematic patterns of counterfactual dependence' (Woodward [2003, 192]). To develop this into a theory of explanation, we need to say more precisely *which* patterns of counterfactual dependence matter for explanation. On the W-H account this is done by focusing on modal information that allows one to answer questions about how the explanandum would have been *different* under a special type of change in the explanans [Woodward, 2003, 192]. The changes in the explanans that are relevant are those that result from interventions on the explanans variable(s) (with respect to the explanandum variable). The notion of intervention is technical; it plays the role of ruling out changes in the explanans variable that are brought about by (i) changes in the explanandum variable, and (ii) changes to some other ('common cause') variable that changes the explanandum variable independently of the explanans variable.

This approach has a number of attractive features. First, it is natural to take relations (whether these are nomological, causal, or mathematical) that are explanatory to be capable of providing some sort of modal information. Second, by requiring more specifi-

[13] As we noted above, Lange [2013, 493–496] explictily considers and rejects this option. Pincock [2015] is sympathetic to the general idea of dependence accounts but explicitly rejects W-H dependence as a viable candidate.

[14] In the terminology of Woodward and Hitchcock [2003a,b] an explanatory generalisation must be invariant under at least one *testing* intervention.

cally that the modal information relates to different possible states of the explanandum—how the actual explanandum would have been different had the explanans been different—the account captures the natural idea that explanatory information is information about worldly dependences. In the W-H account the critical modal notion of dependence gets cashed out in causal terms through the notion of a (testing) intervention. This allows one to rule out backtracking counterfactuals—for example, 'had the period of the pendulum been different, then the length of the pendulum would have been different'—as providing the right dependences. The focus on what-if-things-had-been-different questions that concern *changes* in the target explanandum also provides a clear contrast to mere subsumption under modal regularities in the spirit of the DN account of explanation. Third, this focus on interventions squares well with much of our experimental practice. Fourth, the account is, as Woodward and Hitchcock [2003a,b] discuss at length, very well suited to capture not only what it takes to have an explanation, but also what makes an explanation better or worse. Having an explanation is a matter of having the right information about the dependences, and *the more the better*.

The W-H account separates an explanans into two parts: a specification of an invariant explanatory generalisation and a specification of the actual values of the variables in the explanatory generalisation.[15]

For a simple illustration, let us look at the explanation of the period of a simple gravitational pendulum. Let us assume that we want to explain the fact that the period T takes the value t_1 (our explanandum M) by using the simple pendulum law. In terms of the W-H account, we have an explanatory generalisation $T \approx 2\pi\sqrt{\frac{l}{g}}$. The explanans consists of this generalisation and in a specification of the actual values of l and g. In order to have an explanation on the W-H account, the simple pendulum law has to (at least approximately) correctly give the actual value of the period as t_1 under an intervention that fixes the values of l and g to the actual values. Moreover, the simple pendulum law must capture (in at least one case) how the period would *change* under at least one (and ideally more) interventions that change(s) the length or the gravitational acceleration. It thus captures the *dependence* of the explanandum on these variables.

Now we are in a position to make the contrast between the different dimensions of abstraction more precise. Let us take the first dimension of abstraction first. In the context of the simple pendulum explanation, we can ask whether the explanandum $T = t_1$ is independent of particular background conditions which are in some sense part of the system but not represented as variables in the explanatory generalisation $T \approx 2\pi\sqrt{\frac{l}{g}}$. We can think of the microphysical features that make the pendulum cord inextensible, for example, or the mass of the bob, and so on. The degree of *independence* of the explanation from such background conditions was the focus of the first dimension of abstraction. (In contrast to such 'internal' background conditions there are other background conditions that are external to the system and clearly irrelevant, such as, the bob being made in Japan, the exchange rate of the US dollar to the Malaysian ringgit, the colour of the shirt of the person setting the bob in motion, etc.)

In the second dimension of abstraction, we considered the degree of independence of the explanation from the actual laws of nature. In the context of the pendulum example,

[15] See Woodward [2003, 203] for a detailed account.

we can ask whether we can vary the actual laws of nature and still expect the simple pendulum explanation to apply. We can alter many of the laws of, for example, electromagnetism, without affecting the explanation, provided that those alterations do not run afoul of the assumptions presupposed in the context of the why-question, for example, the cord being nearly inextensible. As mentioned earlier, we can also alter many of the fundamental aspects of gravitational acceleration, such as whether its nomological basis is an action at a distance effect or a manifestation of curved space-time.

Finally, in the third dimension of abstraction, we consider the range of conditions directly relevant to changes in the variables in the explanatory generalisation, $T \approx 2\pi\sqrt{\frac{l}{g}}$ such that the simple pendulum explanation of the period holds. For example, we could decrease g by, say, moving the pendulum to a higher altitude or to the moon, and we could change the length of the pendulum, and yet, the explanation of the period would work in just the same way. On the W-H account, abstractness in this dimension is just a matter of the invariance of the explanatory generalisation, 'measured' by the range of alternative values of variables l and g for which the generalisation holds and is change relating.

So far we have simply described the W-H account. Before we can apply these ideas to the paradigmatic examples of abstract explanations, we need to extend and refine some of them. In particular, the W-H account was developed as an account of *causal* explanation, while the abstract explanations we are interested in are plausibly non-causal. As we move beyond causal explanation, we can no longer appeal to the exact understanding of the relevant class of explanatory counterfactuals that Woodward and Hitchcock use. They delineate the class of relevant counterfactuals in terms of the causal notion of intervention: explanatory counterfactuals describe the effect of a testing intervention. These counterfactuals are underwritten by contingent explanatory generalisations, and the effects of surgical testing interventions are naturally understood as concerning causal dependences. In the non-causal cases this notion of causal dependence becomes inapplicable or unnatural, because the notion of causal intervention is inapplicable, or because the counterfactuals are not underwritten by contingent laws of nature (but, rather, by logic or mathematics, or metaphysical truths). Is it still possible in these cases to characterise explanatory dependences in counterfactual terms?

The answer is yes; we have analogous explanatory dependences in the situations involving Mother's strawberries, Königsberg's bridges, etc.[16] The challenge is how to select the *right* counterfactuals without assuming that causal notions are applicable. In responding to this challenge we note, first of all, that regardless of whether or not we can apply the idea of a causal intervention to the explanans variable, we can still have a good grasp of what it means to *change* the explanans: e.g. to change the number of strawberries, or the bridge configuration in Königsberg. Furthermore, regardless of whether or not the explanatory generalisation is a contingent law of nature or a stronger modal truth, we can reason about how the explanandum depends on these changes, and whether the dependence is non-symmetric in a way that supports explanatory what-if-

[16][Woodward, 2003, 221] notes passingly that it seems natural to extend the account to non-causal explanations, but he does not develop it further. For more discussion on how to decouple the counterfactual aspect and the causal aspect of Woodward's account, see Saatsi and Pexton [2013], Saatsi [forthcoming], and Rice [2015].

things-had-been-different reasoning.

For example, in the case of Mother we have a very good grasp of what changing the explanans variable amounts to. Indeed, changing the number of strawberries seems as straightforward an intervention as any. We also know how this changes the system's divisibility-by-three, and consequently the possibility of Mother's success or failure. We thus grasp how Mother's success or failure explanatorily depends on the number of strawberries that she has (and the number of children). In order for the dependence to be regarded as explanatory it needs to be appropriately directed. Intuitively, it is. As we indicated earlier, the notion of intervention plays the role of blocking changes to the explanans that go through changes to the explanandum (or through a common cause variable, etc.). Similarly, changes to the number of strawberries do not seem to go through changes to the system's divisibility-by-three, or through changes to Mother's success or failure. After all, changing the system's divisibility by three does not mandate any particular number of strawberries. However, changes to Mother's success or failure do seem to go through changes to her number of strawberries (or children).

This intuition can be supported more precisely as follows. On the W-H account a successful explanation makes use of an explanatory generalisation for two main tasks (as outlined above). First, to show that setting the value of the explanans variable to its actual value fixes the value of the explanandum variable to its actual value. Second, to show that changes to the value of the explanans variable away from its actual value will, in some cases, change the value of the explanandum variable. Now, consider the intuition that changes to the system's divisibility-by-three go through changes in the number of strawberries, but not vice versa. This is underwritten by the fact the explanatory generalisation regarding the number of strawberries and their divisibility-by-three supports the first task above asymmetrically. Fixing the explanans variable to its actual value should fix the explanandum variable to its actual value, but fixing the system's (non-)divisibility-by-three does not fix the number of strawberries to any particular value. In contrast, the number of strawberries being twenty-three does fix the system's (non-)divisibility-by-three. This asymmetry underlies the intuition that changes to the number of strawberries do not go through changes in the system's divisibility-by-three.[17] The asymmetry does not seem to be a causal matter, however, since the intervention on the number of strawberries can have an 'effect' (e.g. rendering the set divisible-by-three) that is rather more intimately related to it. To this extent Lange [2013] is absolutely right to stress the independence of this connection from any causal laws.

Very similar considerations apply to the case of Königsberg's bridges.[18] In this case, it's not clear whether the explanans variable—concerning the global configuration of the whole bridge system—can be changed in a way that necessarily satisfies our intu-

[17]The explanatory generalisation regarding the number of strawberries and their (non-)divisibility-by-three supports some counterfactuals running in the opposite direction, of course. For instance, it's true that had the system's (non-)divisibility-by-three been different, then the number of strawberries would have been different. So why is this counterfactual not explanatory of the number of strawberries? Answer: the explanatory generalisation, with the supposed 'explanans' variable fixed, does not say what the actual number of strawberries is.

[18]The full story here is somewhat more complicated since we are modelling the physical bridge system and there are considerations of the applicability of the model. The importance of considerations of applicability for directionality of explanation is stressed by Jansson [2015].

itions about what counts as a causal intervention. But be that as it may, we can surely have a very good grasp of what it means to change the explanans variable (by strategically building and destroying bridges so as to achieve a particular overall configuration). And again, the explanatory generalisation tells us how this changes the bridge system's traversability that depends on the configuration being in a certain way. (We will lay this out in more detail in the next section.) This dependence is again appropriately directed: while fixing the bridges to be a particular way fixes the (non-)traversability of the bridge system, the reverse does not hold. We cannot fix the bridges to a *particular* configuration merely by ensuring that the bridge system is (say) non-traversable.[19]

There is more to be said about the nature of explanatory dependence in abstract explanations, of course. One can wonder about the metaphysics of the dependence relation, for example, as we have not said anything about it. Our hope is that one doesn't have to get into metaphysics of dependence in supporting the broader idea of there being such explanatory dependencies in the world, transcending those readily incorporated into the W-H account as it stands. It is enough to show how we can pick out the right counterfactuals. (This is in tune with Woodward's own attitude to this issue.)

Or one may wonder what to make of the fact that in the cases at hand the explanatory relation does not seem to connect two distinct events; it is not that 'we change one thing and another change follows.'[20] For instance, the event of destroying some of Königsberg's bridges can *ipso facto* be the event of rendering the bridges Euler-tourable. In response, we would like point out that it is critical to the W-H account to leave behind the idea that explanatory relations hold between events (or event-types); rather, explanatory relata are treated as *variables*. (See e.g. Woodward and Hitchcock [2003a, 10-11]) In causal cases this amounts to a notion of explanatory causal relevance between such variables. Similarly, in the cases at hand it is natural think of explanatory variables as designating different aspects of the system in question. We can make sense of such variables standing in an explanatory dependence relation, even when changes in the variables' values involve one and the same event. (See Saatsi [2016] for related discussion of explanatory variables in connection with non-causal explanations from 'geometry of motion'.)

There is also more to be said about the directionality of explanation. Here we refer the reader to work one of us has done elsewhere, as exemplifying a possible way of developing further the ideas we have only briefly sketched here.[21] But we hope to have said enough to show that the prospects for a more detailed account are no worse—and in our opinion they are better—when extending a dependence account in the W-H spirit

[19]The true counterfactual 'Had the traversability of the bridge system been different, then the system of bridges would have been different' does not explain the actual bridge configuration for reasons analogous to those discussed in footnote 17.

[20]We thank an anonymous reviewer for raising this question. (S)he put it in these terms.

[21]See [Removed for anonymous review] for a suggestion for a nomological dependence account and [Removed for anonymous review] for a dependence account for metaphysical cases. There is also more to say about how we should understand optimality explanations, for example, the honeycomb case. We believe that a similar account to the one that we have given here holds for the honeycomb case too, but this has to be argued on a case-by-case basis. The discussion of many of these cases is made more complicated by the existence of closely related *causal* explanations (as Lange [2013, 500] notes for the honeycomb case).

than when focusing on independence along the lines of Lange and Pincock.[22]

Our suggested diagnosis of the original W-H account is that it correctly identifies explanatory counterfactuals as those that are appropriately change relating and directed. From now on, let us call the explanatory counterfactuals simply *change relating counterfactuals* (adapting Woodward's use of the term) to keep in mind that they are a strict subset of all counterfactuals and a broader class than interventionist causal counterfactuals. Before we can move on to consider a detailed application of the account, we need to consider one final aspect of extending the W-H account.

Since the relevant counterfactuals are selected by focusing on *changes* to some object or system, the modal information that we are interested in is tied to the system or object in question. They are what Woodward calls "same object" counterfactuals":

> Suppose that we wish to explain the behavior of some object or system o. As the standard view is usually understood, it claims that generalizations of form "All As are Bs" are explanatory of the behavior of o if they support counterfactuals of the following form: ...If some object o^*, different from o and that does not possess property A, were to be an A, then it would be a B.
>
> Call such counterfactuals "other object" counterfactuals: they describe what the behavior of objects other than o would be under the counterfactual circumstances in which they are A. By contrast, according to the view I have been defending, to count as invariant and hence explanatory with respect to o, a generalization must support "same object" counterfactuals that describe how the very object o would behave under an intervention. [Woodward, 2003, 281]

At first glance this may seem to suggest that all of the explanations that we have in mind must be about specific, *particular* systems or objects in order for the distinction—central to the W-H account—between "same object" (SO) and "other object" (OO) counterfactuals to apply. Pincock [2015] takes this to rule out the application of the W-H account to abstract explanations, since some highly abstract explanations do not seem to provide information about any particular object at all (much less information about changes to any particular object).

In our view, one should not think of the contrast between SO counterfactuals and OO counterfactuals in the way Pincock does. First, note that even in the case of causal explanations, most scientific explanations do not have a particular, individual object as the explanatory target. The explanandum is typically generic (and general even when couched in language such as 'the period of a simple pendulum'). The importance of the distinction between SO and OO counterfactuals lies in the fact that it focuses our attention on the right kind of *change relating* counterfactuals. When we make use of the

[22] In Pincock [2015, 865] the directionality comes from an otherwise unanalysed relation of instantiation, and in Lange [2013, 508] the directionality comes from an otherwise unanalysed notion of some features being constitutive of a system. In our view, while these are potential avenues for identifying the source of non-causal directionality, they are neither more plausible nor better understood than the alternative provided by change-relating counterfactuals.

simple pendulum law to explain the period of a simple pendulum, the right counterfactuals to have in mind are those that ask how the period of a generic kind of dynamical system, viz. simple pendulum, is affected by changing, say, its length. Counterfactuals concerning objects other than simple pendulums do not come into play, even when they are reasonable and well defined. (It may be true that had a hammer been suspended around an appropriate pivot with the head down, then we could have used the simple pendulum law to explain its period, but such OO counterfactuals are not required in order to explain the period of a simple pendulum.)

Although the W-H account does not apply directly to the case of Königsberg's bridges, the central notion of change relating (SO) counterfactuals can be carried over to this case. This is the work of the next section.

5 Königsberg – Encore!

We will now revisit the Königsberg case to illustrate the above counterfactual account and its virtues. So far we have mainly criticised Pincock's and Lange's motivations for thinking that there is a sharp distinction between causal explanations and abstract non-causal explanations, maintaining that one can instead approach both types of explanations in the fundamentally same spirit. We now push for a stronger claim, arguing that the core claim of the counterfactual account (as we understand it), that explanatoriness is associated exclusively with the third dimension of abstraction, can be *tested* against the alternative viewpoints. This will provide a clear reason to prefer the counterfactual account.

Let us go back to the 18th c. Königsberg, and ask Q_K: Why is it impossible to make a round-tour of Königsberg crossing each of its seven bridges exactly once? An intuitive explanation-sketch response goes as follows. Clearly each visit of a landmass ('island') requires the use of two bridges: one in, and one out. Else you get stuck. Therefore, in order for a network of bridges to allow for a round tour—to be 'tourable'—each island must have a number of bridges to/from it that is some multiple of two. On the other hand, if there is one (or more) island(s) with an odd number of bridges to/from it, the system is not tourable. 18th-century Königsberg had four such troublesome junctions, rendering a round-tour of Königsberg impossible.

Euler initiated a famed graph-theoretic explanation that makes the above sketch precise. This is standard material in graph-theory textbooks, illustrating the explanatory use of mathematical notions (such as *connected graph*, its *vertices*, and their *degrees*).[23] But before we get to the graph-theoretic explanation in its full generality, it is worth attempting to answer Q_K *without* graph theory. In the counterfactual framework explanations must involve an invariant, change-relating generalisation that supports counterfactuals indicating an explanatory dependence of the explanandum on the explanans. A generalisation that thus underwrites an answer to Q_K could be markedly less abstract and less general than the graph-theoretic explanation, without thereby being unexplanatory.

[23] It is unsurprising that many authors refer to this explanation as a paradigmatic case of a (distinctly) mathematical explanation of an empirical fact. In addition to the authors already covered, see e.g. Lyon [2012].

As a matter of fact it is easy to find a simple invariant generalisation that furnishes a non-mathematical answer to Q_K. Let us focus our attention on the kind of bridge system that connects *exactly* four islands, with *at most* two bridges between any two islands. Euler's Königsberg (represented below) is one of these systems.

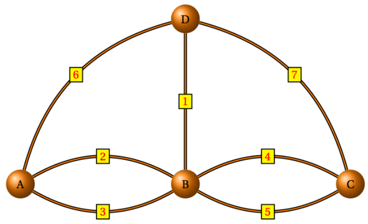

Figure 1: Königsberg's bridge system.

There are 395 such bridge systems altogether. We can classify them as follows. Call a bridge system *'even'* iff each island has an even number of bridges to/from it. In this particular case, this means that each island must have 2, 4, or 6 bridges to/from it. If a bridge system is not even, call it *'odd'*. Considering the specific type of bridge system exhibited by Königsberg, we ask: why is it not tourable? In answering this question we naturally look for an explanatory generalisation capable of providing suitable modal information by supporting appropriate, change relating counterfactuals. Focusing our attention, for now, only on the bridge systems consisting of four connected islands with at most two bridges between any two islands, a fitting generalisation is not hard to find: of all these bridge systems, all and only the *even* ones are tourable.[24] This is a *true* generalisation about this set of 395 different types of bridge systems. It is also a generalisation that can afford us with a degree of explanatory purchase on Q_K. That is, by reference to this generalisation we can begin to answer Q_K simply by noting that Königsberg's bridge system is not tourable, because it is not *even*; it would be tourable, if it were *odd*. Königsberg's tourability—the feature that is our explanandum—*depends on* this high-level physical property of the system.[25]

More formally, we can define the following binary variables X and Y:

[24]This fact about these bridge systems can be in principle established by a variety of means, e.g. by attempting to draw every such system without lifting your pen, or by playing with a comprehensive collection of miniature models of such systems. Getting epistemic access, or representing that fact, in principle need not involve mathematics. (cf. Saatsi [2011])

[25]Why do we regard this as a (high-level) *physical* property? Because adding or subtracting a bridge between any two islands plainly makes for a physical difference, and we can get from an *even* configuration to an *odd* configuration simply by adding and/or subtracting a sufficient number of bridges. Furthermore, if we are dealing with a finite number of possible bridge systems (as above, focusing on 395 particular configurations), we can in principle do without the mathematical concepts of even and odd in delineating the two kinds of configurations, since the difference between *even* and *odd* is expressible logically without mathematics. So, we see little reason to regard these as mathematical or purely formal properties of the bridges. Having said that, we should stress that we are not hereby claiming that the Königsberg explanation is a causal one, and nothing in our analysis hangs on whether or not the explanation counts as a mathematical one.

$X =$ *even/odd* system
$x = 1$: *even*
$x = 0$: *odd*

$Y =$ (non-)tourable system
$x = 1$: tourable
$x = 0$: (non-)tourable

The simple explanatory generalization at stake then states that for bridge systems of this kind—four islands, maximum of two bridges between any pair—it holds that

$$X = Y$$

This equation should be read left to right, as indicating an asymmetric dependence of Y-variable on the X-variable. The tourability (or otherwise) of a bridge system depends on it being *even* (or *odd*). Being *even* (or *odd*) does *not* depend on tourability; it only depends on the number of bridges. All in all, the explanation naturally fits the counterfactual framework presented above (§4).

The explanatory generalisation, while narrow, is explanatory nevertheless, even if minimally so. Undoubtedly, the explanation provided immediately raises further questions. Why exactly is this generalisation true, for example? What if we start adding or subtracting 'islands'? What if we relax the restriction that there are at most two bridges between any pair of islands? These are obvious further questions, and this clearly renders the explanation shallow and somewhat contrived, especially in comparison to a full-blown graph-theoretic account. But none of this diminishes the philosophical significance of this explanation. For however minimal and shallow the toy explanation is, we deem it to have *some* explanatory power nevertheless, and we maintain that this is due to the explanation providing modal information of the right sort.[26]

The shallow explanation above clearly has a degree of abstractness along the three dimensions introduced in §2. Indeed, the explanation is *highly* abstract with respect to the different specific material realizations of the bridges (the first dimension), and also with respect to the underlying laws of physics (the second dimension). As a matter of fact, the explanation is as abstract along these dimensions as the full-blown graph-theoretic explanation! Yet the explanation is shallow. This clearly speaks against the idea that the (minimal) explanatoriness in question springs from abstraction along these lines, as Pincock and Lange would have it.

[26] Is minimal explanation of this sort ever completely *satisfactory*? That depends on the context of the why-question. Imagine Königsberg without Euler—or any other mathematician for that matter. The King of Königsberg, annoyed by his failure to do a round-tour of the city, wishes to understand the situation. "This city must be made thus tourable, at whatever cost!", he commands, putting his best people to work. "Well, actually, the less it costs the better—and we certainly cannot afford more than two bridges between any two islands!" The King's minions get to work, and painstakingly demonstrate, by non-mathematical means, that for any financially feasible set-up of bridges it holds that it is tourable if and only if it is *even*. Equipped with this knowledge, they can explain the situation to the King: "What ever you do with the bridges, make sure the whole system is *even*, and you will be able to tour it, since tourability depends on this feature alone (at least for the kinds of bridge systems we can afford)."

With regard to the third dimension of abstractness, the explanatory generalisation does allow us to answer some change relating what-if-things-had-been-different questions. We have counterfactual information of the right, SO, kind. We can answer counterfactual questions about how the tourability of a generic kind of structural system, viz. bridge system, is changed by changes to the oddness or evenness of the system (at least as long as we stay within the constraints of four connected islands and a maximum of two bridges between any two islands). This is what makes it explanatory, even if shallowly so. Yet, its degree of abstractness along this dimension is very limited. Although the explanatory generalisation never delivers the *wrong* answer, the generalisation is simply silent on what happens in cases of more than four islands or more than two bridges between some islands. Thus, the explanation using the generalisation breaks down in these cases. [27]

In order to understand precisely how the shallow explanation compares in its abstractness to the deeper graph-theoretic explanation, we need to pay attention to the fact that the variable X concerns a *determinable* property of the system: it being *even* or *odd*. This determinable property is determined by the bridges' configuration. There are various ways for a bridge system to be *even*. One way is for each island to have (say) 4 bridges to/from it. This is still a determinable property of the system, determined by a specific way of having 4 bridges to/from each island. (Cf. Figure 2)

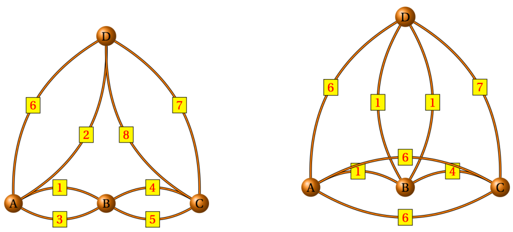

Figure 2: Two determinate 4-4-4-4 configurations.

Here is another way for a bridge system to be *even*: one of the islands has 8 bridges to/from it, and the other three islands have 2, 2, and 4 bridges to/from them, respectively. While our generalisation supported some explanatory what-if-things-had-been-different questions, it is simply silent on the system's tourability if we change X from *odd* to *even* by allowing one island to have as many as 8 bridges to/from it. Similarly, it is silent on what happens if we change X from *even* to *odd* by including 7 bridges to/from some island. To answer what-if-things-had-been-different questions corresponding to these cases, we need a broader generalisation, $X = Y$, that applies to systems as rich in bridges as these.

[27] Here two differences between our third dimension of abstraction and (even the extended) notion of invariance in Woodward and Hitchcock [2003a,b] are important. First, we take the question to be whether the variations destroy the explanation (not merely the generalisation). The generalisation itself does not break down. It just does not apply. We take this focus on the whole explanation from Potochnik [2010] and Weslake [2010]. Second, the range of variations are understood as the range of cases in which one of the variables in the explanans can be changed without destroying the explanation. While this is also the idea of the W-H account, in Woodward and Hitchcock [2003a,b] it is often natural to interpret the range of invariance as the range of changes in the *variable values* for which the generalisations continue to hold.

Differences in the determinate configurations do not explicitly feature as a variable in the explanatory generalisation $X = Y$. Indeed, on the face of it the explanatory generalisation looks the same regardless of the range of determinate what-if-things-had-been-different questions it is taken to support. Differences in the range of determinate realizations of evenness/oddness do get into play through the restriction in application. When we consider *changes* to the variable (*even*, *odd*) these changes have to go through changes in the bridge system (there is no way to change the evenness of the system that does not go through changing the specific configurations of the bridges). If we are justified in taking the generalisation $X = Y$ to apply to systems that have (say) a maximum of three bridges between any two islands, then our explanation covers a wider range of conditions directly relevant to changing the value of the explanans variable X. In particular, we can now also consider what-if-things-had-been-different situations with 7, 8 or 9 bridges to/from some island(s).

In this way we can straightforwardly compare different Königsberg explanations with respect to their degree of abstraction in the third dimension. The explanation where we restrict the generalisation to a maximum of three bridges between any two islands applies to all the ways of varying the variable in the explanatory generalisation that the explanation restricted to a maximum of two bridges covers—and then some! The full-blown graph-theoretic explanation is, of course, maximally abstract along this third dimension of abstractness. The explanatory generalisation now covers any (connected) bridge system of arbitrary many islands and bridges. The graph-theoretic explanation has considerable depth in contrast to the shallow, non-mathematical explanation. This is solely due to increased abstraction along the third dimension. The explanatory dependence of the shallow explanation is subsumed under a more general explanatory dependence, enabling us to answer a much wider range of change relating *what-if-things-had-been-different* questions. The shallow explanation does not contain irrelevant information about the nature of the bridges, their material constitution or their length, say, or about the underlying nomological features, such that the increase in explanatory depth could be due to abstracting away from such information. Indeed, as already noted, the explanans of these two explanations are already maximally abstract along the first and second dimensions, and the considerable increase in the explanatory power should be attributed solely to the third dimension of abstraction.

6 Conclusion

We have argued that paradigmatic abstract (plausibly) non-causal explanations can be naturally accommodated with an account that associates explanatoriness with suitable information about dependence. This improves the prospects of subsuming many explanations under a unified counterfactual framework, opposing the current trend that emphasizes the explanatory value of abstraction as a sui generis source of explanatoriness. We found the motivations for this trend questionable (§3), leaving room for a more unified account.

More unified theories are often better, ceteris paribus, but we are not just expressing this kind of a prima facie preference for unification. Rather, we motivate the unified account by the following considerations. First, we argued that the unified account better captures intuitions about 'explanatory depth'. Part of the force behind the W-H framework comes from the fact that it provides a natural starting point for capturing intuitions about explanatory goodness (Woodward and Hitchcock [2003a,b] and Ylikoski and Kuorikoski [2010]). We argued, in the same spirit, that our viewpoint can be tested by varying dependence information—the hypothesized source of explanatoriness—keeping other things fixed (§5). We maintain that our intuitions about radically varying explanatory power naturally correspond to radically varying amounts of W-H dependence (but not independence) information, in a way that is difficult to accommodate from the alternative viewpoints.

Secondly, the unified account has a further virtue worth flagging. This has to do with *dissolving* difficult questions facing the more disjunctive accounts that regard abstract explanations fundamentally different from causal explanations in their source of explanatoriness. Noting that causal explanations also typically abstract away from a huge amount of physical detail raises a question about Pincock's point of view, for example: What explains the distinct qualitative difference between causal explanations that exhibit a degree of abstraction along each of the three dimensions, on the one hand, and the *sui generis* 'abstract explanations', on the other? Is there a 'threshold' of abstraction above which the counterfactual conception fails, despite capturing abstract causal explanations so well?

A similar question can be raised for Lange's account. Many causal explanations also abstract away from a huge amount of underlying nomological detail. Thus, all causal explanations that incorporate a degree of abstraction along the second dimension of abstraction also show how the explanandum is, to a corresponding degree, 'necessary to a stronger degree of necessity' by virtue of showing the irrelevance of some of the actual laws involved. So why is it that such modal information about independence becomes explanatory *in a sui generis way* in connection with abstract explanations? Or is it the case that such modal information always contributes to the explanatory power, but it contributes in a different way when an explanation abstracts away from (almost) *all* nomological information, as in the case of 'mathematical' explanations?[28] Why do abstract explanations explain so differently—in Salmon's 'modal' mode—from causal explanations that incorporate a degree of similar abstraction?

In our view Pincock and Lange have not provided satisfactory answers to these questions. In particular, their arguments that abstract explanations do not work by providing information about *causal* dependence do not show that they work in a radically different way instead (as opposed to providing information about non-causal W-H dependence). The more unified account we advocate has the virtue of sidestepping these issues entirely. According to this account the paradigmatic abstract explanations, despite their non-causal character, are explanatory for the fundamentally same reason as causal explanations are. In both cases the explanatory power springs from counterfactual information of the same sort, and the paradigmatic exemplars of abstract explanations need not be regarded as *sui generis*.

[28] *Almost* all, since these explanations still contain information associated with the various contingent facts presupposed by the why question.

References

Robert Batterman and Collin Rice. Minimal Model Explanations. *Philosophy of Science*, 81:349—376, 2014.

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

Alisa Bokulich. Can Classical Structures Explain Quantum Phenomena? *British Journal for the Philosophy of Science*, 59(2):217–235, 2008b.

Alisa Bokulich. Distinguishing Explanatory from Non-Explanatory Fictions. *Philosophy of Science*, 79(5):725–737, 2012.

David Braine. Varieties of Necessity. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 46:139–70, 1972.

Lina Jansson. Explanatory Asymmetries: Rehabilitating Laws of Nature. *Journal of Philosophy*, CXII(11), 2015.

Marc Lange. *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*. Oxford University Press, 2009.

Marc Lange. What Makes a Scientific Explanation Distinctively Mathematical? *British Journal for the Philosophy of Science*, 64(3):485–511, 2013.

Marc Lange. How to Explain the Lorentz Transformations. In S. Mumford and M. Tugby, editors, *Metaphysics and Science*. Oxford University Press, Oxford, 2015.

Aidan Lyon. Mathematical Explanations of Empirical Facts, and Mathematical Realism. *Austrasian Journal of Philosophy*, 90(3):559–578, 2012.

Aidan Lyon and Mark Colyvan. The Explanatory Power of Phase Spaces. *Philosophia Mathematica*, 16(3):227–43, 2008.

Christopher Pincock. A Role for Mathematics in the Physical Sciences. *Noûs*, 42:253–275, 2007.

Christopher Pincock. On Batterman's 'On the the Explanatory Role of Mathematics in Empirical Science'. *The British Journal for the Philosophy of Science*, 62:211–217, 2011.

Christopher Pincock. Abstract Explanations in Science. *British Journal for the Philosophy of Science*, 66(4):857–882, 2015.

Angela Potochnik. Levels of Explanation Reconceived. *Philosophy of Science*, 77(1):59–72, 2010.

Alexander Reutlinger. Is There A Monist Theory of Causal and Non-Causal Explanations? *Philosophy of Science*, forthcoming.

Collin Rice. Moving Beyond Causes: Optimality Models and Scientific Explanation. *Noûs*, 49(3):589–615, 2015.

David-Hillel Ruben. *Explaining Explanation*. Routledge, 1990.

Juha Saatsi. The Enhanced Indispensability Argument: Representational versus Explanatory Role of Mathematics in Science. *The British Journal for the Philosophy of Science*, 62(1):143–154, 2011.

Juha Saatsi. On explanations from geometry of motion. *The British Journal for the Philosophy of Science*, page axw007, apr 2016. doi: 10.1093/bjps/axw007. URL http://dx.doi.org/10.1093/bjps/axw007.

Juha Saatsi. On Mathematics' 'Indispensable Explanatory Role'. *Mind*, forthcoming.

Juha Saatsi and Mark Pexton. Reassessing Woodward's Account of Explanation: Regularities, Counterfactuals, and Noncausal Explanations. *Philosophy of Science*, 80(5): 613–624, 2013.

Wesley Salmon. Scientific Explanation: Three Basic Conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Volume Two: Symposia and Invited Papers (1984):293–305, 1985.

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

Brad Weslake. Explanatory Depth. *Philosophy of Science*, 77:273—294, 2010.

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

James Woodward and Christoper Hitchcock. Explanatory generalizations, part I: A counterfactual account: A counterfactual account. *Noûs*, 37(1):1–24, 2003a.

James Woodward and Christoper Hitchcock. Explanatory generalizations, part II: Plumbing explanatory depth. *Noûs*, 37(2):181–199, 2003b.

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