Logique & Analyse **237** (2017), 47-66

A QUASI-INTERVENTIONIST THEORY OF MATHEMATICAL EXPLANATION

Victor Gijsbers

Abstract

Explanations in mathematics are not yet well understood. I discuss Steiner's theory of mathematical explanation, then attempt to improve it by assimilating its core intuitions to Woodward's counterfactual theory of explanation. The theory that results deals successfully with many cases, but it fails to handle certain types of explanatory asymmetry. To fix this, I draw on Woodward's interventionist theory of causation and develop a quasi-interventionist theory of mathematical explanation. According to this theory, the asymmetry of mathematical explanations is subjective in the sense that it does not depend on the objective structure of mathematics itself; but I argue that this is not a problem, since the same subjectivity can be found in causal explanation.

Keywords: mathematical explanation, causal explanation, counterfactual, quasi-interventionism, explanatory asymmetry

1. Introduction

Philosophers of mathematics have been interested in mathematical explanation for a long time (see the overview in Mancosu 2001), but the topic has often been ignored by theorists working on scientific explanation. Given that most theories of scientific explanation depend either on the concept of law of nature or on that of causation, neither of which has an obvious application in mathematics, this lack of interest is easy to understand. The conceptual gap suggests that if there is explanation in mathematics, it must be something very different from explanation in the physical sciences. Thus, in developing what is perhaps the most influential theory of mathematical explanation, Steiner (1978) does not attempt to reuse any of the central ideas of the already existing philosophy of scientific explanation, but comes up with the new fundamental notion of a 'characterising property'.

An important exception to this rule is Kitcher (1989), who argues that his unificationist theory of explanation captures both causal and mathematical explanation. However, Kitcher's unificationism faces severe technical problems (see Gijsbers 2007; for problems with its application to mathematics, see also Tappenden 2005, Hafner & Mancosu 2008). The unificationist

doi: 10.2143/LEA.237.0.3197358.

theory defended by Schurz (1999) overcomes some of the technical problems, but it is far from clear that it can be applied to mathematics, since it is based on a distinction between data and hypotheses that (again) has no obvious application in the *a priori* science of mathematics. Unificationism has thus not (yet) brought us the sought-after unification of mathematical and scientific explanation.

Kitcher's general strategy is nevertheless commendable. For it would surely be surprising if there are two things called explanation – scientific explanation and mathematical explanation – and they then turn out to be wholly distinct. The unity of all explanation is an excellent working hypothesis; and a good strategy for making sense of mathematical explanation is to attempt to generalise or adapt one of the existing theories of scientific explanation in such a way that it comes to encompass mathematics as well.

Among these theories, there are three reasons to single out Woodward's (2003) theory, which consists of a counterfactual account of causal explanation and an interventionist account of causation. First, Woodward's theory has been at the centre of philosophical attention during the past decade, and is often seen as the most promising research project in the philosophy of explanation. Second, there is – as will become clear later in this article and as was already hinted at by Woodward himself (2003, p. 220-221) – a suggestive similarity between Steiner's theory of mathematical explanation and Woodward's theory of causal explanation. The intuitions underlying these two theories turn out to be of the same ilk and therefore might be easy to assimilate to each other. Third, several other authors, including Rice (2015) and Reutlinger (2015), have already suggested ways of applying or adapting some of Woodward's ideas to forms of non-causal explanation, including, in Reutlinger's case, to mathematical explanation.

The paper will proceed as follows. I will start by presenting Steiner's theory in section 2, and will discuss why it is not by itself satisfactory. In section 3, using the work of Weber and Verhoeven (2002) and Woodward (2003), I will lay out a counterfactual theory of mathematical explanation. This theory, which is similar to the proposals of Reutlinger (2015) and Frans & Weber (2014), is a good step on the way to characterising mathematical explanation.

However, I then argue in section 4 that this counterfactual theory does not capture all aspects of mathematical explanation. In particular, it fails to capture certain forms of explanatory asymmetry. I show how we can remedy this by adapting Woodward's interventionist theory of causation in such a way that it becomes applicable to mathematics. The resulting quasi-interventionist theory implies that mathematical explanations are subjective in a sense that causal explanations do not seem to be. In section 5, I will argue that this does not invalidate the idea of mathematical explanation: the same kind of subjectivity, I contend, can also be found in certain unproblematic types of non-mathematical explanation.

2. Steiner's Theory of Mathematical Explanation

Steiner (1978) develops what is probably the most discussed theory of mathematical explanation. He recognises that while all proofs of a theorem establish its truth, some proofs also give insight into *why* the theorem holds, while others do not. The aim of his theory is to elucidate this distinction between explanatory and non-explanatory proofs. (Steiner does not claim that only proofs can explain, but he offers no account of other types of mathematical explanation.)

We will look at Steiner's main example. Let $S(n)$ be the sum of all natural numbers from 1 to n, that is, $S(n) = 1 + 2 + 3 + \ldots + n$. Then the following theorem holds for all n:

$$S(n) = \frac{n(n+1)}{2}.$$

We can give two different proofs of this theorem. The first is a proof using mathematical induction. We first notice that the theorem holds for $n = 1$, since $1 = 1 * (1 + 1) / 2$. We then show that if the theorem holds for $n = k$, it also holds for $n = k + 1$:

$$S(k+1) = S(k) + (k+1) = \frac{k(k+1)}{2} + (k+1) = \frac{k(k+1)}{2} + \frac{2(k+1)}{2} = \frac{(k+1)(k+2)}{2}.$$

By mathematical induction, the theorem thus holds for all n.

The second proof proceeds in a radically different way. By the commutativity of addition, we know that

$$1 + 2 + \ldots + (n - 1) + n$$

equals

$$n + (n - 1) + \ldots + 2 + 1.$$

So if we add these two sums, we get $S(n) + S(n) = 2\,S(n)$. Now, notice that the first term on the top and the first term on the bottom add up to $n + 1$, that the second terms on both sides also add up to $n + 1$, and that in fact all n pairs of terms add up to $n + 1$. This means that the total sum of the two sums is $n(n + 1)$, and the theorem follows immediately.

According to Steiner, the first proof is not explanatory, while the second is. He offers the following theory to explain the difference:

> My proposal is that an explanatory proof makes reference to a characterizing property of an entity or structure mentioned in the theorem, such that from the proof it is evident that the result depends on the property. It must be evident, that is, that if we substitute in the proof a different object of the same domain, the theorem collapses; more, we should be able to see as we vary the object how the theorem changes in response. (Steiner 1978, p. 143.)

Applied to our example, Steiner's theory yields the following judgements:

> [The explanatory proof] that the sum of the first n integers equals $n(n + 1)/2$ proceed[s] from characterizing properties: … by characterizing the symmetry properties of the sum $1 + 2 + \ldots + n$. … By varying the symmetry … we obtain new results conforming to our scheme. The proof by induction [on the other hand] does not characterize anything mentioned in the theorem. Induction, it is true, characterizes the *set* of all natural numbers; but this *set* is not mentioned in the theorem. (Steiner 1978, pp. 144-145.)

In this passage, Steiner nicely captures the intuitive reason that the second proof is explanatory: it makes us aware of a symmetry of $S(n)$, and of how this symmetry is in a sense "responsible" for the truth of the theorem. The competent reader will be able to see how the theorem will change "as we vary" this property: e.g., what would happen if we construct a series such that the terms form pairs that do not add up to $n + 1$ but to, say, $n + 2$ or to 17.

Steiner's reason to dismiss the inductive proof as non-explanatory is less convincing. His claim that this proof does not explain because the set of all natural numbers is not mentioned in the theorem cannot be accepted, since a more formal version of the proof would have started with "$\forall n \in \mathbb{N}$: …". In this presentation, then, the set of natural numbers *would* have been mentioned. But writing down the proof in a more formal way would not make it explanatory.

Luckily, Steiner's theory can account for the non-explanatory nature of the first proof without making this dubious claim. What we need to emphasise is the fact that the first proof does not show us how we can vary the result by varying a property of the sequence. The formula $n(n + 1) / 2$ is taken as a given; we are not shown how we ourselves could have reached it, if we had just been given the sequence and not the theorem to be proven. But this means that when we come to understand the first proof, we do not thereby learn how to handle similar cases.

Take the following sequence:

$$R(n) = 1 + 3 + 5 + \ldots + (2n - 1).$$

What is the general formula for $R(n)$? Grasping the first proof does not tell us. But grasping the second proof certainly does; for the method that underlies the second proof can be easily applied to this sequence as well. Make a pair-wise addition of the terms of $R(n)$ and its reverse, and we end up with n terms all equal to $2n$. Hence, $R(n) = n^2$.

If "the object" is the sequence of which we are determining the sum, then the second proof beautifully exemplifies Steiner's idea that we have an explanation when we are "able to see as we vary the object how the theorem changes in response". Or does it? For it seems that we need something more than just the proof. We need to grasp a general *recipe* of which this

proof is only an example; but the recipe itself is not part of the proof. We will return to this issue shortly.

Is Steiner's theory of mathematical explanation satisfactory? Not according to Resnik & Kushner (1987) and Hafner & Mancosu (2005), who propose counterexamples to the theory. A detailed discussion of these would take us beyond the limits of this paper.[1] And that discussion would be bound to be inconclusive anyway, since everything depends on what exactly we take Steiner to mean with a "characterizing property". Steiner defines it as

> a property unique to a given entity or structure within a *family* or domain of such entities or structures (p. 143),

but that definition is both vague – what counts as the relevant family or domain? – and also apparently too strict. For instance, the sequence of the first n natural numbers is not the *unique* sequence that has the symmetry property used in the second proof.

I propose that we drop all talk about characterising properties, and focus on the more successful part of Steiner's theory: the idea that an explanatory proof shows us how the theorem changes as we vary the object. We can explicate this further using the ideas of Weber & Verhoeven (2002) and Woodward (2003), and that is what we will do in the next section.

3. The Counterfactual Theory

As indicated in the previous section, a proof by itself does not have explanatory power. (Steiner himself seems to have recognised this, since he speaks of the necessity of a "proof-idea" (p. 147); but he doesn't explicate this

[1] But it may nevertheless be useful to say at least a little about the counterexample proposed by Hafner and Mancosu, since it might also serve as a counterexample to the theory of mathematical explanation that will be developed below. Hafner and Mancosu argue that the arbitrary sequences of positive numbers which feature in Kummer's convergence test have (because of their very arbitrariness) no characterising properties that could be varied. Therefore, Steiner cannot account for explanations of the validity of Kummer's convergence test.

Against this, I hold that it is not true that Kummer's sequences have no characterising properties. What is true is that no varying of any finite set of numbers in a sequence would make a difference to its usability in Kummer's convergence test. But there *is* a property of the sequence which *can* be varied and which *does* make a difference: the property of consisting (after a certain point) of only positive numbers. Sequences that have this property can be used to test convergence, but other sequences, like 1, –1, 1 –1, …, cannot. So the answer to "Why can we use 1, 2, 3, 4, … in Kummer's convergence test?" will mention that this sequence belongs to the class of sequences that, after some arbitrary number of elements, consist of only positive numbers. Steiner's theory can capture this, and so can the theory to be developed below.

If, in some context, the question were to arise why *all* (now *completely* arbitrary) sequences share a certain property, then Steiner and I would have to admit that the only possible explanations of this would lie in the difference between sequences (as such) and some class of non-sequences. But this too seems reasonable enough.

notion.) This has also been noticed by Weber and Verhoeven (2002), who argue that "proofs can become parts of answers, but they don't answer the [explanatory] question [of why the theorem is true] in their own right" (p. 300). After all, a proof by itself doesn't show how varying the mathematical object that the proof is about will change the theorem. For that, we need a *contrast*; we need not only the current proof, but also something else that shows what happens when the object is varied.

Weber and Verhoeven propose an improved version of Steiner's theory. The explanandum consists of the two contrasting claims that (a) mathematical objects of class X have property Q, and (b) mathematical objects of class Y have an incompatible property Q'. The explanation itself then consists of a proof of (a) and a proof of (b). Furthermore, the two proofs should 'run parallel' in the following way: they use the same axioms, premises and logical rules; and one of the proofs should use a defining property of X while the other uses a defining property of Y. (The term "defining property" seems to inherit some of the vagueness of Steiner's "characterizing property," but I take it that these properties should at least distinguish X's from Y's.) That the proofs are parallel in this way assures that the explanatory proof can be "deformed … into a proof of a related theorem" (p. 303), or, to remain closer to Steiner's words, that we can see how the theorem changes as we vary the object from an X to a Y.

One might object to Weber and Verhoeven that in Steiner's example of an explanatory proof, only a single proof was given, so that this example shows that one does not need two proofs. But any competent reader was able to come up with the contrastive proof herself; and the single proof would not have been explanatory if the reader had not had this competence. In other words, the contrasting proof will often be left implicit, but that doesn't mean that it is not an essential part of the explanation.

Weber and Verhoeven (perhaps for reasons of space) limit their theory to two proofs that contrast with each other. This is too narrow a conception of mathematical explanation. We saw in Steiner's example that grasping his explanatory proof allows us to set up a limitless number of new theorems. Rather than having a class X of sums whose outcome is $n(n + 1)/2$ and contrasting it with a class Y of sums with some other outcome, we gain insight into sums with all kinds of outcomes. When we grasp Steiner's example, we do not merely grasp two proofs; we rather grasp a *proof recipe* with an unlimited number of applications.

In Steiner's example the proof recipe is something like this:

> **Theorem.** Let O be a sequence of n terms with the property that the sum of term k and term $n + 1 - k$ is C for all n. Let S be the sum of all terms of O. Then $S = nC/2$.
>
> **Proof.** We add O term-to-term to its reverse O'. Each sum of two terms will be C, and there will be n such sums. Therefore, the total sum of O and O' is nC.

> Since the total sum of O' is equal to the total sum of O, it follows immediately that $S = nC/2$.

By "grasping" the recipe I mean not only that we come to know its validity, but also that we gain the ability to apply it to simple cases. If someone were to claim that he had understood Steiner's explanatory proof recipe, but then turned out to have no clue how to find the sum of the first n odd numbers $(1 + 3 + 5 + \ldots + (2n-1))$, we would come to the conclusion that he had in fact not understood the recipe.

Our generalised version of Steiner's theory, then, claims that to explain a mathematical theorem, we need to give a proof recipe that allows us to prove the theorem, and that also allows us to see how the theorem would change if the input of the proof recipe is changed. To those who have studied theories of non-mathematical explanation, this will sounds much like a *counterfactual* theory, such as the theory defended by Woodward (2003). Woodward writes:

> "[E]xplanation is a matter of exhibiting systematic patterns of counterfactual dependence. [Explanations include generalisations that] locate their explananda within a space of alternative possibilities and show us how which of these alternatives is realized systematically depends on the conditions cited in the explanans. They do this by enabling us to see how, if these initial conditions had been different or had changed in various ways various of these alternatives would have been realized instead." (2003, p. 191)

Although some of the terms used by Woodward – such as "initial conditions" – are not directly applicable to mathematics, Woodward's underlying intuition about explanation is identical to Steiner's: to explain is to show how the explanandum would have changed if the input of the explanation had been different. We may therefore hope to use the technical details of Woodward's counterfactual theory to flesh out our theory of mathematical explanation. To do so, let us start by looking at the first half of Woodward's definition of explanation:

> Suppose that M is an explanandum consisting in the statement that some variable Y takes the particular value y. Then an explanans E for M will consist of (a) a generalization G relating changes in the value(s) of a variable X (where X may itself be a vector or n-tuple of variables X_i) and changes in Y, and (b) a statement (of initial or boundary conditions) that the variable X takes the particular value x. (Woodward 2003, p. 203)[2]

Crucially, for Woodward an explanandum is not a single fact taken in complete isolation, but a fact of the form $Y = y$, where Y is a variable that can

[2] Weber and Verhoeven use the letters X and Y to denote two classes of mathematical objects. Woodward uses X and Y to denote variables. From this point onwards, we will adopt Woodward's usage.

also take values different from y within a certain specified range. The explanandum thus comes with a dimension of variation, and the explanans has to invoke a generalisation G that relates variation in another variable (or set of variables) X to that variation in Y. G shows us how the value of Y changes as we vary the value of X.

Thus, causal explanations and mathematical explanations both use generalisations: causal generalisations in the first case, proof recipes in the second. And while these may seem to be rather different things – causal generalisations are claims about the world, proof recipes are sets of instructions – their role in explanations is in fact exactly similar. What they do is link two properties to each other: the speed of the ball to the breaking of the window; the symmetry property of the series to its sum. Both tell us how to deduce the value of one variable from the value of another variable.

We nevertheless do need to make some changes to Woodward's theory if we want to adapt it to mathematical explanation. The most crucial difference between physical objects and mathematical objects may be that physical objects have many contingent properties, whereas mathematical objects have none. If a ball hit a window and broke it, it makes sense to think about what would have happened to the breaking if we had varied the speed of the ball. But it makes no sense to wonder what would happen to the sum of the series $S = 1 + 2 + 3 + \ldots + n$ if we had varied its symmetry property; for the series cannot have any other symmetry properties than those that it has and the sum cannot be different than it is. In Woodward's causal formalism, it makes sense to think of a variation of the property of an object. But for mathematical explanation, this doesn't make sense. Instead, we need to talk – as Steiner does – about varying the object itself. We cannot change the symmetry property of S, but we can choose a different series S'.

In a mathematical explanation, then, the explanandum is that for object o, variable Y takes on value y; that is, $Y(o) = y$. To explain this, we first identify a domain of objects, O, on which variable Y and a variable (or set of variables) X take a value for every object. We then give a proof recipe G that allows us to prove, given that an object is in O and that X takes a certain value for that object, that Y takes a certain value for that object. Finally, we perform the proof for object o, showing that $X(o) = x$, and that $X(o) = x$ implies $Y(o) = y$.

In Steiner's example, o is the sum $1 + 2 + 3 + \ldots + n$. Y is the outcome of a sum, and $y = n(n + 1)/2$. The domain O is the set of all sums with the symmetry property that pair-wise addition of the sum to its reverse gives the same result for every pair. X is the value of this result, with x being $n + 1$. The proof recipe G is the proof recipe given above, which shows that if $X = x$, $Y = nx/2$. The explanation is completed by showing that $X(o) = n + 1$ and then using G to prove that $Y(o) = n(n + 1)/2$.

The main structure of Woodward's counterfactual theory thus remains intact; the mathematical case is just slightly more complicated because we need to mention both a variation of the object and a variation of the explaining properties of that object. Let us now look at the second half of Woodward's definition of explanation, which specifies several further conditions:

> A necessary and sufficient condition for E to be (minimally) explanatory with respect to M is that (i) E and M be true or approximately so; (ii) according to G, Y takes the value y under an intervention in which X takes the value x; (iii) there is some intervention that changes the value of X from x to x' where $x \neq x'$, with G correctly describing the value y' that Y would assume under this intervention, where $y' \neq y$. (Woodward 2003, p. 203)

Here (i) guarantees that the explanans is (at least approximately) true, which Woodward requires because he believes we cannot explain the world with false theories. (There have been some challenges to this assumption recently, for instance by De Regt (2015), but for reasons that probably do not extend to the realm of mathematics.) Woodward's requirement (ii) guarantees that generalisation G actually links up the actual initial condition $X(o) = x$ with the explanandum $Y(o) = y$; while his (iii) guarantees that the value of X is relevant to the value of Y, since intervening on X can change Y.

How do these conditions relate to mathematics? We need to change condition (i) in two ways. First, we require strict truth, not approximate truth – a notion that is out of place in mathematics. Second, since a proof recipe is not true or false, we do not require that G is true, but instead that all applications of G are logically valid.

Requirements (ii) and (iii), while important, have to be reworded in such a way that the term 'intervention' drops out, since Woodward's causal notion of intervention cannot be applied to mathematics. Requirement (ii) then becomes "for all p in O such that $X(p) = x$, G generates a proof that $Y(p) = y$". We need this requirement to ensure that G allows us to prove what we need to prove.

Stripped of the notion of intervention, requirement (iii) becomes "there is at least one o' in O such that $X(o') = x'$ where $x \neq x'$, and G generates a valid proof that $Y(o') = y'$, where $y' \neq y$." To illustrate the need for (iii), let us look at an example. Suppose that we try to explain why 17 is divisible by 1. Here the object o is the number 17, and Y is a boolean variable with the possible values "is divisible by 1" and "is not divisible by 1". Now let O be the domain of all natural numbers, and X be a boolean variable taking the values "is a prime" and "is not a prime". We can then set up a proof recipe which allows us to deduce that if an object is a prime, it is divisible by 1; and that if it is not a prime, it is (also) divisible by 1. But of course, being a prime doesn't explain why 17 is divisible by 1. Varying X doesn't change Y, that is, varying whether a number is prime doesn't change whether it is divisible by 1. Requirement (iii) excludes such cases, and

ensures that *X* is relevant to *Y*, that it is a difference maker for *Y*, and that therefore the actual value of *X* can explain the actual value of *Y*.

With these simple changes, Woodward's counterfactual theory of causal explanation can be adapted to mathematical explanation. It incorporates the intuitions of Steiner's theory as well as the further explications of Weber and Verhoeven; while adding the technical precision and detail needed to exclude certain types of non-explanations.

In the next section, I will argue that the one major respect in which we did not follow Woodward's lead – namely, by removing the notion of intervention from the theory of explanation – leads to problems for our counterfactual theory, problems which we need to solve by reintroducing something akin to the notion of intervention. But before we go there, it will be useful to compare our counterfactual theory with two other recent attempts at formulating a theory of mathematical explanation that explicitly invoke Woodward: Reutlinger (2015) and Frans & Weber (2014).

Reutlinger (2015) defends the claim that a counterfactual theory of explanation can encompass both causal and non-causal explanation; but more specifically he argues that a counterfactual theory can deal with two mathematical explanations – Euler's explanation of the Königsberg bridge problem and renormalisation group theory explanations of universality. For this purpose, Reutlinger reformulates Woodward's counterfactual theory while dropping the notion of intervention, just like we have done. While he gives slightly less technical detail than I have given above, it is fair to say that any differences between Reutlinger's theory and the counterfactual account we have just developed are bound to be small.

Frans & Weber (2014) develop a theory of mathematical explanation that takes its cue not from Woodward's counterfactual theory, but from the theories of mechanistic explanation descended from Machamer, Darden & Craver (2000). Frans and Weber define the basic ideas as follows: "A *mechanistic explanation* of a capacity is a description of the underlying mechanism. […] A *mechanism* is a collection of entities and activities that are organized such that they realize the capacity" (p. 6). They then proceed to apply this idea to mathematics by identifying capacities with mathematical dependencies and activities with difference-makers; the result being that one explains a mathematical dependency by showing that the difference-makers are organised in such a way that the theorem stating the dependency is established.

At this point, Frans and Weber appeal to Woodward's notion of an intervention, a notion which they, in distinction to the counterfactual account developed in the current section, include in their theory of mathematical explanation. They identify intervention with the imaginary manipulation of a mathematical entity, writing:

> The proof also identifies which entities and properties are relevant in order to explain why the theorem holds, since imaginary manipulation of an entity

> shows how the system, of which a dependency between an input and output is described in the theorem, changes in response. An entity and a property are relevant if changing the property would change the outcome of the proof. (p. 12).

The theory Frans and Weber thus arrive at has obvious similarities with the counterfactual theory proposed here. Both claim that a mathematical explanation explains by allowing us to see how the explanandum would be different if some properties of the mathematical objects were different. Frans and Weber believe it is important that they take the bottom-up approach of starting with entities that can be manipulated, rather than the top-down approach of starting with generalisations; but whether that is a difference that makes a difference – and thus, more generally, whether there is a deep difference between mechanisms and systems of linked regularities – is something we cannot go into here.

Frans and Weber's use of Woodward's terms of manipulation and intervention in the context of mathematics is interesting, but they say little about how exactly we have to think of these concepts. Suppose, for instance, that I have an equilateral triangle with sides 1. Presumably, we can explain the fact that it has area $\sqrt{3}/4$ by pointing to the angles and the length of the sides. But what happens to the area of the triangle when I manipulate one of the sides such that its length becomes 2? There doesn't seem to be an obvious answer to that, since it depends on what other things we keep fixed – the angles, the length of the other sides, or even the area itself. Before we can use Woodward's idea of an intervention in mathematics, we would need to say more about how to understand such scenarios – and I hope to provide that something more in later parts of this paper. In this way, the current article could be read as a friendly extension of the account of Frans and Weber. (Though, because of the previous point, Frans and Weber would probably not read it that way).

4. Asymmetry and the Quasi-Interventionist Theory

Let us return to the counterfactual theory without interventions, as presented in the previous section. Is anything wrong with it? More specifically, did we sacrifice anything of importance when we dropped interventions from Woodward's account?

In Woodward's theory of causal explanation, interventions are needed to deal with certain explanatory asymmetries. Why does the length of the flagpole explain the length of the shadow, but not the other way around? Because intervening on the pole will change the length of shadow, while intervening on the shadow will leave the pole unchanged. Introducing interventions into our theory of explanation allows us to handle this kind of case.

Do such asymmetries exist in mathematical explanation? Reutlinger (2015) writes:

> Finally, having the notorious flagpole-shadow scenario in mind […], one may wonder whether [this theory] accounts for the explanatory asymmetry in the case of non-causal explanations. This is one of the deepest puzzles of the current philosophy of explanation […] and it is an open research question as to how one can capture the explanatory asymmetry in the non-causal cases. It is even possible that non-causal explanations do not generally display such an asymmetry. (p. 7)

It is my aim in this section to argue that mathematical explanations do indeed display explanatory asymmetries; to show that the counterfactual theory outlined above fails to do justice to these asymmetries; and to indicate how we can deal with them using mathematical quasi-interventions that are analogous to Woodward's interventions.

Consider the question why $\mathbb{N}^2$ is countable. We can explain this with a proof recipe that shows, using Cantor's pairing function, that if a set has the property of being countable, then its Cartesian product with itself is also countable. More formally:

> Let O be a sufficiently rich class of sets (in particular, we will assume that O contains both $\mathbb{N}$ and $\mathbb{R}$). Let X be the boolean variable that tracks whether an object p in O is countable; and let Y be the boolean variable that tracks whether p^2 is countable. Let there also be a proof recipe G that allows us to prove, using Cantor's pairing function, for any countable set that its Cartesian product with itself is countable; and that allows us to prove for any uncountable set that its Cartesian product with itself is uncountable. (We here omit the details of the proofs.)
>
> Then we can explain why $\mathbb{N}^2$ is countable by pointing out that $\mathbb{N}$ is countable, and using G to prove that $\mathbb{N}^2$ is countable as well.

This is a completely unobjectionable mathematical explanation, and probably the one given by a mathematician when she is asked to explain the countability of $\mathbb{N}^2$. But the following is also a good explanation according to the counterfactual theory outlined above:

> Let O be a sufficiently rich class of sets of the form Q^2 (in particular, we will assume that O contains both $\mathbb{N}^2$ and $\mathbb{R}^2$). Let X be the boolean variable that tracks whether an object p in O is countable; and let Y be the boolean variable that tracks whether another set s, with $s^2 = p$, is countable. Let there also be a proof recipe G that allows us to prove that if X is true, Y is true; and that allows us to prove that if X is false, Y is false. (We here omit the details of the proofs.)
>
> Then we can explain why $\mathbb{N}$ is countable by pointing out that $\mathbb{N}^2$ is countable, and using G to prove that $\mathbb{N}$ is countable as well.

But no mathematician asked to explain why $\mathbb{N}$ is countable would give us this story. It doesn't make sense to explain the countability of $\mathbb{N}$ from the countability of $\mathbb{N}^2$. To explain why there are countably many natural

numbers, we instead need to point out that by definition any subset of $\mathbb{N}$ is countable, and that $\mathbb{N}$ is of course a subset of itself. $\mathbb{N}^2$ and its properties are irrelevant.

Why is $\mathbb{N}^2$ countable? Because $\mathbb{N}$ is countable, and because $\mathbb{N}^2$ can be brought in one-to-one correspondence with $\mathbb{N}$, and because any set that can be brought in one-to-one correspondence with a countable set is countable. Reversing that order, and pointing out that $\mathbb{N}$ must be countable because it can be brought in one-to-one correspondence with the countable set $\mathbb{N}^2$, gets something wrong – even though the argument is logically valid.

We here have an example of asymmetry that seems to be akin to the flagpole and the shadow. Properties of the flagpole can be invoked to explain properties of the shadow, but not the other away around, even though there are valid arguments in both directions. That same relation holds between $\mathbb{N}$ and $\mathbb{N}^2$. How can we capture this in our theory of mathematical explanation? In the flagpole and shadow example, we introduce a *causal* relation that can be recognised by considering the effects of *physical interventions*. My suggestion is that we attempt to follow Woodward as closely as possible; and that in our example, we introduce a *quasi-causal* relation that can be recognised by considering the effects of *mathematical quasi-interventions.* These quasi-interventions reveal asymmetries inherent not in the mathematical proofs, but in our mathematical practice. It is the choices that we make when we are doing mathematics that generate explanatory asymmetries, especially where choices about definition are concerned.

As an example, let us look at the definition of countability. The standard definition is that a countable set is any set with the same cardinality as some subset of $\mathbb{N}$. However, one can also define a countable set as any set with the same cardinality as some subset of $\mathbb{N}^2$. These definitions are equivalent, so the mathematician – or the community of mathematicians – can freely choose one or the other.

But this choice immediately generates an explanatory asymmetry. If the first definition is chosen, then $\mathbb{N}$'s countability is trivial; it is countable by definition. The countability of $\mathbb{N}^2$, on the other hand, is far from trivial, and is to be explained by showing that $\mathbb{N}^2$ has the same cardinality as $\mathbb{N}$. On the other hand, if we had chosen the second definition of countability, then the countability of $\mathbb{N}^2$ would have been trivial; while the countability of $\mathbb{N}$ would then have been explained by showing that we can bring $\mathbb{N}$ in one-to-one correspondence with $\mathbb{N}^2$ or one of its proper subsets.

Thus, the definitions we choose can be crucial to the order of explanation. In particular, definitions which grant a property P to a certain set of objects O directly and then recursively grant it to other objects indirectly if they bear relationship R to an object that has P, set up an explanatory asymmetry where:

- if the object o is in O, we explain $P(o)$ by pointing out that o is in O;

- if the object o is not in O, we explain $P(o)$ by showing that o is related by R (perhaps in several steps) to some object in O.

It is useful to compare this proposal to Marc Lange's (2009) argument that inductive proofs of $\forall n : P(n)$ are generally not explanatory. Lange argues that in inductive proofs, we can almost always construct both an 'upward' and a 'downward' proof, that is, both a proof that shows that $P(i) \rightarrow P(i + 1)$ and a proof that shows that $P(i) \rightarrow P(i–1)$ (for $i > 0$). He then shows that this allows us to take any natural number as the starting point for our inductive 'explanations'. For instance, $P(0)$ can be used to explain $P(5)$, but $P(5)$ can also be used to explain $P(0)$. Since Lange believes that explanation has to be asymmetric, and so no two facts can appear in each other's explanans, it follows that inductive proofs cannot count as real explanation: they fail the test of asymmetry.

According to our proposal, however, at least *some* inductive explanations, namely, explanations of properties whose definition is recursive, will have the asymmetry Lange requires. To take a simple example, we might wonder why in Peano arithmetic zero is a natural number; and also why S(S(S(S(0)))) – that is, four – is a natural number. Now in this case, there is a clear asymmetry. Zero's status as a natural number must be explained by pointing to the axiom which states that zero is a natural number. Four's status as a natural number must be explained through a recursive argument starting from zero and using the axiom that the successor of a natural number is a natural number. The fact that zero is a natural number is thus part of the explanation of the fact that four is a natural number, but not the other way around. Of course, with different axioms this asymmetry might flip; yet that doesn't show that there is no inductive explanation, but only that mathematical explanations have to be evaluated against the background of the definitions used by the mathematician.

In general, if a mathematical object or property A is part of the definition of mathematical object or property B, this generates an explanatory asymmetry from A to B: facts about B can be explained by facts about A, but not the other way around. Thus, if we define $\mathbb{N}$ in terms of the Peano axioms, we can then explain properties of $\mathbb{N}$ in terms of properties of the axioms, but not the reverse. If we define $\mathbb{N}^2$ as the Cartesian product of $\mathbb{N}$ with itself, this means that properties of $\mathbb{N}^2$ are to be explained in terms of properties of $\mathbb{N}$, and not the other way around (unless the property itself is given to $\mathbb{N}^2$ by definition).

Now as soon as we have any such asymmetry in mind, we can start looking at mathematical objects or properties in a way that is analogous to the way that the Woodwardian interventionist looks at objects or properties that are causally related. Thus, when the Woodwardian thinks of this asymmetry

height of flagpole → length of shadow

she interprets this as indicating that an intervention on the height of the flagpole changes the length of the shadow, but not the other way around. Analogously, when faced with this asymmetry

$$\text{Peano axioms} \rightarrow \mathbb{N}$$

we can interpret it as indicating that an 'intervention' on the axioms will 'change the properties' of $\mathbb{N}$. Of course, we cannot really intervene on the axioms, nor can the properties of $\mathbb{N}$ really change. But we can keep the connection between the axioms and the set constant (it is, in this case, the connection of "being the set that satisfies the axioms") and consider what set we end up with when slightly different axioms are chosen. What we are doing, then, is we keep the asymmetrical links laid down by our definitions fixed, then vary one of our mathematical objects or properties, and observe what must change downstream if the definitions remain fixed. Let us call this process mathematical *quasi-intervention*; it is the mathematical analogue of causal intervention. We call it a 'quasi' intervention because it is not, of course, an intervention in Woodward's causal sense.

Introducing the concept of a quasi-intervention allows us to accommodate the asymmetry of mathematical explanation in a way analogous to how Woodward accommodates the asymmetry of causal explanation. Remember that the two specifically interventionist requirements in his theory were the following:

> (ii) according to G, Y takes the value y under an intervention in which X takes the value x; (iii) there is some intervention that changes the value of X from x to x' where $x \neq x'$, with G correctly describing the value y' that Y would assume under this intervention, where $y' \neq y$. (Woodward 2003, p. 203)

In the counterfactual theory of section 3, we reformulated these requirements so that they no longer spoke of interventions, as follows:

> (ii) for all p in O such that $X(p) = x$, G generates a proof that $Y(p) = y$; (iii) there is at least one o' in O such that $X(o') = x'$ where $x \neq x'$, and G generates a valid proof that $Y(o') = y'$, where $y' \neq y$.

But if we want to do justice to the asymmetries of mathematical explanation, this does not quite work, as we have seen. These two requirements are too permissive; we need to make sure that mathematical quasi-interventions on $X(p)$ can lead to changes in $Y(p)$. Condition (iii) captures this idea in so far that it identifies a change of object which changes the value of X and which allows us to prove that the value of Y is then also changed; but by appealing to proof rather than quasi-intervention, it fails to ensure that the $X \rightarrow Y$ direction follows the direction required by the explanatory asymmetries of the definitions. The easiest way to fix this is to simply strengthen the requirement:

(iii) there is at least one o' in O such that (a) $X(o') = x'$ where $x \neq x'$, and (b) G generates a valid proof that $Y(o') = y'$, where $y' \neq y$, and (c) the quasi-intervention of changing x to x' leads to the change of y to y'.

We thus end up with a quasi-interventionist theory of mathematical explanation that is as close as possible – given the inherent differences between causal and mathematical objects – to Woodward's interventionism. I submit that this theory leads to the correct results for the examples we have been discussing. Of course, whether it also works for other examples of mathematical explanation needs to be investigated further; and this is especially needful in the light of Hafner & Mancosu's (2005) argument that mathematical explanation is a highly varied phenomenon, which may lead us to worry that the examples discussed in this case – all of them rather simple cases from number theory – might not represent that full variety.

However, instead of embarking on that project now, I want to raise and dispel a worry of a different kind: the worry that the quasi-interventionist theory turns mathematical explanation into something that is too subjective to really be about mathematics.

5. Explanatory Asymmetry and Subjectivity

The quasi-interventionist theory introduces the subject or the community into assessments of mathematical explanation. Quasi-interventions cannot be defined purely in terms of the mathematical objects themselves; we also need the definitions chosen by the mathematician. This suggests that the asymmetry of mathematical explanation does not reflect an objectively real asymmetry in the subject matter of the explanation.

In this, mathematical explanations would seem to differ sharply from causal explanations. Causal asymmetries are objectively real: there are causal laws that ensure that we really cannot manipulate the shadow by intervening on the flagpole. (Van Fraassen 1980 famously disagreed, but few have followed him). But the asymmetries of mathematical explanation only exist when we first choose certain definitions and then keep them fixed, *as if* they were unbreakable laws. But this is only an *as if*. Nothing stops us from choosing other definitions.

Why would this be a problem? On the one hand, mathematical explanations are mirroring the asymmetry of a practice, not that of the mathematical objects themselves. But on the other hand, they purport to explain facts about the objects, *not* about the practice: we explain the countability of $\mathbb{N}^2$, and not some fact about ourselves as mathematicians. There seems to be a tension here. If what we want to understand are the mathematical objects, how can considerations about ourselves as mathematicians influence which explanations are acceptable? Do such explanations really explain mathematical facts?

Within the context of the quasi-interventionist theory, there are two ways to deal with these questions; one of them, which I prefer, leaves what has been said up till now intact and argues that causal explanations too can have asymmetries which are practice-dependent and that this is therefore a feature of explanation in general; while the other argues that, contrary to what I have been claiming, the asymmetries of quasi-intervention are in fact objectively real facts about mathematics. This second way can claim support from a recent trend in the philosophy of mathematics – discussions of grounding – and is thus worth a brief look before I develop my own argument.

It has been essential for my claim that asymmetries are practice-dependent that there is no objective way to stipulate which mathematical objects or properties or facts are prior to others; that this is a matter of choice for the practising mathematician. However, several authors have recently attempted to provide us with an objective explanatory order in mathematics by connecting it with Bolzano's concept of grounding (Tatzel 2002; Betti 2010; Rumberg 2013). Can this literature be invoked to defend the objectivity of the asymmetries in mathematical explanation? I don't believe so. Some explanatory asymmetries are clearly not objective, and no account of grounding should claim them to be so, as I want to illustrate now with a simple example.

Consider the following case. We start with the sequence of positive numbers S: {1, 5, 28, 3, 2, 47, 3}. From this, we define a new sequence, T, where T_i is the sum of the first i terms of S. Thus, T is {1, 6, 34, 37, 39, 86, 89}. Why is the sequence T positive and strictly increasing? Because S consists of only positive numbers (for brevity's sake, we leave much of the explanation implicit). That is a fine explanation. And the opposite explanation fails: it is not the case that S consists of only positive numbers *because* T is positive and strictly increasing.

Now consider the following case. We start with a sequence of positive, strictly increasing numbers T: {1, 6, 34, 37, 39, 86, 89}. From this, we define a new sequence S such that $S_1 = T_1$ and for all $i > 1$, $S_i = T_i - T_{i-1}$. This means that S is {1, 5, 28, 3, 2, 47, 3}. Why does S contain only positive numbers? Because T is positive and strictly increasing. And again, the explanation the other way around does not work.

But the two cases are of course identical, except that the order of definition was different. So if my assessment of what explains what is correct, it turns out that the explanatory order is dependent on entirely contingent and subjective choices in how we define our mathematical objects. If we get T from S, then the properties of S explain those of T. If we get S from T, then the explanations go the other way around.

I take it that we *do* have intuitions about explanatory asymmetry in these cases, and that it is also obvious that there is no objective way to prefer one of the definitions over the other. If that is so, then at least some of the

asymmetries in mathematical explanation are wholly dependent on subjective choices. So even if the grounding strategy would work in some cases, we'd still need another strategy for dealing with the remaining ones.

The other way of dealing with the tension between the objective and the subjective aspect of mathematical explanation is to notice that some causal explanations also exhibit an asymmetry that depends on our practices. Take, for instance, the ideal gas law, $PV=NkT$. Given the values of three of the four variables in the law, we can deduce the value of the fourth. What is more, in most systems we could intervene on any of the four variables, keep two others fixed, and see the changes in the system. E.g., given a container of gas, we can heat it (changing T), crush it (changing V), pump more gas into it (changing N), or open it and let it come into equilibrium with a (much larger) container with a different pressure (changing P), all while keeping one or two of the other variables fixed. So this is not a case like the flagpole and shadow, where we just cannot change the flagpole by intervening on the shadow. There are no objective causal asymmetries here.

And yet, in certain circumstances, there are explanatory asymmetries. Take a pressure cooker.[3] It makes sense to explain the pressure in the pressure cooker by pointing out how it would change under interventions on the temperature when the volume and the number of particles are fixed; that is what we would do when we were asked to explain the basic workings of a pressure cooker. But would it make sense to explain the volume of the pressure cooker by pointing out how it would change under interventions on any of the other three variables? No. Instead we would explain the volume by talking about the manufacturing process in the factory, or the intentions of the manufacturers. Why? Perhaps the reason is that interventions that allow the volume of the cooker to change while two of the other variables are held fixed, are rather exotic and far from everyday life; they would, for instance, involve weakening an area of the pan in such a way that it could bulge out when under sufficient pressure. But note that the exoticness of an intervention can only be assessed against the background of our practices. It is not an objective fact about the physics of pressure cookers.

We would also never explain the pressure in the pressure cooker by pointing out how the pressure would change under interventions on the volume. This is an especially interesting case because we don't have to depart from everyday life. Indeed, nothing is easier than increasing the pressure by decreasing the volume: all one needs is a firm blow with a hammer to the side of the cooker while it is cooking. But because deforming pressure cookers with hammers is not one of our usual practices, while heating pressure cookers on a stove is, we consider it explanatory to explain

[3] I am indebted to an anonymous referee for this example.

the pressure from interventions on the temperature but non-explanatory to 'explain' the pressure from interventions on the volume.

There is of course more to be said about the role that practices play in causal explanatory asymmetries. For our current purposes, however, it is enough to notice that practice-dependence is not a special feature of mathematical explanations. It exists even in the paradigmatic realm of causal explanations. This means that it can hardly threaten the cogency of our theory of mathematical explanation.

6. Conclusion

I have argued that mathematical explanations should be understood as quasi-interventionist explanations. By adapting Woodward's counterfactual theory of causal explanations and his interventionist criteria for causation, we were able to not only improve on the accounts of Steiner and Weber & Verhoeven, but to also tackle the problem of explanatory asymmetry in mathematics, which until now had been mostly neglected. The resulting theory is, I hope, a serious contender for the best theory of mathematical explanation currently available.

Several questions and areas of research remain, however. First, it is necessary to test the quasi-interventionist theory against a wider array of examples of mathematical explanation. Second, it is necessary to assess in more detail whether theories of grounding can account for explanatory asymmetries in mathematics and, if so, for how many of them. And finally, the practice-dependence of explanatory asymmetries bears further investigation. We have our work cut out for us.

References

[1] Betti, A. (2010). Explanation in metaphysics and Bolzano's theory of ground and consequence. Logique et Analyse, 211, 281–316.

[2] Frans, J. & E. Weber (2014). Mechanistic Explanation and Explanatory Proofs in Mathematics. Philosophia Mathematica, advanced access March 9, 2014.

[3] Gijsbers, V. (2007). Why Unification is neither necessary nor sufficient for explanation. Philosophy of Science, 74, 481-500.

[4] Hafner, J. & P. Mancosu (2005). The Varieties of Mathematical Explanation. (In P. Mancosu, K. Jørgensen & S. Pedersen (Eds.), Visualization, Explanation and Reasoning Styles in Mathematics (pp. 215–250). Berlin: Springer).

[5] Hafner, J. & P. Mancosu (2008). Beyond Unification. in P. Mancosu (ed.), *The Philosophy of Mathematical Practice*, Oxford: Oxford University Press, 151–178.

[6] Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. (In P. Kitcher & W. Salmon (Eds.), Scientific Explanation, Minnesota

Studies in the Philosophy of Science, vol. XIII (pp. 410-505). Minneapolis: University of Minnesota Press).

[7] Lange, M. (2009). Why proofs by mathematical induction are generally not explanatory. Analysis, 69, 203-211.

[8] Machamer, P., L. Darden, and C. Craver (2000). Thinking about mechanisms, Philosophy of Science, 67, 1–25.

[9] Mancosu, P. (2001). Mathematical Explanation: Problems and Prospects. Topoi, 20, 97-117.

[10] De Regt, H. W. (2015). Scientific Understanding: Truth or Dare? Synthese, 192, 3781–97.

[11] Resnik, M. D. & D. Kushner (1987). Explanation, Independence and Realism in Mathematics. The British Journal for the Philosophy of Science, 38, 141-158.

[12] Reutlinger, A. (2015). Is There A Monist Theory of Causal and Non-Casual Explanations? The Counterfactual Theory of Scientific Explanation. Philosophy of Science, accepted December 10, 2015.

[13] Rice, C. (2015). Moving Beyond Causes: Optimality Models and Scientific Explanation. Noûs, 49, 589-615.

[14] Rumberg, A. (2013). Bolzano's Concept of Grounding (*Abfolge*) against the Background of Normal Proofs. The Review of Symbolic Logic, 6, 424-459.

[15] Schurz, G. (1999). Explanation as Unification. Synthese, 120, 95-114.

[16] Steiner, M. (1978). Mathematical Explanation. Philosophical Studies, 34, 135-151.

[17] Tappenden, J. (2005). Proof Style and Understanding in Mathematics I: Visualization, Unification and Axiom Choice. In P. Mancosu, K. Jørgensen and S. Pedersen (eds.), Visualization, Explanation and Reasoning Styles in Mathematics, Berlin: Springer, 147–214.

[18] Tatzel, A. (2002). Bolzano's Theory of Ground and Consequence. Notre Dame Journal of Formal Logic, 43, 1-25.

[19] Van Fraassen, B. (1980). The Scientific Image, Oxford: Oxford University Press.

[20] Weber, E. & L. Verhoeven (2002). Explanatory Proofs in Mathematics. Logique & Analyse, 179-180, 299-307.

[21] Woodward, J. (2003). Making Things Happen. (Oxford: Oxford University Press).

Victor Gijsbers
Leiden University
V.Gijsbers@phil.leidenuniv.nl