Challenges Facing Counterfactual Accounts of Explanation in Mathematics

Marc Lange, *Philosophia Mathematica* 2022

Please cite only the published version.

*1. Introduction.* The past two decades have witnessed an explosion of interest in counterfactual accounts of causal scientific explanation. Most influentially, Woodward (2003) has portrayed causal scientific explanations as operating by providing information about systematic patterns of counterfactual dependence – about how the explanandum would have been different, had facts in the explanans been different in various specific ways. More recently, philosophers have devoted increased attention to non-causal scientific explanations. Attracted by the prospect of giving a unified picture of scientific explanation, some philosophers have worked to extend the counterfactual account of causal explanation to non-causal scientific explanation. (This work includes Baron, Colyvan, and Ripley 2017; Bokulich 2008; French and Saatsi 2018; Jansson 2015; Kistler 2013; Pexton 2014; Reutlinger 2016; Rice 2015; Saatsi 2018; Saatsi and Pexton 2013; and Woodward 2018.)

Philosophers have recently also paid increased attention to explanation in mathematics – for instance, to the way that certain mathematical proofs not only prove theorems, but also thereby explain why those theorems hold. (Mancosu 2018 surveys the recent literature.) In light of the strong interest in counterfactual accounts of scientific explanation, it was inevitable that some philosophers would propose extending this approach to explanations in mathematics. Indeed, this has already begun (Frans and Weber 2014; Gijsbers 2017; Baron, Colyvan, and Ripley (2020); and Reutlinger, Colyvan, and Krzyżanowska (forthcoming).)

The purpose of the present paper is to identify some important challenges that any counterfactual account of explanation in mathematics must face – challenges that are not faced by counterfactual accounts of scientific explanation and that have been underappreciated in the literature.[[1]](#endnote-1) Although none of these challenges constitutes a knock-down argument against any possible counterfactual account of explanation in mathematics, these challenges pose significant obstacles to any such account. At the very least, any such account must in some way address them.

Any account that sees explanations in mathematics as answering what-if-things-had-been-different questions will have to regard as nontrivial some counterfactual conditionals with mathematically impossible antecedents (i.e., “countermathematicals”) such as “Had been an algebraic rather than a transcendental number, then…”, “Had 13 been divisible by 6, then…” and “Had there existed a planar map for which four colors do not suffice, then…”. Advocates of counterfactual accounts of explanation in mathematics have acknowledged that countermathematicals are vacuously true according to standard possible-worlds semantics (Stalnaker 1968, Lewis 1973) since there is no possible world in which their antecedents are true. Accordingly, these advocates have appealed to the burgeoning literature proposing that we modify the standard semantics so that some counterfactuals with impossible antecedents bear nontrivial truth-values in some contexts. (For references, see Baron and Colyvan 2016: 73.) Some advocates (e.g., Baron, Colyvan, and Ripley 2017) have sketched their own semantics of countermathematicals.

At least for argument’s sake, I will embrace this response. No challenge posed in this paper will depend on regarding all countermathematicals as trivial or on denying that countermathematicals have any place in mathematical practice. (Jenny 2018 has shown that some do.[[2]](#endnote-2)) I will argue that counterfactual accounts of explanation in mathematics are threatened not by the existence of too few nontrivially true countermathematicals, but rather by the existence of too many. The paper will argue that counterfactual dependence fails to track explanatory relations in mathematical practice.

In section 2, I will briefly elaborate the core idea behind counterfactual accounts of explanation and how its advocates have proposed extending that idea to explanation in mathematics. In the subsequent sections, I will turn to various challenges this effort faces. In section 3, I will argue that a given mathematical fact has too many explanations, according to the counterfactual approach. In section 4, I will argue that the existence of so many explanations conflicts with mathematical practice because pedestrian mathematical facts have no explanations there, whereas the counterfactual approach often deems them to have explanations. In section 5, I will argue that the existence of so many explanations on the counterfactual approach prevents its capturing the asymmetry of explanation in mathematics. In section 6, I will argue that the capacity to answer what-if-things-had-been-different questions does not correlate with explanatory power in mathematics: the reason that certain facts explain, and others do not, often has nothing to do with relations of counterfactual dependence. In fact, some features enabling certain proofs to reveal the explanandum to be counterfactually dependent on certain facts actually *deprive* those proofs of explanatory power. I will conclude in section 7.

*2. Counterfactual accounts.* As Woodward (2003: 11) says, the core idea motivating counterfactual accounts of explanation is that explanations tell us about systematic patterns of counterfactual dependence and so enable “us to see what sorts of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways.” Woodward (2003: 221) suggests that the idea that explanations “must answer what-if-things-had-been-different questions” could be “the common element in many forms of explanation, both causal and noncausal.” He even briefly contemplates explanations in mathematics fitting this picture, although “the dependence of a theorem on the assumptions from which it is derived is not any sort of causal dependence.”

Reutlinger (2016: 737; 2018: 78-9) develops this core idea in a simple and natural way that is sufficiently general to encompass explanations in mathematics:

* An explanans consists of generalizations, initial conditions, and auxiliary assumptions (all of which are true) that entail (or confer some probability on) the explanandum.
* The generalizations and auxiliary assumptions would still have held, had the initial conditions been different in some specific way, and they thereby underwrite at least one counterfactual conditional specifying how the explanandum (or its probability) would then have been different.
* No proper subset of the explanans is capable of satisfying the above conditions.

Reutlinger, Colyvan, and Krzyżanowska (forthcoming) take these purportedly necessary and sufficient conditions for explanation and apply them to explanation in mathematics. That conditions along roughly these lines carve out explanations in mathematics likewise seems to be intended by the other advocates of counterfactual accounts of explanation in mathematics. Both Frans and Weber (2014) and Baron, Colyvan, and Ripely (2020) take the difference between explanatory and non-explanatory proofs in mathematics to lie in whether the proof identifies a pattern of counterfactual dependence revealing what factors, were they changed, would make a difference to the theorem being explained.

Consider a brief example. Reutlinger, Colyvan, and Krzyżanowska (forthcoming) maintain that the reason why one cannot “square the circle” – i.e., the reason that a square of the same area as a given circle cannot be constructed (in a finite number of steps, using only straightedge and compass) – is that is a transcendental number (i.e., is not the root of a polynomial equation with rational coefficients). That is transcendental is the “initial condition” in a proof that the construction cannot be carried out; the generalizations in the proof are that the area of any circle (with radius r) equals r2 (so that constructing a square of the same area as the circle requires constructing a line of length ) and various generalizations about what can be constructed (from which it follows by abstract algebra that all constructible lengths must be solutions of certain polynomial equations with rational coefficients). The proof supplies the resources to show that had not been transcendental (while meeting certain further conditions), then one could square the circle.

I will set aside many of the specific details of various proposed counterfactual accounts of explanations in mathematics (such as the arguments by Frans and Weber (2014: 242) and Gijsbers (2017: 51-2) that an explanatory proof is only part of a theorem’s explanation, since the explanation must go on to use that proof to address some what-if-things-had-been-different questions about the explanandum). These details will make no difference(!) for my purposes, since my concern is to pose challenges for any account along roughly these lines, whatever its details.[[3]](#endnote-3) Likewise, only the most general features of any proposed semantics of countermathematicals (e.g., “Had not been transcendental…, then one could square the circle”) will matter for my purposes. The general sort of semantics suggested by advocates of counterfactual accounts of explanation in mathematics is inspired by Woodward’s picture of the way that “interventions” are understood to hypothetically bring about the states of affairs (e.g., the match’s being struck) posited by the antecedents of the counterfactual conditionals associated with causal explanations. Roughly speaking, an intervention is a hypothetical change that cuts any causal influence exerted on the fact being “twiddled” by the counterfactual’s antecedent while retaining the twiddled fact’s capacity to causally influence other facts (e.g., the match’s being lit). To apply this picture to countermathematicals, anything causal or temporal (as the term “initial conditions” suggests) must be removed; the role of causal connections is played by connections of logical entailment. According to Baron, Colyvan, and Ripley (2017; 2020), the mathematics (such as ’s not being transcendental) posited by a countermathematical antecedent is hypothetically brought about by a “twiddle” that cuts any logical connection from other, “upstream” mathematical facts to the antecedent’s negation (a truth, such as ’s being transcendental) while retaining all (contextually relevant) logical connections from the posited mathematics to other, “downstream” matters. In this way, the mathematical facts “upstream” from the antecedent (as Baron, Colyvan, and Ripley put it) are held fixed under the countermathematical antecedent, whereas the mathematical facts “downstream” are allowed to vary; those mathematical facts can counterfactually depend on the antecedent.[[4]](#endnote-4) Baron, Colyvan, and Ripley do not endorse any particular analysis of what it constitutes this “stream” flowing in one direction rather than another; though they keep open the possibility of such an analysis, they “aim to leave the distinction between upstream and downstream facts” (the direction of “mathematical dependence”) “at an intuitive level”. Thus, countermathematicals are supposed to operate in much the same way as do ordinary counterfactuals (e.g., “Had I struck the match, it would have lit”) even though for ordinary counterfactuals, the “upstream” facts are in the past (or, at least, are causally prior to the facts that the antecedent “twiddles”) and the “downstream” facts are in the future (or, at least, causally posterior to the twiddled fact).

As I mentioned in the previous section, I will presuppose for argument’s sake that countermathematicals operate in a roughly “twiddle-and-ramify” manner. In particular, I will presuppose that contextual considerations prevent counterfactual changes to mathematics from ramifying indefinitely, thereby trivializing countermathematicals. I will now identify some challenges to our using countermathematical dependence to understand explanation in mathematics.

*3. Too many explanations.* A given mathematical fact has too many explanations on a counterfactual account; it is too easy to satisfy the necessary and sufficient conditions given above. For example, consider the fact that 123321 is divisible by 37. This fact seems to have a great many explanations, according to the counterfactual account. Here are just four:

* That 123321 is 37 more than 123284 (i.e., 123284 + 37 = 123321), that 123284 is divisible by 37, and that (for any natural number *a*) *a* + 37 is divisible by 37 iff *a* is divisible by 37.
* That 123321 is 37 less than 123358 (i.e., 123358 – 37 = 123321), that 123358 is divisible by 37, and that (for any natural number *a*) *a* – 37 is divisible by 37 iff *a* is divisible by 37.
* That 12332 – 11 = 12321, that 12321 is divisible by 37, and that (for any natural number *a*) *a* is divisible by 37 iff the number obtained by subtracting 11 times *a*’s unit digit from the number obtained by removing *a*’s unit digit (in base 10) is divisible by 37.
* That 444 is divisible by 37, that 123 + 321 = 444, and that (for any natural number *a*) *a* is divisible by 37 iff the number obtained by taking *a*’s digits (in base 10) in groups of 3, beginning from the right, and adding those numbers together, is divisible by 37.

More “explanations” could be given besides these.

All of these “explanations” satisfy the necessary and sufficient conditions above (presuming that countermathematical conditionals can be nontrivial and are governed by something like the semantics proposed by advocates of counterfactual accounts of explanation in mathematics). For instance (from the first example), had 123284 not been divisible by 37, then 123321 (being 37 more than 123284) would also not have been divisible by 37. (Under this countermathematical twiddle to the “initial condition” that 123284 is divisible by 37, the cited generalization (*a* + 37 is divisible by 37 iff *a* is divisible by 37) and “auxiliary assumption” (that 123321 is 37 more than 123284) are held fixed.) This countermathematical antecedent seems no more exotic than those entertained by advocates of counterfactual accounts of explanation in mathematics (such as Baron, Colyvan, and Ripley’s (2017: 6) “Had 13 been divisible by 2 and by 6…”).

Perhaps an advocate of a counterfactual account would embrace the conclusion that 123321’s divisibility by 37 can be explained in the above four ways (and more). After all, in many cases, a fact having one causal explanation has a great many such explanations: some of them cite its more proximate causes, others its more distant causes (in principle, all the way back to the Big Bang), and perhaps some cite its more macroscopic, coarse-grained causes and others its more microscopic, fine-grained causes. This plethora of causal explanations of (say) a match’s lighting is unproblematic. So why should a mathematical fact’s having so many explanations be problematic?

I have no knock-down reply to this response. One possible reply is that we already had a grip on causal relations independent of both explanatory and logical relations, so we already knew that oftentimes many causal chains (some longer and some shorter, some coarser-grained and some finer) lead to the same explanandum. Therefore, the multiplicity of causal explanations of the same explanandum does not concern us. By contrast, we do not have as secure of a prior grip on explanation in mathematics, so there is no independent confirmation that so many explanations exist of 123321’s divisibility by 37. Therefore, the counterfactual account’s endorsing so many explanations suggests that its sufficient conditions are simply too low.

Another reply to citing the multiplicity of causal explanations of the same fact is that this multiplicity is illusory: those various causal explanations have subtly different targets because they involve different contrast classes, reflect different interests, and have different presuppositions. If we want to know why the match lit at 7:54 pm on Thursday rather than remaining unlit, then that it was struck a moment before (considering the oxygen in the room…) sometimes explains why. Some information about earlier causal relations explains why the match existed in the first place at 7:54 pm on Thursday, but that the match existed then is presupposed by the question “Why did the match light at 7:54 pm on Thursday rather than remaining unlit?”; the match’s existence at that time is entailed by both members of the contrast class. Therefore, information about causes of the match’s existence does not help to explain why the match lit at 7:54pm on Thursday rather than remaining unlit. Likewise, a microlevel causal account of the match’s lighting might explain why it flamed up in a certain very specific way rather than differently, but the microaccount would not explain why the match lit at that time rather than remaining unlit. Furthermore, the question “Why was the match lit?” might be asked in a conversational context focusing on the behavior and psychology of the match-striker. In that case, an answer would have to contain information about the match-striker’s motives; the reply “The match lit because it was struck” would entirely miss the question’s point. Distinct causal explanations of the match’s lighting thus respond to different interests.

Likewise, in canonical philosophical examples of multiple scientific explanations of the same fact, different explanations respond to different interests or involve different kinds of explanation. In Salmon’s (1989: 180-82) “friendly physicist” case, for instance, the same fact purportedly receives both a “bottom-up”, “causal-mechanical” explanation and a “top-down”, “unificationist” explanation (cf. Lange 2014; 2017: 119-20). Similarly, in Hanson’s (1958: 34) famous example, a death in an automobile accident is explained differently by a physician, a barrister, and a traffic engineer considering their different interests.

However (to return to the mathematical case), nothing analogous occurs in the case of the various “explanations” of 123321’s divisibility by 37. Their plurality does not reflect different contrast classes, different interests, or different presuppositions. Explanation in mathematics seems to come very cheaply, on the counterfactual approach. (A mathematical fact may indeed have multiple explanations, but not as cheaply as the counterfactual approach allows.)

A third reply to the objection “What harm is there in a mathematical fact’s having multiple explanations when a natural fact may have multiple causal explanations?” is that plenty of harm results, as the next two challenges illustrate.

*4. Explanations of pedestrian mathematical facts.* As we have just seen, there are many ways to demonstrate 123321’s divisibility by 37: by appealing to 123284’s divisibility, or to 123358’s, or to 12321’s, or to 444’s. I argued that for each of these numbers, its divisibility by 37 qualifies on the counterfactual account as figuring in an explanation of 123321’s divisibility by 37. This seems like too many explanations. So (you might reasonably ask) which of them does in fact explain 123321’s divisibility by 37? The answer is: none of them. In fact, there is no answer to the question “Why is 123321 divisible by 37?” outside of some special context (and insofar as this question is not just a clumsy way of asking merely for a proof that 123321 is divisible by 37).

For a pedestrian mathematical fact such as 123321’s divisibility by 37, mathematical practice recognizes no reason why it obtains. I hereby wager that none of the “explanations” given above of 123321’s divisibility by 37 is ever cited in mathematical practice as explaining why (rather than merely demonstrating that) 123321 is divisible by 37. By a “pedestrian” mathematical fact, I mean one having no feature that is especially striking (surprising, salient, remarkable, unexpected,…) in a given context, such as the fact’s symmetry, its standing in some striking pattern with certain other mathematical facts, or its unexpected simplicity.[[5]](#endnote-5)

Even for a mathematical fact that exhibits some striking feature and so could have an explanation, mathematicians routinely acknowledge that it may have no explanation, merely proofs that it holds. For instance, the Taylor series for and the Taylor series for have exactly the same convergence behavior (namely, converging only if |x| < 1).[[6]](#endnote-6) This is a striking similarity, considering that the graphs of the two functions are so dissimilar. (Only the latter function goes undefined at x = 1; the former behaves quite soberly there.) Having highlighted this striking similarity between the two otherwise dissimilar functions, the mathematician Michael Spivak (1980: 482) asks why the two functions’ Taylor series have the same convergence behavior. Then he cautions: “Asking this sort of question is always dangerous, since we may have to settle for an unsympathetic answer: it happens because it happens – that’s the way things are!”[[7]](#endnote-7)

If these two Taylor series’ displaying the same convergence behavior had no explanation, then it would have been described by mathematicians as a “mathematical coincidence.” There are plenty of mathematical facts that mathematicians deem to be mathematical coincidences, such as that the 7th-10th digits of the decimal expansion of *e* (= 2.71828182845904…, the base of the natural logarithms) are the same as its 3rd-6th digits. (The mathematician Timothy Gowers (2007: 34) describes this fact as “universally regarded as an amusing coincidence, a fact that does not demand an explanation.”[[8]](#endnote-8)) Likewise, it is thought to be a mathematical coincidence that these two Diophantine equations (that is, equations where the variables can take only integer values)

2*x*2(*x*2 – 1) = 3(*y*2 – 1)

and

*x*(*x*–1)/2 = 2*n* – 1

have exactly the same five positive solutions for *x* -- namely, 1,2,3,6, and 91 (Guy 1988: 704). Each of these facts has two components (*e*’s two stretches of four digits or the two Diophantine equations’ *x*-solutions, respectively) bearing a striking similarity. What makes this similarity a “mathematical coincidence” is that has no explanation; it simply obtains.

The distinction between mathematical coincidences and non-coincidences thus figures in mathematical practice and rests on the distinction between mathematical facts having and those lacking any explanation (while displaying some striking feature). But on the counterfactual account, explanations come so cheaply that the distinction between mathematical coincidences and non-coincidences is erased. There are mathematical facts (e.g., a fact used to evaluate one or the other side of one of the Diophantine equations with some value substituted for *x*) such that had it been different in some specific way, then the two Diophantine equations would not have had exactly the same positive solutions for *x*.

Advocates of the counterfactual approach might reply that what makes a fact qualify as mathematically coincidental is not its having no explanation, but rather its components having no *common* explainer (e.g., that no fact helps to explain both the first Diophantine equation’s *x*-solutions and the second equation’s *x*-solutions). However, since explanation comes so easily on the counterfactual approach, it will turn out to be too easy to find facts that the counterfactual approach deems to be common explainers. For the two stretches of four digits in *e*’s decimal expansion (for instance), any of the algorithms for deriving *e*’s value (e.g., that *e* = (1/0!) + (1/1!) + (1/2!) +…) will involve facts common to deriving both stretches of digits. It will likewise be easy to find “common explainers” of 123321 and 123358 both being divisible by 37: each of these divisibility facts follows from (and is counterfactually dependent on) 123284’s divisibility by 37 (considering that 123321 = 123284 + 37, 123358 = 123284 + 37x2, and the generalization that for any natural numbers *a* and *n*, *a* + 37*n* is divisible by 37 iff *a* is divisible by 37).

This returns us to the original point of this section: that pedestrian math facts have explanations on the counterfactual approach, whereas in mathematical practice they do not. There is nothing striking about 123321 and 123358 both being divisible by 37 (outside of a context giving these two numbers some special significance). Yet their divisibility by 37 has an explanation (even a common explainer) on the counterfactual approach. It seems to me that (outside of some special context) mathematical practice recognizes nothing that it would even *be* to explain why (beyond proving that) these two numbers are both divisible by 37, since there is nothing striking about this similarity. By contrast, we immediately have at least a rough idea of what we are asking for in asking why 123321 and 753357 are both divisible by 37, since there is indeed something strikingly similar in these two numbers. (I will pick up this thread in the next section.) The counterfactual approach is ill-equipped to capture the phenomena in mathematical practice that depend on many mathematical facts having no explanations.

*5. Explanatory asymmetry.* It is widely believed that at least typically, explanation is asymmetric: if *F* helps to explain why *G* obtains, then *G* does not help to explain why *F* obtains. Perhaps there are exotic circumstances (e.g., time travel) or other special cases (such as positive-feedback loops described in atemporal and coarse-grained ways) where scientific explanations are symmetric. But ordinarily, explanations do not run in circles. (If circular explanation is permitted, then there would seem to be nothing to preclude self-explanation by going completely around the circle – except for an *ad hoc* prohibition on it.)

Counterfactual accounts of *causal* explanation capture explanatory asymmetry by appealing to counterfactuals that are asymmetric because their antecedents are posited as brought about by “interventions”. As I mentioned earlier (and will discuss below), an intervention has an asymmetry built into it. But the source of that asymmetry (namely, causation) is absent from mathematics. In place of causal relations, counterfactual accounts of explanation in mathematics appeal to entailment relations. Obviously, those relations can be symmetric. Therefore, these accounts will fail to yield an asymmetry of counterfactual dependence and hence an asymmetry in explanation in mathematics. Such explanatory symmetry violates mathematical practice and fails to respect a core feature of explanation.

For instance (building on the example from section 3), a counterfactual approach would seem to endorse not only that 123284’s divisibility by 37 explains 123321’s divisibility by 37, but also the reverse. Likewise, it is the case not only that a square’s area is counterfactually dependent on the length of its side, but also that the length of its side is counterfactually dependent on its area. A counterfactual approach would seem to countenance both directions as explanatory. Similarly, a planar graph is traversable iff it is Eulerian (i.e., either zero or exactly two of its nodes have an odd number of edges), and a counterfactual account presumably endorses using a graph’s possessing either one of these properties to explain why it possesses the other.[[9]](#endnote-9)

To offer this objection, I have no need to presuppose that all explanations in mathematics are asymmetric. The challenge presupposes only that cases like the above are not symmetric. Perhaps when dealing with the foundations of mathematics, some explanations run in circles; for example, perhaps Zorn’s lemma, the axiom of choice, and the well-ordering theorem are mutually explanatory. (Each entails the others within Zermelo-Fraenkel set theory.) But such an example does not suggest that there are circular explanations all over workaday mathematics.

An advocate of a counterfactual approach could simply bite the bullet and deny that explanation in mathematics is typically asymmetric.[[10]](#endnote-10) But I am unaware of any case in the workaday mathematics literature where a mathematician has presented *F* as explaining why *G* and also (in the same conversational context, with the same contrast classes, and using the same axioms and definitions) has cited *G* as explaining why *F*.

Another reply for the counterfactual approach’s advocate would be to add to her necessary and sufficient conditions for explanation a further condition to the effect that if the other conditions are satisfied both for *F* explaining *G* and for *G* explaining *F*, then neither of them qualifies as an explanation. This requirement would prevent explanatory symmetry. However, it is terribly *ad hoc*. Furthermore, I see nothing to prevent those previous conditions from being satisfied in both directions but where exactly one direction is genuinely explanatory.

Another reply for the counterfactual approach’s advocate would be to emphasize that the relevant countermathematicals reflect not merely entailment, but what Baron, Colyvan, and Ripley (2020) call relations of “mathematical dependence.” However, as I mentioned earlier, these authors offer no account of what “mathematical dependence” consists in beyond entailment. Therefore, it is difficult to place much weight on their calling mathematical dependence “asymmetric” (p. 6) since they say nothing about where this asymmetry comes from. Their metaphor of “upstream” and “downstream” presupposes this asymmetry rather than explicating it.

One advocate of a counterfactual approach has fully recognized the threat posed by explanatory asymmetry (while noting (2017: 65) that the problem has “been mostly neglected”) and has spelled out the notion of mathematical dependence: Gijsbers (2017: 57-62). He acknowledges that (as I have just argued) entailment cannot fund the distinction between “upstream” and “downstream” for the counterfactuals underwriting explanation in mathematics. Accordingly, he appeals to another aspect of mathematics: certain facts are taken as axioms and definitions whereas others are not. Gijsbers proposes that the antecedent of the relevant sort of counterfactual posits some mathematics as brought about by a “quasi-intervention” that is “the mathematical analogue of causal intervention” (2017: 61). To see how this is supposed to work, let’s recall how Woodward’s notion of “(causal) intervention” secures asymmetry in causal explanation.

According to Woodward (2003: 98), an “intervention” *I* on physical variable *X* with respect to physical variable *Y* is a counterfactual operation that would cause *X* to change (while holding fixed the laws governing *X*’s causal impact) and would result in a change to *Y*, if at all,

(i) not by *I*’scausing *Y* to change as a cause of changing *X*

*I* 🡪 *Y* 🡪 *X*

(ii) not by *I*’scausing *Y* to change by a separate causal pathway from the pathway by which *I* affects *X*

*Y* 🡨 *I* 🡪 *X*

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

*X* 🡨*I* ----(statistical correlation) ---- *Y*’s causes 🡪 *Y*

but (iv) as an effect of changing *X*

*I* 🡪 *X* 🡪 *Y*

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

For instance, this approach achieves explanatory asymmetry when a pendulum’s length *L* helps to causally explain its period *T* (Woodward 2003: 197-8). There is a possible intervention on *L* with respect to *T* – roughly, a way to vary *L* that would change *T* (if at all) only as an effect of varying *L*. Such an intervention could consist in shortening the cord by which the bob is suspended. Under such an intervention, *T* would have been different. Because of this counterfactual dependence, *L* can help causally explain *T.* Now let’s see the explanatory asymmetry arise when we try to use *T* to help causally explain *L.* There is a possible intervention on *T* with respect to *L* – roughly, a way to vary *T* that would change *L* (if at all) only as an effect of varying *T.* This intervention cannot consist in shortening the cord and thereby causing *T* to change, since that operation violates condition (i) above on such an intervention (because the operation causes *L* to change as a cause of changing *T*). Rather, such an intervention could consist in moving the pendulum to a location with a different gravitational acceleration. But under this intervention on *T*, *L* would remain unchanged. Thus the explanatory asymmetry is generated. We cannot manipulate *L* by intervening on *T* but we can do the reverse.

Gijsbers proposes to use a similar idea to pick out the kind of counterfactual dependence that tracks explanatory dependence in mathematics. In place of causal priority, Gijsbers proposes that we appeal to the priority of axioms and definitions to theorems. A “quasi-intervention” on some mathematical fact (or entity) *X* with respect to another mathematical fact (or entity) *Y* is a counterfactual change to *X* that would bring about a change in *Y* (if at all) only by virtue of *Y* being defined in terms of *X* or governed by axioms that include *X*. Just as a causal intervention holds fixed the causal laws (if any) by which *X* has a causal impact on *Y*, so a quasi-intervention holds fixed any role that *X* may play in the definition or axioms for *Y*. In this way, Gijsbers says, the explanatory asymmetry arises. In his example, a countermathematical of the form “Had the natural numbers been…, then the ordered pairs of natural numbers would have been…” can be true when the antecedent posits some mathematics as brought about by quasi-intervention on the set of natural numbers (by altering the Peano axioms). The quasi-intervention will ramify to the ordered pairs of natural numbers since those ordered pairs are defined in terms of the natural numbers (and the quasi-intervention holds fixed this definition, along with the natural numbers’ definition as the set satisfying the Peano axioms – while altering what the Peano axioms say). By contrast, Gijsbers says, no countermathematical “Had the ordered pairs of natural numbers been…, then the natural numbers would have been…” will be true when the antecedent posits some mathematics as brought about by quasi-intervention on the ordered pairs of natural numbers relative to the natural numbers – since there is no such quasi-intervention. Any change to those ordered pairs would have to be brought about by a change to the naturals, considering that the ordered pairs are defined in terms of the naturals.

In short, the countermathematicals allowing counterfactual dependence to track explanatory priority involve “keep[ing] the asymmetric links laid down by our definitions fixed, then vary[ing] one of our mathematical objects or properties, and observ[ing] what must change downstream if the definitions remain fixed” (Gijsbers 2017: 61). Like Baron, Colyvan, and Ripley, Gijsbers uses the “downstream” metaphor. But Gijsbers spells out how it is more than merely entailment relations that go into fixing the stream’s direction. Gijsbers’s proposal is that the direction of explanation “follows the direction required by the explanatory asymmetries of the definitions” (2017: 61): “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” (2017: 60).

Gijsbers’s approach faces forthrightly the problem of finding some mathematical asymmetry to generate asymmetry in mathematical explanation. However, even where *Y* is defined or axiomatized in terms of *X* and so the relevant countermathematical holds when its antecedent posits mathematics brought about by quasi-intervention, counterfactual dependence may fail to track explanatory dependence. For instance, 123321 is defined in terms of lower natural numbers by recursive application of the successor relation. But I am still not inclined to accept that 123321’s divisibility by 37 is explained by 123284’s (and also by 444’s), on pain of too many explanations.

Perhaps this is a difficult case to judge, since I deny that *anything* explains 123321’s divisibility by 37 (when this pedestrian mathematical fact is simply taken in isolation, in a context where there is nothing remarkable about it). So let’s switch to a closely related explanandum (which I foreshadowed at the previous section’s close) that mathematicians recognize as having an explanation.[[11]](#endnote-11) Take an ordinary calculator keyboard (without the zero key).

7 8 9

4 5 6

1 2 3

We can form a six-digit number by taking the three digits on any row, column, or main diagonal on the keyboard in forward and then in reverse order. For instance, the bottom row taken from left to right, and then right to left, yields 123321. There are sixteen such “calculator numbers”: 123321, 321123, 456654, 654456, 789987, 987789, 147741, 741147, 258852, 852258, 369963, 963369, 159951, 951159, 357753, and 753357. As you can easily verify (with a calculator!), each of these numbers is divisible by 37. Is this (as the title of a recent *Mathematical Gazette* article asks) a coincidence?[[12]](#endnote-12) Why does the calculator-number theorem hold?

A proof that simply takes each calculator number in turn, showing it to be divisible by 37, treats the result as if it were a coincidence. Mathematicians do not regard such a proof as explaining why the result holds. In contrast, here is a proof of the calculator-number theorem (from a later *Mathematical Gazette* article entitled “No Coincidence”) that mathematicians recognize as explaining why it holds:

Let a, a + d, a + 2d (which could be the three digits forming the calculator number) be any three integers in arithmetic progression. Then (forming a calculator number)

a.105 + (a + d).104 + (a + 2d).103 + (a + 2d).102 + (a + d).10 + a.1

= a(105 + 104 + 103 + 102 + 10 + 1) + d(104 + 2.103 + 2.102 + 10)

= 1111111a + 12210d = 1221(91a + 10d).

So not only is the number divisible by 37, but by 1221 (= 3 x 11 x 37) (Nummela 1987: 147)

An account of mathematical explanation should identify what makes the latter proof explanatory, unlike the brute-force proof that checks each calculator number individually.

However, both of these proofs satisfy all of the requirements of the counterfactual approach, even Gijsbers’s additional requirement that the counterfactual’s antecedent posit mathematics brought about by quasi-intervention. Both proofs use “calculator numbers” defined as the numbers produced by a certain kind of sequence of six key-pressings (namely, by taking any row, column or main diagonal on the keyboard forward and then backward). Had (countermathematically) a different kind of sequence – say, six key-pressings of the same key – instead been associated with being a calculator number, then every calculator number would still have been divisible by 11 but would not have been divisible by 37. This counterfactual could be demonstrated either by a case-by-case proof or by a proof like the above explanatory proof that treats all of the calculator numbers in a unified, uniform way. Thus, both of the original proofs tell us how to ascertain this counterfactual, and mathematics posited by its antecedent is understood to be brought about by varying a definition while holding fixed the asymmetric dependence of other facts on that (altered) definition.[[13]](#endnote-13)

In other words, both the case-by-case proof and the unified proof begin with the definition of “calculator number” and go downstream from there. Both help to reveal how divisibility by 37 would have been different had the definition been different (or a property figuring in it, such as the property of being an ordinary calculator keyboard, been different). Both proofs tell us how to answer what-if-things-had-been-different questions, where the “things” in question involve the way that the calculator numbers are generated (which is the definition of “calculator number”). Therefore, on the Woodwardesque picture, both proofs should supply explanatory power.

I will extract further lessons from this example in the following section. For now, let me turn to a different problem facing Gijsbers’s proposal. For some explanations in mathematics, the axioms and definitions of the terms in the explanandum are not the source of the explanatory priority of the explanans over the explanandum (see Lange 2019). The most noteworthy examples are explanations depending upon “impure” proofs: proofs using some concepts figuring neither in the theorem being proved nor in the definitions of the concepts figuring in that theorem nor in the definitions of the concepts figuring in those definitions… nor in some other way helping to determine the theorem’s content (Detlefsen 2008: 193). For instance, Spivak’s (1980: 528) explanation of the two Taylor series’ having the same convergence behavior (from section 4) proceeds through facts about complex numbers, even though the two Taylor series involve only real numbers. By introducing imaginary numbers, the explanation introduces concepts exogenous (“not intrinsic”, mathematicians say (Detlefsen and Arana 2011:1)) to the theorem being proved.[[14]](#endnote-14)

There are many similar examples – e.g., Desargues’ theorem (a fact in two-dimensional Euclidean geometry) has an explanation only when it has been situated in three-dimensional projective geometry, thereby introducing (not only an additional spatial dimension but also) points at infinity where parallel lines meet.[[15]](#endnote-15) These additional points (and dimension) are not covered by the axioms for (two-dimensional) Euclidean geometry. Yet a “quasi-intervention” on the explanans relative to the explanandum has to twiddle the explanans so as to bring about a change in the explanandum (if at all) only by virtue of the explanandum being defined in terms of the explanans or governed by axioms that involve the explanans. Therefore, explanatory dependence in such an “impure” case cannot be tracked by counterfactual dependence when countermathematicals are restricted to positing quasi-interventions.

*6. What matters to explanatory power in mathematical practice.* Perhaps the most fundamental challenge that I will offer to the counterfactual approach is that counterfactual relations fail to track explanatory relations in mathematics. Whether a given mathematical fact has explanatory power often has nothing to do with whether the explanandum would have been different, had that fact been different – and whether some proof has explanatory power often has nothing to do with whether the proof reveals the explanandum’s counterfactual dependence on other facts.

Consider, for instance, the explanation (given in the previous section) of the fact that every “calculator number” is divisible by 37. The explanation, in brief, comes from the fact that every calculator number consists (in base 10) of three digits in arithmetic progression, first forwards and then backwards. Is the explanandum counterfactually dependent on this fact -- and does the explanatory proof reveal this counterfactual dependence and derive its explanatory power from doing so?

The explanandum is counterfactually dependent on each calculator number consisting (in base 10) of three digits in arithmetic progression, though not every change to the way that calculator numbers are generated would have made a difference to their divisibility by 37. Had it not been the case that 123321 consists of three digits in arithmetic progression (forwards and then backwards), then 123321 might still have been divisible by 37; obviously, many numbers are so divisible without their digits arising from arithmetic progressions. On the other hand, as I mentioned earlier, had every calculator number consisted (in base 10) of a given digit repeated six times, then it would not have been the case that every calculator number is divisible by 37 (but it would still have been the case that every calculator number is divisible by 11).

But the key question for evaluating whether the counterfactual approach to explanation in mathematics does justice to this example is whether the explanatory proof reveals this counterfactual dependence to us in a way that the non-explanatory, case-by-case proof does not. I don’t see how it does. We could replace 123321 in the case-by-case proof by 111111, for instance, and (by discovering that 111111 is indivisible by 37) learn the explanandum’s counterfactual dependence on the key-strokes displaying the one pattern rather than the other. So the counterfactual approach fails to capture the difference in explanatory power between the two proofs.

Consider the fact (to which the explanatory proof appeals) that any calculator number’s digits form an arithmetic progression (forwards and backwards). What makes this fact explanatorily relevant to the fact that all calculator numbers are divisible by 37? The most obvious difference between the explanatory and non-explanatory proofs in this example is that the explanatory proof treats all of the calculator numbers together and in the same way, whereas the non-explanatory proof treats them one-by-one. It seems to me that this is ultimately the reason for the proofs’ difference in explanatory power. The explanatory proof treats all of the calculator numbers in a unified, uniform way throughout, deriving their divisibility by 37 from something else that they all have in common by virtue of being calculator numbers. This proof (unlike the case-by-case proof) thereby reveals it to be no coincidence that every calculator number is divisible by 37.

Mathematicians themselves say about many proofs that they are non-explanatory precisely because they fail to give a unified, uniform treatment to all of the cases that the explanandum reveals to possess some striking similarity. One notorious example is Appel and Haken’s computer-assisted proof of the four-color theorem. Their proof works by showing that a set of local arrangements of countries is “unavoidable” (roughly, at least one of these arrangements must be part of any planar map) and that each of the set’s arrangements is “reducible” (i.e., cannot occur in the smallest counterexample to the theorem because if it appears in a non-4-colorable map, then there is a map with fewer countries that is non-4-colorable; any *n*-colorable map will still be *n*-colorable after the addition of a reducible configuration).[[16]](#endnote-16) The proof thus works by showing that there is no smallest non-4-colorable map. Does this proof reveal the answer to what-if-things-had-been-different questions? Perhaps it reveals that had the given set of reducible configurations not managed to be unavoidable – had unavoidability required the addition of another configuration – then the theorem would not have held. The unavoidable set used in the proof was arrived at roughly by the process of taking a small unavoidable set of *prima facie* reducible configurations, replacing any irreducible configuration in it with several reducible ones, then adding configurations to that set in order to restore its unavoidability, and so forth in a process of mutual adjustment. Perhaps had the set used in the proof been unable to end this process, then any additional configurations required to fashion an unavoidable set would have failed to be reducible, allowing a map to be non-4-colorable. On the other hand, perhaps any required additional configurations might have been reducible but also might not have been. I cannot tell which countermathematical obtains and I know of no mathematician who even discusses the matter. Yet a counterfactual approach regards such matters as crucial to the proof’s explanatoriness.

On the contrary, whether or not such counterfactual dependence obtains makes no difference(!); in either event, the proof would fail to explain why the four-color theorem holds. As Stewart (1975: 304) says, the proof “does not give a satisfactory explanation of why the theorem is true. … The answer appears as a kind of monstrous coincidence.” That is, its failure to explain arises from the fact that the unavoidable set of reducible configurations contains enormously many elements (1476 in the original proof; a more recent proof uses 633) and each element must be proved separately to be reducible. That they are all reducible is thereby portrayed as a giant coincidence.

Such tremendous disunity is what mathematicians and philosophers cite as making the proof non-explanatorily. For example, Baker (2009: 148; cf. 2008: 342) remarks: “There are 1476 different sub-cases that are individually considered. Thus, the proof is very unexplanatory.” The graph-theorist Paul Seymour (2016: 417) agrees, strongly suspecting that the four-color theorem is no coincidence – that it has a unified, uniform proof (as yet unknown) that explains why four colors suffice for any planar map. Seymour believes that “the massive case analysis of the computer proof” fails to identify “the ‘real’ reason the [theorem] is true; what exactly is it about planarity that implies that four colours suffice?”[[17]](#endnote-17)

Plenty of other proofs are recognized by mathematicians as explanatory but do *not* acquire that power by giving a unified, uniform treatment to various cases that the explanandum reveals to possess some striking similarity. But those proofs also do not derive their explanatory power by providing information about patterns of counterfactual dependence. For example, D’Alessandro (2020) describes the mathematician Tim Gowers (2008: 261) as asking for an explanation of a famous identity from Euler:

Gowers demands an explanation, not merely a proof, by asking “What on earth, one might wonder, has to do with adding up reciprocals of squares?” In this context, an explanation over and above a proof of this identity must derive the identity from considerations where has an obvious and natural place. Gowers gives such a derivation using Fourier analysis featuring a periodic function with period 2. Gowers (2008: 162) says:

Now we have a reason for the appearance of : it comes up in the formula for the Fourier coefficents. What is more, its appearance there can be explained as well. A periodic function on is more naturally thought of as a function defined on the unit circle. The Fourier coefficient …is a certain average defined on the unit circle, so we have to divide by the length of the circle, which is 2

What makes this proof explanatory in this context has nothing to do with this proof’s providing information about patterns of counterfactual dependence or answers to what-if-things-had-been-different questions. The proof is explanatory by virtue of tracing the salient feature of the explanandum (its featuring ) to a similar feature of the problem’s setup.

An advocate of a counterfactual account might reply that what makes Gowers’s derivation explanatory is that it indeed reveals a counterfactual dependence: it shows that had not been something to do with the unit circle, then it would not have figured in the identity. Of course, this counterfactual says little about precisely how the identity would have been different under the countermathematical antecedent – only that it would not have had in it. But if that crude sort of counterfactual dependence suffices for explanation,[[18]](#endnote-18) then many other sorts of proofs of the same identity will also count as explanatory. The identity has many proofs (e.g., Euler’s original proof using the sine function’s Taylor series expansion, another proof squeezing the above sequence between one convergent sequence involving the cotangent and another involving the cosecant).[[19]](#endnote-19) For each proof, there will be some corresponding countermathematical capturing the way that enters the derivation, such as “Had sin(x) not had zeros at for *k* = 1, 2,…, then would not have figured in the identity.” For that matter, Euler generalized his original proof to show that and . Thus, the same proof recipe generates answers to many specific what-if-things-had-been-different questions regarding the original identity (e.g., “What if the sequence were the sum of reciprocal fourth powers rather than reciprocal squares?”) along with telling us that would not have figured in the sequence’s sum if had lacked some or another property under whose auspices it entered the proof (e.g., as the smallest positive number *x* for which sin(*x*) = 0). Nevertheless, Gowers does not cite Euler’s proof as explanatory. Rather, it provokes the why question.

Another way to see that explanatory power and counterfactual dependence come apart is that the very same features that make certain proofs (and facts) *non*-explanatory in mathematics nevertheless often enable those proofs to reveal the explanandum’s counterfactual dependence on the cited facts. For example, I mentioned earlier that the Taylor series for and for have exactly the same convergence behavior (namely, converging only if |x| < 1) and that this similarity can be explained (Spivak 1980: 528) only by placing this explanandum (which exclusively concerns real numbers) in the context of complex numbers. However, using the root test[[20]](#endnote-20) (which identifies a quantity such that the sequence converges if this quantity is less than 1 but diverges if it exceeds 1), we could give separate proofs of the two infinite series’ convergence behavior. This pair of proofs exposes many knobs that can be countermathematically twiddled, where those twiddles would ramify to impact the explanandum. For instance, the pair of proofs reveals many countermathematicals such as “Had the root test for ’s Taylor series yielded a quantity greater than 1 at *x* = , then the series would have diverged there.” But all of these revelations about the result’s counterfactual dependence do not enhance the explanatory power of the root test’s separate application to the two series. Although the separateness of these proofs increases the number of knobs available to twiddle, their separateness is precisely what deprives these proofs of explanatory power in this case; an explanatory proof must trace the two functions’ common convergence behavior to another feature that the two functions share.

Here is another common kind of example where the same consideration that yields greater informativeness about patterns of counterfactual dependence also yields less explanatory power. Some proofs arrive at their results by the grace of “fortuitous” algebraic cancellations and other algebraic “miracles”. Those coincidences supply knobs to be twiddled; the proofs expose that had those “miracles” not occurred, then the explanandum would not have held. But those coincidences also deprive the proof of explanatory power. An explanation would rely on no such convenient miracle (and would even explain where any apparent “miracle” came from).

Here is an example. The explanandum is the theorem that the members of the Somos-4 sequence are all integers (rather than fractions). The Somos-4 sequence is defined by the following recursion formula:

*a*0 = *a*1 = *a*2 = *a*3 = 1

for *n* > 3, *a*n = (*a*n−3 *a*n−1 +*a*n−22)/*a*n−4.

As successive values of the sequence were first computed (to beyond *a*100), there was widespread surprise as each was discovered to be an integer. Ultimately, a general proof was sought and a proof by mathematical induction was found. The base case was that the sequence’s first 8 members (1,1,1,1,2,3,7,23) are all integers. The inductive step aims to show that if 8 successive members *B*(0),…,*B*(7) are integers, then *B*(8) = [*B*(5)*B*(7) + *B*(6)2]/*B*(4) is an integer, too – i.e., that *B*(5)*B*(7) + *B*(6)2 0 (mod *B*(4)). A typical approach (Malouf 1992: 258) to proving the inductive step is first to show that

*B*(5)*B*(7) + *B*(6)2*B*(4) *B*(1)*B*(2) [*B*(5)*B*(7) + *B*(6)2] = *B*(1)*B*(2)*B*(5)*B*(7) + *B*(1)*B*(2)*B*(6)2

and then to substitute *B*(2)*B*(4) + *B*(3)2 for *B*(1)*B*(5) in the first term by appealing to the recursion formula for *B*(5):

*B*(5) = [*B*(2)*B*(4) + *B*(3)2]/*B*(1).

The substitution yields

*B*(4) *B*(2)2*B*(4)*B*(7) + *B*(2)*B*(3)2*B*(7) + *B*(1)*B*(2)*B*(6)2.

But since *B*(2) and *B*(7) are integers (by the inductive supposition that *B*(0),…,*B*(7) are integers), the first term is a multiple of *B*(4) and so is equivalent (mod *B*(4)) to 0, contributing nothing to the sum:

*B*(4) *B*(2)*B*(3)2*B*(7) + *B*(1)*B*(2)*B*(6)2.

In the same way, we then use the recursion formula for *B*(7) to replace *B*(3)*B*(7) in the first term

*B*(4) *B*(2)*B*(3)*B*(5)2 + *B*(1)*B*(2)*B*(6)2

and the recursion formula for *B*(6) to replace *B*(2)*B*(6) in the second term

*B*(4) *B*(2)*B*(3)*B*(5)2 + *B*(1)*B*(3)*B*(5)*B*(6)

and the recursion formula for *B*(6) again to replace *B*(3)*B*(5) in the first term

*B*(4) *B*(2)2*B*(5)*B*(6) + *B*(1)*B*(3)*B*(5)*B*(6)

*B*(4) *B*(5)*B*(6) [*B*(2)2 + *B*(1)*B*(3)].

Fortuitously, the bracketed expression is just the numerator of the recursion formula for B(4), so

*B*(4) *B*(0)*B*(5)*B*(6)*B*(4),

which (since *B*(0), *B*(5), and *B*(6) are supposed integral) is just a multiple of *B*(4), yielding

*B*(4) 0.

As Gale (1991a: 41) says, “[a]lthough the proof is very simple, it depends on the fortuitous fact that the factor [in brackets above] turns up.” It was later found that the same strategy works for the Somos-5 sequence

*a*0 = *a*1 = *a*2 = *a*3 = *a*4 = 1

for *n* > 4, *a*n = (*a*n−4 *a*n−1 + *a*n−3*a*n-2)/*a*n−5

using the first 10 members of the sequence as the base case. Hickerson then used the same strategy (with the first 12 members as the base case) for the Somos-6 sequence

*a*0 = *a*1 = *a*2 = *a*3 = *a*4 = *a*5= 1

for *n* > 5, *a*n = (*a*n−5 *a*n−1 + *a*n−4 *a*n−2 +*a*n−32)/*a*n−6

and even for any generalization thereof for arbitrary *a*0,…, *a*5. However, the Somos-*k* sequences for *k* ≥ 8 do not consist exclusively of integers; the requisite fortunate bracketed expression fails to emerge.

The above proof can help to answer many what-if-things-had-been-different questions. It reveals that if the sequence in question had instead been Somos-5, the result would have been the same, whereas had it instead been Somos-8, then the outcome would have been different. Furthermore, the proof reveals that the theorem about Somos-4 depends on the factor in brackets turning up as it did. If it hadn’t arisen, then Somos-4 would not have consisted entirely of integers.

But the proof’s revealing this counterfactual dependence is precisely what deprives the proof of explanatory power; this counterfactual dependence on the bracketed factor’s turning up does not make its turning up a reason why the theorem holds. Gale (1991b: 49) notes that despite our having this proof, we lack an explanation of the theorem; the above sequences “for unexplained reasons always … yield integer terms.” The revealed counterfactual dependence on such a “fortuitous fact” is precisely what deprives the proof of explanatory power, as Gale (1991a: 41) writes regarding this proof:

But what have we learned? As Hickerson puts it, “The thing I dislike about my proof is that it doesn’t explain why the result is true. It depends primarily on the fact that when you compute a12, there’s an unexpected cancellation. But why does this happen?” Indeed, the proof, rather than illuminating the phenomenon, makes it if anything more mysterious. … Perhaps, if and when we find the “right” proof, the situation will become clarified, but must there necessarily be a right proof? One is reminded of the proof of the four-color theorem.

The explanandum is counterfactually dependent on the miraculous cancellation but is not explained by this coincidence. The miraculous cancellation is what the proof reveals can be twiddled countermathematically to alter the explanandum, but it is also what makes the proof non-explanatory.

*7. Conclusion.* The various challenges that I have identified as facing counterfactual approaches to explanation in mathematics may not ultimately be the most intractable obstacles they encounter. For example, perhaps pragmatic considerations that I have not discussed can help counterfactual accounts to circumvent some of these challenges. Consider a fact (or proof) that mathematicians regard as non-explanatory, but that a counterfactual account must regard as explanatory. Perhaps the reason that mathematicians say that it is non-explanatory is not that it cannot explain, but rather that it is not the kind of explanation that those particular mathematicians, in that particular conversational context, are seeking. The same phenomenon routinely occurs in connection with scientific explanation, as in Hanson’s (1958: 34) famous example (mentioned earlier) of explaining a death in an automobile accident. Of course, such a response must avoid ad hocery and cannot address examples involving a fact (or proof) that mathematicians regard as (highly) explanatory, but that a counterfactual account must regard as non-explanatory (or merely minimally explanatory).

One theme common to many of the above challenges is that the counterfactual dependence of a proof’s conclusion on its premises concerns what happens when those premises are twiddled, whereas the explanatory power of a fact (or proof) depends on the details of the inferential path between the premises and the conclusion.[[21]](#endnote-21) For instance, the difference between a unified, uniform proof and a massively case-by-case proof – or between a proof that does and does not depend on a fortuitous cancellation – is evident only from its path, not merely from the relation between the two ends of that path (that is, from whether we can modify its conclusion by twiddling with its premises).

In the face of the challenges that I have explored, an advocate of the counterfactual approach might agree that mathematical practice recognizes many factors contributing to (or detracting from) explanatory power but having nothing to do with counterfactual dependence. Nevertheless, she might insist, there is a special kind of mathematical explanation supplied by information about the explanandum’s counterfactual dependence on other facts – a kind of explanation recognized by (some) philosophers who are interested in explanation in mathematics. It seems to me that this would constitute a Pyrrhic victory for the counterfactual account. It would not show that counterfactual dependence is connected to explanation as it figures in actual mathematical practice – the sort of explanation that has been my exclusive concern here.

Alan Baker 2008. Experimental Mathematics. *Erkenntnis* 68: 331-344.

Alan Baker 2009. Mathematical Accidents and the End of Explanation. In Otávio Bueno and Øystein Linnebo (eds.), *New Waves in Philosophy of Mathematics*. Houndmills: Palgrave Macmillan, pp. 137-59.

Sam Baron and Mark Colyvan 2016. Time Enough for Explanation. *Journal of Philosophy* 113: 61-88.

Sam Baron, Mark Colyvan, and David Ripley 2017. How Mathematics Can Make a Difference. *Philosophers’ Imprint* 17(3): 1-19.

Sam Baron, Mark Colyvan, and David Ripley (2020). A Counterfactual Approach to Mathematical Explanation. *Philosophia Mathematica* 28(1):1-34.

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

Robin Chapman 2003. Evaluating (2). http://empslocal.ex.ac.uk/people/staff/rjchapma/etc/zeta2.pdf

Mark Colyvan 2012. *An Introduction to the Philosophy of Mathematics*. Cambridge: Cambridge University Press.

William D’Alessandro 2020. Viewing-as Explanations and Ontic Dependence. *Philosophical Studies* 177:769-92.

Michael Detlefsen 2008. Purity as an Ideal of Proof. In Paolo Mancosu (ed.), *The Philosophy of Mathematical Practice*. Oxford: Oxford University Press, pp. 179-197.

Michael Detlefsen and Andrew Arana 2011. Purity of Methods, *Philosophers’ Imprint* 11(2): 1-20.

Katie Elliot and Marc Lange forthcoming. Running it up the Flagpole to see if Anyone Salutes: A Response to Woodward on Causal and Explanatory Asymmetries. *Theoria*.

Joachin Frans and Erik Weber 2014. Mechanistic Explanation and Explanatory Proofs in Mathematics. *Philosophia Mathematica* 22: 231-248.

Steven French and Juha Saatsi 2018. Symmetries and Explanatory Dependencies in Physics. In Alexander Reutlinger and Juha Saatsi (eds.), *Explanation Beyond Causation*. Oxford: Oxford University Press, pp. 185-205.

David Gale 1991a. Mathematical Entertainments. *Mathematical Intelligencer* 13(1): 40-42.

David Gale 1991b. Mathematical Entertainments. *Mathematical Intelligencer* 13(4): 49-52.

Victor Gijsbers 2017. A Quasi-Interventionist Theory of Mathematical Explanation. *Logique & Analyse* 60(237): 47-66.

Timothy Gowers 2007. Mathematics, Memory, and Mental Arithmetic. In Mary Leng, Alexander Paseau, and Michael Potter (eds.), *Mathematical Knowledge*. Oxford: Oxford University Press, pp. 33-58.

Timothy Gowers 2008. . In Timothy Gowers, June Barrow-Green, and Imre Leader (eds.), *The Princeton Companion to Mathematics*. Princeton: Princeton University Press, pp. 261-63.

Richard K. Guy 1988. The Strong Law of Large Numbers. *American Mathematical Monthly* 95: 697-712.

Norwood Russell Hanson. 1958. *Patterns of Discovery*. Cambridge: Cambridge University Press.

Marc Lange 2014. Did Einstein Really Believe that Principle Theories are Explanatorily Powerless? *Perspectives on Science* 22: 449-463.

Marc Lange 2017. *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. New York: Oxford University Press.

Marc Lange 2019. Ground and Explanation in Mathematics. *Philosophers’ Imprint* 19.33: 1-18.

Marc Lange 2021. Asymmetry as a Challenge to Counterfactual Accounts of Non-Causal Explanation. *Synthese* 198: 3893-3918.

Lina Jansson 2015. Explanatory Asymmetries: Laws of Nature Rehabilitated. *Journal of Philosophy* 112: 577-599.

Matthias Jenny 2018. Counterpossibles in Science: The Case of Relative Computability. *Noûs* 52: 530-560.

Max Kistler 2013. The Interventionist Account of Causation and Non-Causal Association Laws. *Erkenntnis* 78(1): 1-20.

Jaakko Kuorikoski forthcoming. There are no Mathematical Explanations. *Philosophy of Science*.

David Lewis. 1973. *Counterfactuals*. Oxford: Blackwell.

Janice Malouf 1992. An Integer Sequence from a Rational Recursion. *Discrete Mathematics* 110: 257-61.

Paolo Mancosu 2018. Explanation in Mathematics, in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, (Summer 2018 Edition), URL = <https://plato.stanford.edu/archives/sum2018/entries/mathematics-explanation/>.

Eric Nummela. 1987. No Coincidence. *The Mathematical Gazette* 71(456): 147.

Mark Pexton 2014. How Dimensional Analysis Can Explain. *Synthese* 191(10): 2333-2351.

Alexander Reutlinger 2016. Is There a Monist Theory of Causal and Non-Causal Explanations? The Counterfactual Theory of Scientific Explanation. *Philosophy of Science* 83(5): 733-745.

Alexander Reutlinger 2017. Does the Counterfactual Theory of Explanation Apply to Non-Causal Explanations in Metaphysics? *European Journal for Philosophy of Science* 7: 239-256.

Alexander Reutlinger 2018. Extending the Counterfactual Theory of Explanation. In Alexander Reutlinger and Juha Saatsi (eds.), *Explanation Beyond Causation*. Oxford: Oxford University Press, pp. 74-95

Alexander Reutlinger, Mark Colyvan, and Karolina Krzyżanowska (forthcoming). The Prospects for a Monist Theory of Non-Causal Explanation in Science and Mathematics. *Erkenntnis.*

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

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

Juha Saatsi 2018. On Explanations from Geometry of Motion. *British Journal for the Philosophy of Science* 69(1): 253-273.

Wesley Salmon 1989. Four Decades of Scientific Explanation. In Philip Kitcher and Wesley Salmon (eds.), *Scientific Explanation*, Minnesota Studies in the Philosophy of Science, Vol. 13. Minneapolis: University of Minnesota Press, pp. 3-219.

W.W. Sawyer 1943. *Mathematician’s Delight*. Harmondsworth: Penguin.

W.W. Sawyer 1955. *Prelude to Mathematics*. Harmondsworth: Penguin.

Paul Seymour 2016. Hadwiger’s Conjecture. In John Forbes Nash and Michael Rassias (eds.), *Open Problems in Mathematics*. Berlin: Springer, pp. 417-437.

Michael Spivak 1980. *Calculus*, 2nd ed. Berkeley: Publish or Perish.

Robert C. Stalnaker 1968, A Theory of Conditionals. In Nicholas Rescher (ed.), *Studies in Logical Theory*. Oxford: Blackwell, pp. 98-112.

Ian Stewart 1975. *Concepts of Modern Mathematics*. Baltimore: Penguin.

Timothy Williamson 2018. Counterpossibles. *Topoi* 37:357-368.

Robin Wilson and John J. Watkins (eds.) 2013. *Combinatorics: Ancient and Modern.* Oxford: Oxford University Press.

James Woodward 2003. *Making Things Happen*. New York: Oxford University Press.

James Woodward 2018. Some Varieties of Non-Causal Explanation. In Alexander Reutlinger and Juha Saatsi (eds.), *Explanation Beyond Causation*. Oxford: Oxford University Press, pp. 117-137.

1. D’Alessandro (2020) briefly argues that a counterfactual approach cannot accommodate what he terms “viewing-as” explanations in mathematics. [↑](#endnote-ref-1)
2. This paper thus contrasts sharply with Kuorikoski (forthcoming), which argues against counterfactual accounts of explanation in mathematics on the grounds that the requisite countermathematicals are unavailable.

   However, one should be cautious about maintaining that nontrivial countermathematicals are common in mathematical practice. Reutlinger, Colyvan, and Krzyżanowska (forthcoming) defend the prevalence of such countermathematicals by describing *reductio* proofs in mathematics as involving nontrivial countermathematicals. But if *reductio* proofs involve countermathematicals, then (it seems to me, following Williamson 2018: 363) those countermathematicals are trivial: they show that a blatant contradiction would have held, had a countermathematical antecedent obtained. Therefore, philosophers who argue that nontrivial countermathematicals are common in mathematics cannot cite as examples the countermathematicals in *reductio* proofs and must maintain that countermathematicals are treated very differently in different mathematical contexts: in some contexts, the countermathematical antecedent is allowed to ramify to produce a blatant contradiction, whereas in other contexts, this is not allowed to happen. [↑](#endnote-ref-2)
3. I also wish to acknowledge but set aside the brief remarks made by some advocates of counterfactual accounts of explanation in mathematics (e.g., Frans and Weber 2014: 247; Baron, Colyvan, and Ripley 2020:3) that counterfactual accounts may not cover all explanation in mathematics. These remarks are to some degree belied by the fact that a principal motivation cited for these counterfactual accounts is the possibility of giving a single, unified account of all explanation. In any case, the present paper questions whether a counterfactual account can do justice to explanations in mathematics like those that its advocates discuss. [↑](#endnote-ref-3)
4. Baron, Colyvan, and Ripley (2020: 7-8): “[T]here is a more or less natural division between the upstream and the downstream facts in cases of ordinary counterfactuals. The division is due to the underlying temporal structure of the universe. The mathematical case lacks

   temporal structure; so it is less obvious what the relevant `downstream' and `upstream' facts might be, and thus it is perhaps less clear what we should hold fixed and what we should permit to vary in this case. But while there is no temporal structure, there is an analogous mathematical structure: nodes in the structure correspond to mathematical facts, and the links in the structure

   are asymmetric relations of mathematical dependence: the dependence of one mathematical fact on another. The facts that are ‘upstream’ from a given mathematical fact within such a structure, are the facts that the mathematical fact depends on. The facts that are ‘downstream’ from a given mathematical fact are the facts that depend upon that fact. When we carry forward the implications of a twiddle via the ramification procedure, we carry them through the mathematical facts that are downstream in this sense.” [↑](#endnote-ref-4)
5. I have given examples of these and other salient features, drawn from the mathematics literature, in Lange (2017: 231-346). [↑](#endnote-ref-5)
6. I have previously discussed this example (and some of the others below) in Lange (2017: 290-2, 331, 344-5), where I also cite prior discussions of it in the philosophical literature. [↑](#endnote-ref-6)
7. Gale (1991a: 41) also acknowledges that a theorem, although proved, need not have an explanatory proof. (See section 6 below.) Spivak goes on to point out that “In this case there does happen to be an explanation…” – namely, involving the impure proof appealing to the radius of convergence theorem applied to the complex plane. (I will discuss this briefly below.) [↑](#endnote-ref-7)
8. Baker (2009: 140) discusses this example of a mathematical coincidence and gives a similar analysis of what makes it coincidental. I defend this view of mathematical coincidence more fully in Lange (2017: 276-313). [↑](#endnote-ref-8)
9. This challenge is closely related to a similar challenge facing counterfactual theories of non-causal scientific explanations such as the explanation of the fact that no one ever succeeded in traversing Euler’s famous Königsberg bridges. A counterfactual theory seizes on the truth of “Had all parts been connected to an even number of bridges, then people would not have failed to traverse the bridges” and “Had exactly two parts of town been connected to an odd number of bridges, then people would not have failed to traverse the bridges.” But by the same token, the “reverse” counterfactuals (such as “Had people succeeded in crossing the bridges, then zero or exactly two regions of town would have been connected to an add number of bridges”) threaten to be endorsed by whatever account is given of the “forward” counterfactuals. But the “reverse” counterfactuals would lead to explanatory symmetry in this case. (See note 10.) For more on the challenge that explanatory asymmetry poses for counterfactual accounts of non-causal scientific explanation, see Lange (2021) and Elliot and Lange (forthcoming). [↑](#endnote-ref-9)
10. This gambit is taken by Reutlinger (2017: 253) in connection with non-causal scientific explanations (see note 9): “I think there are good reasons to hold that *some* non-causal explanations are not asymmetric…I believe that *some* (for instance, Euler’s explanation [in connection with the Bridges of Königsberg] … lack such an asymmetry because the counterfactual dependence in question is symmetric.” Likewise, Baron, Colyvan, and Ripley (2020: 8 fn. 8) write: “One way forward is to accept that in such cases of mutual inter-provability [i.e., where mathematical facts can be proved from one another] there is explanation in both directions.” [↑](#endnote-ref-10)
11. This example was first brought to my attention by Roy Sorensen. I discuss this example and Sorensen’s views in Lange (2017: 276-81, 283-6, 353-6, 394-5). [↑](#endnote-ref-11)
12. The article appears (unsigned, as a “gleaning”) on p. 283 of the December 1986 issue. [↑](#endnote-ref-12)
13. A referee suggested that an advocate of the counterfactual approach might reply that the brute-force proof can be excluded from qualifying as explanatory on the grounds that it fails to identify a common explainer for every calculator number’s divisibility by 37. However, requiring a common explainer is no part of Gijbers’s proposal and is not motivated by that proposal’s Woodwardesque motivations; it has nothing to do with counterfactuals, interventions, and so forth. (Woodward’s interventionist account of causal explanations incorporates no such requirement.) It would require an independent motivation. Furthermore, the definition of “calculator number” is common to each of the cases in the brute-force proof; that definition is used to show that 123321 is a calculator number and to show that (for instance) 321123 is a calculator number. The unity of the unified proof lies not merely in its involving a common explainer for each calculator number, but more broadly in its treating every calculator number together and in the same way throughout. [↑](#endnote-ref-13)
14. The passage from Spivak (1980: 482) quoted earlier continues as follows: “In this case there does happen to be an explanation, but this explanation is impossible to give [before the chapter introducing complex numbers]; although the question is about real numbers, it can be answered intelligently only when placed in a broader context.” For another such example, see Sawyer 1955: 180-181. I agree with Sawyer (1943: 232) that complex numbers reveal “the reasons for results which had previously seemed quite accidental.” [↑](#endnote-ref-14)
15. I discuss this example in Lange (2017: 314-46). [↑](#endnote-ref-15)
16. A useful source on this proof’s strategy is Wilson and Watkins 2013. [↑](#endnote-ref-16)
17. Likewise, Colyvan (2012: 81) writes: “Proofs [by cases] lack unity. Different reasons are often offered in the different cases, and it looks as though the theorem itself holds merely by accident. What we would like is a proof that offers the same reason in each case; that would provide an explanation of the theorem in question.” I would say not that this is what we would *like*, but rather that this is (for certain theorems that we have proved by cases) what it would take to explain why the theorem holds (and perhaps what we have some reason to believe exists, though we have not yet found it). [↑](#endnote-ref-17)
18. Perhaps such a crude counterfactual twiddle is intended to be ruled out by Reutlinger’s requirement (given in section 2) that an explanation tell us how the explanandum would have been different, had the explanans’ “initial conditions” been different *in some specific way*. But a counterfactual approach seems to need such a crude counterfactual twiddle in order to find some what-if-things-had-been-different question that the explanatory proof answers and that is connected to what makes it explanatory in mathematical practice. (In any case, this twiddle seems no cruder than “Had π not been transcendental” that (as I mentioned in section 2) some advocates of counterfactual approaches have invoked.) [↑](#endnote-ref-18)
19. Fourteen proofs are given in Chapman 2003. [↑](#endnote-ref-19)
20. https://en.wikipedia.org/wiki/Root\_test [↑](#endnote-ref-20)
21. Gijsbers (2017) also emphasizes the route that the proof takes. But his point is not that many of a route’s features can make a proof (non-)explanatory despite having nothing to do with the counterfactual dependence of the route’s end on its start. Rather, Gijsbers emphasizes that the relevant sort of countermathematical keeps the route fixed under counterfactual twiddles to its start. [↑](#endnote-ref-21)