Counterfactual Scheming
Sam Baron
University of Western Australia
samuel.baron@uwa.edu.au

Abstract
Mathematics appears to play a genuine explanatory role in science. But how do mathematical explanations work? Recently, a counterfactual approach to mathematical explanation has been suggested. I argue that such a view fails to differentiate the explanatory uses of mathematics within science from the non-explanatory uses. I go on to offer a solution to this problem by combining elements of the counterfactual theory of explanation with elements of a unificationist theory of explanation. The result is a theory according to which a counterfactual is explanatory when it is an instance of a generalised counterfactual scheme.

1. Introduction

Call a scientific explanation in which mathematical facts explain physical facts: an extra-mathematical explanation. Philosophers disagree over whether there are any genuine cases of extra-mathematical explanation in science. I will assume that there are. The question then arises as to how such explanations work. Baron, Colyvan and Ripley (2017) suggest one answer to this question, offering a counterfactual approach to extra-mathematical explanation. In this paper, I present and then solve a problem for the counterfactual approach.

The paper is structured as follows. I begin, in §2, by outlining a condition of adequacy that a theory of extra-mathematical explanation should meet. After that, I formulate a basic counterfactual theory of extra-mathematical explanation (§3) and suggest that the basic theory fails to satisfy the adequacy constraint (§4). I proceed to outline a more complex counterfactual theory, which combines elements of a counterfactual approach to explanation with elements of a unificationist theory of explanation (§5). According to the resulting view, a counterfactual is explanatory when it is an instance of a counterfactual scheme. I show that this modified approach is capable of satisfying the condition of adequacy outlined in §2 by applying it to some cases of explanation (§5–§6). I wrap up in §7.

2. Constraints on a Theory

Not every use of mathematics within science is explanatory. Sometimes mathematics plays a weaker role in scientific explanations, as a tool for representing an underlying physical phenomenon, or as a calculating framework. To gain a feel for the difference it is useful to consider a couple of cases. First, a case of scientific explanation in which the mathematics appears to be genuinely explanatory; after that a case in which the mathematics does not play an explanatory role.

The poster-child for extra-mathematical explanation is the case of the North American magicicada, first introduced into the philosophical literature by Baker (2005). The magicicada spends most of its life in the ground in its larval state. When conditions are right, the

magicicada arises in a grand swarm as an adult for around two weeks. The two sub-species possess life-cycles of 13 and 17 years respectively. Why?

Well, assume first that there are ecological constraints that force the cicadas to have a life-cycle in the 12–18-year range. Suppose further that in the past the magicicada was predated by a range of predators with life-cycles of 2, 3, 4, 5, 6, 7, 8 and 9. The optimal strategy for avoiding predation in this scenario is to minimise the frequency of intersection between magicicada swarming and predator swarming. Having a prime-numbered life-cycle achieves this result. For any two numbers $m$ and $n$ their lowest common multiple is maximised when $m$ and $n$ are co-prime. 13 and 17 are both co-prime with each of 2, 3, 4, 6, 7, 8 and 9. Accordingly, compared to other numbers in the 12–18 range, the lowest common multiple between 13 or 17 and numbers in the 2–9 range is maximised. Because the frequency of intersection between the life-cycles of the magicicada and any predators is lowest when the lowest common multiple of the life-cycles is maximised, it follows that the optimal life cycle for predator avoidance in the 12–18-year range is 13 or 17. Cicadas with periodical life-cycles that are being predated by predators that have periodical life cycles should evolve toward this optimum.²

As Baker argues, this explanation for why cicadas have 13 and 17-year life cycles relies on facts about number theory that, intuitively at least, do genuine explanatory work. Contrast the cicada case with Baron’s (2016) train case:

Suppose we want to explain why it is that a train $T$ arrives at a station, $S$, at 3:00 pm. The explanation is as follows: $T$ left another station, $S^*$, 10 kilometres away at 2:00 pm and headed towards $S$ at 10 kph ... Numbers are used to state the distance between stations as well as the speed of the train and a very basic mathematical calculation is deployed, namely: $\frac{10}{10} = 1$. However, the mathematics itself does not do any explanatory work. (Baron, 2016, pp. 459–460)

In this case the mathematics is playing what Melia (2000) calls an indexing role and what Saatsi (2011) calls a representational role. In such cases, mathematics is used to merely represent underlying physical quantities. The mathematics does not add anything, explanation-wise, to our understanding of the target phenomenon. Rather, by representing underlying physical quantities, the mathematics serves as a framework in which the explanandum may be computed.

A theory of extra-mathematical explanation should respect the intuitive difference between cases like the cicada case and cases like the train case. More precisely, it is a constraint on an adequate theory of extra-mathematical explanation that it cleaves the explanatory uses of mathematics from the non-explanatory uses. Call this: the adequacy constraint.

There are three chief respects in which a theory of extra-mathematical explanation may fail to satisfy the adequacy constraint. The first, and most severe, is when the theory simply lacks the capacity to differentiate between the explanatory and non-explanatory uses of mathematics in science; it is blind to the distinction. The second type of failure arises when, despite recognising the distinction between the explanatory and non-explanatory uses of mathematics in science, a theory draws the distinction in such a way that all uses of mathematics in science are deemed explanatory, or none are. While the theory has the capacity to satisfy the adequacy constraint in principle, in practice it fails to be sufficiently sensitive to actual scientific practice.

The third way that a theory might fail to satisfy the adequacy constraint is weaker still. Suppose that a theory can recognise a difference between the explanatory and non-explanatory uses of mathematics in science and, moreover, that the theory does not classify all uses of mathematics in science as explanatory or as non-explanatory. Suppose, however, that the theory mistakenly classifies some uses of mathematics in science as explanatory, or mistakenly classifies some uses of mathematics as non-explanatory. Then the theory fails to fully satisfy the adequacy constraint: it does not fit the target phenomenon.

The degree to which a theory fails to satisfy the adequacy constraint in this third sense is proportional to the number of cases that it misclassifies. Some failures of this third kind are to be expected. The target phenomenon is delineated, in part, by intuitions about which uses of mathematics in science are explanatory and which are not. We shouldn’t expect all of these intuitions to be satisfied by our theoretical ambitions. But we also shouldn’t give up on these intuitions too easily, or else we risk losing a grip on the target phenomenon. The best theory of extra-mathematical explanation correctly classifies every use of mathematics that seems to be explanatory or non-explanatory as such. We should, however, be open to theories that are capable of capturing most but not all cases. Such theories should be accompanied with an explanation as to why our intuitions about a given case are mistaken.

In what follows I will use the combination of the cicada case and the train case as a litmus test for the satisfaction of the adequacy constraint. When a theory cannot properly classify these cases it should be modified in some way. Nothing hangs on these particular examples. The train case could easily be swapped for one of the other examples of purely representational mathematics available. Similarly, the cicada case could be traded for a different case of extra-mathematical explanation. These two cases are, however, broadly indicative of the explanatory and non-explanatory uses of mathematics in science and so lessons learned from these cases can be generalised.

3. A Basic Counterfactual Theory

With the adequacy constraint in hand, I turn now to the task of theory construction. A theory of extra-mathematical explanation is an account of the relation $R$ in virtue of which mathematical facts explain physical ones. There have been a number of attempts to develop such a theory, with little consensus on what the correct theory might be (see, for discussion, Baron et al. (2017); Baron (forthcoming); Jansson and Saatsi (forthcoming); Lange (2016); Lyon (2012); Pincock (forthcoming); Reutlinger (2016); Steiner (1978)). In what follows, however, I will focus on the counterfactual approach developed by Baron et al. (2017). The motivation for doing so is theoretical unification. One of the most successful theories of scientific explanation to date is the counterfactual theory developed by Woodward (2003) and Woodward and Hitchcock (2003). As Nickel (2010) has argued the development of a monistic theory of explanation—a theory that can handle explanation wherever it arises—is a very attractive prospect. The extension of the counterfactual theory of scientific explanation to mathematical cases goes some way toward meeting these monistic ambitions.

Now, Baron, Colyvan and Ripley (2017) never actually state a counterfactual theory of extra-mathematical explanation. They are focused on clearing the way for such a theory by showing how to model extra-mathematical explanations in counterfactual terms. We can, however, state a basic counterfactual theory of extra-mathematical explanation as follows:
A mathematical fact M explains a physical fact P just when if M had not been the case, P would not have been the case.

The basic counterfactual theory of explanation requires counterfactuals with mathematically impossible antecedents. Some philosophers believe that all such counterfactuals are trivially true. I don’t buy that, but I am not going to defend such a claim here. Many such defenses are available. Instead, I will simply assume that there is good sense to be made of such counterfactuals and press on.

While I am not going to defend the non-triviality of counterfactuals with mathematically impossible antecedents, it will be useful in what follows to have a decision procedure for thinking through counterfactuals of this kind. To be clear the procedure is not a full-blown semantics for counterfactuals (viable semantic theories are available, see Bjerring (2014); Brogaard and Salerno (2013)). The procedure is an imaginative procedure by which a reasonable determination of the truth-value of a counterfactual can be made.

The decision procedure I prefer is the one that Baron, Colyvan and Ripley (2017) propose. According to this method, the evaluation of a counterfactual proceeds in three stages. First, we make a judgement about what to hold fixed under counterfactual variation. The facts that we hold fixed are the facts that we take to be true in the scenarios against which the counterfactual is evaluated. Second, of the scenarios picked out by the facts that we are holding fixed, we focus on only those scenarios in which the antecedent of the counterfactual is true. Baron, Colyvan and Ripley (2017) call this stage making a ‘twiddle’. Third, and finally, we ramify. That is, we consider the scenarios that are isolated by the first two stages and then reason through the ‘free facts’ – the facts we are not holding fixed – based on the twiddle that we have made, and based on general principles being held fixed (such as laws or mathematical truths). If the consequent of the counterfactual appears to be true in the relevant scenarios, then we should believe that the counterfactual is true.

According to Baron, Colyvan and Ripley (2017, pp. 9–10), we determine what to hold fixed based on the content of a given counterfactual and the context of evaluation. First, we hold fixed as much of the physical system mentioned in the counterfactual as possible, and as much as we can about the physical world more generally. This includes holding fixed the broad structure of the physical system under consideration, along with the laws of nature and the past up until the explanandum occurs (but not the future). With respect to any mathematical features mentioned in the antecedent, we hold as much of mathematics fixed as possible while allowing enough slack for the twiddle (see the procedure at the end of this section). If there is more than one way to hold fixed in mathematics features which are compatible with realising the antecedent, we let the context of evaluation and the interests of the evaluator select a preferred method. Finally, we hold fixed a representation relation between the mathematics mentioned in the antecedent of the counterfactual, and any physical features that the mathematics represents. This is to provide a ‘bridge’ across which mathematical twiddles can ramify into physical systems (Baron et al., 2017, p. 9).

When it comes to mathematics, it can be difficult to conceptualise the twiddles that we are making. For instance, suppose we want to twiddle 13 by supposing that 13 has 2 and 6 as factors. Such a thing, one might worry, is simply unimaginable. However, Baron, 3

---

3 See Baron et al. (2017); Berto et al. (2018); Bjerring (2014); Brogaard and Salerno (2013); Kment (2014); Mares (1997); Nolan (1997); Restall (1997).
Colyvan and Ripley offer a handy way to imagine twiddles of this kind. We can imagine such twiddles by imagining a relation – multiplication* – which is just like multiplication except when it comes to 2, 6 and 13. Baron, Colyvan and Ripley model the difference between multiplication and multiplication* by graphing it (below). Such graphs play a useful role in scaffolding the imagination. When we twiddle multiplication what we do is imagine a mathematically impossible situation in which the behaviour of multiplication has been changed to match the behaviour of multiplication*.

One might remain worried about twiddling mathematics. Consider the counterfactual that if 1 and 1 had not been 2, then 2 and 1 would not have been 3. In order to realise the antecedent of this counterfactual, we need to suppose that 1 and 1 is not 2. But if 1 and 1 is not 2, then 1 and 1 and 2 is not 4. But 1 and 1 and 2 is 4. So we have a contradiction.

The contradiction is produced by holding too much fixed. If we hold fixed that 1 and 1 and 2 is 4 while, at the same time, twiddling that 1 and 1 is 2, then inconsistency is the result. But we can hold too much fixed for any counterfactual. For instance, suppose that I am wearing a green jumper and that, as a result, I look fine. Suppose that the green jumper is the only item of green clothing that I am wearing. If I had not worn a green jumper, I would not have looked fine. One way that I can evaluate this counterfactual is to hold fixed the clothes that I am wearing aside from the green jumper, and to also hold fixed the fact that I am wearing something green. I could then consider a scenario in which I am wearing a blue jumper. In this scenario, I am wearing a green item of clothing since my clothing is just like it is in the actual world in all non-jumper respects, and yet I am not wearing any item of green clothing, because my green jumper is not blue. That’s a contradiction. I shouldn’t have held fixed both that I am wearing an item of green clothing and that I am wearing what I am wearing actually in all non-jumper respects. So too, in the mathematical case, I shouldn’t have held fixed that 1 and 1 and 2 is 4.

Baron, Ripley and Colyvan (2017) suggest a general method for handling mathematical cases. First, try to find a consistent way of managing the twiddle. To do this, make a decision about what to hold fixed and then twiddle. If a contradiction results, then hold less of the mathematics fixed and then re-twiddle. If a contradiction arises, hold less fixed and twiddle again and so on. Do this until contradictions are cleared from the neighbourhood.
of the counterfactual being evaluated. Contradictions are cleared from the neighbourhood when they have been pushed back far enough into the mathematical structure at issue that the twiddle can be made consistently within the remaining fragment of mathematics.

Of course, when merely pushing contradictions out of the neighbourhood of the counterfactual being evaluated the rest of mathematics may be inconsistent. The point, though, is that we don’t care about all of mathematics. We are only interested in the consistent fragment that we have managed to isolate, since that is sufficient to work through the consequences of a given counterfactual. To handle the remaining inconsistency, we may imagine that the scenarios against which the counterfactual is evaluated are closed under some non-classical consequence relation (such as the consequence relation of LP, see Priest (2002)).

4. Trouble for the Basic Theory

So far, I have outlined a basic counterfactual theory of extra-mathematical explanation and have described a strategy for evaluating counterfactuals with mathematically impossible antecedents. In this section I will argue that the basic counterfactual theory of extra-mathematical explanation fails to satisfy the adequacy constraint outlined in §2.

To see the problem, we need to formulate the cicada case and the train case in counterfactual terms. Let’s start with the cicada case. For this case, we need to look at a counterfactual in which the antecedent is a mathematical fact concerning co-primeness and the consequent is the explanandum: the 13 and 17-year life-cycle lengths of the cicadas. This yields:

\[
\text{[CF1]} \quad \text{If 13 and 17 had not been co-prime with } 2, 3, 4, 5, 6, 7, 8 \text{ or } 9, \text{ then}
\]

North American cicadas would not have had 13 or 17-year life-cycles.

To evaluate [CF1] we begin by making a judgement about what to hold fixed. In this case, we hold fixed (i) the fact that the North American magicicadas are being predated by predators with periodical life-cycles; (ii) facts about the laws of nature including the laws of evolution and natural selection; (iii) facts about the past up until the point at which the cicadas begin to be predated by predators with periodical life-cycles; (iv) ecological facts that constrain the cicadas to have life-cycles in the 12–18-year range; (v) ecological facts that constrain the predators to have life-cycles in the 2–9-year range; (vi) the fact that there are two intervals of years \( L \) and \( L^* \) represented by the numbers 13 and 17 respectively\(^4\) and (vii) as many mathematical facts as possible compatible with realising the antecedent of the counterfactual.

Next, we perform a twiddle and make it the case that 13 and 17 fail to be co-prime with each number in the 2–9-year range. To do this, we can follow Baron, Colyvan and Ripley and imagine that the behaviour of multiplication matches the behaviour of a different function, call it multiplication\(_1\), which behaves oddly with respect to 13 and 17, but is otherwise the same. So, for instance, whereas multiplication does not allow for there to be any number \( n \) such that \( 2 \times n = 13 \), multiplication\(_1\) allows there to be some such number (and so on for each of 3–9, and mutatis mutandis for 17). In making this twiddle, we make

\(^4\)If we don’t hold this fixed, then twiddling 13 and 17 may have no effect on the way that \( L \) and \( L^* \) intersect with other intervals. But if these intersection properties don’t change, then the cicadas may have the same life-cycle lengths as they do actually (namely \( L \) and \( L^* \)) despite the mathematical twiddle.
it the case that the lowest common multiple of 13 and each number in the range 2–9 is, itself, 13. Similarly, the lowest common multiple of 17 and each number in the range 2–9 is 17. That’s because for each such number \( m \) in the 2–9 range, there is some number \( n \) and some number \( o \) such that \( n \times_1 m = 13 \), and \( n \times_1 o = 17 \).

Having made the twiddle to realise the antecedent of [CF1], we then ramify the consequences of that twiddle for the cicadas. In this situation, neither 13 nor 17 will be the optimal life-cycles for avoiding predation. Rather, 13 and 17 will be bettered by other numbers in the 12–18 range. For instance, 15 will substantially outperform 13 and 17, since it is co-prime with 2, 4, 7 and 8, and so a cicada with a 15-year life-cycle will intersect far less frequently with predators possessing those periodic life-cycles compared to 13 or 17 (a 15-year cicada intersects with the 2s every 30 years, the 4s every 60 years, the 7s every 105 years and the 8s every 120 years, while a 13 or 17-year cicada intersects with all predators every 13 or 17 years respectively). So, given that we are holding fixed the selection pressures that drive the cicadas to avoid predation by evolving toward a particular life-cycle length in the scenario under consideration, the cicadas won’t evolve to have 13 or 17-year life-cycles. Rather, they will evolve some other life-cycle in the 12–18-year range, such as 15.

So [CF1] appears to be true. The problem is that if [CF1] is true then [CF2] below also appears to be true when evaluated in the same way. For a train T that leaves station, S*, at 2:00pm and travels to station, S, 10km away at 10kph:

**[CF2]** If \( \frac{10}{10} \) had not equalled 1, T would not have arrived at S at 3:00pm.

We begin our evaluation of [CF2] by holding fixed in the same manner as for [CF1]. In this case, we hold fixed: (i) the fact that the train T leaves station S* at 2:00pm; (ii) the fact that S is 10km away from S*; (iii) the fact that the train travels at a constant speed of 10kph for its entire journey, (iv) the laws of nature including the laws of motion; (v) facts about the past up until just before the train’s journey; (vi) the fact that the distance between stations and the train’s speed are represented by the numbers 10 and 10 respectively and (vii) as many mathematical facts as we can compatible with realising the antecedent of the counterfactual.

Now, we make a twiddle to bring it about that \( \frac{10}{10} \neq 1 \). There are, broadly, two ways to realise the twiddle. Either there is some number \( n \) such that \( n \neq 1 \) and \( \frac{10}{10} = n \) or there is no such number. Suppose we make the twiddle in the first way, and then ramify. If \( \frac{10}{10} = n \) then the train would have taken \( n \) hours to cover the distance. It follows that the train would not have arrived at S* at 3:00pm, since it would have either taken less than an hour to cover the distance, or more than an hour. That’s because \( n \neq 1 \) and so if the train’s total journey time is speed divided by distance, then given its speed and given the distance it had to travel, it would not have taken one hour. Suppose, on the other hand, that we make the twiddle in the second way, and thus that \( \frac{10}{10} \) does not equal any number. If \( \frac{10}{10} \) does not equal any number, then it does not equal 1. Given that it does not equal 1, the train does not take an hour to get from S to S*, since the train only takes an hour when \( \frac{10}{10} = 1 \). So, the train in this scenario would not have arrived at S* at 3pm (it is unclear when the train would have arrived in this second case, because we can’t assign a value to the time it takes to travel the distance, still we know that the value is not 1, which is enough to reason through the situation). Either way, then, [CF2] appears to be true.

Given that [CF1] and [CF2] both appear to be true, the basic counterfactual theory mistakenly classifies the train case as a case of extra-mathematical explanation. Now, con-
ceivably, one may just accept this consequence and concede that our intuitions are wrong: the train case is a genuine extra-mathematical explanation. However, the basic counterfactual theory does not seem to offer much by way of an explanation for why our intuitions are wrong in this case. Our intuitions are wrong given how we have chosen to understand what an extra-mathematical explanation is. But we have not learnt something new about the train case that might reveal why our intuitions have gone astray.

5. Unity and Counterfactuals

In this section I will outline a modified version of the basic counterfactual theory that properly classifies the cicada case and the train case. The idea, in a nutshell, is that the mathematics in the cicada case is explanatory because it plays a unifying role, but not so for the train case. In order to motivate the broad idea I will look briefly at a different case of explanation that uses the same mathematics as the cicada case.

5.1. Grinding Gears

The case is inspired by a recent explanation outlined by Baker (2017) involving gear ratios in fixed speed bicycles. Gears within mechanical systems usually feature a large gear turning a small gear to produce a gear ratio. It is standard for the number of teeth $n$ on the large gear to be prime-numbered. The explanation for this fact exploits the same basic mathematics as the cicada case. Consider a large gear rotating a small gear. The rotation of a gear whereby every tooth on the gear starts and ends in the same position is a gear’s period. The gear’s period is not a temporal period: it is not the time it takes for the gear to return to its original position. It is, rather, a period constructed of angular rotations of the gear’s teeth. One unit of the gear’s period is a rotation of that gear such that every tooth on the gear moves through an angle $m^\circ$ where $m = \frac{360}{n}$. So, for instance, if a gear has 4 teeth, then one unit of the gear’s period is a rotation of every gear’s teeth through $90^\circ$. The period of the gear is 4, since it will take 4 units of rotation to return the teeth back to their original position. The gear’s period, then, equals the number of that gear’s teeth. Thus, if the number of teeth on a gear is prime-numbered, then the gear also completes a rotation in a prime-numbered period.

Now, when producing gears, it is inevitable that gears will feature small imperfections. Suppose, then, that every large gear has a little bump on some tooth. As the large gear rotates the small gear, that bump will rub against the small gear’s teeth. What we want to avoid, as far as possible, is the bump on the large gear repeatedly hitting the same tooth on the small gear. If this is not prevented, then the tooth on the small gear will wear at a greater rate than the other teeth on the gear. This generates a weakness in the small gear which will eventually cause a break down. A system featuring a large gear turning a small gear is thus in an optimal state when, for any given tooth on the small gear, the frequency with which that tooth impacts the bump on the large gear is minimised.

So, for example, suppose that we have a large gear with 8 teeth turning a small gear with 4 teeth. Suppose that the large gear has a bump on one tooth. Then for every rotation of the large gear (8 turns), the tooth with the bump on it will strike the same tooth on the small gear. Compare this with a case in which the large gear has 5 teeth and the small gear has 4 teeth. In this situation, a bump on a tooth of the large gear will strike the same tooth on the small gear once in every 4 full rotations of the large gear (20 turns). More
generally, for any gears $a$ and $b$, where $a$ is larger and $b$ is smaller and where the number of teeth on $a$, $n$, and the number of teeth on $b$, $m$ is such that $n > m$ an imperfection on $a$ will strike the same tooth $b$ once every $c$ turns, where $c$ is the lowest common multiple of $n$ and $m$. For the 8 and 4 gear system, the imperfection strikes the same tooth after 8 turns, because that’s the lowest common multiple of 4 and 8. Similarly, in the 5 and 4 gear system, the imperfection strikes the same tooth after 20 turns, because that’s the lowest common multiple of 5 and 4.

Now, suppose that a company aims to build the longest lasting engine they can – the L-Engine. An engine is long-lasting when it minimises differential gear wear on small gears within the system. The company is constrained by economic and manufacturing constraints to produce the L-Engine using a range of small gears of between 2 and 9 teeth per gear. The same constraints also restrict the company to a choice of using large gears with between 12 and 18 teeth. The company decides to use large gears with either 13 or 17 teeth. Why?

The answer is that these two gear types minimise wear on the small gears inside the system and so maximise the longevity of the engine. The 17-tooth gear will result in an imperfection on the large gear hitting the same tooth on the small gear on average once every 83.1 rotations across all small gears. For the 13-tooth gear, the imperfection on the large gear strikes the same tooth on the small gear on average once every 63.5 rotations across all small gears. The next best is the 16-tooth gear, for which the imperfection on the large gear strikes the same tooth on the small gear on average once every 53.3 rotations across all small gears. 13 and 17 are optimal in this way because the lowest common multiple of 13 or 17 and numbers in the 2–9 range is typically higher than other nearby numbers. It is because 13 and 17 are co-prime with every number in the 2–9-year range that the lowest common multiple between 13 and 17 and each of 2–9 tends to be higher than for the other numbers in the 12–18 range.

This may seem like an artificial example. However, various pieces of machinery, such as the transmission inside a car or the gears inside a clock, make use of gears with prime-numbered teeth precisely because a prime-numbered amount of teeth produces periods of rotation for large gears that tend to be co-prime with a range of possible small gear rotation periods. Indeed, the use of gears with a prime-numbered amount of teeth is an industry standard. Imperfections are an inevitable consequence of the manufacturing process. Of course, large gears don’t always have 13 and 17 teeth: sometimes the gears are bigger, or smaller. Regardless of size, the larger gears tend to have a number of teeth that is co-prime with the number of teeth on the small gears that they turn.

The cicada case and the gear case share a common number-theoretic basis. The commonality is particularly striking because the two physical systems are driven by different physical laws. The cicada case is driven by evolution and natural selection which produces an evolutionary optimisation problem. The gear case is driven by the laws of mechanics plus various economic constraints, which produces a mechanical optimisation problem. In what follows, I will introduce elements of unification into the basic counterfactual theory of extra-mathematical explanation to capture the unity just described. The upshot will be that the cicada case can be cleaved from the train case.
5.2 Counterfactual Schemes

We begin with the concept of a counterfactual scheme. A counterfactual scheme is similar to one of Kitcher’s (1989) argument schemes and has three components. First, it features a generalised counterfactual which is produced by taking a counterfactual such as [CF1] and replacing some or all of the non-logical expressions within the counterfactual with variables. Second, a counterfactual scheme features a set of filling instructions, which explain what can be substituted back into the variables to produce an instance of the counterfactual. Third, a counterfactual scheme features a classification, which explains how an instance of the scheme is to be evaluated. The classification for all of the schemes I will consider is the same: use the Baron, Colyvan and Ripley method of evaluation in combination with a counterfactual semantics. Because the classification is fixed in what follows, I will take it as read for each counterfactual scheme.

Next, we define the notion of an ‘explanatory counterfactual’ as follows:

A counterfactual [CF] featuring a mathematically impossible antecedent is explanatory when [CF] is an instance of a counterfactual scheme [CS] such that:

1. All of the instances of [CS] are true.
2. For at least two instances of [CS], [CF1] and [CF2], [CF1] and [CF2] are nomically distinct.

[CF1] and [CF2] are nomically distinct when the physical laws relevant to the evaluation of those counterfactuals are different. The physical laws needed for the evaluation of a counterfactual are implicated in the ramification stage. When evaluating a counterfactual we ramify a twiddle via some of the laws of nature that we are holding fixed. So, for instance, we might ramify a particular twiddle by rolling the laws of mechanics forward after the twiddle has been made. The train case is an example of this: the laws of mechanics are used to work out what would happen to the train post-twiddle. Alternatively, we might ramify a particular twiddle by rolling the laws of evolution and natural selection forward, which is what we do in the cicada case. Not every law is relevant to ramifying a twiddle; different laws are called for depending on the counterfactual.

By enforcing nomic distinctness, the unification delivered by the model is the kind of unification that Baker (2017, pp. 201–207) associates with topic generality. The core of an explanation displays topic generality when it is not about a designated class of physical facts, allowing the core to apply across a range of such classes. If we designate classes of physical facts via the laws that govern those facts, then explanatory counterfactuals are instances of topic general schemes. Importantly, the topic generality of the schemes and their ability to unify is delivered by their mathematical core.

The first clause in my account of an explanatory counterfactual requires the instances of a counterfactual scheme to be true. These instances can arise within actual physical systems, or within potential physical systems. The scheme thus unifies both actual and physically possible cases of explanation. The second clause makes unity into a necessary condition on explanatory counterfactuals. Explanatory counterfactuals are instances of counterfactual schemes that extend across distinct physical systems. The first clause makes the unification featured in the second clause harder to come by. As a counterfactual scheme is expanded to try and handle nomically distinct explananda, there is a risk of capturing counterfactuals that are not true for the nomically distinct explananda at issue.
The degree to which a counterfactual is explanatory is proportional to the number of nomically distinct counterfactuals found among the instances of its associated counterfactual scheme. The best explanations are the ones that correspond to those counterfactuals possessing the highest degree of explanatoriness. The trade-off between the first and second clauses therefore comes to play a larger role when we are seeking the best explanations. The more we try to generalise a counterfactual scheme to handle a greater range of nomically distinct counterfactuals, the harder it is to ensure that every instance of a counterfactual scheme is true. The more a counterfactual scheme is narrowed by ruling out nomically distinct counterfactuals, the easier it will be to ensure that every instance is true, but the less explanatory the scheme will be.

The above account of an explanatory counterfactual can be used to modify the basic counterfactual theory of extra-mathematical explanation to produce the basic u-counterfactual theory:

[The Basic U-Counterfactual Theory]
A mathematical fact M explains a physical fact P just when the counterfactual ‘if M had not been the case, P would not have been the case’ is an explanatory counterfactual.

5.3. Application
To apply the basic u-counterfactual theory, the cicada case needs to be situated within a counterfactual scheme. One such scheme is this:

[CS1] If \( x_1, x_n \) had not been co-prime with \( y_1, y_2 \ldots \) or \( y_n \), the \( p_1, p_n \) would not have had \( x_n U Cs \).

Filling instructions:

1. The \( p_n \) are periodical phenomena within any actual or physically possible system \( S \) that is under pressure to optimise some feature and where that feature is optimised just when for periodical phenomena \( p*1, p*n \) that are in \( S \) and that are distinct to the \( p_n \), the frequency of intersection between the \( p_n \) and the \( p*n \) is minimised.
2. The \( x_n \) are numbers that are bijectively mapped to the \( p_n \).
3. The \( y_n \) are numbers that are bijectively mapped to the \( p*n \).
4. \( U \) is the unit of the \( p_n \) (e.g., years).
5. The \( Cs \) are the type of period that characterises the \( p_n \) (e.g., life-cycles).

To get back to [CF1] from [CS1], let \( p_n \) be the life-cycles of the North American magicicada; let \( x_n \) be numbers that represent the life-cycle lengths of the cicadas; let \( y_n \) be numbers that represent the life-cycle lengths of the predators of the magicicada, let \( U \) be years and let \( C \) be life-cycles.

[CS1] also has an instance in the gear case. An L-Engine is a physical system that is under pressure to reach an optimal state of longevity. The L-Engine reaches that optimal state when the differential wear on the small gears in that system is minimised. The optimal state is produced by minimising the frequency of the intersection of the large gears with the...
rotation periods of the small gears. The rotation of the large gears, then, can be substituted in for the $p_n$. Let $x_n$ represent the rotational periods of the large gears and let $y_n$ represent the rotational periods of the small gears. Let $U$ be a period of a gear and let $C$ be a rotation of a gear. This yields:

$$[\text{CF3}] \text{ If 13 and 17 had not been co-prime with 2, 3, 5, 6, 7, 8 or 9, the large gears in an L-Engine would not have had 13 or 17 period rotations.}$$

[CF3] is true for the same reasons as [CF1]. To evaluate the counterfactual, we first make a judgement about what to hold fixed. In this case we hold fixed (i) the fact that the large gears are part of an L-Engine being produced by a company aiming to optimise longevity; (ii) facts about the laws of nature including laws of mechanics; (iii) facts about the past up to the point where the production of the L-Engine begins; (iv) economic and mechanical constraints that restrict the large gears inside the L-Engine to lie within the 12–18 tooth range; (v) economic and mechanical constraints that restrict the small gears inside the L-Engine to lie within the 2–9 tooth range; (vi) the fact that there are gears with total amounts of teeth $T$ and $T*$ represented by the numbers 13 and 17 respectively and (vii) as many mathematical facts as possible compatible with realising the antecedent of the counterfactual.

The twiddle on 13 and 17 for [CF3] is performed in the same manner as for [CF1]: by considering a situation in which multiplication behaves like multiplication. Having made the twiddle to realise the antecedent of [CF3], we then ramify the consequences of that twiddle for the gears. In this situation, neither 13 nor 17 will be the optimal rotational periods for avoiding differential wear on small gears. For large gears with 13 teeth, an imperfection on those gears will now strike the same tooth on a small gear on average once every 13 rotations across all small gear types. For large gears with 17 teeth, an imperfection will strike the same tooth on a small gear on average once every 17 rotations. Accordingly, a 16-tooth gear will now be better than either 13 or 17-toothed gears. So, given that we are holding fixed the constraints surrounding the production of L-Engines and given that we are holding fixed the fact that the company is trying to produce L-Engines, it follows that the company won’t produce L-Engines with large gears possessing 13 or 17 teeth. So, the large gears inside L-Engines won’t have 13 or 17 period rotations.

Now, importantly, [CF1] and [CF3] are nomically distinct. The evaluation of [CF1] proceeds via the laws of evolution and natural selection; the evaluation of [CF3] proceeds via the laws of mechanics. So, the second clause in my account of what it is to be an explanatory counterfactual is satisfied by both counterfactuals. In order to satisfy the first clause, [CS1] must have only true instances. Given the filling instructions for [CS1], instances of that scheme only arise within actual or physically possible systems that are under pressure to reach an optimal state, where that optimal state is reached by minimising the frequency of intersection between periodical phenomena within that system. In such systems, prime periods will be required to reach optimality, and so the periodical phenomena in those systems will behave just like cicada life-cycles and gear rotations. The same reasoning that I have used to demonstrate the truth of [CF1] and [CF3] will therefore carry over to other instances of [CS1] that arise. So, there is good reason to suppose that every such instance of [CS1] will be true.

Indeed, for an instance of [CF1] to be false within some actual or physically possible system $S$, there would need to be a relevant disanalogy between $S$ and the physical make-up of the cicada and gear cases. Whatever that disanalogy is, however, it will be possible
to write an exception for that kind of case into the filling instructions for \([CS1]\). Doing so will preserve the truth of \([CF1]\) and \([CF3]\), since the putative disanalogy does not apply to those cases. At worst, then, we end up specify the filling instructions so tightly that only the cicada and gear cases qualify, and thus only \([CF1]\) and \([CF3]\) are instances of \([CS1]\). But that’s fine. In this situation, \([CS1]\) still has only true instances, and still has instances that are governed by distinct physical laws, which is sufficient to qualify \([CF1]\) and \([CF3]\) as explanatory counterfactuals.

5.4. Trains

We return now to the train case. Like the cicada case, the train case can be situated within a counterfactual scheme:

\[\text{[CS2]} \text{ If } \frac{x}{y} \text{ had not equalled } z, A \text{ would not have arrived at } B* \text{ at } t.\]

**Filling instructions:**

1. \(A\) is any actual or physically possible object in motion which leaves a location \(B\) at time \(t_1\) with velocity \(v\) and travels distance \(d\) to a location \(B*\).
2. \(B*\) is any location.
3. \(x\) is any number that represents \(A\)’s velocity.
4. \(y\) is any number that represents the distance that \(A\) travels.
5. \(z\) is any number such that \(t - t_1 = z\).

We can get from \([CS2]\) to \([CF2]\) as follows. Let \(A\) by the train that leaves station \(S\) at 2:00pm and travels 10km to station \(S*\) at 10kph; let \(B*\) be station \(S*\), let \(x\) be 10, let \(y\) be 10 and let \(z\) be 1.

As with \([CS1]\), \([CS2]\) appears to have other true instances. For example, suppose that Suzy’s car leaves Suzy’s house at 8pm travelling 60km/h and arrives at Sara’s house, which is 120km away, at 10pm. This gives us a second instance of \([CS2]\), namely:

\[\text{[CF4]} \text{ If } \frac{120}{60} \text{ had not equalled } 2, \text{ Suzy’s car would not have arrived at Sara’s house at 10pm.}\]

\([CF4]\) is true for the same reasons that \([CF2]\) is: if we hold fixed the distance travelled by the car, the car’s velocity, the departure time, the fact that the distance and speed are represented by 120 and 60 respectively, the laws of nature and the past up until the point of departure, we can then twiddle the mathematics of fractions by bringing it about that \(\frac{120}{60} \neq 2\). If we do this, then no matter what \(\frac{120}{60}\) does equal, Suzy’s car won’t arrive at Sara’s house at 10pm.

The filling instructions for \([CS2]\) isolate only those systems that are relevantly like the physical system featuring a train. It seems plausible, then, that the same reasoning that applies to \([CF2]\) and \([CF4]\) will carry over to each instance of \([CS2]\). Indeed, as in the cicada and gear cases, if the reasoning does not carry over—if one of the instances of \([CS2]\) is false—then that will be because there is some actual or physically possible system that satisfies the filling instructions but that behaves in a manner that is disanalogous to the
train case. Whatever that disanalogy is, we can just write it into the filling instructions for [CS2] to generate an exception for the relevant case. By proceeding in this fashion, we can further narrow [CS2] so that it only has true instances.

So far, the two counterfactual schemes [CS1] and [CS2] appear to be symmetrical. The central difference between the two schemes, however, is that [CS2] does not include at least two instances that are governed by distinct physical laws. Every instance of [CS2] involves an object in motion and a calculation regarding the time it takes the object to get wherever it is going. Accordingly, every instance of [CS2] is evaluated using the same physical laws, namely the laws of mechanics.

In order for [CS2] to possess instances that aren’t all evaluated via the laws of mechanics, the core notions of motion, velocity and distance that appear in the counterfactual scheme need to be generalised. For it is these notions that force the evaluation of any instance of [CS2] to proceed via the laws of mechanics. What we need, in particular, is three broader physical concepts such that motion, velocity and distance are instances of those concepts, and where phenomena that are not purely mechanical can be instances of those concepts as well. Moreover, these concepts cannot be so broad that false instances of [CS1] arise. Whatever those concepts are, the actual or physically possible realisers of those concepts must behave in a manner that is similar enough to motion, velocity and distance that manipulating the mathematics of fractions carries implications for those realisers.

For my part, I cannot see how to generalise [CS2] in the required manner so that [CF2] continues to be an instance of the generalised scheme. There are no physical concepts that fit the bill. It follows, then, that [CF2] is not an explanatory counterfactual, because there is no counterfactual scheme that has [CF2] as an instance, that has only true instances, and that has at least two instances that are evaluated via distinct physical laws. The symmetry between [CF1] and [CF2] is broken.

6. Exchange Rates

This victory is short-lived, however. For there is a counterfactual in the neighbourhood of [CF2] that, unlike [CF2], is an instance of a counterfactual scheme that has nomically distinct instances but, like [CF2], involves counterfactual dependence between mathematical and physical facts. Handling this further case will force the revision of the basic u-counterfactual theory into its final form.

Consider the following scenario. Suppose that Suzy is putting fuel into her car. At the end of the process, Suzy has 70 litres in her tank. Why? The answer is that Suzy started with 20 litres, and then exchanged $50 for fuel at a rate of $1 per litre. Since \( \frac{50}{1} = 50 \), it follows that Suzy ends up with 70 litres in the tank. The mathematics involved in this case is the same as the mathematics involved in the train case. It is thus possible to formulate a counterfactual scheme that has instances in both the train case and the refuelling case. Here it is:

[CS3] If \( \frac{z}{y} \) had not equalled \( z \) then \( c \) would not have ended at \( B^* \).

Filling instructions:

1. \( c \) is any actual or physically possible process involving a change in two

---

\(^5\)I am grateful to Ray Briggs for this exchange-rate example.
metric quantities $Q$ and $Q^*$ at a rate of \( \frac{a}{b} \) with a fixed start point $A$ in metric quantity $Q$ and $B$ in metric quantity $Q^*$.

2. $B^*$ is a fixed end point of change $c$ in metric quantity $Q^*$.

3. $x$ is any number that represents the magnitude of $A$.

4. $z$ is any number such that $B^* - B = z$; $z$ represents the difference in the start and end points of change $c$ in the metric quantity $Q^*$.

5. $y$ is any number such that $y = \frac{a}{b}$.

Using these filling instructions, we can get two counterfactuals. First, for the refuelling case, let $c$ be the process of Suzy’s refuelling in which fuel is exchanged at $1$ per litre, and in which Suzy starts with $50$ and $20$ litres in the tank; let $B^*$ be the end point of $70$ litres; let $x$ be $50$; let $z$ be $50$, and let $y$ be $\frac{1}{1} = 1$. This yields:

[CF5] If $\frac{50}{1}$ had not equalled $50$ then Suzy’s refuelling of her car would not have ended at $70$ litres.

Next, for the train case, let $c$ be the train’s journey from $S$ to $S^*$ in which the train travels at $10$kph, and in which the train starts at $2:00$pm and has a distance of $10$km to travel; let $B^*$ be the end point of the journey at $3:00$pm; let $x$ be $10$; let $z$ be $1$ and let $y$ be $\frac{10}{1} = 1$. This gives us:

[CF6] If $\frac{10}{10}$ had not equalled $1$ then T’s journey would not have ended at $3:00$pm.

The reasoning that results in [CF2] being true works just as well for [CF5] and [CF6] (I’ll spare the reader the details). [CS3] also appears to have only true instances and so the first clause in my account of an explanatory counterfactual is satisfied. But [CS3] also has two instances that are evaluated using distinct physical laws. [CF6] is evaluated using the laws of mechanics, whereas [CF5] is evaluated using principles of economics and laws governing fluids. And so the second clause is also met. So [CF5] and [CF6] are explanatory counterfactuals. So, the train case is classed as a case of extra-mathematical explanation by the basic u-counterfactual theory after all.

As before, one might be willing to accept this result. By reframing the train case in terms of exchange rates, and unifying it with other rate-based cases, it might be thought that we have identified a new case of extra-mathematical explanation. Our intuitions about the train case were, in the first instance, simply wrong. Moreover, the basic u-counterfactual theory has an explanation for why our intuitions were wrong about this case: we were not taking into account the unifying power of the underlying mathematics, and so failed to see the explanatory work that the mathematics was doing.

I am not satisfied with this response, however. The train case continues to seem non-explanatory to me. And so, I recommend a further modification to the basic u-counterfactual theory. The modification is based on the following observation. The unification of the cicada case and the gear case within counterfactual scheme [CS1] is due entirely to the mathematics, and not due to any underlying physical correlate. The unification of the train case and the refuelling case within the counterfactual scheme [CS3], by
contrast, is due to an underlying physical correlate that is being manipulated via the mathematics. This difference in the way that the mathematics achieves unification underwrites the difference in explanatoriness between the two cases.

To make the idea precise, I need to adjust my account of what is to be an explanatory counterfactual as follows:

A counterfactual [CF] featuring a mathematically impossible antecedent is explanatory when:

(i) [CF] is an instance of a counterfactual scheme [CS] such that:

1. All of the instances of [CS] are true.
2. For at least two instances of [CS], [CF1] and [CF2], [CF1] and [CF2] are nomically distinct.

And (ii) there is no other counterfactual scheme [CS*] such that:

1. All of the instances of [CS*] are true.
2. For each instance of [CS] with consequents \(c_1\ldots c_n\) there is a true instance of [CS*] with exactly that consequent.
3. For each instance of [CS*] none of the antecedents of those instances involve a mathematical impossibility.
4. Each instance of [CS] is true because the mathematical twiddles that realise each counterfactual’s antecedent change the physical features in [CS*] that are responsible for unification in that scheme.

In short: a counterfactual with a mathematically impossible antecedent belonging to a given counterfactual scheme fails to be explanatory when there is some other counterfactual scheme that can do the same unifying work as the mathematical one but without appealing to any mathematical facts and, moreover, instances of the mathematical scheme are true because twiddling the mathematics changes the unifying physical feature in the non-mathematical scheme.

Given this modification [CF5] and [CF6] are not explanatory counterfactuals. That’s because there is another counterfactual scheme that satisfies the conditions for being a [CS*] scheme laid out above. The scheme is this:

[CS4] If the rate of \(c\) had not been \(R\) then \(c\) would not have ended at \(B*\).

**Filling instructions:**

1. \(c\) is any actual or physically possible process involving a change in two metric quantities \(Q\) and \(Q*\) with a fixed start point \(A\) in metric quantity \(Q\) and \(B\) in metric quantity \(Q*\).
2. \(B*\) is a fixed end point of change \(c\) in metric quantity \(Q*\).
3. \(R\) is the rate of change between \(Q\) and \(Q*\).
[CS4] yields the following counterfactuals for the train case and the refuelling case (this is easy enough to see, I won’t go through the details):

[CF7] If the rate of exchange for Suzy’s refuelling of her car had not been $1 per litre, then Suzy’s refuelling of her car would not have ended at 70 litres.

[CF8] If the velocity of the train across T’s journey had not been 10kph, then T’s journey would not have ended at 3:00pm.

[CF7] and [CF8] are both true. Consider [CF7]. If the rate of exchange for Suzy had not been a dollar a litre, then she would have been able to put more into the tank or less. Either way, she wouldn’t have ended up with 70 litres (assuming that we hold fixed the laws of nature, any economic principles about exchange, facts about Suzy’s car and facts about the past up to the point at which Suzy begins to refuel). Similar considerations apply to [CF8]. If the train had been travelling at a different velocity, then it would not have ended its journey at 3:00pm. It would have arrived earlier than 3:00pm or later (assuming, again, that we hold fixed the laws of nature along with facts about the train, the stations it is travelling between, the distance between stations, and facts about the past up to the point at which the train begins its journey).

Are all of the instances of [CS4] true? I believe so. Each instance of [CS4] involves a case in which something is exchanged at some rate. When the rate of exchange is changed, then it follows that the end point of the exchange will be different, assuming that we hold fixed the same kinds of facts in each case. The rate of exchange of a given physical process should invariably make a difference to the outcome of the exchange.

Is the consequent of each instance of [CS3] a consequent of some instance of [CS4]? Again, the answer is ‘yes’. [CS4] and [CS3] are really doing the same thing. In both cases, we look at a process involving a rate of exchange and then manipulate the rate. In the case of [CS3] we manipulate the rate by altering the mathematics of fractions. In the case of [CS4] we manipulate the rate by switching to a different rate of exchange. The consequence of each imagined manipulation is, however, the same: we end up altering the end-point of the exchange process. Thus, for any consequent in [CS3], whereby a rate of exchange has a different end point because of some manipulation to the mathematics of fractions, there will be a way to produce that same consequent by simply changing the rate of exchange to a different rate, which is what happens in [CS4].

The mathematical twiddles in [CS3] thus have physical correlates: we can achieve the same outcomes for a physical system by changing the rate of exchange, rather than the mathematics. More than this, however, instances of [CS3] are true because twiddling the mathematics in [CS3] is a way of manipulating the exchange rate in [CS4]. For instance, consider [CF5]. If the exchange rate of fuel does not change when we twiddle the mathematics in [CF5] then, in the counterfactual scenario we end up considering, Suzy will pump the same amount of fuel into her car as in the actual situation, despite the fact that \( \frac{50}{1} \neq 50 \). That’s because, when evaluating this counterfactual, we hold fixed the amount of fuel that Suzy pumps, the amount of money she has and the amount of fuel already in the car (otherwise the counterfactual will be false anyway, since we can simply adjust these other quantities in the counterfactual scenario in which \( \frac{50}{1} \neq 50 \) so that Suzy ends up with 70 litres in the tank). Similarly, if the rate at which distance is exchanged for time does not alter when twiddling the antecedent of [CF3], then even if \( \frac{10}{10} \neq 1 \), the train would have had
the same velocity and so would have arrived at 3pm (assuming that we hold the distance and speed fixed, which, as before, we must to make the counterfactual true). The exchange rate must twist and bend with the mathematics in order to yield the desired outcome.

Accordingly, [CS4] satisfies the conditions for being a [CS*]-type scheme with respect to [CS3]. The mathematics in the train case is therefore not explanatory because the unifying power of [CS3] is not solely due to the mathematics. It is, rather, due to an underlying physical correlate – an exchange rate – that we can get at by twiddling the mathematics.

What of [CS1]? There is no general physical twiddle that we can make to both the cicada system and the L-Engine system that would have the same upshot for both cases as the one produced by altering the co-primeness of 13 and 17. Of course, there are some local physical twiddles that might have the same outcome. For instance, if there had been no predators in the cicada system, then perhaps North American cicadas would not have had 13 or 17-year life-cycles. This counterfactual involves removing the selection pressure that produces an evolutionary optimisation problem for the cicadas. Accordingly, the counterfactual cannot be generalised in such a way that it has an instance in the L-Engine case. The pressures that force an optimisation problem to arise in that case are too different.

But keep in mind that even if some physical correlate can be found that unifies the gear and cicada cases together under a common non-mathematical scheme, it does not follow that the mathematics underlying the cicada case lacks explanatory power. For it must also be shown that instances of [CS1] are true because their corresponding mathematical twiddles alter the unifying feature of the non-mathematical scheme. So far as I know, there is no physical correlate in the cicada case that can play this more specific role. The symmetry between [CF1] and [CF6] is therefore broken. [CF1] is explanatory because the unifying power of [CS1] is not parasitic on any non-mathematical scheme. [CF6] is not explanatory because the unifying power of [CS3] is parasitic.

7. Conclusion

It is time to take stock. We began with the observation that a basic counterfactual theory of extra-mathematical explanation misclassifies some cases of explanation as extra-mathematical ones. I then revised the basic counterfactual theory in two stages.

First, I combined the basic counterfactual theory with elements of a unificationist theory of explanation. According to the basic u-counterfactual theory, a mathematical fact $M$ explains a physical fact $P$ when ‘if $M$ had not been the case $P$ would not have been the case’ is an explanatory counterfactual. An explanatory counterfactual is one that is an instance of a counterfactual scheme that unifies different kinds of actual and physically possible systems together under a common banner.

Second, I modified the basic u-counterfactual theory by refining the notion of an explanatory counterfactual. According to this refinement, a counterfactual with a mathematically impossible antecedent is explanatory when it is (i) an instance of a unifying counterfactual scheme and (ii) there is no other counterfactual scheme that can do the same unifying work without mathematics and is such that instances of the mathematical scheme are true because twiddling the mathematics changes the unifying feature of the non-mathematical scheme.

Call the final view simply: the u-counterfactual theory of extra-mathematical explanation. The u-counterfactual theory is complex and so one might worry that the benefits bought by such complexity are not worth the cost. Though I cannot fully argue the point
here, I believe that the cost is worth paying. For, as noted in §3, a counterfactual approach to extra-mathematical explanation opens up the enticing possibility of developing a theory of scientific explanation that is fully general. Of course there is work to be done. The view needs to be actually integrated with a general counterfactual theory of scientific explanation to demonstrate the benefit that I am claiming for it. In addition, further testing of the theory against a wider variety of cases is needed in order to fully demonstrate the theory’s generality. Nonetheless, the u-counterfactual theory presents a promising way forward.⁶

REFERENCES

— — 2009, ‘Mathematical Explanation in Science’ in British Journal for the Philosophy of Science 60, 611–633


Baron, Sam 2016, ‘Explaining Mathematical Explanation’ in Philosophical Quarterly 66, 458–480
— — forthcoming, ‘Mathematical Explanation By Law’ in British Journal for the Philosophy of Science.

Baron, Sam, Mark Colyvan, and David Ripley 2017, ‘How Mathematics Can Make a Difference’ in Philosophers’ Imprint 17.


Brogaard, Berit and Joe Salerno 2013, ‘Remarks on Counterpossibles’ in Synthese 190, 639–660

⁶I am very grateful to Jamin Asay, Ray Briggs, Tom Donaldson and Jennifer Wang for discussion of earlier drafts of this paper. I am also grateful to audiences at the Hong Kong Metaphysics Workshop in 2018, the Australian Metaphysics Workshop at Kioloa in 2018, and the Australasian Association of Philosophy Conference in Wellington 2018 for comments and discussion. Research on this paper was partly supported by an ARC Discovery Early Career Researcher Award DE180100414.


— — 2010, ‘There is No Easy Road to Nominalism’ in *Mind* 119, 285–306


Nickel, Bernhard 2010, ‘How General Do Theories of Explanation Need To Be?’ in *Noûs* 44, 305–328


