The Ins and Outs of Mathematical Explanation

Mark Colyvan

Gauss referred to mathematics as the queen of sciences. His remark was intended to flag the privileged and exalted position that mathematics occupies within the sciences. And just as royalty is set apart from its subjects, so too is mathematics set apart from the rest of science. Mathematics is certain, its results stand for all time, and it proceeds by a priori methods. The rest of science, by contrast, is uncertain, fallible, and a posteriori. In what follows, I'll highlight another interesting difference between mathematics and the rest of science. This difference concerns the way explanation operates in mathematics and the surprising way mathematical explanations can feature in broader scientific explanations.

First a bit of background on scientific explanation. There are a number of important questions that naturally arise about explanation. When is an explanation required? How do we recognise explanations as explanations? What is an explanation? There are no easy answers to any of these questions, but for present purposes we can take partial answers to the first two questions to be that an explanation is called for when an appropriate "why" question is pressing. Why did the car skid off the road? Why did the anode emit x-rays? An explanation in its most general sense is an answer to such "why" questions, and in answering such questions, explanations reduce mystery. Exactly how an explanation reduces mystery brings us to the third of the aforementioned questions.

There are many competing philosophical accounts of explanation and I won't try to do justice to them here. The topic has a long and distinguished history going back to ancient times. In what follows, I will explore two types of mathematical explanation and demonstrate how they put pressure on one common and intuitively appealing account of explanation. The account in question, the causal account, holds that an explanation consists in providing the causal history of the event in need of explanation. After the causal history is presented, the event in question is no longer thought to be mysterious (i.e., because the posterior probability of the event, conditional on the causal history, is higher than the unconditional prior probability of the event in question). Now I turn to mathematical explanation.

Intramathematical Explanation

It is well known that some, but not all, proofs are explanatory. Explanatory proofs tell us *why* the theorem in question is true, whereas the nonexplanatory proofs merely tell us *that* the theorem in question is true [11]. With an explanatory proof we have one mathematical fact being explained by another (or other) mathematical fact(s). Call such cases of explanation within mathematics *intramathematical explanations.*

An example will help. Consider the well-known Euler result that there is no way of traversing the seven bridges of

https://doi.org/10.1007/s00283-018-9799-1

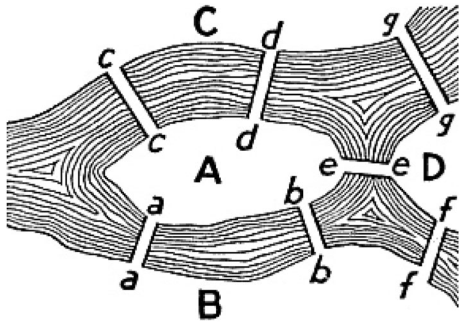
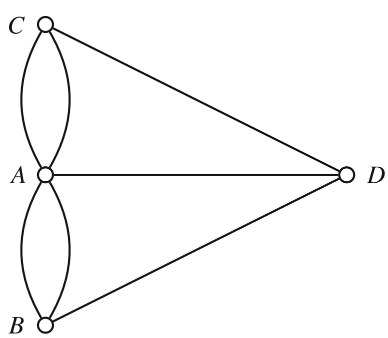


Figure 1. The seven bridges over the River Pregel in Königsberg (in Euler's time) [18].

Königsberg once and only once in a single trip, beginning and ending in the same place.

In its mathematical formulation, this becomes, there is no Eulerian cycle for the following multigraph.



The crucial result here is that a connected graph has an Eulerian cycle iff it has no vertex with an odd degree [17]. That is, there are no vertices (or nodes) with an odd number of edges meeting there.[1] The central idea of the graph-theoretic proof is that on any Eulerian cycle, every arrival at a vertex must be accompanied by a departure; an odd valence for a vertex signifies an arrival without a corresponding departure or a departure without a corresponding arrival. Indeed, this is the core explanatory insight of the graph-theoretic proof.

Compare such a proof with a brute-force, combinatorial proof that there is no Eulerian cycle for the previously mentioned multigraph. The latter would indeed deliver the result, but armed only with such a proof, we would be none the wiser as to why there is no Eulerian cycle for the multigraph in question. We'd just know that all options had been tried and none of them worked. Moreover, the graph-theoretic proof is more general. It shows, for example, that knocking out one of the edges from C to A will not help. The brute-force method would need to start over again to show that there is no Eulerian cycle for the modified graph.[2]

What is the sense of explanation at work here? It is not the causal sense I described earlier, common in science and in everyday life. When we ask, for example, why the car accident occurred, we are typically looking for responses that appeal to driver fatigue, the icy road, the velocity of the vehicle, and so on. Sometimes, however, we might appeal to more general background conditions such as "it is a dangerous road," or "he's not a very good driver," or the like. These more general conditions offer interesting complements to causal explanations and, in some contexts, it is the more general conditions that we are interested in [3]. Still, in many legal, scientific, and everyday contexts when we seek an explanation, what we are after is the causal history of the accident: all the temporally-prior events or conditions that ultimately resulted in the accident. But mathematical explanation is not in the business of providing causal histories. Talk of causes is completely out of place in mathematics. The existence of the number i *caused* the polynomial $X^2 + 1$ to have two roots? No! The existence of i is the reason that $X^2 + 1$ has two roots, but it is a category mistake to think of i as causing anything. Numbers and other mathematical objects do not seem to be the right kind of things to be in the causal nexus of the universe.[3]

It might be tempting to suggest that explanatory proofs don't provide causal explanations but something stronger: deductive explanations. After all, the conclusion of a chain of deductive reasoning, as we have in the proof of a mathematical theorem, follows of necessity. It doesn't merely follow because of contingent events in the past. Although this is right—mathematical consequence is typically deductive and this is stronger than causal consequence—deduction cannot be the key to mathematical explanations. If it were, all proofs would be explanatory, but this is clearly not the case. For example, the brute-force, combinatorial proof mentioned earlier also delivers a deductive result. But, as we have seen, this proof is not explanatory. The source of mathematical explanation must lie elsewhere.

Perhaps mathematical explanation relates to the structure of the proof.[4] We have a variety of different proof structures in mathematics: for example, there is conditional proof, reductio ad absurdum, finite induction, transfinite induction, disjunctive syllogism, universal generalisation, and proof by cases, to name a few. But there are problems with the view that the explanatoriness of a proof rests entirely on its structure. Apart from anything else, it is often relatively easy to transform the proof structure without any

[1] Or to revert to the original problem: the walk is possible so long as it does not involve an odd number of bridges with one end on the same land mass.

[2] The graph-theoretic proof explains the *mathematical fact* that there is no Eulerian cycle for the above multigraph. That's the intramathematical explanation arising from the explanatory proof. But the proof also plays a crucial role in explaining why the previously described walk around Königsberg cannot be completed. The latter is the related extramathematical fact. Shortly I'll have more to say about this quite different kind of explanation.

[3] Indeed, this is one of the reasons many philosophers have misgivings about the existence of mathematical objects. The causal inertness of mathematical objects makes it difficult to understand how we can have mathematical knowledge [4].

[4] There is also an interesting question about the status of pictures in proofs: can pictures serve as proofs and do they, in some cases, deliver genuine understanding [5]?

substantial change to the main body of the proof. Think, for example, of how, with a small change in the setup, a reductio proof of the infinitude of the primes can be turned into a proof by universal generalisation [8].

A more promising strategy is to look below the level of the structure of the proof and at the details of the proof. Here we find two quite distinct lines of thought arising. One is that a proof is explanatory because it proceeds via the "right kind" of paths. What are the right kinds of paths? Perhaps they are paths that connect results to other results in the same domain of mathematics. This might be why elementary proofs are valued in number theory: they deliver number-theoretic results via number-theoretic means, without excursions into complex analysis. On the other hand, sometimes a proof is seen to be explanatory because it builds bridges between different areas of mathematics, often by showing that a result in one area is a special case of a more general result. Category theory proofs of results in group theory (such as the Free Group Theorem) may be instances of this [15, p. 123]. If this is right, there are two quite distinct kinds of explanation operating in mathematics: one local and the other more global and unifying.

Extramathematical Explanation

So far we have been considering the explanation of one mathematical fact by another. Contrast this intramathematical explanation with another kind of explanation in which mathematics can feature: the explanation of some physical phenomenon by appeal to mathematics. Call the latter *extramathematical explanation*. This is more controversial. Although our understanding of intramathematical explanation leaves a lot to be desired, its existence is widely accepted. But there is a rather heated philosophical debate surrounding the very existence of extramathematical explanation.[5]

What's all the fuss about? The main reason that some philosophers object to the extramathematical explanation is that it would seem to commit them to Platonism: the view that mathematical objects, such as sets, numbers, functions, and the like, exist. Why? If you subscribe to a principle known as *inference to the best explanation*, you're committed to all the entities that play an explanatory role in your best science: electrons, quarks, black holes, gravitational waves, and so on. Add mathematical objects to this list and it just gets too weird for some philosophers' tastes. Unless they are willing to give up on the principle of inference to the best explanation, those who find Platonism unacceptable are forced to deny that there are any extramathematical explanations. Be that as it may, there seems to be no shortage of examples of mathematics apparently playing crucial explanatory roles in science. Here I'll quickly look at a couple of these.

The first involves the Honeycomb Theorem. Charles Darwin observed the incredible regularity of hexagons in honeybee hives and conjectured that the hexagonal shape was in some sense optimal [9]. Darwin was right, but vindication took some time. It finally came in 2001 from a mathematical theorem: the Honeycomb Theorem [12]. The Honeycomb Theorem states that a hexagonal grid represents the optimal way to divide a surface into regions of equal area with the least total perimeter. Of course, the mathematics on its own doesn't explain why honeybees build hexagonal hives. We also need some biology: wax is expensive so bees need to minimise its use (hence minimising total perimeter) while maximising the area for honey storage (there's no wasted space in the hive). It's also worth noting that this is a tiling problem, not a sphere-packing problem. This is because bees need access to the cells. Now throw in some evolutionary theory to explain why less-efficient bees have been selected against, and we are left with hexagonal hives. Although the full explanation involves a mixture of biology and mathematics, arguably it's the mathematics that's doing the heavy lifting [14].

Another phenomenon whose explanation involves mathematics appears in the life cycle of cicadas. There are seven species of North American cicadas in the genus *Magicicada*, each with prime-number life cycles. These cicadas emerge from the ground en masse once every 13 or 17 years, depending on the species. The explanation biologists offer for these unusual life cycles is in terms of evolutionary advantage. A periodic organism trying to avoid a periodic predator needs to minimise the number of years of overlap between itself and the predator. It is rather straightforward to prove that having a prime life cycle does this [2]. If a life cycle has a prime period, predators need very specific periods to overlap on a regular basis. For example, the predator would need the same prime life cycle, suitably coordinated. Again we see that the explanation relies on mathematics—in this case some elementary number theory. Moreover, the mathematics seems to be doing the important work in the explanation.

This last case also highlights an interesting general point about extramathematical explanations. These explanations typically tell us that not only is the world thus and so, but, in a very important sense, it *had to be thus and so*. The cicada life cycles are apparently squeezed from above by (not well-understood) biological considerations and from below by the fact that small primes are not as effective for predator avoidance. The remaining window, arguably, has four primes: 7, 11, 13, and 17. The periods found in nature are the largest primes in this set. These represent the optimal predator-avoidance strategies, subject to various biological constraints.

There are many other cases of such extramathematical explanations discussed in the literature.[6] What is common to them all is that the mathematics is crucial to the success of the explanation, and if there is any mathematics-free alternative available, such explanations are impoverished in various ways. In particular, the mathematics-free, causal explanations get bogged down in the contingent detail and fail to reveal the big picture. For example, a causal account

[5]The main proponents of extramathematical explanation are Alan Baker [2], Mary Leng [13], and me [6, 7].

[6]Other examples are the Euler graph-theory explanation of why the bridges of the Königsberg walk cannot be completed, why the physical act of squaring the circle is impossible, and why the Kirkwood gaps in the asteroid belt have the specific locations they do. See [16] for discussions of some of these examples.

of the Kirkwood gaps in the asteroid belt would give the causal histories of all the asteroids in the vicinity and would show why each asteroid fails to be orbiting in one of the Kirkwood gaps. Without the mathematical analysis, it looks as though it's a mere coincidence that there are no asteroids in the Kirkwood gaps. But it is no coincidence. The Kirkwood gaps are unstable orbits, and this crucial piece of the story is delivered by the mathematical explanation [8, chap. 5]. Part of the power of mathematics is that it enables abstraction away from the often-irrelevant, contingent details, and it goes straight to the core of the explanation.

What Is Explanation?

It appears that mathematics can deliver what we intuitively recognise as both intra- and extramathematical explanations, and, as we have seen, neither of these conform to the causal-history model of explanation. This suggests that further investigation of mathematical explanation could profitably contribute to a more general understanding of the nature of explanation. Mathematical explanation is also important for debates over Platonism and its rivals. After all, if an entity plays an indispensable explanatory role in one of our best scientific theories, this is taken by many scientists and philosophers alike to be a sure sign that the entity in question exists. The existence of extramathematical explanation would thus seem to lend support to Platonism.

Philosophers of mathematics have been largely interested in extramathematical explanation, and mathematicians have long been interested in intramathematical explanation. But the ease with which we accept the application of intramathematical explanations to physical phenomena, as in the Honeycomb Theorem case, suggests that these two kinds of mathematical explanations are closely related. Philosophers of mathematics and mathematicians really ought to hang out together more (and they would both benefit from discussions with philosophers of science). For a start, it would be extremely useful for philosophers to have a good stock of proofs considered by mathematicians to be explanatory and proofs not considered to be explanatory.[7] It would also be good to have mathematicians' thoughts on what distinguishes explanatory proofs from the others. It would be interesting to explore whether proof is the only locus of explanation. After all, if one kind of intramathematical explanation lies in unifying branches of mathematics, perhaps domain extensions and even generalised definitions could be thought to facilitate explanation. These are all issues on which this philosopher of mathematics would welcome the opinions of mathematicians.

ACKNOWLEDGMENTS

Thanks to Clio Cresswell and Greg Restall for very helpful discussions on the topic of this article. Thanks also to Jim Brown, Mary Leng, and an anonymous *Mathematical Intelligencer* referee for detailed and constructive comments on earlier versions of the article. This work was funded by an Australian Research Council Future Fellowship (Grant Number FT110100909) and the Alexander von Humboldt Stiftung.

Department of Philosophy
University of Sydney
A14 Main Quadrangle
Sydney, NSW, 2006
Australia
e-mail: mark.colyvan@sydney.edu.au

REFERENCES

[1] Aigner, M., and Ziegler, G. M. 2010. *Proofs from the Book*, 4th edition, Heidelberg: Springer.
[2] Baker, A. 2005. ''Are There Genuine Mathematical Explanations of Physical Phenomena?'' *Mind*, 114: 223–238.
[3] Baron, S., and Colyvan, M. 2016. ''Time Enough for Explanation,'' *The Journal of Philosophy*, 113(2): 61–88.
[4] Benacerraf, P. 1973. ''Mathematical Truth,'' *The Journal of Philosophy*, 70(19): 661–679.
[5] Brown, J. R. 2008. *The Philosophy of Mathematics: A Contemporary Introduction to the World of Proofs and Pictures*, 2nd edition. London: Routledge.
[6] Colyvan, M. 2001. *The Indispensability of Mathematics*, New York: Oxford University Press.
[7] Colyvan, M. 2010. ''There is No Easy Road to Nominalism,'' *Mind*, 119(474): 285–306.
[8] Colyvan, M. 2012. *An Introduction to the Philosophy of Mathematics*, Cambridge: Cambridge University Press.
[9] Darwin, C. 2006. *On the Origin of Species by Means of Natural Selection*. Mineola, NY: Dover. (First Published in 1859.)
[10] Davis, P. J., and Hersh, R. 1981. *The Mathematical Experience*, Boston: Birkhäuser.
[11] Gowers, T., and Neilson, M. 2009. ''Massively Collaborative Mathematics,'' *Nature*, 461 (15 October 2009), 879–881.
[12] Hales, T. C. 2001. ''The Honeycomb Conjecture,'' *Discrete Computational Geometry*, 25: 1–22.
[13] Leng, M. 2009. *Mathematics and Reality*. Oxford: Oxford University Press.
[14] Lyon, A., and Colyvan, M. 2008. ''The Explanatory Power of Phase Spaces,'' *Philosophia Mathematica*, 16(2): 227–243.
[15] Mac Lane, S. 1998. *Categories for the Working Mathematician*, 2nd edition, New York: Springer.
[16] Mancosu, P. 2008. ''Explanation in Mathematics,'' in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), URL = plato.stanford.edu/archives/fall2008/entries/mathematics-explanation/.
[17] Trudeau, R. J. 1993. *Introduction to Graph Theory*. New York: Dover.
[18] Weisstein, E. W., 2016. 'Königsberg Bridge Problem,'' *MathWorld—A Wolfram Web Resource*. URL = http://mathworld.wolfram.com/KoenigsbergBridgeProblem.html (accessed August 2016).

[7] Some steps toward this goal have been made in [1, 10].

