

Explanation in Mathematics: Proofs and Practice (forthcoming in *Philosophy Compass*)

William D'Alessandro

Abstract

Mathematicians distinguish between proofs that explain their results and those that merely prove. This paper explores the nature of explanatory proofs, their role in mathematical practice, and some of the reasons why philosophers should care about them. Among the questions addressed are the following: what kinds of proofs are generally explanatory (or not)? What makes a proof explanatory? Do all mathematical explanations involve proof in an essential way? Are there really such things as explanatory proofs, and if so, how do they relate to the sorts of explanation encountered in philosophy of science and metaphysics?

1 Introduction

A mathematical explanation occurs when a fact of some sort is explained by a piece of mathematics—a theorem, a diagram or a proof, for example. Many philosophers think there are mathematical explanations in science. These are cases where a piece of math explains an empirical fact. (For instance, certain species of North American cicada have synchronized, periodic life cycles lasting 17 years. The fact that 17 is prime may be part of the reason why.¹) Other philosophers say there are explanations in pure mathematics, where one piece of math explains another. (For instance, it's widely agreed that Galois's work in algebra explains why polynomial equations of degree 5 don't have a general solution analogous to the quadratic formula.) Both phenomena have been topics of recent work, but this paper deals only with the second. So from here on I'll use the term 'mathematical explanation' (or 'ME' for short) to mean "mathematical explanation in pure mathematics".²

There are many interesting issues surrounding ME—too many to canvass in one paper. So rather than attempting a general survey, I'll be focusing on a specific set of questions about proofs and their role in explanation. The next section gets the ball rolling with an overview and some examples. Section 3 discusses the explanatory value (or lack thereof) of several particular types of proof, and section 4 asks what makes a proof explanatory in general. Section 5 considers whether all mathematical explanations involve proof in an essential way. Finally, section 6 addresses skepticism about the notion of explanatory proof.

The notion of explanatory proof has a long and interesting history, going back at least to Aristotle's distinction between mere demonstrations and those that provide an *aitía* (cause, reason or explanation).³ Unfortunately the reader won't learn much of this story here. Although I've tried to point out some major historical touchstones, this article is meant mostly as a guide to recent work.

¹For more on the cicada example, see [Baker 2005]. For a recent *Philosophy Compass* guide to mathematical (and other noncausal) explanations in science, see [Reutlinger 2017].

²Some authors use the terms 'extra-mathematical explanation' and 'intra-mathematical explanation' to distinguish these two phenomena.

³The distinction is from Book I of the *Posterior Analytics*. For a discussion of Aristotle on mathematical explanation, see [Mancosu 2000].

2 Explanatory proof and mathematical practice

Let me start with an example of what I mean (and don't mean) by “explanatory proof”. Suppose you want to show that the sum of the first n odd natural numbers is n^2 , i.e. that

$$1 + 3 + 5 + \cdots + n = \sum_{k=1}^n (2k - 1) = n^2$$

for all natural numbers $n \geq 1$. One approach is to use mathematical induction. Clearly the statement is true for $n = 1$, and applying the induction hypothesis⁴ gives

$$\begin{aligned} \sum_{k=1}^{n+1} (2k - 1) &= \sum_{k=1}^n (2k - 1) + 2(n + 1) - 1 \\ &= n^2 + 2n + 1 \\ &= (n + 1)^2, \end{aligned}$$

as needed.

This gets the job done, logically speaking; the proof is certainly sound. But it seems not to explain the result. The fact that we can move 1s and 2s and n s around until they assume the right form gives no clue about *why* sums of odd numbers might have anything to do with squares.

Here's a different sort of proof, the key idea of which is to *view numbers as arrangements of dots*. Start with 1, the first odd number, which can be regarded as a square array of side 1 (and hence of area 1^2). The second odd number is 3, and we can think of adding 3 to 1 as augmenting the original square array so as to make a new one of side 2 (and hence of area 2^2). Similarly, adding 5 to $1 + 3$ gives a square array of side 3, and so on, as in the diagram.

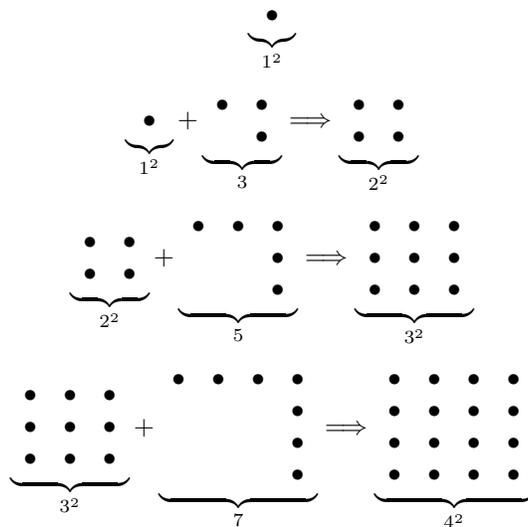


Figure 1: Dot-diagram proof of $1 + 3 + 5 + \cdots + 2n - 1 = n^2$.

It's easy to see that the pattern will hold in general—for any $n - 1$, the sum of the first $n - 1$ odd numbers corresponds to a square array of side $n - 1$ and area $(n - 1)^2$, and adding the next odd number can be viewed as augmenting this array to produce a new one of side n and area n^2 . Hence the diagram shows that the

⁴That is, the assumption that the statement to be proved holds for all integers less than $n + 1$, which implies in particular that $\sum_{k=1}^n (2k - 1) = n^2$.

sum of the first n odd natural numbers is n^2 . What’s more, this sort of proof is plausibly explanatory.⁵ One can readily understand from the proof *why it is* that the identity holds.

The above case is simple, but ME isn’t limited to the sorts of toy examples beloved by philosophers. On the contrary, mathematicians from antiquity to the present have often been deeply concerned about explanation. And this concern has shaped how mathematics is carried out, evaluated, interpreted and taught. So understanding ME is part of understanding the norms and goals of mathematical practice—a project that philosophers of mathematics have turned to with increasing interest over the last couple decades.⁶

The 2015 Polymath project “Explaining Identities for Irreducible Polynomials” provides a recent example of explanatory concerns in action.⁷ The problem involves a certain infinite sum of reciprocals of polynomials.⁸ As per the project description:

It was numerically observed [in earlier work] that one appears to have the remarkable cancellation

$$\sum_{P \in \mathcal{P}} \frac{1}{1+P} = \frac{1}{t} + \frac{1}{t+1} + \frac{1}{t^2+t} + \dots = 0.$$

...The Polymath proposal is to investigate this phenomenon further (perhaps by more extensive numerical calculations) and supply a theoretical explanation for it. ([Tao 2015])

It’s often thought that mathematical inquiry begins and ends with proving new theorems, but this kind of example shows otherwise. Even though Thakur and his collaborators already knew the “remarkable cancellation” result to be true, they found it important to understand why such an identity should hold, and they considered the search for an explanation worthy of a major research effort.⁹ As the number theorist Fernando Gouvêa writes:

It’s often said that proofs serve as the criterion for truth in mathematics: we prove things in order to establish that they are true. This is certainly true, but it doesn’t explain something else we do, namely, provide new proofs of old results. We already know those theorems are true, so in giving new proofs we are not seeking to establish that. What we are seeking is understanding. We want to know *why* the theorem is true, and a proof can (sometimes) tell us that. ([Gouvêa 2015]; emphasis in original)

Another reason to study ME, then, is because doing so promises to enrich (and perhaps correct) our ideas about mathematical practice. See for instance [Detlefsen 1988], [Mancosu 1999], [Tatzel 2002], [Harari 2008], [Mancosu & Hafner 2008], and [D’Alessandro 2018] for more on the role of explanatory concerns in historical and contemporary mathematics.

3 Particular proof-types: exhaustion, induction, abstraction and mechanisms

Many aspects of proofs are drawn from a repertoire of standard techniques and forms. This repertoire includes general patterns of reasoning that appear throughout mathematics (e.g. proof by contradiction), as well as domain-specific styles of argument (e.g. showing that a series converges by comparing it to another series). Proof methods aren’t all created equal; different techniques have various epistemic, cognitive and practical strengths and weaknesses. Consequently, mathematicians value some methods more highly than

⁵A number of authors have made claims to this effect. For instance, [Gullberg 1997] gives a similar dot-diagram proof that the sum of the first n natural numbers is $n(n+1)/2$, and claims that “the figure shows why” the identity holds (289). [Hanna 1990] contrasts the inductive proof of this identity with the dot-diagram proof, claiming that the latter but not the former is explanatory (10-11). [Steiner 1978a] makes the same comparison with the same conclusion (136-137). See chapter 8 of [Giaquinto 2007] for a discussion of the epistemology of dot-diagram arguments, including an extensive defense of the claim that the images used in such arguments count as genuine proofs.

⁶For a sample of recent work on the philosophy of mathematical practice, see [Mancosu 2008b], as well as many of the essays in [Bueno & Linnebo 2009].

⁷The Polymath projects are a series of collaborative online efforts to solve outstanding mathematical problems, headed by Timothy Gowers and Terence Tao. See <https://polymathprojects.org> for a list of current and previous projects.

⁸Specifically, these are the irreducible polynomials in $F_2[t]$, the ring of polynomials over the two-element finite field F_2 .

⁹This project, by the way, culminated successfully in a paper by David Speyer ([Speyer 2016]).

others. One factor in this assessment (among many others) is the relative explanatory power of different kinds of proof.¹⁰

Consider proof by exhaustion, for instance. Here the strategy is to prove a general result by showing that the statement holds for each of a finite number of cases, which jointly exhaust the relevant possibilities. The only known proofs of some famous theorems (e.g. the classification of finite simple groups and the four-color theorem) are exhaustion arguments. Infamously, some of these proofs rely on computers to check very large numbers of cases.

Philosophers and mathematicians have often expressed dissatisfaction with proofs by exhaustion. This isn't surprising, since "brute-force" arguments in science and elsewhere are often deprecated as unexplanatory. (If you want to know why metals conduct electricity, you won't be impressed with the answer "Because gold conducts electricity, and so does silver, and so does copper, ...".) The problem is that a good explanation should show what the cases have in common, in virtue of which they're all similar in the relevant way. As Mark Colyvan writes, "Proofs [by exhaustion] lack unity. There are often different reasons offered in the different cases and it looks like 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" ([Colyvan 2012], 81).

Mathematicians frequently look for better alternatives to exhaustion proofs. For instance, according to the graph theorist Paul Seymour, the current proof of the four-color theorem

is still not satisfying, requiring as it does the extensive use of a computer. ...We would very much like to know the "real" reason the 4CT is true; what exactly is it about planarity that implies that four colours suffice? Its statement is so simple and appealing that the massive case analysis of the computer proof surely cannot be the book proof. ([Seymour 2016], 417)¹¹

It's sometimes suggested that the explanatoriness of an exhaustion argument decreases with the number of cases considered. Alan Baker writes, for instance, that the current proof of the four-color theorem "is highly disjunctive: There are 1476 different sub-cases that are individually considered. Thus, the proof is very unexplanatory" ([Baker 2009], 148)

A more controversial case is proof by induction. Proofs of this type standardly proceed as follows: first, the statement in question is proved for the "base case"—the smallest natural number to which the statement applies, often $n = 1$ —and then it's shown that if the statement holds for a given natural number n , it must also hold for $n + 1$. From these two facts it follows that the statement holds for all natural numbers (greater than or equal to the base case).

Many inductive proofs don't *seem* very explanatory. Take the proof that the sum of the first n odd numbers is n^2 , given in the previous section—although the argument is convincing, it strikes me as providing little insight about why the identity should hold. But some authors don't share this intuition. For instance, [Kitcher 1975] and [Brown 1997] claim that inductive proofs typically are explanatory; according to Brown, this is because "induction—the passage from n to $n + 1$ —more than any other feature, best characterizes the natural numbers" (177). In the absence of a good theory about the relationship between explanation and characteristicness, however, the force of Brown's suggestion is unclear.¹²

Marc Lange has proposed to "end this fruitless exchange of intuitions" with a principled argument that inductive proofs are never explanatory ([Lange 2009], 205). Lange's strategy is to show that, if any such proofs were explanatory, then there would exist explanatory circles, which are presumably impossible. The reasoning is as follows. Suppose that there's an explanatory inductive proof of some general fact about the natural numbers. Call this fact $\forall n P(n)$. As it turns out, a typical inductive proof can be converted into an alternative argument that starts from an arbitrary natural number, say 5, and then proceeds inductively

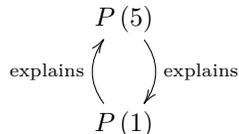
¹⁰Other considerations include depth ([Gray 2014]), beauty ([Inglis & Aberdein 2015]), simplicity ([Mizrahi 2016]), surveyability ([Bassler 2006]), abstractness ([Pincok 2015]), generalizability ([Steiner 1978a]), transferability ([Easwaran 2009]), transparency, computer checkability and constructiveness. Some of these factors may contribute to a proof's explanatoriness, or vice versa, but in general the relationship between explanation and other proof features is up for debate.

¹¹"The book" is God's book, which contains the best possible proof of every theorem. (The idea is Paul Erdős's.) For some current best guesses about the contents of the book, see [Aigner & Zeigler 2010].

¹²The theory presented in [Steiner 1978a] might provide support for Brown's claim, since it identifies explanatory proofs as those that make essential use of "characterizing properties". I don't know whether Brown had Steiner's work in mind. In any case, Steiner's account has been widely criticized and is generally considered unpromising; see section 4 below.

“downward” and “upward”.¹³ Lange claims that both proofs are explanatory if either one is. (“There is nothing to distinguish them, except for where they start” (209).)

If the ordinary proof is explanatory, however, then the truth of the base case $P(1)$ presumably helps explain the truth of $P(5)$. Similarly, if the “downward and upward from 5” proof is explanatory, then the truth of the base case $P(5)$ helps explain the truth of $P(1)$. If both proofs are explanatory, then, we get a circle



But this is presumably impossible, since explanation is an asymmetric relation.¹⁴ It follows that the original inductive proof wasn’t explanatory.

As [Baker 2010] points out, the situation isn’t quite as simple as Lange suggests. There’s reason to question the crucial claim that the two kinds of inductive argument can’t differ in explanatory power. For instance, the “downward and upward” proof is more disjunctive than the standard proof, and hence perhaps less explanatory. (For further criticism and discussion, see [Baldwin 2016], [Dougherty 2017], [Hoeltje et al. 2013], [Lehet 2019], [Salverda 2018], [Wysocki 2017].)

Other types of proof are noteworthy for their positive explanatory value. For instance, [Frans & Weber 2014] argue that some proofs—notably, certain kinds of geometric arguments—explain by identifying the mechanisms responsible for their results. Here, mechanisms are understood in terms of dependence relations, which we can discover by carrying out interventions on the entities and properties appearing in the proofs.

If this vocabulary sounds familiar, it’s because Frans and Weber’s work is an adaptation of the popular New Mechanist approach to scientific explanation.¹⁵ On this view, “explanation is a matter of elucidating the causal structures that produce, underlie, or maintain the phenomenon of interest” ([Craver & Tabery 2017]). This sort of causal dependence, in turn, is often understood along “interventionist” or “manipulationist” lines. Roughly speaking, on a manipulationist account like James Woodward’s ([Woodward 2003]), a causal (and hence explanatory) relation holds between two variables when a change in one variable would change the value of the other. So causal explanations answer “what-if-things-had-been-different” questions—they tell us whether and how various possible interventions would make a difference to the explanandum phenomenon. (Cutting off the oxygen supply leads to the fire going out, but cutting off the nitrogen supply doesn’t. So the presence of oxygen, but not of nitrogen, partly explains the occurrence of the fire.)

This is essentially Frans and Weber’s picture, with the proviso that their dependence relations are non-causal and their interventions are “imaginary”.¹⁶ Their main example is the Butterfly Theorem from plane geometry, pictured above. The theorem is as follows: Let PQ be a chord of a given circle with midpoint M , and let AD and BC be two other chords, intersecting PQ at X and Y respectively. Then M is also the midpoint of the segment XY .

¹³The idea here is to run two separate inductive arguments, one that establishes the result for $1 \leq n \leq 5$ (the “downward” part) and another that proves it for $n \geq 5$ (the “upward” part).

¹⁴One might think—as an anonymous referee suggests—that some axioms may be self-explaining, and hence that the asymmetry claim isn’t true in general. I’m skeptical; I suspect that few if any axioms possess the foundational metaphysical and epistemic properties that are sometimes claimed for them (cf. [Easwaran 2008], [Maddy 2011]). In any case, Lange’s example doesn’t involve axioms (or other plausible candidates for self-explanatoriness), so it’s not crucial for him that the claim be true in its full generality.

¹⁵See for instance [Machamer et al. 2000] or [Bechtel & Abrahamsen 2005]. For a *Compass* guide to mechanisms, see [Andersen 2014a] and [Andersen 2014b].

¹⁶It’s worth noting that, even in the empirical setting, some interventions are imaginary in the sense that they’re impossible to carry out. For instance, we want to say that the gravitational attraction of the moon explains the tides. For a manipulationist like Woodward, this entails that an intervention on the former would change the latter. But there may be no physically possible process that alters the moon’s mass or distance from Earth without itself directly affecting the tides; this disqualifies such a process from counting as a proper intervention. See [Woodward 2003], Chapter 3 for discussion of this sort of case.

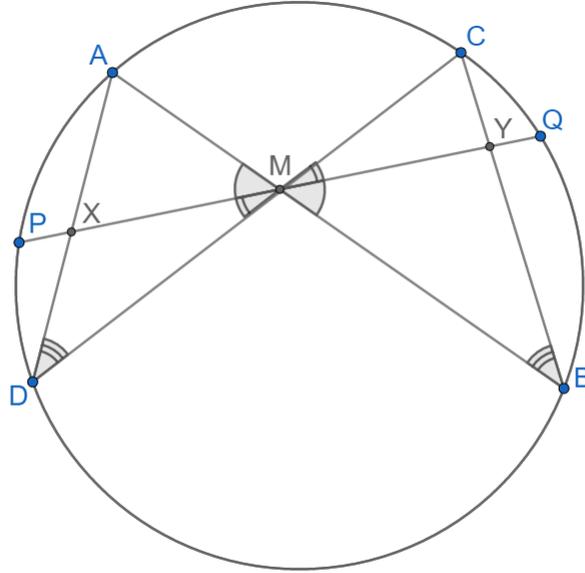


Figure 2: The Butterfly Theorem: M , the midpoint of PQ , is also the midpoint of XY .

Frans and Weber present a proof of the Butterfly Theorem that “identifies which entities and properties are relevant in order to explain why the theorem holds” (12). Their proof is supposed to show, for instance, that the result depends on properties of the triangle $\triangle ADM$. A key step in the proof uses the similarity of $\triangle ADM$ and $\triangle CBM$; if we imagine deforming $\triangle ADM$ so that this similarity is lost, the reasoning no longer goes through. By the manipulationist criterion, this shows that the theorem depends on the similarity of $\triangle ADM$ and $\triangle CBM$. By contrast, an intervention that tilts the chord PQ —so that the left endpoint is moved higher than the right endpoint, say—won’t change the midpoint of XY . Thus the theorem doesn’t depend on (and isn’t explained by) the angle of PQ .

It’s interesting to compare the contemporary mechanist approach to ME with an episode from classical mathematics. In the 5th century, the Neoplatonist philosopher Proclus wrote an influential commentary on Euclid’s *Elements* which criticized a number of Euclid’s proofs. One of Proclus’s complaints was that certain proofs, although sound, fail to provide Aristotelian *aities* (causes or explanations). Among the arguments that Proclus found inadequate was Euclid’s proof of *Elements* I.32, which asserts that the sum of a triangle’s interior angles is equal to two right angles. Figure 3 shows a diagram for this proof. (See [Harari 2008] for a study of Proclus’s views on ME.)

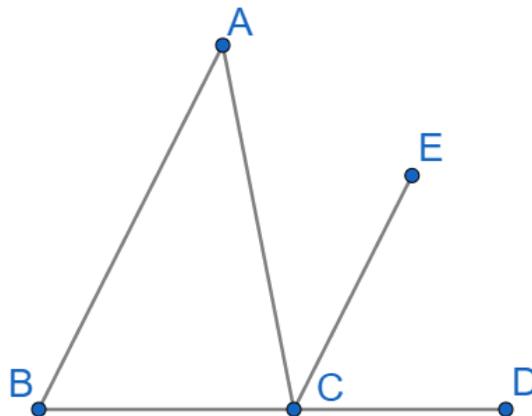


Figure 3: *Elements* I.32: the sum of $\triangle ABC$ ’s internal angles is equal to $\angle BCA + \angle ACE + \angle ECD = 180^\circ$.

Euclid’s strategy is to show that the sum of the internal angles, $\angle ABC + \angle BCA + \angle CAB$, is equal to the sum $\angle BCA + \angle ACE + \angle ECD$. Since the second sum is clearly 180° , this proves the proposition. Proclus has this to say about Euclid’s reasoning:

[W]hen it is proved that the interior angles of a triangle are equal to two right angles from the fact that the exterior angle of a triangle is equal to the two opposite interior angles, how can this demonstration be from the cause? ...For the interior angles are equal to two right angles even if there are no exterior angles, for there is a triangle even if its side is not extended. ([Friedlein 1873], 206.12-26, quoted and translated in [Harari 2008], 138-139.)

Here Proclus seems to be saying that Euclid’s proof “has no explanatory worth because triangles have the sum of their angles equal to two right angles by virtue of being triangles and not [by virtue] of having an external angle” ([Harari 2008], 139).

Proclus’s objection, or something closely related, can be stated in mechanist terms. A diagnosis in this spirit might run as follows: Euclid’s proof of I.32 makes essential use of the external angle $\angle ACD$. But the truth of the theorem doesn’t depend on the measure of this angle, since an intervention that makes $\angle ACD$ smaller or larger wouldn’t affect the sum of $\triangle ABC$ ’s internal angles. Thus Euclid’s proof fails to pinpoint the mechanisms responsible for the theorem, and so it isn’t properly explanatory.

Applying the mechanist approach is fairly natural in the context of synthetic geometry, where proofs deal with spatially structured systems whose parts are easy to act on by imaginary intervention. It remains to be seen whether explanatory proofs in other areas of mathematics can be usefully analyzed in mechanist terms. (For some evidence that they can, see [Lange 2017], which proposes to identify the mechanism responsible for the failure of the infinitary sum rule for derivatives.)

Finally, modern mathematics has often made progress by climbing higher on the ladder of generality; [Pincock 2015] sets out to show how this process can have explanatory value. His account deals with proofs that explain “by invoking a more abstract kind of entity than the topic of the theorem” (1). On Pincock’s view, such proofs are explanatory because they identify a kind of metaphysical dependence relation between the relatively concrete objects in the explanandum and the relatively abstract objects in the explanans. Pincock calls this relation “abstract dependence”. By linking ME with metaphysics in this way, Pincock aims to “generaliz[e]... [the] ontic conception of explanation to those cases where causal relations no longer apply” (7).¹⁷

The first task is to clarify the intuitive idea of one object being more abstract than another. Pincock analyzes relative abstractness in terms of instancehood, in roughly the type-token sense. For example, the word-type *cat* has particular token inscriptions of the word ‘cat’ as instances. So the word-type is more abstract than the word-token. Similarly, the “concrete” groups $(\{0, 1\}, +)$ and $(\{1, -1\}, \times)$ are instances of the “abstract” cyclic group C_2 . (An abstract group is “a group characterized only by its abstract [algebraic] properties and not by the particular representations chosen for elements” ([Weisstein 2018]). The algebraic properties of C_2 in particular are given by the rules $a \circ a = b \circ b = a$ and $a \circ b = b \circ a = b$. That is, C_2 is the unique abstract group with two elements, one of which is the identity and the other of which is self-inverse.)

Pincock’s motivating example is the classical problem of finding general solution formulas for polynomial equations.¹⁸ It turns out that such formulas exist only for polynomials of degree less than 5—a fact which puzzled mathematicians for many years, but which was explained around 1830 by the work of Évariste Galois. The key insight is that a polynomial equation admits a solution formula just in case its Galois group is solvable.¹⁹ The proof of this fact—and, more specifically, of the unsolvability of the general quintic equation—relates a concrete group associated with a specific polynomial to an abstract Galois group. This, according to Pincock, is a case of “abstract dependence”: the concrete group mentioned in the theorem metaphysically depends on the abstract group appearing in the proof. So the unsolvability proof counts

¹⁷The ontic conception of explanation, originating with the work of Alberto Coffa and Wesley Salmon around 1980, is the idea that explanations are grounded in—and provide information about—worldly dependence relations. Salmon focused mostly on causation, but recent authors in the ontic tradition have increasingly countenanced non-causal forms of dependence.

¹⁸More specifically, the problem is to find solution formulas “in radicals”, i.e. in terms of the coefficients of the polynomial, basic arithmetical operations and n th roots only. The quadratic formula $x = -b \pm \frac{\sqrt{b^2 - 4ac}}{2a}$ accomplishes this for a second-degree polynomial $ax^2 + bx + c$.

¹⁹The Galois group of a polynomial $p(x)$ is, roughly, a certain set of permutations of the roots of $p(x)$. Solvability is a technical condition having to do with the composition of a group’s subgroups. An introduction to Galois theory, including precise definitions of the terms mentioned here, can be found in almost any undergraduate abstract algebra textbook.

as an abstract mathematical explanation.²⁰ ([D’Alessandro forthcoming a] gives a different analysis of this case, arguing that Galois’s theorems, rather than their proofs, are his main explanatory achievement.)

Note that one can invoke a more abstract *entity* without appealing to a more general *principle* (and conversely). So Pincock’s approach shouldn’t be viewed as a kind of deductive-nomological theory, according to which “[a phenomenon] is explained by subsuming it under general laws” ([Hempel & Oppenheim 1948], 136). Laws, arguments and inferences play no essential role in Pincock’s account; as with other proponents of the ontic conception, his goal is to tie explanation directly to mathematical objects and the dependence relations in which they stand. For more on the relationship between ME and the ontic conception—including an argument that some mathematical explanations can’t be understood along these lines—see [D’Alessandro forthcoming b].

4 Explanatory proof in general

The last section looked at the explanatory value of some particular types of proof. This section asks a broader question: what set of features makes a proof either explanatory or unexplanatory in general?

The most influential account in the literature is due to Mark Steiner, from his pioneering paper “Mathematical Explanation” ([Steiner 1978a]). Steiner’s view is motivated by a familiar thought: namely, “the idea that to explain the behavior of an entity, one deduces the behavior from the essence or nature of the entity” (143). In mathematics, however, it’s no good to think of an essence as a special property possessed by an object in every possible world. (On the standard view, mathematical objects have all their properties necessarily.) Instead of appealing to essences in the modal sense, then, Steiner prefers to speak of “characterizing properties”: that is, “propert[ies] unique to a given entity or structure within a family or domain of such entities or structures” (143). (Compare [Fine 1994] and Kit Fine’s subsequent work exploring a non-modal conception of essence and its role in explanation.)

According to Steiner, the hallmark of explanatory proofs is the use they make of characterizing properties. For him, a proof is explanatory just in case it depends (in an evident way) on a characterizing property of some object mentioned in the theorem. Such a proof should also be “deformable”, in the sense that varying the characterizing property yields similar proofs of related results. Instead of giving precise definitions for these terms, Steiner motivates and clarifies his account by way of examples.

Here’s one. Consider the dot-diagram proof from Figure 1. The proof evidently uses a characterizing property of the odd natural numbers, since it’s all and only these numbers that correspond to dot arrays of the appropriate configuration. What’s more, the diagram can be easily modified to prove other identities involving sums of natural numbers. Steiner himself discusses the identity $1 + 2 + 3 + \dots + n = \frac{1}{2}(n(n+1))$, proved in Figure 4 below.²¹

²⁰The details of Pincock’s account are more somewhat involved than this. For instance, not every proof that invokes a more abstract entity counts as explanatory: a further condition is that this be the *least more abstract* entity that can account for the fact to be explained.

²¹The proof given here is slightly different from Steiner’s version.

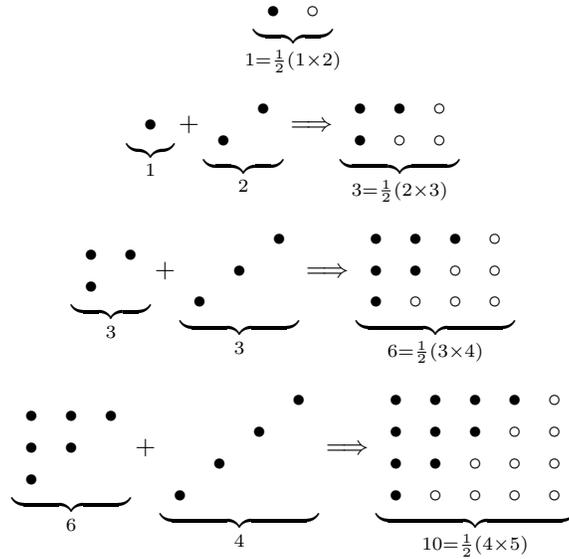


Figure 4: Dot-diagram proof of $1 + 2 + 3 + \dots + n = \frac{1}{2} (n(n + 1))$.

The idea here is to view 1 as half of a 1×2 rectangular array, $1 + 2$ as half of a 2×3 array, $1 + 2 + 3$ as half of a 3×4 array, and so on.

On Steiner’s account, then, both of these dot-diagram proofs are explanatory. (As he notes, his view has the somewhat strange consequence that “explanation is not simply a relation between a proof and a theorem; rather, a relation between an array of proofs and an array of theorems, where the proofs are obtained from one another by [deformation]” (143).)

Steiner’s paper is full of suggestive examples and ideas, many of which have been picked up by later authors. But his account has been much criticized and is now widely agreed to be unsatisfactory, at least as a general theory of explanatory proof.

One difficulty with the theory is the lack of clarity surrounding its key concepts. Steiner doesn’t try to define ‘characterizing property’ or ‘family’, and he gives no criteria for deciding whether an object is mentioned in a theorem, or whether a proof counts as deformable. It’s far from clear how to make these ideas more precise in a principled way that gives the right verdicts about cases. There are other issues too. For instance, some seemingly explanatory proofs concern “arbitrary” objects that lack characterizing properties (cf. [Hafner & Mancosu 2005]). And Steiner’s deformability criterion is doubtful. Marc Lange has argued that “we can appreciate a proof’s explanatory power (or impotence) just from examining the details of that proof itself, without considering what else could be proved by instantiating the same scheme” ([Lange 2014], 523; see also [Pincock 2015], 8-9). [Resnik & Kushner 1987] is an early response to Steiner that makes several of the above points.

Perhaps Steiner’s theory is most charitably viewed as an account of one particular way that proofs can explain. Weber and Verhoeven suggest, for example, that it only applies to contrastive questions of the form “Why do mathematical objects of class X have property Q , while those of class Y have property Q' ?” ([Weber & Verhoeven 2002], 300).²²

Another “classical” account of explanatory proof is that of Philip Kitcher. Although Kitcher devoted no single publication to ME, his unificationist theory of scientific explanation was also meant to handle cases from pure mathematics ([Kitcher 1989]). On Kitcher’s approach, a proof counts as explanatory just in case it instantiates an argument pattern from the “explanatory store”, that is, the set of argument patterns that most efficiently systematizes our knowledge in a given domain. The unificationist theory of explanation has been much discussed elsewhere, so I won’t expand on its subtleties here; see §5 of [Woodward 2017] for a start.

Kitcher’s account has had a smaller impact on the ME literature than that of Steiner, and those who

²²They also hold, however, that “even within this restricted domain, Steiner’s theory must be corrected and completed” (300).

discuss it have mostly found it wanting. For instance, [Hafner & Mancosu 2008] and [Pincock 2015] each present apparent counterexamples: a case where the best systematization yields proofs that mathematicians reject as unexplanatory (for Hafner and Mancosu), and a case where the discovery of an explanatory proof failed to produce a better systematization (for Pincock). [Tappenden 2005] is a sympathetic exploration of the idea that unification promotes explanation in mathematics. But, Tappenden argues, what counts as unifying in the relevant sense depends on more than mere economy of proof schemata. “[S]uccessfully identifying unifying generalities is assessed not by counting the total number of patterns but rather by the quality of the patterns themselves: Are they the right ones (are they deep or fruitful or revealing or whatever?)” (169).

Most work on ME since Steiner has been relatively narrow in focus, directed at the explanatory qualities of particular proofs or proof-types. But Lange has recently developed a more ambitious theory of explanatory proof. Lange’s view is that, “in many cases, at least, ...[an explanatory proof is] a proof that exploits a certain kind of feature in the problem: the same kind of feature that is outstanding in the result being explained” ([Lange 2014], 489). One type of “outstanding feature” is *symmetry*: when a theorem shows some phenomenon to be surprisingly symmetrical, we turn for an explanation to a proof that turns on a similar invariance. (Consider the Butterfly Theorem discussed in the previous section. The fact that M is the midpoint of XY as well as PQ is noteworthy, and the similarity of $\triangle ADM$ and $\triangle CBM$ would seem to be implicated somehow; a proof that fails to exploit this obvious symmetry would strike us as missing the point. See [Frans & Weber 2014], 14-15 for an example of this sort of “bad” proof.) Another feature that calls for explanation is *unity*: when a theorem shows different cases to exhibit a striking commonality, an explanatory proof will be one that exposes their underlying sameness. (This is why “brute-force” methods like proof by exhaustion often seem unsatisfying.) Finally, the *simplicity* of a result is often salient, and in this case we consider a proof explanatory when it reveals some correspondingly simple feature of the problem situation. (The “remarkable cancellation” theorem mentioned in section 2 displays a provocative simplicity: why should you get something as nice as 0 when adding together the reciprocals of these polynomials? Ditto for Euler’s identity $e^{i\pi} + 1 = 0$, whose simple explanation involves viewing $e^{i\theta}$ as a rotation of 1 around the origin by an angle of θ radians.) Although Lange holds that symmetry, unity and simplicity are among the features that most often call out for explanation, he allows that other qualities can also be salient in this way ([Lange 2014], 524).

Lange marshals an impressive set of examples to support his case, and I think he leaves little doubt that explanatory proofs often work in the way he describes. But the scope of his theory remains unclear: he suggests that it accounts for “many cases, at least”, but does this mean “quite a few”, “most” or “all”? Lange sometimes seems to characterize his view as a fully general theory of explanatory proof²³, although he doesn’t argue directly for this.

Perhaps we should be hesitant to accept that Lange has had the last word. After all, by its very nature, Lange’s account only applies to cases with conspicuous qualities that seem to call out for explanation. (As he writes, “if [a] result exhibits no noteworthy feature, then to demand an explanation of why it holds, not merely a proof that it holds, makes no sense” (507).) But in many settings, it’s possible to find ourselves with explanations which we weren’t looking for and whose existence we didn’t suspect. Some facts seem like banal happenstance *until* you see the (surprisingly deep and illuminating) reason why they’re true. If this sort of case occurs in science and elsewhere, it’s unclear why mathematics should be any different.

A very different approach to explanatory proof has been taken by Matthew Inglis and Juan Pablo Mejía-Ramos ([Inglis & Mejía-Ramos 2019]; see also [Delarivière et al. 2017] for a related view). Following a proposal of Daniel Wilkenfeld’s ([Wilkenfeld 2014]), Inglis and Mejía-Ramos argue that a proof is explanatory precisely when and because it generates understanding (“in an appropriate manner and at an appropriate time” (1)). Of course, it’s hardly controversial to suggest that explanation and understanding are closely related. What’s distinctive about this view is that, while most authors see understanding as a *byproduct* of a more fundamental explanation-making feature, Inglis and Mejía-Ramos are epistemicists who take understanding as the *criterion* for explanation.

In order to make this idea precise, it’s necessary to say what understanding is and how a proof can generate it (or not). Here Inglis and Mejía-Ramos draw on recent work in philosophy and psychology.

²³For instance, he describes himself as having “tried to identify *the* basis on which certain proofs but not others are explanatory” (524, emphasis mine), and he claims unqualifiedly that “an explanatory proof requires some feature of the result to be salient and requires the proof to exploit a similar noteworthy feature in the problem” (524).

Particularly important is the notion of a schema—that is, “a cognitive structure that permits us to treat multiple elements of information as if [they] were a single element” (10), and which thus helps us recognize, remember and reason about complex data. For Inglis and Mejía-Ramos, the possession of relevant schemas is constitutive of understanding. So an explanatory proof is one that facilitates the creation of such schemas and their consolidation in long-term memory. As they write:

Our account suggests that the archetypal explanatory proof would have at least three properties. First, it would have features that make it easy, or at least as easy as possible, to select the information from sensory memory into working memory that is necessary for a successful processing stage. ...Second, it would have features that make it easier to coordinate the new knowledge contained in the proof with existing schemas retrieved from long-term memory, and therefore to reorganise the new and existing information into coherent new schemas. Finally, it would be likely to split the working memory load it gives to its readers between their visual and verbal/auditory channels so that the chances of their working memory capacity being exceeded during the schema-organisation process is minimised. (13)

As Inglis and Mejía-Ramos point out, it follows that explanatoriness is partly though not totally subjective. The subjective element derives from cognitive and epistemic differences between agents. The same proof might successfully stimulate schema formation in an expert but not in a novice, and so the proof would count as explanatory for the expert only. On the other hand, since most humans share substantial cognitive similarities, many features of proofs will tend to boost or diminish understanding across the board. (For instance, a proof with both visual and verbal elements will typically be less cognitively taxing than a purely verbal argument.) As mathematics educators have long realized, there are plenty of useful general principles of this sort. Contrary to an often-voiced complaint about psychologism, then, this sort of view doesn't render explanation hopelessly and uninterestingly subjective.

I consider Inglis and Mejía-Ramos's cognitivist approach promising, but some of the details are questionable. For instance, their account implies that a person without access to long-term memory couldn't possibly understand anything. (“[O]ne can be said to have understood something when a sufficiently well-organised schema... has been encoded into long-term memory” (13).) This seems implausible; surely people can achieve a transitory understanding that fails to consolidate. Perhaps Inglis and Mejía-Ramos could retain the schema formation requirement while dropping the condition on long-term storage.

5 Do all mathematical explanations involve proofs?

It should be evident by now that proofs are a key ingredient of many mathematical explanations. But it's less obvious how the two are related in general. Do mathematical explanations necessarily consist of proofs, or involve proofs in some other essential way?

Mark Steiner seems to have thought so. Although [Steiner 1978a] acknowledges that explanation by proof is only one type of ME (47), it has little to say about alternative cases, and proofs appear to play a supporting role even in these examples. What's more, Steiner's other work on ME ([Steiner 1978b]) makes no distinction between mathematical explanations and explanatory proofs. Most later authors have followed Steiner in this respect. The literature continues to focus on issues of proof, and even if other types of ME are sometimes mentioned, the prevailing attitude seems to be that mathematical explanations are always proof-based in some respect.

Recent work has questioned this mindset. [D'Alessandro forthcoming a] and [Lange 2016], for instance, argue that many cases of ME don't consist of explanatory proofs. Both think that other bits of mathematics—notably theorems, but perhaps also things like theories and diagrams—can serve as explanantia. Mathematical practice seems to bear out such a view, as mathematicians often recognize theorems as explanatory.

Even so, the question remains whether a theorem can genuinely explain in its own right, or only in virtue of its relationship to a proof. D'Alessandro holds the former view, and argues that studies like [Pincock 2015] have gone wrong in looking for explanatory proofs when there are none to be found (as in Pincock's analysis of the Galois theory case). One piece of evidence for this claim is the tendency of mathematicians to judge that a statement *would* be explanatory if it turned out to be true, even when nobody is in a position to guess

what a proof of the theorem would look like. This suggests that explanatory theorems need not “borrow” or “inherit” their status from explanatory proofs.

On Lange’s view, however, “[a] theorem can explain [one of its instances] only if the theorem is no coincidence and hence only if it has a certain kind of proof” ([Lange 2016], 345). Here Lange is talking about a “common proof” of all the theorem’s cases—that is, a proof that reveals some respect in which the cases are alike, and which “proceed[s] from there to arrive at the result by treating all of the (classes of) cases in exactly the same way” ([Lange 2016], 287).²⁴ According to this picture, the explanatory power of a theorem remains tied to the characteristics of its proofs.

Besides appealing to theorems and the like, mathematicians also explain by pointing to arguments that aren’t intended as proofs. [Lange 2017] studies this phenomenon. He shows how a non-proof argument is often given to explain why $(f_1 + f_2 + f_3 + \dots)'(a)$, the derivative of an infinite sum, can differ from $f_1'(a) + f_2'(a) + f_3'(a) + \dots$, the sum of the derivatives of its terms. The explanation is that a certain simple proof of the finite version of the identity breaks down in the infinite case. (Roughly, this happens because one can always choose the smallest δ_n in the epsilon-delta expansions of the relevant limits when there are finitely many terms, but not when there are infinitely many.) This argument doesn’t show that the infinitary statement is *false*, but only that it can’t be proved in a certain way. Still, mathematicians accept the argument as an explanation.

Lange’s analysis of this type of case is an extension of the theory discussed in the previous section. On this view, non-proofs and proofs alike are explanatory when they match some noteworthy feature of a problem with a corresponding feature in the setup. In Lange’s example, the difference between the finite case (where the sum rule holds) and the infinite case (where it breaks down) is the salient feature to be accounted for. The non-proof argument succeeds as an explanation, then, because it reveals “another difference between the [two] cases and show[s] how that difference turns out to make a difference to the sum rule’s holding” (15).

6 Skepticism about explanatory proof

In spite of the many apparent examples of explanatory proof and the philosophical interest they’ve generated, some authors have expressed doubts about the very existence of the phenomenon. [Zelcer 2013] is the most sustained skeptical broadside; [Weber & Frans 2017] gives a response. This section presents some of the reasons for and against taking explanatory proof seriously.

Zelcer’s claims fall into two general categories. The first type of complaint is that mathematicians don’t often talk about explanation or ascribe it much importance. ([Resnik & Kushner 1987] makes the same charge.) The second is that our core commitments about the nature of explanation rule out there being such things as explanatory proofs.

I think the first claim is indefensible. If anyone doubts that mathematicians talk much about explanation, they should peruse the works in this paper’s bibliography, where they’ll find hundreds of examples. A casual internet search will turn up many more. Some of these are offhanded remarks, but plenty are embedded in serious, thoughtful and careful discussions about the goals of mathematics and the merits of different kinds of proofs. Indeed, as the Polymath example from §2 illustrates, the desire for explanation often serves as an explicit impetus for mathematical research.

So I don’t think there’s any serious doubt that mathematicians countenance a relation they call “explanation”, that they believe this relation sometimes holds between proofs and theorems, or that they value and pursue proofs they deem explanatory (in this sense). Unless these practices are some kind of systematic mistake, the term “explanation” as used in such contexts presumably refers to something. The only question is about the nature of this relation. Is it really a type of explanation? Or is it something else, misleadingly called by the same name?

Zelcer holds the latter view. According to him, those who claim that proofs can genuinely explain “are equivocating on ‘explanation’ and are using the word in a way that significantly diverges from the scientific meaning” (174). To show this, Zelcer presents several features that he takes to be hallmarks of scientific explanation, and then argues that proofs lack these features. The apparent differences include

²⁴See [Lange 2010], or Chapter 8 of [Lange 2016], for more on Lange’s notions of mathematical coincidences and common proofs.

the following: (1) explanation and prediction are closely related in science, but there are no predictions in mathematics; (2) every natural fact is presumably explicable, but some mathematical facts can't be explained; (3) scientific explanation is associated with surprise reduction, but mathematical truths are necessary and hence can't be surprising; (4) to grasp a scientific explanation is to gain new scientific knowledge, but mathematical explanations couldn't be informative in this way. Zelcer concludes that mathematics involves nothing like explanations in the standard (objective, scientific) sense. Instead, when mathematicians call proofs explanatory, they may be referring to "mere stylistic features that communicate mathematics more clearly or in a psychologically more satisfying or pedagogically more useful way" (176).

These claims are open to criticism on an individual basis. (For instance, as [Weber & Frans 2017] point out, many accounts of scientific explanation don't posit any particular connection with prediction or surprise reduction. Nor do they entail that everything can be explained.) One can also dispute whether (1)-(4), even if true, would establish Zelcer's skeptical conclusion. After all, mathematical and scientific explanation have common features that are arguably more important than these alleged differences. Explanations of both types generate understanding, for instance, and this accounts for much of their distinctive value. (It's sometimes suggested that this is a definitive feature of explanation in general; cf. [Grimm 2010], [Khalifa 2013], [Strevens 2013], [Wilkenfeld 2014], [Turri 2015], [Waskan et al. 2015].) Moreover, as we've seen, ideas about scientific explanation can often be fruitfully applied to mathematical cases. So there's little reason to insist that proofs can't be explanatory in the standard sense.

Still, we haven't ruled out the possibility that there are *some* interesting general differences between scientific and mathematical explanation. Would we be forced to agree with Zelcer if this turned out to be the case? I don't think so. Zelcer writes as though there are just two options: either ME is just the same sort of thing as scientific explanation, or else it's something entirely different and the label "explanation" is a misnomer. But there's another possibility. Perhaps scientific and mathematical explanation are distinct species falling under a common genus (which may have other members, e.g. metaphysical explanation). In this case, we should expect to find both fundamental similarities and significant differences. (For instance, many philosophers think scientific explanations must be underwritten by ontic or counterfactual dependence relations; [D'Alessandro forthcoming b] argues that some mathematical cases don't work this way. And [Morris 2019] describes some differences in instrumental value between explanatory proofs and scientific explanations.) If this is right, then Zelcer's pessimism is unwarranted; there's nothing equivocal or otherwise inappropriate about calling some proofs explanatory.

7 Conclusion

Explanatory proof is a fascinating phenomenon. Those interested in mathematical practice certainly can't ignore it, but it would be wrong to think of ME as a niche topic in the philosophy of mathematics. If proofs can genuinely explain, then anyone who wants to understand the nature and function of explanation should stand up and take notice.

Much work on the subject remains to be done. The accounts of [Lange 2014] and [Inglis & Mejía-Ramos 2019] are the only general theories currently on the table, and they have yet to receive much critical attention. In any case, the literature to date leaves important questions unanswered about the relationship between ME and other elements of the theory of explanation. For example: Can explanatory proof ultimately be reconciled with the ontic or counterfactual conceptions of explanation, which have dominated metaphysics and philosophy of science in recent years? If so, how? If not, what general notion of explanation can cover all of these cases? What should we make of persistent disagreements among mathematicians about how best to explain some theorems? I hope this paper will help encourage students and researchers in a variety of fields to start taking up these issues.

References

- [Aigner & Zeigler 2010] Aigner, Martin and Günter M. Zeigler. 2010. *Proofs from the Book* (4th edition). Berlin: Springer.

- [Andersen 2014a] Andersen, Holly. 2014. "A field guide to mechanisms: Part I." *Philosophy Compass* 9, 274-283.
- [Andersen 2014b] Andersen, Holly. 2014. "A field guide to mechanisms: Part II." *Philosophy Compass* 9, 284-293.
- [Baker 2005] Baker, Alan. 2005. "Are there genuine mathematical explanations of physical phenomena?" *Mind* 114, 223-238.
- [Baker 2009] Baker, Alan. 2009. "Mathematical accidents and the end of explanation." In Otávio Bueno and Øystein Linnebo (eds.), *New Waves in Philosophy of Mathematics*, New York: Palgrave Macmillan, 137-159.
- [Baker 2010] Baker, Alan. 2010. "Mathematical induction and explanation." *Analysis* 70, 681-689.
- [Baldwin 2016] Baldwin, John T. 2016. "Foundations of mathematics: Reliability and clarity: The explanatory role of mathematical induction." In Jouko Väänänen, Åsa Hirvonen and Ruy de Queiroz (eds.), *Logic, Language, Information and Computation: Proceedings of the 23rd International Workshop*, Berlin: Springer-Verlag, 68-82.
- [Bassler 2006] Bassler, O. Bradley. 2006. "The surveyability of mathematical proof: A historical perspective." *Synthese* 148, 99-133.
- [Bechtel & Abrahamsen 2005] Bechtel, William and Adele Abrahamsen. 2005. "Explanation: A mechanist alternative." *Studies in History and Philosophy of the Biological and Biomedical Sciences* 36, 421-441.
- [Bennett 2011] Bennett, Karen. 2011. "Construction area (no hard hat required)." *Philosophical Studies* 154, 79-104.
- [Brown 1997] Brown, James R. 1997. "Proofs and pictures." *British Journal for the Philosophy of Science* 48, 161-180.
- [Bueno & Linnebo 2009] Bueno, Otávio and Øystein Linnebo (eds.). 2009. *New Waves in Philosophy of Mathematics*. New York: Palgrave Macmillan.
- [Colyvan 2012] Colyvan, Mark. 2012. *An Introduction to the Philosophy of Mathematics*. Cambridge: Cambridge University Press.
- [Craver & Tabery 2017] Craver, Carl and James Tabery. 2017. "Mechanisms in science." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 edition), URL = <<https://plato.stanford.edu/archives/spr2017/entries/science-mechanisms/>>.
- [D'Alessandro 2018] D'Alessandro, William. 2018. "Arithmetic, set theory, reduction and explanation." *Synthese* 195, 5059-5089.
- [D'Alessandro forthcoming a] D'Alessandro, William. Forthcoming. "Mathematical explanation beyond explanatory proof." *British Journal for the Philosophy of Science*. DOI: 10.1093/bjps/axy009.
- [D'Alessandro forthcoming b] D'Alessandro, William. Forthcoming. "Viewing-as explanations and ontic dependence." *Philosophical Studies*. DOI: 10.1007/s11098-018-1205-5.
- [Delarivière et al. 2017] Delarivière, Sven, Joachim Frans and Bart Van Kerkhove. 2017. "Mathematical explanation: A contextual approach." *Journal of the Indian Council of Philosophical Research* 34, 309-329.

- [Detlefsen 1988] Detlefsen, Michael. 1988. “Fregean hierarchies and mathematical explanation.” *International Studies in the Philosophy of Science* 3, 97-116.
- [Dougherty 2017] Dougherty, John. 2017. “What inductive explanations could not be.” *Synthese* 195, 5473-5483.
- [Easwaran 2008] Easwaran, Kenny. 2008. “The role of axioms in mathematics.” *Erkenntnis* 68, 381-391.
- [Easwaran 2009] Easwaran, Kenny. 2009. “Probabilistic proofs and transferability.” *Philosophia Mathematica* 17, 341-362.
- [Fine 1994] Fine, Kit. 1994. “Essence and modality.” *Philosophical Perspectives* 8, 1-16.
- [Frans & Weber 2014] Frans, Joachim and Erik Weber. 2014. “Mechanistic explanation and explanatory proofs in mathematics.” *Philosophia Mathematica* 22, 231-248.
- [Friedlein 1873] Friedlein, Gottfried. 1873. *Procli Diadochi in Primum Euclidis Elementorum Librum Commentariū*. Leipzig: B.G. Teubner.
- [Giaquinto 2007] Giaquinto, Marcus. 2007. *Visual Thinking in Mathematics: An Epistemological Study*. New York: Oxford University Press.
- [Gouvêa 2015] Gouvêa, Fernando Q. 2015. Review of *The Quadratic Reciprocity Law: A Collection of Classical Proofs*, by Oswald Baumgart, ed. and trans. by Franz Lemmermeyer. *MAA Reviews*, June 30, 2015, online at <<https://www.maa.org/press/maa-reviews/the-quadratic-reciprocity-law-a-collection-of-classical-proofs>>.
- [Gray 2014] Gray, Jeremy. 2014. “Depth—A Gaussian tradition in mathematics.” *Philosophia Mathematica* 23, 177-195.
- [Grimm 2010] Grimm, Stephen R. 2010. “The goal of explanation.” *Studies in the History and Philosophy of Science* 41, 337-344.
- [Gullberg 1997] Gullberg, Jan. 1997. *Mathematics: From the Birth of Numbers*. New York: Norton.
- [Hanna 1990] Hanna, Gila. 1990. “Some pedagogical aspects of proof.” *Interchange* 21, 6-13.
- [Harari 2008] Harari, Orna. 2008. “Proclus’ account of explanatory demonstrations in mathematics and its context.” *Archiv für Geschichte der Philosophie* 90, 137-164.
- [Hafner & Mancosu 2005] Hafner, Johannes and Paolo Mancosu. 2005. “The varieties of mathematical explanation.” In Paolo Mancosu, Klaus Frovin Jørgensen and Stig Andur Pedersen (eds.), *Visualization, Explanation and Reasoning Styles in Mathematics*, Berlin: Springer, 215-250.
- [Hafner & Mancosu 2008] Hafner, Johannes and Paolo Mancosu. 2008. “Beyond unification.” In Paolo Mancosu (ed.), *The Philosophy of Mathematical Practice*, Oxford: Oxford University Press, 151-178.
- [Hempel & Oppenheim 1948] Hempel, Carl and Paul Oppenheim. 1948. “Studies in the logic of explanation.” *Philosophy of Science* 15, 135-75.
- [Hoeltje et al. 2013] Hoeltje, Miguel, Benjamin Schnieder and Alex Steinberg. 2013. “Explanation by induction?” *Synthese* 190, 509-524.
- [Inglis & Aberdein 2015] Inglis, Matthew and Andrew Aberdein. 2015. “Beauty is not simplicity: An analysis of mathematicians’ proof appraisals.” *Philosophia Mathematica* 23, 87-109.

- [Inglis & Mejía-Ramos 2019] Inglis, Matthew and Juan Pablo Mejía-Ramos. 2019. “Functional explanation in mathematics.” *Synthese*. DOI: 10.1007/s11229-019-02234-5.
- [Khalifa 2013] Khalifa, Kareem. 2013. “The role of explanation in understanding.” *British Journal for the Philosophy of Science* 64, 161-187.
- [Kitcher 1975] Kitcher, Phillip. 1975. “Bolzano’s ideal of algebraic analysis.” *Studies in History and Philosophy of Science* 6, 229-269.
- [Kitcher 1989] Kitcher, Phillip. 1989. “Explanatory unification and the causal structure of the world.” In Phillip Kitcher and Wesley Salmon (eds.), *Scientific Explanation* (Minnesota Studies in the Philosophy of Science, Volume XIII), Minneapolis: University of Minnesota Press, 410-505.
- [Lange 2009] Lange, Marc. 2009. “Why proofs by mathematical induction are generally not explanatory.” *Analysis* 69, 203-211.
- [Lange 2010] Lange, Marc. 2010. “What are mathematical coincidences (and why does it matter)?” *Mind* 119, 307-340.
- [Lange 2014] Lange, Marc. 2014. “Aspects of mathematical explanation: Symmetry, unity, and salience.” *Philosophical Review* 123, 485-531.
- [Lange 2015] Lange, Marc. 2015. “Explanation, existence and natural properties in mathematics—A case study: Desargues’ theorem.” *Dialectica* 69, 435-472.
- [Lange 2016] Lange, Marc. 2016. *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. New York: Oxford University Press.
- [Lange 2017] Lange, Marc. 2017. “Mathematical explanations that are not proofs.” *Erkenntnis*, DOI: 10.1007/s10670-017-9941-z.
- [Lehet 2019] Lehet, Ellen. 2019. “Induction and explanatory definitions in mathematics.” *Synthese*, DOI: 10.1007/s11229-019-02095-y.
- [Machamer et al. 2000] Machamer, Peter, Lindley Darden and Carl F. Craver. 2000. “Thinking about mechanisms.” *Philosophy of Science* 67, 1-25.
- [Maddy 2011] Maddy, Penelope. 2011. *Defending the Axioms: On the Philosophical Foundations of Set Theory*. New York: Oxford University Press.
- [Mancosu 1999] Mancosu, Paolo. 1999. “Bolzano and Cournot on mathematical explanation.” *Revue d’Histoire des Sciences* 52, 429-455.
- [Mancosu 2000] Mancosu, Paolo. 2000. “On mathematical explanation.” In Emily Grosholz and Herbert Breger (eds.), *The Growth of Mathematical Knowledge*, Dordrecht: Kluwer, 103-119.
- [Mancosu 2001] Mancosu, Paolo. 2001. “Mathematical explanation: Problems and prospects.” *Topoi* 20, 97-117.
- [Mancosu 2008a] Mancosu, Paolo. 2008. “Mathematical explanation: Why it matters.” In Paolo Mancosu (ed.), *The Philosophy of Mathematical Practice*, New York: Oxford University Press, 134-150.
- [Mancosu 2008b] Mancosu, Paolo (ed.). 2008. *The Philosophy of Mathematical Practice*. New York: Oxford University Press.

- [Mancosu 2018] Mancosu, Paolo. 2018. “Explanation in mathematics.” *The Stanford Encyclopedia of Philosophy* (Summer 2018 edition), ed. Edward N. Zalta, URL = <<https://plato.stanford.edu/archives/sum2018/entries/mathematics-explanation/>>.
- [Mancosu & Hafner 2008] Mancosu, Paolo and Johannes Hafner. 2008. “Unification and explanation: A case study from real algebraic geometry.” In Paolo Mancosu (ed.), *The Philosophy of Mathematical Practice*, New York: Oxford University Press, 151-178.
- [Mizrahi 2016] Mizrahi, Moti. 2016. “Why simpler arguments are better.” *Argumentation* 30, 247-261.
- [Morris 2019] Morris, Rebecca Lee. 2019. “Do mathematical explanations have instrumental value?” *Synthese*. DOI: 10.1007/s11229-019-02114-y.
- [Pincock 2015] Pincock, Christopher. 2015. “The unsolvability of the quintic: A case study in abstract mathematical explanation.” *Philosophers’ Imprint* 15, 1-19.
- [Resnik & Kushner 1987] Resnik, Michael D. and David Kushner. 1987. “Explanation, independence and realism in mathematics.” *British Journal for the Philosophy of Science* 38, 141-158.
- [Reutlinger 2017] Reutlinger, Alexander. 2017. “Explanation beyond causation? New directions in the philosophy of scientific explanation.” *Philosophy Compass*, DOI: 10.1111/phc3.12395.
- [Ruben 1990] Ruben, David-Hillel. 1990. *Explaining Explanation*. New York: Routledge.
- [Salmon 1984] Salmon, Wesley. 1984. “Scientific explanation: Three basic conceptions.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2, 293-305.
- [Salmon 1989] Salmon, Wesley. 1989. *Four Decades of Scientific Explanation*. University of Pittsburgh Press: Pittsburgh.
- [Salverda 2018] Salverda, Josephine. 2018. “Can proofs by mathematical induction be explanatory?” In Mario Piazza and Gabriele Pulcini (eds.), *Truth, Existence and Explanation: FilMat 2016 Studies in the Philosophy of Mathematics*, Cham, Switzerland: Springer, 163-188.
- [Sandborg 1998] Sandborg, David. 1998. “Mathematical explanation and the theory of why-questions.” *British Journal for the Philosophy of Science* 49, 603-624.
- [Seymour 2016] Seymour, Paul. 2016. “Hadwiger’s conjecture.” In John Forbes Nash and Michael Rassias (eds.), *Open Problems in Mathematics*, Berlin: Springer, 417-437.
- [Speyer 2016] Speyer, David E. 2016. “Some sums over irreducible polynomials.” arXiv:1608.03014.
- [Steiner 1978a] Steiner, Mark. 1978. “Mathematical explanation.” *Philosophical Studies* 34, 135-151.
- [Steiner 1978b] Steiner, Mark. 1978. “Mathematics, explanation, and scientific knowledge.” *Noûs* 12, 17-28.
- [Strevens 2013] Strevens, Michael. 2013. “No understanding without explanation.” *Studies in History and Philosophy of Science Part A* 44, 510-515.

- [Tappenden 2005] Tappenden, Jamie. 2005. “Proof style and understanding in mathematics I: Visualization, unification and axiom choice.” In P. Mancosu, K. Jørgensen and S. Pedersen (eds.), *Visualization, Explanation and Reasoning Styles in Mathematics*, Berlin: Springer, 147-214.
- [Tatzel 2002] Tatzel, Armin. 2002. “Bolzano’s theory of ground and consequence.” *Notre Dame Journal of Formal Logic* 43, 1-25.
- [Tao 2015] Tao, Terence. December 28, 2015. “Polymath proposal: explaining identities for irreducible polynomials.” *The Polymath Blog*. URL = <<http://polymathprojects.org/2015/12/28/polymath-proposal-explaining-identities-for-irreducible-polynomials/>>.
- [Turri 2015] Turri, John. 2015. “Understanding and the norm of explanation.” *Philosophia* 43, 1171-1175.
- [Waskan et al. 2015] Waskan, Jonathan, Ian Harmon, Zachary Horne, Joseph Spino and John Cleverger. 2014. “Explanatory anti-psychologism overturned by lay and scientific case classifications.” *Synthese* 191, 1013-1035.
- [Weber & Frans 2017] Weber, Erik and Joachim Frans. 2017. “Is mathematics a domain for philosophers of explanation?” *Journal for General Philosophy of Science* 48, 125-142.
- [Weber & Verhoeven 2002] Weber, Erik and Liza Verhoeven. 2002. “Explanatory proofs in mathematics.” *Logique et Analyse* 179-180, 299-307.
- [Weisstein 2018] Weisstein, Eric W. May 30, 2018. “Abstract group.” *MathWorld*—A Wolfram Web Resource. URL = <<http://mathworld.wolfram.com/AbstractGroup.html>>.
- [Wilkenfeld 2014] Wilkenfeld, Daniel A. 2014. “Functional explaining: A new approach to the philosophy of explanation.” *Synthese* 191, 3367-3391.
- [Woodward 2003] Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- [Woodward 2017] Woodward, James. 2017. “Scientific explanation.” In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), URL = <<https://plato.stanford.edu/archives/fall2017/entries/scientific-explanation/>>.
- [Wysocki 2017] Wysocki, Tomasz. 2017. “Explanatory circles, induction, and recursive structures.” *Thought* 6, 13-16.
- [Zelcer 2013] Zelcer, Mark. 2013. “Against mathematical explanation.” *Journal for General Philosophy of Science* 44, 173-192.