Comparing Mathematical Explanations

Isaac Wilhelm

(please cite version in *The British Journal for the Philosophy of Science*)

Abstract

Philosophers have developed several detailed accounts of what makes some mathematical proofs explanatory. Significantly less attention has been paid, however, to what makes some proofs more explanatory than other proofs. That is problematic, since the reasons for thinking that some proofs explain are also reasons for thinking that some proofs are more explanatory than others. So in this paper, I develop an account of comparative explanation in mathematics. I propose a theory of the 'at least as explanatory as' relation among mathematical proofs.

1 Introduction

Some mathematical proofs are more explanatory than others. Euclid's proof of the Pythagorean theorem, for example, is more explanatory than the proof from the law of cosines.[1] The double-sum proof of the fact that $1 + 2 + \cdots + n = n(n + 1)/2$, for each natural number n,[2] is generally taken to be more explanatory than the inductive proof (Steiner, 1978,

[1]For the latter proof, substitute 90° into the angle variable in that law.

[2]The double-sum proof is as follows. Let S equal $1 + 2 + \cdots + n$. Then

$$\begin{array}{rcccccccc} & S & = & 1 & + & 2 & + \cdots + & n \\ + & S & = & n & + & (n-1) & + \cdots + & 1 \\ \hline & 2S & = & (n+1) & + & (n+1) & + \cdots + & (n+1) \end{array}$$

where the expression below the line features n copies of $n + 1$. Therefore, $2S = n(n + 1)$, so $1 + 2 + \cdots + n = n(n + 1)/2$.

p. 136; Lange, 2017, pp. 281-282).[3] More generally, brute force proofs are less explanatory than non-brute force proofs. And reductio proofs are generally less explanatory than non-reductio proofs.

Contemporary accounts of mathematical explanation do not focus on cases like these. They do not focus on *comparative* mathematical explanation: they do not focus on what makes some proofs more explanatory than other proofs. Instead, they focus on *non-comparative* mathematical explanation: they focus on what makes some proofs explanatory and others not (Kitcher, 1989; Lange, 2014a; Steiner, 1978). Of course, proponents of these accounts generally recognize that proofs can fall somewhere between being explanatory and being non-explanatory: some proofs are borderline cases of explanations; some proofs are only somewhat explanatory (Lange, 2017, p. 298). But these borderline, comparative cases are not the primary theoretical targets.

In this paper, I focus on explanatory comparisons in mathematics. I posit a relation—call it the 'comparative explanatory relation'—that obtains between two proofs whenever one is at least as explanatory as the other. After discussing some formal features of the comparative explanatory relation, I present two accounts of it: one takes the relation to be primitive, and the other analyzes it in terms of simplicity and structural depth. I also provide some reasons for favoring the non-primitive account.

As will become clear, it is worth switching our attention from non-comparative mathematical explanation to comparative mathematical explanation. For as I discuss later on, non-comparative explanation can be analyzed in terms of comparative explanation. It is hard to see how comparative explanation could be analyzed in terms of non-comparative explanation, however. So one of the upshots of this paper is a recommended shift of focus: the literature on mathematical explanation would do well to focus on comparative explanation first, and non-comparative explanation second.

[3] One might disagree: perhaps the inductive proof is more explanatory because (i) the inductive proof relies on the property of being a successor, and (ii) that property is extremely explanatory when it comes to explaining number-theoretic facts. My account of mathematical explanation is compatible with this alternative judgment.

In addition to contributing to the literature on mathematical explanation, this paper helps fill a gap in the literature on explanation in general. Comparative explanation is a relatively undertheorized phenomenon: accounts of causal explanation, mathematical explanation, and so on, generally focus on what it takes for something to explain, rather than what it takes for something to be more explanatory than something else.[4] In this paper, I take some initial steps towards a more general theory of comparative explanation.

In Section 2, I explain why we need a comparative account of explanation in mathematics. In Section 3, I present the primitive and non-primitive accounts of the comparative explanatory relation, and I argue for the latter. In Section 4, I discuss some attractive features of these accounts. Finally, in Section 5, I discuss some problems that arise for an alternative account based on ideas discussed by Lange (2014a; 2014b).

2 Motivation

There are several reasons for thinking that some mathematical proofs are more explanatory than others. Each is analogous to—or exactly the same as—a reason for thinking that some mathematical proofs explain. So insofar as one thinks that some proofs are explanatory, one should think that some proofs are more explanatory than other proofs.

The first reason concerns examples: just as there are intuitively compelling cases of explanatory proofs, there are intuitively compelling cases of some proofs being more explanatory than others. Consider the 'calculator number theorem' (Lange, 2017, pp. 276-277). A calculator number is a six-digit number obtained by taking three digits from any row, column, or main diagonal on a calculator keypad in forward and then reverse order.

[4] For some exceptions, see (Hempel & Oppenheim, 1948) and (Hitchcock & Woodward, 2003).

7	8	9
4	5	6
1	2	3

So for example, 123321 is a calculator number, as are 852258 and 357753. The calculator number theorem says that every calculator number is divisible by 37. One proof simply goes through all sixteen calculator numbers and shows, by brute force calculation, that 37 divides each. Another proof exploits a simple structural property of calculator numbers: the three digits used to construct a given calculator number form an arithmetical progression.[5] The latter proof is more explanatory than the former proof. And the lesson here seems to generalize: in general, brute force proofs are less explanatory than non-brute force proofs.

There are other reasons to think that some proofs are more explanatory than others. Perhaps most importantly, just as mathematical practice provides a reason to think that some proofs explain, mathematical practice also provides a reason to think that some proofs are more explanatory than other proofs. For mathematicians frequently claim as much. While discussing the relationship between equalizers and monomorphisms in category theory, for example, Goldblatt writes "the next theorem gives a somewhat deeper explanation of the situation just described" (1984, p. 57). After pointing out that the nth triangular number is a binomial coefficient, Alsina and Nelson say that "[o]ne explanation for this is that each [nth triangular number] is equal to $n(n+1)/2$, but this answer sheds little light on why it is true. Here is a better explanation using the Cantor principle" (2010, p. 5).

Mathematical education researchers often make similar claims: for example, Hanna writes that "[s]ome proofs are by their nature more explanatory than others" (2000, p. 8). And so do philosophers. Lange writes that "[t]he proof of the differentiation expansion theorem from the three laws of combination arguably provides a *deeper* explanation of the theorem than its proof from premises concerning only differentiation" (2014b, p. 12). And Brown

[5] Since those three digits can be represented as the arithmetical progression a, $a+d$, and $a+2d$ for some a and d, each calculator number can be written as $10^5a + 10^4(a+d) + 10^3(a+2d) + 10^2(a+2d) + 10(a+d) + a$, which reduces to $(3 \cdot 11 \cdot 37)(91a + 10d)$. Hence, each calculator number is divisible by 37.

writes that "[i]n the two number theory cases above, a proof by induction is probably more insightful and explanatory than the picture-proofs" (1999, p. 34).[6]

To summarize: the sorts of reasons philosophers usually give for thinking that mathematical proofs explain—intuitive examples of explanatory proofs, observations of mathematical practice, and so on (Lange, 2014a; Mancosu, 2000)—are *also* reasons for thinking that some proofs are more explanatory than others. So to capture this comparative feature of proof-based explanation in mathematics, we need more than accounts of what distinguishes explanatory proofs from non-explanatory proofs. We also need accounts of what makes some proofs more explanatory than others.

Before presenting some such accounts, however, I should make three clarificatory remarks. First, two kinds of mathematical explanation are discussed at length in the philosophical literature: mathematical explanations of physical facts, and mathematical explanations of mathematical facts. In this paper, I investigate the latter.

Second, some mathematical explanations of mathematical facts may not be proofs. Some mathematical explanations may involve the conceptual recasting of an entire discipline (Mancosu, 2008, p. 142). And some mathematical explanations cite entire theories (D'Alessandro, 2018, p. 4). In this paper, I do not explore explanations like these. I restrict my attention to the explanatory capacities of proofs.

Third, in this paper, I focus on the objective basis for comparative explanation in mathematics. I do not explore the more subjective elements of comparative mathematical explanation. For instance, I do not explore the ways in which agents' background knowledge, cognitive capacities, cultural contexts, and so on, may affect their perception of the explanatoriness of a proof. Instead, I explore an objective worldly relation—the comparative explanatory relation—which backs, or licenses, or underwrites, agents' subjective judgments about whether one proof is more explanatory than another. This objective worldly relation obtains between proofs and theorems, where proofs and theorems are taken to be objec-

[6]The first number theory case is the theorem that for any natural number n, $1 + 3 + \cdots + (2n - 1) = n^2$. The second number theory case is the theorem that for any natural number n, $1 + 2 + \cdots + n = n(n + 1)/2$.

tive, worldly entities: abstract objects, concrete structures, or perhaps something else. Of course, there are many interesting issues concerning the connection between the objective comparative explanatory relation and our subjective judgments. But they deserve their own paper.

3 Two Comparative Accounts of Mathematical Explanation

In this section, I present two different accounts of the comparative explanatory relation. First, I give a rigorous definition of the kind of proof that this relation compares. Second, I discuss some of this relation's formal features. Third, I present the two accounts. One takes the comparative explanatory relation to be primitive. The other analyzes that relation in terms of the relative simplicity and structural depth of mathematical proofs. As will become clear, I prefer the latter account.

Here is the definition of a proof. Let t be a theorem and let Γ be a collection of formulas. Then $\mathfrak{p}$ is a 'proof of t from Γ' just in case $\mathfrak{p}$ is a finite sequence of formulas $\langle f_1, f_2, \ldots, f_n \rangle$ such that the following two conditions hold.

1. The last formula, f_n, is t.
2. For all i, either f_i is an axiom, or f_i is in Γ, or there are $j, k < i$ such that f_i is the result of applying modus ponens to f_j and f_k.

So a proof of a theorem is just a bunch of formulas which satisfy the standard constraints of a Hilbert-style deductive system. In what follows, I will often refer to Γ as a collection of premises for the proof in question.

One might worry that the above definition of a proof is overly restrictive: for instance, many informal proofs do not fit the criteria of a Hilbert-style deductive system. I think that this is indeed a constraint of the above definition, but for two reasons, it is not problematic here. First, many informal proofs can be rewritten so that they conform to the conditions of a Hilbert-style deductive system. So the above definition does not classify all informal derivations as non-proofs: many count as proofs, on the above definition. Second, in this paper, not much hangs on this particular definition of what a proof is. I adopt this definition because (i) it is clear and precise, and (ii) it keeps the forthcoming discussion rigorous. But a different definition could be used instead.

For the purposes of this paper, the axioms are those of first-order logic: I study the explanatory capacities of first-order logical proofs. Many of my claims may apply to proofs in different axiomatic systems: proofs which, for instance, invoke the axioms of intuitionistic logic, or the axioms of second-order logic. But some of my claims may not. So in this paper, I remain neutral as to what makes for more explanatory proofs in other logical systems. I remain neutral, for example, on what makes some second-order logical proofs more explanatory than others.[7] I will say a bit more about this, however, later on.

The comparative explanatory relation obtains between proofs of the same theorem, from the same collection of premises.[8] For a given theorem t and a given collection of premises Γ, denote this relation by $\succeq_{t,\Gamma}$.[9] So the expression $\mathfrak{p} \succeq_{t,\Gamma} \mathfrak{q}$ is shorthand[10] for the following: proof $\mathfrak{p}$ of theorem t from collection of premises Γ is at least as explanatory as proof $\mathfrak{q}$ of t from Γ.[11]

[7]There is not much literature on whether explanatoriness is invariant across different logical systems. A sequence of formulas only counts as a proof relative to some specified logic: a proof in classical first-order logic, for instance, may not be a proof in intuitionistic logic. So a question arises: are proofs in different systems explanatory in virtue of the same sorts of things? For instance, suppose that a first-order logical proof is explanatory just in case that proof has property P. Is P what makes other proofs, in other logics, explanatory? Is a second-order logical proof explanatory just in case that proof has P? Or are the standards of explanatoriness different, for proofs in second-order logic? These questions are interesting, and they deserve investigation. But for brevity's sake, I do not pursue them here.

[8]For reasons of space, I do not explore the question of whether a proof of one theorem can be more explanatory than a proof of some other theorem.

[9]Because of the relativity of this relation to theorems and premises, it is more accurate to describe this relation as "the comparative explanatory relation relative to t and Γ" rather than "the comparative explanatory relation". For brevity, however, I use the latter expression.

[10]Note that for ease of exposition, in this paper, I often slide between using a formula and mentioning it.

[11]For brevity, I assume that the comparative explanatory relation is not context-sensitive: any particular instance of it obtains, or fails to obtain, independently of context. I also assume that the comparative explanatory relation is not vague: there is always a non-vague fact of the matter as to whether any particular instance of it obtains or fails to obtain. But it is not difficult to incorporate context-sensitivity, or vagueness, into this account. Just add a context-dependent parameter c to the subscript of the $\succeq$ symbol: on the new formalism, the expression $\mathfrak{p} \succeq_{t,\Gamma,c} \mathfrak{q}$ is shorthand for the claim that relative to context c, proof $\mathfrak{p}$ of theorem t from collection of premises Γ is at least as explanatory as proof $\mathfrak{q}$ of t from Γ. This account of the comparative explanatory relation allows for vagueness: perhaps there is a context c such that in c, it is vague whether or

One might worry about the fact that, on this approach to comparative mathematical explanation, the explanatory capacities of proofs are only ever compared relative to a common collection of premises. For that seems to imply the following: if $\mathfrak{p}$ and $\mathfrak{q}$ are proofs of the same theorem t, but $\mathfrak{p}$ and $\mathfrak{q}$ use different premises, then the explanatory capacities of $\mathfrak{p}$ and $\mathfrak{q}$ cannot be compared. In other words, this approach to comparative explanation seems to imply that if $\mathfrak{p}$ is a proof of t from a collection of premises Γ_1, and $\mathfrak{q}$ is a proof of t from a different collection of premises Γ_2, then $\mathfrak{p}$ and $\mathfrak{q}$ cannot stand in the comparative explanatory relation.

For a somewhat subtle reason, however, that is not quite right. Perhaps the explanatory capacities of $\mathfrak{p}$ and $\mathfrak{q}$ cannot be compared relative to Γ_1, and perhaps the explanatory capacities of $\mathfrak{p}$ and $\mathfrak{q}$ cannot be compared relative to Γ_2. Nevertheless, the explanatory capacities of $\mathfrak{p}$ and $\mathfrak{q}$ may still be comparable relative to some other collection of premises: for instance, $\Gamma_1 \cup \Gamma_2$. In other words, suppose that neither $\mathfrak{p} \succeq_{t,\Gamma_1} \mathfrak{q}$, nor $\mathfrak{q} \succeq_{t,\Gamma_1} \mathfrak{p}$, nor $\mathfrak{p} \succeq_{t,\Gamma_2} \mathfrak{q}$, nor $\mathfrak{q} \succeq_{t,\Gamma_2} \mathfrak{p}$, obtains. That is perfectly compatible with $\mathfrak{p} \succeq_{t,\Gamma_1 \cup \Gamma_2} \mathfrak{q}$ obtaining. So the explanatory capacities of proofs from different premises can, indeed, be comparable: they are comparable relative to all those premises taken together.

For example, consider Saidak's proof (2006) and Euler's proof (Ribenboim, 1989, pp. 7-8) of Euclid's theorem: the theorem that there are infinitely many primes. Saidak's proof is based on principles of elementary arithmetic.[12] Euler's proof—or rather, a straightforward modification of it which eliminates the appeal to *reductio ad absurdum*—is based on principles of elementary analysis.[13] Relative to just the elementary arithmetic principles, or just the elementary analysis principles, the explanatory capacities of these proofs cannot be compared. But relative to the collection consisting of both the elementary arithmetic principles *and* the

not $\mathfrak{p} \succeq_{t,\Gamma,c} \mathfrak{q}$ obtains. In contexts like that, the details of the context do not settle whether or not proof $\mathfrak{p}$ of t from Γ is at least as explanatory as proof $\mathfrak{q}$ of t from Γ.

[12]The proof consists in constructing an infinite sequence of numbers, each divisible by a prime that does not divide any of the prior elements of the sequence; it follows that there are infinitely many primes.

[13]This version of Euler's proof shows that the product of a set of numbers all greater than one, each of which is a function of a distinct prime, is infinite; it follows that this set of numbers—and thus, the number of primes—is infinite too.

elementary analysis principles, the explanatory capacities of these proofs may be comparable.

I take this to be a feature, not a bug, of the present approach to comparative mathematical explanation. For it has the following nice implication: the explanatory capacities of proofs can vary, depending on how the mathematics is axiomatized. When a mathematical theory is axiomatized one way—that is, if one collection of formulas Γ is used to axiomatize it—then proof $\mathfrak{p}$ may be more explanatory than proof $\mathfrak{q}$. But if that same mathematical theory is axiomatized differently—using, say, a different collection of formulas Γ'—then proof $\mathfrak{q}$ may be more explanatory than proof $\mathfrak{p}$. In other words, this approach to comparative mathematical explanation respects the following possibility: the explanatoriness of a proof can vary, depending on what is taken to be axiomatic. And theories of comparative mathematical explanation should respect that. Axioms, after all, are like fundamental explainers: they are the basic elements from which many mathematical explanations are built. So plausibly, the explanatory capacities of proofs can indeed vary, depending on what those fundamental explainers are.

For those who disagree, however, I offer the following. Let Γ be some particular, natural, explanatorily privileged collection of formulas. Perhaps Γ consists of all axioms for some particular collection of mathematical theories: number theory, analysis, topology, and so on.[14] Or perhaps Γ is empty: so only proofs which establish conditional theorems—that is, theorems of the form "If thus-and-so, then such-and-such"—have comparable explanatory capacities.[15] Then hold that all explanatory comparisons must be made relative to that particular choice of premises. In other words, hold that the comparative explanatory relation only ever obtains between proofs whose premises are in Γ; call this the 'fixed premises' view of the comparative explanatory relation. Throughout the rest of this paper, none of my

[14] Thanks to an editor for suggesting this.

[15] Note that if Γ is empty, then the above definition of a proof implies that the only assumptions which proofs can invoke are axioms of first-order logic. This might sound problematic, but it is not. By the deduction theorem, any given theorem—that follows from certain premises—can be reformulated as a conditional which contains those premises in its antecedent. For instance, let ϕ be the conjunction of the axioms of number theory which are used to establish Euler's theorem (including all the requisite instances of whichever axiom schemas are needed). Then Euler's theorem can be reformulated like this: if ϕ, then there are infinitely many primes.

claims require any variation in Γ. So everything to follow is perfectly compatible with the fixed premises view.[16] Readers who dislike my claim that Γ can vary, from one explanatory comparison to another, are welcome to endorse the fixed premises view instead.

The comparative explanatory relation is a preorder: it is transitive and reflexive. It is transitive because if one proof is at least as explanatory as a second, and the second proof is at least as explanatory as a third, then the first proof is at least as explanatory as the third too. In other words, for any proofs $\mathfrak{p}$, $\mathfrak{q}$, and $\mathfrak{r}$ of theorem t from collection of formulas Γ, the following holds: if $\mathfrak{p} \succeq_{t,\Gamma} \mathfrak{q}$ and $\mathfrak{q} \succeq_{t,\Gamma} \mathfrak{r}$, then $\mathfrak{p} \succeq_{t,\Gamma} \mathfrak{r}$.

The comparative explanatory relation is reflexive because each proof is at least as explanatory as itself. As it turns out, however, the reflexivity of the comparative explanatory relation can be derived from a different—but equally intuitive—constraint. The constraint relates the explanatory capacities of proofs to the explanatory capacities of proofs which they contain. In particular, say that proof $\mathfrak{p}$ of t from Γ is a 'subproof' of proof $\mathfrak{q}$ of t from Γ just in case $\mathfrak{p}$ is a subsequence of $\mathfrak{q}$. If $\mathfrak{p}$ is a proper subproof of $\mathfrak{q}$,[17] then $\mathfrak{q}$ contains unnecessary fluff. Perhaps $\mathfrak{q}$ contains several lines which are totally unrelated to the theorem in question: maybe $\mathfrak{q}$ invokes axioms that are completely unnecessary for the proof of t, for instance; or maybe $\mathfrak{q}$ first establishes a totally unrelated theorem $t^\star$, and then goes on to establish t from scratch. Then here is another constraint to which the comparative explanatory relation conforms.

> SUBPROOF
>
> Let Γ be a collection of formulas. Let $\mathfrak{p}$ be a proof of theorem t from Γ, and let $\mathfrak{q}$ be a proof of t from Γ. If $\mathfrak{p}$ is a subproof of $\mathfrak{q}$, then $\mathfrak{p} \succeq_{t,\Gamma} \mathfrak{q}$.

Since every proof $\mathfrak{p}$ of t from Γ is a subproof of itself, SUBPROOF implies that every proof is at least as explanatory as itself. That is, SUBPROOF implies that the comparative explanatory relation is reflexive.

[16] If one adopted the fixed premises view, of course, then parts of the rest of this paper would have to be rephrased: descriptions of some examples would need rephrasing, for instance. But those rephrasings would be merely cosmetic. The philosophical upshots of the examples would remain the same.

[17] Say that $\mathfrak{p}$ is a 'proper subproof' of $\mathfrak{q}$ just in case $\mathfrak{p}$ is a subproof of $\mathfrak{q}$ and $\mathfrak{q}$ is not a subproof of $\mathfrak{p}$.

The comparative explanatory relation may not be total. That is, there may be a theorem t, a collection of premises Γ, and proofs $\mathfrak{p}$ and $\mathfrak{q}$ of t from Γ, such that $\mathfrak{p}$ is not at least as explanatory as $\mathfrak{q}$ and $\mathfrak{q}$ is not at least as explanatory as $\mathfrak{p}$. I think this is might happen sometimes: perhaps some proofs' explanatory capacities are incomparable. But I will not argue for that here. I remain neutral with respect to whether, for each t and each Γ, $\succeq_{t,\Gamma}$ is total.

I have postulated the comparative explanatory relation, and I have described some of its features. But I have not yet analyzed that relation. That is, I have not specified the conditions under which one proof is more explanatory than another. So in the remainder of this section, I propose and evaluate two accounts of the comparative explanatory relation: a primitivist account, and an account based on the notions of simplicity and structural depth.

The first account rejects the need for any analysis at all: it takes the comparative explanatory relation to be primitive. The comparative explanatory relation does not reduce to—it is not analyzable in terms of—more basic notions. As a primitive, irreducible matter of fact, for each theorem t and each collection of premises Γ, some proofs stand in the $\succeq_{t,\Gamma}$ relation and others do not. Call this account 'Explanatory Primitivism'.

The main problem for primitivist accounts, of any posit, is their potential obscurity. Since the primitive posit is not analyzed in terms of something more basic or more familiar, it might seem utterly mysterious. So does this problem arise for Explanatory Primitivism? Would a primitive relation of comparative explanatoriness be mysterious and obscure?

I think not. There are other ways of explicating a posit besides analyzing it: one can give clear, compelling examples of the posit; one can envelop the posit in a formalism; and one can describe the theoretical roles that the posit plays, connecting it to other notions in a larger, more encompassing philosophical theory. I have already done the first two: I have given many examples of proofs which stand in the comparative explanatory relation, and I have described that relation's formal features. Neither the examples, nor the formal features, relied on an analysis of the comparative explanatory relation. In the next section, I discuss several theoretical roles which that relation plays. Among other things, it yields analyses of (i) non-comparative explanation in mathematics, and (ii) comparative understanding. So because there are many good examples of proofs which stand in the $\succeq_{t,\Gamma}$ relation, because there is a clear formalism for $\succeq_{t,\Gamma}$, and because $\succeq_{t,\Gamma}$ can be used to analyze other theoretical notions, the $\succeq_{t,\Gamma}$ relation is neither mysterious nor obscure. Explanatory Primitivism about $\succeq_{t,\Gamma}$ is viable.

Whenever possible, however, it is better to analyze than to primitively posit. So consider a second account, which analyzes the comparative explanatory relation in terms of two desiderata for proofs: structural depth, and simplicity. Some proofs cite deep facts about the mathematical structures at issue; others cite relatively shallow facts. Some proofs are simple and concise; others are long and convoluted. These desiderata generally pull in opposite directions: the more deep facts a proof cites, the longer it tends to be; the shorter the proof, the fewer facts—including deep facts—it invokes. Some proofs strike a better balance between depth and simplicity than others. Perhaps these proofs cite comparatively deeper facts about the relevant mathematical structures, and they do so in approximately the same amount of space. According to the second account, more explanatory proofs strike a better balance. More precisely, given a theorem t and a collection of premises Γ, proof $\mathfrak{p}$ of t from Γ is at least as explanatory as proof $\mathfrak{q}$ of t from Γ—that is, $\mathfrak{p} \succeq_{t,\Gamma} \mathfrak{q}$—if and only if $\mathfrak{p}$ strikes at least as good a balance between structural depth and simplicity as $\mathfrak{q}$. Call this account 'Explanatory Balance'.[18]

My notion of structural depth is different from Lange's notion of depth (2014b). For Lange, a theorem's explanatoriness determines its depth (2014b, p. 1), and depth need not have anything to do with structure. For me, depth determines explanatoriness, and the relevant notion of depth is structural.

Simplicity is a quantifiable feature of purely formal proofs.[19] A proof's simplicity is

[18]In addition to structural depth and simplicity, there may be other desiderata which determine the explanatory capacities of proofs. For lack of space, I will not explore that here.

[19]Brenner (2017) and Scorzato (2013) discuss several characterizations of simplicity in the context of formal scientific theories. Those characterizations can be extended to account for the simplicity of a formal proof.

determined by its length, the lengths of its lines, and the number of non-logical symbols that it features.[20] So strictly speaking, on this account, simplicity is a feature of formal proofs rather than informal proofs.

One might worry about that. Mathematicians and philosophers make judgments about simplicity on the basis of informal proofs, since proofs are rarely presented in their fully rigorous, first-order regalia. So is this the right notion of simplicity for capturing mathematicians' and philosophers' judgments about the explanatory capacities of proofs? Why think that the perceived simplicity of informal proofs has anything to do with the simplicity of formal proofs?

I think that these two notions of simplicity are correlated: the perceived simplicity of informal proofs tracks the simplicity of formal proofs. In other words, the notion of simplicity at play in mathematical practice tracks the notion of simplicity which can be measured by the lengths of proofs, the lengths of lines in proofs, the number of non-logical symbols featured, and so on. Intuitive judgments about simplicity, made on the basis of informal mathematical proofs, are generally reliable indicators of the simplicity of corresponding formal proofs.

This seems pretty plausible to me. But in addition to its intuitive plausibility, there are two other reasons to accept it. First, mathematicians and philosophers routinely make judgments about the simplicity of formal proofs, and those judgments track facts about the number of lines in those proofs, the lengths of those lines, and the number of nonlogical symbols that those proofs feature. For example, consider a standard proof of the propositional logic formula $p \to p$:[21]

(1) $p \to \big((p \to p) \to p\big)$

(2) $\Big(p \to \big((p \to p) \to p\big)\Big) \to \Big(\big(p \to (p \to p)\big) \to (p \to p)\Big)$

(3) $\big(p \to (p \to p)\big) \to (p \to p)$

[20] Just as in Lewis's best system account of laws (1973), this account of simplicity may have to invoke naturalness in order to avoid complications relating to proofs which seem extremely simple, but only because they invoke highly disjunctive, highly non-natural predicates.

[21] Lines (1), (2), and (4) are axioms of propositional logic. Lines (3) and (5) follow from prior lines by modus ponens.

(4) $\big(p \rightarrow (p \rightarrow p)\big)$

(5) $p \rightarrow p$

Intuitively, this proof is simple. It is also simple according to the criteria just mentioned: it is just five lines long, its lines are all quite short, and it features only one non-logical symbol. Though this is just one example, it illustrates a more general pattern: in cases of formal proofs, simplicity judgments generally track facts about objective measures of simplicity. In other words, when it comes to formal proofs, simplicity judgments are reliable. And that provides a defeasible, but compelling, reason to think that simplicity judgments about an informal proof are reliable indicators of facts about the simplicity of any corresponding formal proof.

Second, mathematicians' intuitive judgments about the simplicity of informal proofs must be backed by something objective, if those judgments are made true by objective mathematical facts. Otherwise, those judgments have no objective basis whatsoever; so otherwise, those judgments seem false. The simplicity of a formal proof provides a natural, reasonable, objective basis for the perceived simplicity of any corresponding informal proofs.

More generally, when trying to characterize the simplicity of an informal proof, formal proofs can be helpful guides. Compared to informal derivations, formal derivations allow for better ways of identifying exactly what the proof in question is. The identity conditions for formal proofs, that is, are much cleaner than the identity conditions for informal proofs. Because of that, it is easier to give a rigorous account of simplicity for formal proofs than to give a rigorous account of simplicity for informal proofs. And so the simplicity of a formal proof can be quite useful, when trying to determine the simplicity of an informal proof.

Three more clarifications about Explanatory Balance are in order. First, the notion of depth—though left unanalyzed here—is clarified by examples. A fact about the distribution of primes, for instance, is deeper than a fact about a particular prime: whereas the former says a great deal about the structure of the naturals, the latter only says something about a particular place in that structure. The fact about arithmetical progressions among the digits of calculator numbers, cited in Lange's proof of the calculator number theorem, is deeper than the fact that 123321 is divisible by 37: whereas the former says a great deal about the decimal structure of calculator numbers, the latter only says something about 123321. In general, some facts invoke more mathematical structure than other facts. Some facts say more about the structures at issue. And more structurally evocative facts—the ones which say more—are deeper than less structurally evocative facts.

Second, only *some* structures matter, for the purposes of determining the depth of a mathematical fact in a proof. In particular, if $\mathfrak{p}$ is a proof of theorem t from a collection of premises Γ, then the depth of a fact in $\mathfrak{p}$ is determined by the degree to which it illuminates the structures specified by both Γ and t. For example, let $\mathfrak{q}$ be a number-theoretic proof, so that Γ is the collection of Peano's axioms. And let t be a theorem of number theory. Then the depth of a fact in $\mathfrak{q}$ is determined by the degree to which that fact illuminates the number-theoretic structures that Γ and t describe. Or take Saidak's proof and Euler's proof of Euclid's theorem, so that Γ is the collection consisting of both the principles of elementary arithmetic and the principles of elementary analysis. Then the depth of any given fact in Saidak's proof, or the depth of any given fact in Euler's proof, is determined by the degree to which that fact illuminates (i) the arithmetical, analytical structures that Γ describes, and (ii) the number-theoretic structures that Euclid's theorem describes.[22]

Third, and relatedly, note that the depth of a fact is determined by the degree to which it illuminates mathematical *structure*, rather than the degree to which it illuminates the intrinsic natures of mathematical objects, fictionalist mathematical discourse, or any other such thing. Because of that, Explanatory Balance presupposes that mathematical structures exist. But Explanatory Balance does not presuppose any particular ontology of those structures. It is obviously compatible with ante rem structuralism (Shapiro, 1997): the relevant structures for the notion of depth, in ante rem structuralism, would be abstract objects. But Explanatory Balance is also compatible with Field's nominalism (1980): the

[22]For some proofs, Γ is the empty set. Facts in those proofs illuminate, among other things, first-order logical structure.

relevant structures are physical structures, and they inhere in physical systems. Explanatory Balance is even compatible with fictionalist approaches to mathematics, so long as those approaches countenance some sort of fictionalist structure.

Think of Explanatory Balance as a quasi-'best system' account of mathematical explanation. It is like a best system account, insofar as it balances simplicity against something like strength. It is only a *quasi*-'best system' account, however, because depth and strength are only superficially similar. The depth of a fact is not determined by its capacity to summarize. Depth is more metaphysically weighty: the depth of a fact is determined by the degree to which it says something about structure. So ultimately, the source of a mathematical fact's explanatory power does not lie in the summary that it provides. The source of its explanatory power lies in the structure that it reveals.

Plausibly, the ideas underlying Explanatory Balance can be used to account for the explanatory capacities of proofs in other logical systems. For instance, consider a comparative explanatory relation for second-order logical proofs. And consider the following quasi-'best system' account of that relation: more explanatory second-order logical proofs strike a better balance between simplicity and depth.

This observation—that the above ideas can be extended to other logical systems—is a point in favor of Explanatory Balance. In fact, this observation is a point in favor of the quasi-'best system' approach to comparative explanation more generally. For plausibly, Explanatory Balance and quasi-'best system' approaches support accounts of explanatory comparisons in many different logical systems.

Explanatory Balance gets lots of cases right. For instance, let t be the calculator number theorem: every calculator number is divisible by 37. Let $\mathfrak{p}$ be the proof of t from the axioms of number theory which invokes the fact that the digits in a calculator number always form an arithmetical progression. Let $\mathfrak{q}$ be the proof of t from the axioms of number theory which invokes facts about specific calculator numbers being divisible by 37. Then $\mathfrak{p}$ is simpler than $\mathfrak{q}$, since $\mathfrak{q}$ features a lengthy litany of computations and $\mathfrak{p}$ does not. In addition, the facts invoked in $\mathfrak{p}$ are deeper than the facts invoked in $\mathfrak{q}$: the fact that calculator number digits form an arithmetical progression is deeper than facts about specific calculator numbers being divisible by 37. So $\mathfrak{p}$ strikes a better balance between depth and simplicity than $\mathfrak{q}$. Therefore, according to Explanatory Balance, $\mathfrak{p}$ is the more explanatory proof.

A good account of the comparative explanatory relation should imply that in general, brute force proofs are less explanatory than non-brute force proofs (Lange, 2014a, p. 499). Explanatory Balance gets that right. Brute force proofs are generally no simpler than non-brute force proofs, since brute force proofs feature long lists. And the facts invoked in brute force proofs are generally much less deep, since a series of facts about particular items—about numbers, or groups, or whatever—does a poor job of illuminating structure. So non-brute force proofs generally strike the better balance between depth and simplicity. And therefore, according to Explanatory Balance, non-brute force proofs are generally more explanatory than brute force proofs.

A good account of the comparative explanatory relation should also imply that generally, reductio proofs are less explanatory than non-reductio proofs (Mancosu, 2000, p. 111). Explanatory Balance gets that right too. There are two reasons for the diminished explanatory capacities of proofs by reductio: such proofs are often less simple than non-reductio proofs, and such proofs are often less structurally deep than non-reductio proofs. Let us briefly consider both.

First, consider simplicity. Some reductio proofs can be transformed, quite easily, into non-reductio proofs that are extremely similar to their reductio counterparts. In these situations, the reductio proof tends to be less explanatory, since it tends to be less simple. This often happens when the reductio proof proceeds by (i) assuming the negation of the theorem to be proved, and from that, deriving the negation of a sentence which the non-reductio proof derives, and then (ii) using part (or all) of that non-reductio proof to derive the unnegated sentence, thus reaching a contradiction. For in these cases, the lines in the reductio proof often run parallel to the lines in the non-reductio proof, except that the reductio proof derives an additional formula: the negated sentence whose non-negated counterpart is invoked in the non-reductio proof.[23] Because of that, the reductio proof ends up being slightly longer, and thus slightly less simple.

Second, reductio proofs tend to be less structurally deep than non-reductio proofs. To see why, let t be a theorem and let Γ be a collection of premises. Each line in a non-reductio proof of t from Γ describes a feature of Γ's structure: the first line describes a feature F_1, the second line describes a feature F_2, and so on, up to the final line t which describes a feature F. So altogether, the non-reductio proof says that Γ has the structure F, and *also* the structure F_1, *and* F_2, and so on. In contrast, a reductio proof shows that if Γ has the structure of $\neg F$—the structure described by the negation of t—then Γ must have various other structural features which ultimately lead to a contradiction: the structural feature F_1' described by one line of the proof, the structural feature F_2' described by another line, and so on. So altogether, the reductio proof says that Γ has the structure of F, and Γ has the structure of *either* $\neg F_1'$, *or* $\neg F_2'$, and so on. Therefore, the lines in the non-reductio proof say that Γ has a big conjunction of structural features, while the lines in the reductio proof say that Γ has a big disjunction of structural features. Conjunctions generally say more than

[23]Here is an illustrative example: the theorem is that $\sqrt{2}$—in particular, the positive square root of 2—is irrational, and the proofs use the Fundamental Theorem of Arithmetic (FTA).

Non-Reductio Proof

(1) Let p and q be any positive integers.
(2) There exist non-negative integers $\alpha_1, \alpha_2, \ldots$ and $\beta_1, \beta_2, \ldots$ such that $p = 2^{\alpha_1}3^{\alpha_2}\cdots$ and $q = 2^{\beta_1}3^{\beta_2}\cdots$ (FTA).
(3) $2\alpha_1 \neq 2\beta_1 + 1$.
(4) Therefore, $2^{2\alpha_1}3^{2\alpha_2}\cdots \neq 2^{2\beta_1+1}3^{2\beta_2}\cdots$.
(5) Therefore, $p^2 \neq 2q^2$.
(6) Therefore, $\frac{p}{q} \neq \sqrt{2}$.
(7) Therefore, the positive $\sqrt{2}$ is irrational.

Reductio Proof

(i) Suppose that the positive $\sqrt{2}$ is rational.
(ii) Then there exist positive integers p and q such that $\frac{p}{q} = \sqrt{2}$.
(iii) Therefore, $p^2 = 2q^2$.
(iv) There exist non-negative integers $\alpha_1, \alpha_2, \ldots$ and $\beta_1, \beta_2, \ldots$ such that $p = 2^{\alpha_1}3^{\alpha_2}\cdots$ and $q = 2^{\beta_1}3^{\beta_2}\cdots$ (FTA).
(v) Therefore, $2^{2\alpha_1}3^{2\alpha_2}\cdots = 2^{2\beta_1+1}3^{2\beta_2}\cdots$.
(vi) Therefore, $2\alpha_1 = 2\beta_1 + 1$.
(vii) $2\alpha_1 \neq 2\beta_1 + 1$.
(viii) Therefore, the positive $\sqrt{2}$ is irrational.

Notice the clear correspondence between various lines in these two proofs. Lines (2), (3), and (7) are identical to lines (iv), (vii), and (viii) respectively. Lines (3), (4), and (5) are negated versions of lines (vi), (v), and (iii) respectively. Line (1) corresponds to line (i): each is a supposition which starts off its respective proof. And line (6) corresponds to line (ii): the identity asserted in (6) is the negation of the identity asserted in (ii). Note that each line in the non-reductio proof corresponds to a separate line in the reductio proof, but two distinct lines in the reductio proof correspond to one line in the non-reductio proof: in particular, both (vi) and (vii) correspond to (3). As a result, the reductio proof has one more step.

disjunctions. So in general, the non-reductio proof says more about the structure of Γ than the reductio proof.

These are just rough observations about how brute force proofs generally compare to non-brute force proofs, and how reductio proofs generally compare to non-reductio proofs. I would not be surprised if some brute force proofs are more explanatory than some non-brute force proofs, or some reductio proofs are more explanatory than some non-reductio proofs. In those cases, however, I conjecture that Explanatory Balance would still get the right result: if a reductio proof is more explanatory than a non-reductio proof, for instance, then that is because the reductio proof strikes a better balance between depth and simplicity. But those exceptions aside, the above observations show that Explanatory Balance makes the correct *ceteris paribus* prediction regarding comparisons between brute force proofs and non-brute force proofs, and the correct *ceteris paribus* prediction regarding comparisons between reductio proofs and non-reductio proofs.

Because Explanatory Balance has these implications, I prefer it to Explanatory Primitivism. To have the same implications, Explanatory Primitivism would need to be supplemented with two additional postulates: the postulate that in general, brute force proofs are less explanatory than non-brute force proofs; and the postulate that in general, reductio proofs are less explanatory than non-reductio proofs. So to get the same results, in the case of Explanatory Primitivism, one would have to build those results in 'by hand'. Explanatory Balance requires no such supplementation. Because of that—and because non-primitivist accounts of a posit are generally preferable to primitivist accounts—Explanatory Balance is better than Explanatory Primitivism.

But in what follows, the choice between Explanatory Balance and Explanatory Primitivism does not really matter. Both accounts avoid problems, discussed in Section 5, that alternative accounts face. And as discussed in Section 4, both reap the benefits enjoyed by any comparative account of mathematical explanation.

4 The Benefits of Comparative Accounts

Comparative accounts of mathematical explanation, such as Explanatory Primitivism and Explanatory Balance, are attractive for many reasons. In Section 4.1, I discuss a particularly important one: comparative accounts can be used to analyze *non*-comparative explanation in mathematics. In Section 4.2, I discuss four more.

4.1 Analyzing Non-Comparative Explanation

Comparative accounts of mathematical explanation support a straightforward, elegant account of non-comparative explanation in mathematics. To see how, let Γ be a collection of formulas. Let $\mathfrak{p}$ be a proof of theorem t from Γ. The 'degree of explanatoriness' of $\mathfrak{p}$ may be represented by a real number in the interval $[0,1]_{t,\Gamma}$.[24] Denote this real number by $d_{\mathfrak{p}}$. These degrees should be assigned in a way that preserves the ordering: that is, require that if $\mathfrak{p} \succeq_{t,\Gamma} \mathfrak{q}$ then $d_{\mathfrak{p}} \geqslant d_{\mathfrak{q}}$.[25] Then say that a proof is explanatory just in case its degree of explanatoriness exceeds a particular threshold. More precisely, here is an account of non-comparative explanation in terms of all this machinery.

> Non-Comparative Explanation
>
> Let Γ be a collection of formulas. Let $\mathfrak{p}$ be a proof of theorem t from Γ. Let $d_{\mathfrak{p}}$ be the degree of explanatoriness of $\mathfrak{p}$, and let τ be the explanatoriness threshold. Then $\mathfrak{p}$ explains t (relative to Γ) if and only if $d_{\mathfrak{p}} > \tau$.

Note that though Non-Comparative Explanation does not explicitly invoke any ex-

[24]The subscript 't, Γ' indicates that the degree of explanatoriness of a proof is always relative to a particular theorem t and a particular collection of premises Γ.

[25]I do *not* require that, in addition, if $d_{\mathfrak{p}} \geqslant d_{\mathfrak{q}}$ then $\mathfrak{p} \succeq_{t,\Gamma} \mathfrak{q}$. To see why, note that if I did, then the following biconditional would obtain: $\mathfrak{p} \succeq_{t,\Gamma} \mathfrak{q}$ if and only if $d_{\mathfrak{p}} \geqslant d_{\mathfrak{q}}$. It would follow, of course, that the comparative explanatory relation is total. But as mentioned in Section 3, I want to remain neutral on that. So I only require that proofs are assigned numerical degrees in a way which satisfies the conditional in the main text. In other words, I assume that certain facts about degrees of explanatoriness—for instance, the fact that those degrees are totally ordered under the usual ordering of the reals—do not correspond to any facts about the associated proofs' explanatory capacities.

planatory comparisons of the form $\mathfrak{p} \succeq_{t,\Gamma} \mathfrak{q}$, it implicitly relies on such comparisons. For as mentioned above, the comparative explanatory relation is invoked in the characterization of the degrees of explanatoriness.

There is a choice point for NON-COMPARATIVE EXPLANATION, regarding the threshold τ. Perhaps τ varies from one context to another; call this the 'contextual account' of τ. Or perhaps τ is the same across contexts; call this the 'non-contextual account' of τ. I adopt the contextual account in what follows. For as I show in Section 4.2, the contextual account allows me to resolve a disagreement in the literature on mathematical explanation. In addition, the contextual account provides an elegant theory of how context seems to affect the perceived explanatory capacities of proofs. Aspects of context—background knowledge, cognitive capacities, culture, and so on—affect those perceptions by influencing the value of τ.[26]

As NON-COMPARATIVE EXPLANATION shows, comparative mathematical explanation can be used to account for non-comparative mathematical explanation: a proof explains a theorem just in case that proof's degree of explanatoriness exceeds some threshold. In other words, non-comparative mathematical explanation reduces to (i) comparative mathematical explanation, along with (ii) facts about degrees and thresholds. It is not clear, however, whether non-comparative mathematical explanation can be used to account for comparative mathematical explanation. Those who discuss both comparative and non-comparative ex-

[26]To measure τ in a given context, one would conduct various psychological, linguistic, and social science experiments. For example, one would have to ask various people—in a given context—a series of questions about (i) which proofs of a given theorem t are more explanatory than which other proofs, and (ii) which proofs of t count as explanations of t. The answers to these questions would guide one in assigning numerical degrees of explanatoriness to the proofs. And those degrees, in turn, could be used to estimate τ. In this way, the value of τ should be understood as fixed by context in much the same way that other contextually-determined thresholds are. For instance, according to standard linguistic theories, features of context determine the thresholds for gradable adjectives like 'full'. What counts as a 'full' cup of water varies from one context to the next: in some contexts, the cup need only be filled up 90% of the way, in order to count as full; in other contexts, the cup would have to be filled to the brim, in order to count as full. The best way to determine whether the threshold in a given context is .9 (90% filled), or 1 (100% filled), or something else, is to conduct psychological, linguistic, and social science experiments. The contextual account says similarly for τ: in order to figure out the cut-off that determines whether or not a proof counts as an explanation *full stop*, one would have to conduct a series of experiments (thanks to an anonymous referee for raising this concern about measurement).

planation (Lange, 2014a; 2014b; 2017) do not account for either in terms of the other. In other words, there seem to be no reductions of comparative mathematical explanation to (i) non-comparative mathematical explanation, along with (ii) something else.

For these reasons, I think that theories of mathematical explanation ought to focus on explanatory comparisons first, and only later address non-comparative explanation. Accounts of comparative mathematical explanation facilitate, in a rather direct way, accounts of non-comparative mathematical explanation. But account of non-comparative mathematical explanation do not seem to facilitate accounts of comparative mathematical explanation. So I think that much of the literature on mathematical explanation, though very fruitful and interesting, has emphasized the wrong cluster of phenomena. The literature has focused on non-comparative mathematical explanation first, and comparative mathematical explanation second. Better, I think, to focus on comparative mathematical explanation first, and non-comparative mathematical explanation second.

4.2 Four More Benefits

In this section, I discuss four more benefits of comparative accounts of mathematical explanation. First, they respect mathematical practice. Second, they can be used to analyze epistemic notions which are intimately related to explanation, such as understanding. Third, they imply that proofs of logical axioms, or of formulas in a collection of premises, are unexplanatory. Fourth, they allow different proofs to count as explanations in different contexts, and so they a resolve disagreement—which has recently emerged in the literature—over whether most proofs explain.

First, comparative accounts respect the actual practice of mathematicians. As the list of examples in Section 2 makes clear, mathematicians often compare the explanatory capacities of proofs. They seek more explanatory proofs for theorems that have already been derived, and they favor more explanatory proofs over less explanatory ones. Comparative accounts of mathematical explanation respect these practices. For according to comparative accounts, there really are facts of the matter about the relative explanatory capacities of proofs.

Second, comparative accounts can be used to analyze the epistemic state of understanding. Different proofs often provide a basis for different degrees of understanding of a theorem: some proofs provide a basis for more understanding of the theorem they establish than others. These degrees of understanding can be characterized as follows. Let $\mathfrak{p}$ be a proof of theorem t from Γ, and let $d_{\mathfrak{p}}$ be the degree of explanatoriness of $\mathfrak{p}$. Then the degree of understanding of t, for which $\mathfrak{p}$ provides a basis, is equal to $d_{\mathfrak{p}}$. That is, the degree of understanding that a proof can provide for a theorem equals that proof's degree of explanatoriness.

Note that proofs only provide a *basis* for understanding. They do not always provide understanding itself. In order for a proof to help an agent understand a theorem, various context-sensitive conditions must obtain: the agent must be sufficiently familiar with mathematics, the agent must be comfortable with the specific notation used in the proof, and so on. So explanatory proofs—or sufficiently explanatory proofs—merely provide the scaffolding from which understanding can be built.

In short, comparative explanation can be used to illuminate comparative understanding. For understanding is the subjective correlate of objective explanation. As discussed in Section 2, the comparative explanatory relation is objective and mind-independent: it is a part of the non-mental world. Comparative understanding, in contrast, is subjective and mind-dependent: it is a kind of mental state. Nevertheless, comparative explanation and comparative understanding are intimately related. Subjective, mind-dependent states of comparative understanding are supported by objective, mind-independent facts about comparative explanation.

Third, comparative accounts imply that if t is an axiom, or if t is a member of a collection of premises Γ, then given Non-Comparative Explanation—and given an extremely plausible assumption about the explanatoriness of one particular proof of t—it follows that no proof of t is explanatory. The particular proof of t is the proof $\langle t \rangle$ which proves t simply by citing it; call this the 'reflexive' proof of t from Γ. The extremely plausible assumption is that reflexive proofs have explanatoriness degree 0; that is, $d_{\langle t \rangle} = 0$ for any t such that t is an axiom or t is a member of the collection Γ of premises at issue.[27] Note that $\langle t \rangle$ is a subproof of each proof $\mathfrak{q}$ of t from Γ, since t always appears in the last line of any such proof. So by SUBPROOF, for any proof $\mathfrak{q}$ of t from Γ, $\langle t \rangle \succeq_{t,\Gamma} \mathfrak{q}$. By the account of degrees of explanatoriness given above, it follows that $d_{\langle t \rangle} \geqslant d_{\mathfrak{q}}$. Since the extremely plausible assumption says that $d_{\langle t \rangle} = 0$, and since $d_{\mathfrak{q}}$ is in $[0,1]_{t,\Gamma}$, it follows that $d_{\mathfrak{q}} = 0$. So for any threshold τ in $[0,1]_{t,\Gamma}$, it is not the case that $d_{\mathfrak{q}} > \tau$. Therefore, by NON-COMPARATIVE EXPLANATION, $\mathfrak{q}$ does not explain t (relative to Γ). And this, of course, is the right result: since axioms are fundamental explainers, they have no explanations.

Fourth, comparative accounts—in conjunction with NON-COMPARATIVE EXPLANATION—imply that different proofs are explanatory in different contexts. It all depends on the contextually-specified value of the explanatoriness threshold τ. In some contexts, τ might be quite low; and so most proofs are explanations. In other contexts, τ might be pretty high; and so fewer proofs are explanations.

Because of this, comparative accounts can be used to resolve a disagreement in the literature on mathematical explanation. According to some, pretty much all proofs explain (Hersh, 1993, p. 397; Bell, 1976, p. 29). According to others, many proofs are not explanations (Mancosu, 2001, pp. 107-108; Lange, 2017, p. 310). Comparative accounts of mathematical explanation, in conjunction with NON-COMPARATIVE EXPLANATION, reconcile these competing views. In some contexts, the threshold τ is so low that most proofs count as explanations. Pretty much any proof has a degree of explanatoriness which exceeds τ. So pretty much all proofs explain. But in other contexts, τ is so high that many proofs are not explanations at all. The threshold τ exceeds many proofs' degrees of explanatoriness. So many proofs are not explanations.

[27] This is extremely plausible because such proofs seem utterly unexplanatory. It is reasonable to assume that nothing can explain itself, and that is exactly what reflexive proofs would purport to do.

5 Extending Non-Comparative Accounts

One might wonder whether an account of non-comparative mathematical explanation, such as Lange's account (2014a; 2014b), can be extended to cover explanatory comparisons in mathematics. That is, one might wonder whether a comparative account of mathematical explanation can be extracted from Lange's non-comparative account.

In this section, I suggest not. Since my discussion will be brief, however, there may well be solutions to the issues I raise. So the reader may consider this an invitation to show, in detail, how such an extraction should go.

According to Lange (2014a, p. 507), proof $\mathfrak{p}$ of theorem t is explanatory just in case the following three conditions obtain:

1. a certain feature of the result—that is, of the conclusion—of t is salient,
2. a feature of the same kind is in the set-up—that is, in the assumptions—of t, and
3. $\mathfrak{p}$ exploits that feature in the set-up.

Lange defines the terms 'result' and 'set-up' for theorems of the form 'all Fs are Gs': the set-up of such a theorem is an instantiation of F, and the result of such a theorem is an instantiation of G (2014a, pp. 487-488). The notions of one feature being 'of the same kind' as another, of a proof 'exploiting' a feature, and of a feature being 'salient', are primitive.

In order to extend this account to cover comparative mathematical explanation, the non-degreed notions 'of the same kind', 'exploits', and 'salient' must be replaced by counterpart notions that admit of degrees. One might think that this project will succeed: Lange, for instance, suggests that his account can accommodate proofs that fall somewhere in between being explanatory and being utterly unexplanatory (2017, p. 298), perhaps because of degreed versions of the notions that his account invokes. But I am not sure: I worry about the prospects of using degreed versions of the notions 'of the same kind', 'exploits', and 'salient' in an account of comparative mathematical explanation. Let us consider the problems that arise for each.

First, consider 'of the same kind'. It is not clear, to me, what a degreed version of this notion would look like. A feature either is, or is not, of the same kind as another. Kind membership seems all or nothing. It does not seem like it comes in degrees.

Second, consider 'exploits'. This notion seems like it can come in degrees: it seems like some proofs can exploit a feature more than other proofs. But I worry that neither the degreed version of 'exploits', nor the non-degreed version, is a sufficiently clear primitive on which to base an analysis of mathematical explanation. In order for proof $\mathfrak{p}$ to exploit a feature of the set-up, $\mathfrak{p}$ must do more than cite that feature. What more is required, for that feature to count as exploited by $\mathfrak{p}$? And what is required for $\mathfrak{p}$ to exploit that feature more than some other proof $\mathfrak{q}$ exploits it?

Third, consider 'salient'. Like 'exploits', this notion seems like it can come in degrees: one feature can be more or less salient than another. But that will not help extend Lange's account to cover comparative mathematical explanation. For on Lange's account, salience is not attributed to features of proofs: it is attributed to features of theorems. So a degreed conception of salience does not yield a degreed conception of the explanatory capacities of *proofs*. A degreed conception of salience only tells us whether a certain feature of one *theorem* is more or less salient than a certain feature of another *theorem*.[28]

Even if Lange's non-comparative account does not extend to cover explanatory comparisons, however, one might think that a comparative account can be extracted from remarks that Lange makes elsewhere. Lange often talks about the explanatory *power* of proofs: some proofs, he says, are more explanatorily powerful than others (2014a, p. 519). Since explanatory power is a comparative, degreed notion, one might think that an account of Lange's explanatory power is an account of the comparative explanatory relation.

But I suspect not. Lange does not provide an account of explanatory power, since for his

[28]In other words, the problem is this: one wants to say something like "proof $\mathfrak{p}$ of theorem t is more explanatory than proof $\mathfrak{q}$ of t just in case, among other things, the salience of some feature in $\mathfrak{p}$ is greater than the salience of that feature in $\mathfrak{q}$". But on Lange's account, only the salience of certain features of *theorems* ultimately matters for determining whether a proof is explanatory. The salience of certain features of *proofs* does not enter the picture.

purposes—that is, for the purposes of analyzing non-comparative mathematical explanation—he does not need to. He occasionally claims that one proof has more explanatory power than another. But he neither analyzes the notion of explanatory power, nor posits it as a primitive, nor directly connects it to his account of non-comparative explanation. So I doubt that Lange's notion of explanatory power is meant to support an account of the comparative explanatory relation. It seems more like a helpful, heuristic way of tracking—but not analyzing—the explanatory comparisons that mathematicians and philosophers often make.[29]

6 Conclusion

I have argued that there is a comparative explanatory relation in mathematics, and I have presented two accounts of it: Explanatory Primitivism and Explanatory Balance. Explanatory Primitivism takes that relation to be primitive. Explanatory Balance analyzes it in terms of the balances that proofs strike between structural depth and simplicity.

Both accounts describe comparative explanation in mathematics. Because Explanatory Balance has additional attractive features relating to brute force proofs and reductio proofs, which need not be put in 'by hand', I prefer it to Explanatory Primitivism. But regardless, both are quite attractive. Both can be used to reduce non-comparative mathematical explanation to comparative mathematical explanation, for instance. And both are more plausible than various extensions of prominent accounts of non-comparative mathematical explanation.

[29]Similarly for Lange's notion of explanatory depth. Two accounts of depth can be extracted from Lange's comments: one analyzes depth in terms of mathematical coincidences (Lange, 2014b, pp. 200-201), and the other analyzes depth in terms of answers to why-questions (Lange, 2014b, p. 201). The former analysis is non-comparative, and so cannot be used to give an account of the comparative explanatory relation. The latter is comparative, but like explanatory power, it seems more like a helpful, heuristic piece of terminology for talking about explanatory comparisons than a full-fledged analysis.

Acknowledgements

[removed for anonymous review]

References

Alsina, C., & Nelsen, R. B. (2010). *Charming Proofs*. Washington D.C.: The Mathematical Association of America.

Bell, A. W. (1976). A Study of Pupils' Proof-Explanations in Mathematical Situations. *Educational Studies in Mathematics, 7*(1/2), 23–40.

Brenner, A. (2017). Simplicity as a criterion of theory choice in metaphysics. *Philosophical Studies, 174*, 2687–2707.

Brown, J. R. (1999). *Philosophy of mathematics: an introduction to the world of proofs and pictures*. New York, NY: Routledge.

D'Alessandro, W. (2020). Mathematical Explanation Beyond Explanatory Proof. *The British Journal for the Philosophy of Science, 71*, 581–603.

Field, H. (1980). *Science Without Numbers*. Princeton, NJ: Princeton University Press.

Goldblatt, R. (1984). *Topoi: The Categorical Analysis of Logic*. Amsterdam: Elsevier.

Hanna, G. (2000). Proof, Explanation and Exploration: An Overview. *Educational Studies in Mathematics, 44*(1/2), 5–23.

Hempel, C. G., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science, 15*(2), 135–175.

Hersh, R. (1993). Proving is Convincing and Explaining. *Educational Studies in Mathematics, 24*(4), 389–399.

Hitchcock, C., & Woodward, J. (2003). Explanatory Generalizations, Part II: Plumbing Explanatory Depth. *Noûs, 37*(2), 181–199.

Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher & W. C. Salmon (Eds.), *Scientific Explanation* (pp. 410–505). Minneapolis, MN: University of Minnesota Press.

Lange, M. (2014a). Aspects of Mathematical Explanation: Symmetry, Unity, and Salience. *Philosophical Review, 123*(4), 485–531.

Lange, M. (2014b). Depth and Explanation in Mathematics. *Philosophia Mathematica,* *23*(2), 196–214.

Lange, M. (2017). *Because Without Cause.* New York, NY: Oxford University Press.

Lewis, D. (1973). *Counterfactuals.* Malden, MA: Blackwell.

Mancosu, P. (2000). On Mathematical Explanation. In E. Grosholz & H. Breger (Eds.), *The Growth of Mathematical Knowledge* (pp. 103–119). Dordrecht: Springer.

Mancosu, P. (2001). Mathematical Explanation: Problems and Prospects. *Topoi, 20*, 97–117.

Mancosu, P. (2008). Mathematical Explanation: Why it Matters. In Mancosu (Ed.), *The Philosophy of Mathematical Practice* (pp. 134–150). New York, NY: Oxford University Press.

Ribenboim, P. (1989). *The Book of Prime Number Records* (2nd ed.). New York, NY: Springer-Verlag.

Saidak, F. (2006). A New Proof of Euclid's Theorem. *The American Mathematical Monthly, 113*(10), 937–938.

Scorzato, L. (2013). On the role of simplicity in science. *Synthese, 190*(14), 2867–2895.

Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology.* New York, NY: Oxford University Press.

Steiner, M. (1978). Mathematical Explanation. *Philosophical Studies, 34*(2), 135–151.