Philosophia Mathematica Advance Access published September 12, 2014

1–19
10.1093/philmat/nku022
Philosophia Mathematica

Depth and Explanation in Mathematics†

Marc Lange

Department of Philosophy, University of North Carolina at Chapel Hill, Chapel Hill, N.C. 27599-3125 U.S.A.
E-mail: mlange@email.unc.edu

ABSTRACT

This paper argues that in at least some cases, one proof of a given theorem is deeper than another by virtue of supplying a deeper explanation of the theorem — that is, a deeper account of why the theorem holds. There are cases of scientific depth that also involve a common abstract structure explaining a similarity between two otherwise unrelated phenomena, making their similarity no coincidence and purchasing depth by answering why questions that separate, dissimilar explanations of the two phenomena cannot correctly answer. The connections between explanation, depth, unification, power, and coincidence in mathematics and science are compared.

1. INTRODUCTION

Perhaps not a single one of the mathematical examples in this paper involves truly *deep* mathematics. Indeed, some of them are drawn from recreational mathematics. Nevertheless, each of them illustrates how certain bits of mathematics are *deeper* than others. In this way, I will emphasize that depth in mathematics is a matter of degree. How deep some mathematics must be to qualify as 'deep' likely depends on the conversational context. Oftentimes some mathematics qualifies as 'deep' in a given context only by implicit contrast with another, shallower bit of mathematics. (In like manner, how tall something must be to qualify as 'tall' depends on the context, and oftentimes something qualifies as 'tall' in a given context only by implicit contrast with another, shorter thing).

I will focus not merely on the comparative depth of various *theorems*, but also on the comparative depth of various *proofs* of the same theorem. In at least some cases, I will argue, one proof of a given theorem is deeper than another by virtue of supplying a (deeper) *explanation* of that theorem — that is, a (deeper) account of why the theorem holds.[1] (It thus supplies 'insight' into or 'understanding' of the theorem. I take all of

†My thanks to all of the participants at the UCI Workshop on Mathematical Depth for their helpful comments, and especially to Pen Maddy for organizing the workshop.

[1]I include the qualification 'in at least some cases' because I have no general argument that the depth of all deep proofs arises from their explanatory power, and it may be that 'depth' is used in

these terms to be roughly synonymous, though none does much to explicate what it is to 'explain' a theorem over and above simply proving it.)

How much mathematical depth a theorem (or proof) in fact possesses seems to be treated in mathematics as a fact holding independent of our epistemic state; we may discover how deep a theorem (or proof) is, and our making this discovery does not change its depth. However, a theorem's having considerable mathematical depth seems often to be associated with its being difficult to prove or deeply hidden or unexpected, and these features certainly do depend on our epistemic state. How, then, can a deep theorem continue to count as deep even after it is no longer hidden from us or unexpected by us or even difficult for us to prove?

I think that we can make some progress toward resolving this puzzle if we understand the mathematical depth of at least some theorems in terms of their roles in mathematical explanations, since presumably the reason why a theorem is true does not change with our changing epistemic circumstances. (However, as we will see, we may well ask different why questions in different epistemic circumstances, and an answer to one why question may well provoke others.) Mathematical explanation, in turn, is connected to mathematical coincidence, since what initially appears to be a coincidence turns out not in fact to be coincidental if its various components unexpectedly turn out to have a formerly hidden common explainer.

Another reason that I will look at the relation between explanation and depth in mathematics is that this relation may help us to bear in mind that depth is a notion that arises not only in mathematics, but also in science. One scientific explanation of a given fact — one account of why that fact holds — can be deeper than another. (Once again, depth is a matter of degree.) I will compare depth in mathematics to depth in science.

I will also look briefly at the relations between depth, explanation, and unification in mathematics and in science. When a mathematical result captures a respect in which various disparate cases turn out to be alike, a case-by-case proof of the result tends to be a shallow proof compared to a proof that unifies the various cases. As G.H. Hardy [1967, p. 113] memorably said, ' "enumeration by cases" . . . is one of the duller forms of mathematical argument', and although Hardy did not fully elaborate his thought, it seems that what he had in mind is not that a case-by-case approach is relatively unreliable or lacks some other pragmatic virtue. Rather, I would say, a case-by-case proof does not explain why the given theorem holds when the theorem is in fact no coincidence. A case-by-case proof then fails to identify the real reason that the theorem holds, whereas a unifying proof supplies this explanation. The same phenomenon occurs in science.

Finally, I would like to explore the connection between mathematical depth and mathematical power or importance. A deeper theorem presumably has the potential,

different cases to refer to different virtues in a proof. (For this reason, I focus throughout on specific examples.) The same qualification applies to my remarks about 'depth' as used to characterize theorems rather than proofs. (Gödel's incompleteness theorems may be deep without their depth being associated with a deep explanation.) I regard most of my remarks about depth in mathematics as tentative and exploratory. Note, however, that I am not suggesting that depth and explanatory power just happen to coincide in the case of some proofs; I am claiming that in those cases, depth is constituted (at least partly) by explanatory power.

1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	
Total	

Fig. 1. The blank table.

at least, to be mathematically more fruitful — perhaps in having greater explanatory power. As we will see, a theorem's importance may itself become a fact for which there is an explanation.

2. DEEPER AND SHALLOWER PROOFS OF THE SAME (SHALLOW) THEOREM

My first example of comparative depth may well appear initially to epitomize the complete absence of mathematical depth. To see how depth nevertheless becomes a natural concept to use in characterizing this example, I will introduce the example in the manner common in books of mathematical 'magic', parlor games, and wonders. In fact, I first encountered this example in such a book: Martin Gardner's *Mathematical Circus* [1979, pp. 101–104, 167–168].

Here is a way to amaze and amuse your friends. Tell them that you can add numbers much more rapidly than they can — even if they use a calculator. Demonstrate your addition wizardry by asking them to select any two numbers and to insert one in the first row and one in the second row of the blank table in Figure 1. Then ask them to fill out the rest of the table by inserting the sum of those first two rows in row 3, the sum of rows 2 and 3 in row 4, the sum of rows 3 and 4 in row 5, and so on through row 10 — and finally to complete the table by summing all of the numbers in rows 1 through 10. While your friends are furiously filling in the table, you look over their shoulders and wait until they fill in row 7. Then you boldly announce the grand total long before they can reach it. They will be astonished. Figure 2 shows the completed table for the case where the initial numbers are 4 and 7.

Your friends may well try having you begin with different numbers, but they will encounter the same result. They may see how far your wizardry extends by having you begin with fractions, negative numbers, and other exotics. Eventually, they will ask you to divulge your mathematical trick — which is that for any initial two numbers, the grand total equals 11 times the entry in row 7. (It is easy to multiply any number by 11: stick a zero on the end of the original number and then add this result to the original number.) For instance, in Figure 2, the entry in row 7 is 76, and 76 times 11 is 836, the grand total.

1.	4
2.	7
3.	11
4.	18
5.	29
6.	47
7.	76
8.	123
9.	199
10.	322
Total	836

Fig. 2. One way to fill in the table.

1.	x		
2.			y
3.	x	+	y
4.	x	+	$2y$
5.	$2x$	+	$3y$
6.	$3x$	+	$5y$
7.	$5x$	+	$8y$
8.	$8x$	+	$13y$
9.	$13x$	+	$21y$
10.	$21x$	+	$34y$
Total	$55x$	+	$88y$

Fig. 3. A proof that the mathemagic trick will always work.

Figure 3 displays the argument that Martin Gardner uses to prove that for any two initial numbers x and y, the sum of the ten rows is 11 times the 7^{th} row.

As I mentioned at the outset, this theorem is *not* a deep mathematical result; it is much more of a mathematical curiosity than a mathematical fact having any great importance or depth. Nevertheless, one proof of this fact can be deeper than another.

Looking at the table used in Martin Gardner's proof, we cannot help but think of the xs and ys as forming separate sequences. We then recognize that the coefficients of the x terms in lines 3 to 10 are the first 8 members of the Fibonacci sequence (*i.e.*, the sequence beginning $1, 1, 2, \ldots$ where each further term is equal to the sum of the two immediately preceding terms) and that the coefficients of the y terms in lines 2 to 10 are the first 9 members of the Fibonacci sequence. This proof, then, reveals that the result's holding on the x side consists of the fact that the sum of the first 8 Fibonacci numbers (the x-coefficients for lines 3–10) plus 1 (the x-coefficient from line 1) equals 11 times the 5^{th} Fibonacci number (line 7's x-coefficient), and that the result's holding on the y side consists of the fact that the sum of the first 9 Fibonacci numbers equals 11 times the 6^{th} Fibonacci number (line 7's y-coefficient). In other words (where F_i is the

i^{th} Fibonacci number):

$$F_1 + F_2 + \cdots + F_8 + 1 = 11F_5;$$
$$F_1 + F_2 + \cdots + F_9 = 11F_6.$$

My point is that having proved the theorem by using the table filled in with xs and ys, we find ourselves looking at the original result as having two separate components: an x-result and a y-result. And now we want to know why the x-sum and the y-sum end up having the very same property: the x-sum works out so that the coefficient in the grand total is 11 times the coefficient on the 7$^{\text{th}}$ line and the y-sum turns out to be like that, too. Why are the x-sum and the y-sum alike in this regard?

The proof using Gardner's table prompts this why question but fails to answer it. However, this why question can be answered by another proof of the same mathematical fact — a proof that goes somewhat deeper. Think of the Fibonacci sequence as *doubly* infinite: with $F_1 = 1$ and $F_2 = 1$, F_0 must equal $F_2 - F_1 = 0$, F_{-1} must equal $F_1 - F_0 = 1$, and so forth. Notice that F_0 and F_{-1} are the x-coefficients in lines 1 and 2 of the table. So in place of using the relation $F_1 + F_2 + \cdots + F_8 + 1 = 11F_5$ to capture the x side, we can use

$$F_{-1} + F_0 + \cdots + F_8 = 11F_5.$$

Likewise (since $F_0 = 0$), in place of using the relation $F_1 + F_2 + \cdots + F_9 = 11F_6$ to capture the y side, we can use

$$F_0 + F_1 + \cdots + F_9 = 11F_6.$$

Thus, the x-side result and the y-side result are both instances of the generalization that the sum of any 10 consecutive members of the doubly infinite Fibonacci sequence equals 11 times the 7$^{\text{th}}$ member. We can prove this generalization (somewhat laboriously) as follows:

For any integer n,

$$\begin{aligned}
&F_{n+1} + F_{n+2} + F_{n+3} + F_{n+4} + F_{n+5} + F_{n+6} + F_{n+7} + F_{n+8} + F_{n+9} + F_{n+10}\\
&\quad = 2F_{n+3} + 2F_{n+4} + 2F_{n+5} + 0F_{n+6} + F_{n+7} + 2F_{n+8} + 2F_{n+9}\\
&\quad = 4F_{n+5} + 0F_{n+6} + 3F_{n+7} + 4F_{n+8}\\
&\quad = 4F_{n+5} + 4F_{n+6} + 7F_{n+7}\\
&\quad = 11F_{n+7}.
\end{aligned}$$

Although the steps of this proof are pretty boring, it seems to me a somewhat deeper proof of the same relatively inconsequential mathematical result. What makes this proof a somewhat deeper piece of mathematics than the first proof?

One possibility is simply that by connecting the result being proved to the Fibonacci sequence, the second proof connects the result to a lot of other mathematics. But I do not think that this is all there is to it; the greater depth here, I suspect, comes from the connection to the Fibonacci sequence allowing the deeper proof to do something that Gardner's proof cannot do. The deeper proof unifies the x side of Gardner's table with the y side, showing it to be *no coincidence* that they are alike in that the coefficient in the grand total is 11 times the coefficient on the 7$^{\text{th}}$ line. The deeper proof reveals that the reason why the two sides share this property is because they share another property: each is a 10-member fragment of the doubly infinite Fibonacci sequence, and any 10 successive members of that sequence have the property of summing to 11 times the 7$^{\text{th}}$ member.

Of course, by using the notion of a 'mathematical coincidence' to explicate the second proof's greater depth, I have appealed to an idea that itself requires explication. Elsewhere [Lange, 2010] I have emphasized the importance of the notion of mathematical coincidence to mathematical practice, and I have argued that the notion of mathematical coincidence is best explicated in terms of the notion of mathematical explanation. In particular, part of what makes a given mathematical result coincidental is that its components *have no common mathematical explanation*.[2]

The notion of a mathematical explanation was already implicit in my earlier remark that Gardner's proof of the mathemagic trick's working for any two initial numbers prompts a why question that the proof fails to answer. Accordingly, another way of thinking about what makes the second proof deeper than Gardner's proof is that the second proof answers more why questions than Gardner's proof does. In particular, it answers why questions that Gardner's proof provokes but leaves unanswered. Admittedly, Gardner's proof successfully explains why you were able to get the grand total so rapidly — that is, why your addition trick worked in all of the cases that your friends presented to you. But Gardner's proof also ends up provoking *another* why question: why are the x and y sides alike? (Without Gardner's proof, we would never have decomposed the addition-trick theorem into an x side and a y side in the first place.) Gardner's proof fails to answer that why question. By contrast, the second proof answers the question and thereby reveals it to be no coincidence that the x and y sides both possess the property that allows the trick to work (namely, having as their sum 11 times their 7$^{\text{th}}$ member). In particular, the second proof shows that the reason why the two sides have in common the property that allows the trick to work is because they have in common another property: each consists of 10 consecutive Fibonacci numbers.

The notion of an 'explanation' of some mathematical result (such as a proof of the result that reveals why the result holds) requires explication. Elsewhere [Lange, forthcoming] I argue that when a mathematical result exhibits some salient feature (such as a symmetry), a proof explains why that result holds if and only if the proof exploits the same kind of feature in the setup (*e.g.*, the same symmetry) as is salient in the result. Furthermore, I argue that one kind of feature that is striking in many mathematical results is their unity: the result strikingly identifies some property as common to every case in some class. An explanation of such a result is a proof that not only exploits another property that is common to each of these cases, but also proceeds in the same way for each of them.

[2] On my view, not all mathematical explanations reveal that the result being explained is no coincidence — that is (roughly speaking), that its components (ascribing a common property to the cases they cover) have a common mathematical explanation. For instance, when Galois theory explains why the quintic is not generally solvable in radicals, the fact being explained does not (in a typical context) fall into several components (and so perforce there is no common explanation of its components).

Whether a property of some result is salient depends on the context. On my view of mathematical explanation, then, a proof's explanatory power may shift in association with shifts in the salience of particular features of the result being explained. When we first encountered the mathemagic trick, its salient feature was that it worked each time your audience asked you to try it, no matter what the first two numbers used. In other words, the theorem's salient feature was that it identifies a property common to every one of the cases that the audience tried (and, indeed, to every case that they might have tried), the common property being that in each case, the grand total equals 11 times row 7. Your audience wanted to know whether or not this similarity among the various cases in which you actually performed your trick is a coincidence. That is, in asking why the trick worked, your audience was asking whether or not there is a proof that derives this common feature from some other feature that is common to every setup of this kind. Such a proof would show the theorem (that for any initial two numbers in the table, the grand total equals 11 times the 7^{th} number) to be no coincidence. In asking why the trick worked, your amazed audience was demanding exactly the kind of proof that Gardner presented: one that unifies all of the cases falling under the theorem, treating them all in the same way. However, once we have seen Gardner's proof, the result's salient feature shifts to the fact that the x-sum and the y-sum have a property in common: each works out so that the coefficient in the grand total is 11 times the coefficient on the 7^{th} line. In this new context, Gardner's proof is no longer accurately characterized as explaining why the result holds, since his proof treats the x side separately from the y side. Indeed, it is only because the proof treats the two sides separately that we decomposed the result into separate x and y sides and so noticed their similarity. In answering the original question 'Why does this theorem hold?', this proof provoked another question (which we might also express as 'Why does this theorem hold?') that this proof cannot answer. The second proof is deeper by virtue of answering this question, thereby showing that it is no coincidence that the x side and y side are alike.[3]

That the deeper proof gets at least some of its depth by answering why questions provoked but not answered by the shallower proof is like the way in which some

[3] This shift of context and salience is sometimes visible in mathematics textbooks. For instance, one textbook [Benjamin and Quinn, 2003, pp. 30–31] begins in the familiar sort of mathemagical context by presenting 'A Gibonacci Magic Trick', a title that indicates the kind of proof that would be correctly characterized as explanatory: one that reveals the trick's success in all possible cases to be no coincidence. The textbook then gives the proof using the table filled in with xs and ys: '[t]he explanation of this trick involves nothing more than high-school algebra'. However, once that table decomposes the result into the x side and the y side, the context shifts; a previously unrecognized feature of the result becomes salient. The text records that the explanation being given cannot help raising why questions that it cannot answer (though answers do exist):

> ... the total of Rows 1 through 10 will sum to $55x + 88y$. As luck would have it, (actually by the next identity), the number in Row 7 is $5x + 8y$.

The result's x-side and y-side components are depicted by the proof as if their joint holding were a matter of 'luck': a mathematical coincidence. But in fact, it is not — as the teaser 'by the next identity' hints. (Of course, all of the facts involved are mathematically necessary. The only kind of 'luck' that might be present is mathematical coincidence.)

scientific results count as deeper than others by virtue of supplying deeper explanations — that is, by answering more why questions, especially why questions provoked by shallower explanations. For example, phenomena involving falling bodies can be explained (according to classical physics) by Galileo's law of falling bodies but they can be explained more deeply by Newton's laws of motion and gravity. Some phenomena involving gases can be explained by gas laws but more deeply by statistical mechanics and the kinetic-molecular theory of gases.[4] Why are those explanations deeper? Presumably because they answer not only the why questions answered by the shallower explanations, but also some more why questions besides — especially why questions that were prompted but left unanswered by the shallower explanations (such as questions about why the explainers in those explanations obtain). In classical physics, Newton's laws of motion and gravity explain why Galileo's law of falling bodies holds and so give deeper explanations than Galileo's law does. Statistical mechanics and the kinetic-molecular theory of gases explain why the gas laws hold insofar as they do and so give deeper explanations than the gas laws provide. The proof that exploits the doubly infinite Fibonacci sequence not only explains why the mathemagic trick succeeded, but also explains why the x and y sides of Gardner's table turn out to work in the same way.

Of course, there are also differences between the mathematical and scientific examples. Greater depth in the scientific examples that I have just given involves describing *causal* processes at a more fundamental level, whereas there are presumably *no* causal processes at work in a purely mathematical case. But both kinds of examples may involve the deeper explanation answering more why questions — in particular, why questions provoked naturally by the shallower explanation.

I deny that greater depth in *all* scientific cases involves describing causal processes at a more fundamental level. For instance (as I will mention in Section 4), I believe that (according to classical physics, at least) we can use conservation laws to explain why certain fundamental interactions (such as gravity or electromagnetism) are alike in conserving a given quantity (such as energy or linear momentum). Moreover, we can use spacetime symmetries (within a Hamiltonian dynamical framework) to account for those conservation laws and thereby give a deeper explanation of why gravity and electromagnetism, for instance, both conserve momentum. None of these explanations works by describing the world's network of causal relations (as I argue in my [2011a; 2012; 2013a; 2013b]). Thus, a scientific explanation that appeals to spacetime symmetries does not acquire its greater depth by virtue of describing causal processes at a more fundamental level. In scientific practice, spacetime symmetries are called upon to explain the conservation laws, but not *vice versa*; symmetry meta-laws answer why questions (such as why the conservation laws hold) that are naturally provoked but not answered by non-causal explanations appealing to conservation laws.

In the mathemagical example, I have contrasted a shallower proof with a deeper proof of the same theorem. I do not mean to suggest that greater depth requires lesser

[4] 'The known laws of thermodynamics were essentially just generalizations from experience; by penetrating to the underlying causal mechanisms, physicists hoped to gain deeper insights into what heat really is and why it behaves as it does.' [Hunt, 2010, p. 46]

depth to make it deeper. But I suspect that it is often helpful to have something shallower to which to compare something deeper in order to throw its greater depth into sharp relief.

3. DEEPER AND SHALLOWER THEOREMS

In the example that I have been examining, what is deeper or shallower is a *proof* of a given theorem. In other examples, a *theorem* itself is termed 'deep' or 'shallow' — especially by contrast with another theorem. For instance, a theorem may be termed 'deep' by way of highlighting the fact that it has other, 'shallower' theorems as special cases. A proof of the more general theorem (as long as it is a unified proof affording uniform treatment to all of these special cases) can then be a comparatively deep way of proving and thereby unifying the various narrower theorems.[5] If a proof proves all of the special cases in a uniform way by proving the more general theorem, then this proof contrasts with other, shallower proofs each of which proves only one or another of the special cases, where these various proofs fail to unify the special cases because these proofs differ significantly among themselves.

Here is an example. Consider this result concerning expansion in powers — the binomial theorem:

If f and g are numbers and n is a natural number, then

$$(f+g)^n = f^n + \binom{n}{1} f^{n-1} g + \binom{n}{2} f^{n-2} g^2 + \cdots + \binom{n}{n-1} f g^{n-1} + g^n$$
$$= \sum_{k=0}^{n} \binom{n}{k} f^{n-k} g^k,$$

where $\binom{n}{k} = n!/k!(n-k)!$.

Now consider this result concerning expansion in derivatives — the general rule expressing the n^{th} derivative of a product in terms of the product of derivatives ('general Leibniz rule'):

If $f(x)$ and $g(x)$ are n-times differentiable functions of real numbers x, and if $f^{(n)} = \frac{d^n f}{dx^n}$ is the n^{th} derivative of f (and f's 0^{th} derivative is f), then

$$(fg)^{(n)}(x) = \frac{d^n f}{dx^n} g + n \frac{d^{n-1} f}{dx^{n-1}} \frac{dg}{dx} + \frac{n(n-1)}{(1)(2)} \frac{d^{n-2} f}{dx^{n-2}} \frac{d^2 g}{dx^2} + \cdots + f \frac{d^n g}{dx^n}$$
$$= \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)}(x) g^{(k)}(x).$$

As early as 1695, Leibniz noticed the striking analogy between these two results; he even argued in a 1697 letter to John Wallis that his notation was better than Newton's because it makes this analogy more salient [Koppelman, 1971, pp. 157–158]. Regarding this analogy, Johann Bernoulli wrote to Leibniz in 1695: 'Nothing is more elegant than the agreement you have observed . . . [D]oubtless there is some underlying secret'

[5] In [Lange, unpublished], I examine what it is for a proof to give uniform treatment to various special cases.

[Leibniz, 2004, p. 398].[6] The 'underlying secret' being sought might be understood as something hidden that a deeper theorem would reveal — perhaps an underlying structure common to exponentiation and differentiation that is responsible for this analogy, where 'responsibility' here is to be cashed out in terms of explanation.

In fact, the similarity of these two results is no coincidence; it is not 'founded on accidental analogy', as Duncan Gregory [1841, p. iv] put it. Rather, exponentiation and differentiation are alike in *this* respect because they are alike in obeying the same three 'laws of combination', as Gregory called them: the laws of 'commutativity', 'distributivity', and 'repetition'. Here they are:

	Exponentiation (a, f, and g are numbers)	Differentiation ($f(x,y)$, $g(x)$, and $h(x)$ are functions)
Commutative law:	$fg = gf$	$\frac{\partial}{\partial x}\frac{\partial}{\partial y}f = \frac{\partial}{\partial y}\frac{\partial}{\partial x}f$
Distributive law:	$a(f+g) = af + ag$	$\frac{d}{dx}(g+h) = \frac{d}{dx}g + \frac{d}{dx}h$
Law of repetition:	$f^n f^m = f^{n+m}$	$\frac{d^n}{dx^n}\frac{d^m}{dx^m}g = \frac{d^{n+m}}{dx^{n+m}}g$

From the fact that exponentiation obeys these three laws, the binomial theorem follows, and from the fact that differentiation obeys these three laws, the product rule follows in mathematically the same way. Thus, Gregory said, both of these results 'depend only on the laws of combination to which the symbols are subject, and are therefore true of all symbols, whatever their nature may be, which are subject to the same laws of combination' [1841, p. 237].[7] There is a *deep* analogy between these two operations that is responsible for the observed analogy between these two expansion results.

So we have a deep theorem: that any operation subject to these laws of combination obeys an analogue of the binomial theorem. Today we would say that the binomial theorem in this broad sense holds in any commutative ring. What makes this theorem deep? I suggest that we can account for its depth by unpacking the idea that the theorem reveals a similarity at a very abstract level between two unrelated, apparently quite dissimilar operations. It shows how the two separate expansion theorems, one for exponentiation and one for differentiation, can be generalized. The general theorem has the differentiation and exponentiation expansions as special cases.

[6] 'Nihil elegantius est quam consensus quem observasti inter numeros potestatum a binomio et differentiarum rectangulo; haud dubie aliquid arcane subset.'

[7] In a moment, I will propose understanding Gregory's talk of 'dependence' here in terms of explanation. Gregory [1837, p. 32] also argued that the reason why certain differential equations are so much more difficult to solve than others is because they cannot be solved by replacing derivatives with powers because these laws of combination do not apply: 'the second law of combination does not hold with regard to these symbols of operation ... It is this peculiarity with regard to the combinations of the symbols $\langle x\rangle$ and d/dx which gives rise to the difficulty in the solution of linear equations with variable coefficients.'

To say that the general theorem has the differentiation and exponentiation expansions as special cases is not to say merely that the differentiation and exponentiation expansions follow from the general theorem (together with the fact that differentiation and exponentiation are subject to these laws of combination). The two expansions follow from their conjunction, too, but the general theorem is obviously far deeper than their conjunction. The general theorem, unlike the conjunction, identifies a respect in which the differentiation and exponentiation theorems are genuinely similar; the properties captured by the three 'laws of combination' (being commutative, distributive, *etc.*) are respects in which exponentiation and differentiation are alike. These properties are mathematically *natural* properties, not mere shadows of predicates.[8]

There is still more to the relation between the general theorem and the differentiation and exponentiation 'special cases' that helps to make the general theorem deeper: exponentiation and differentiation obey the same laws of expansion precisely *because* they obey the same three laws of combination. This is the 'because' of mathematical explanation. What is responsible for an operation's obeying the given law of expansion is just that it obeys the three laws of combination; no further details of the operation are responsible. Gregory emphasized this point, and François-Joseph Servois [1814–15, p. 142] was more explicit in putting this point in terms of explanation; regarding this example, he wrote: 'It is necessary to find the cause, and everything is very happily explained.'[9] That the two operations obey the same laws of combination is, he said, the true origin (*la véritable origine*, p. 151) of the analogy between the two results. Separate, unrelated derivations of the two results would prove them but would not explain why they hold. Regarding the analogy deployed to solve a linear differential equation by solving an algebraic equation and then exchanging powers for derivatives, George Boole said:

> The analogy . . . is very remarkable, and unless we employed a method of solution common to both problems, it would not be easy to see the reason for so close a resemblance in the solution of two different kinds of equations. But the process which I have here exhibited shows, that the form of the solution depends solely on . . . processes which are common to the two operations under considerations, being founded only on the common laws of the combination of the symbols. [Boole, 1841, p. 119]

[8]I am invoking the distinction between what Armstrong [1978, pp. 38–41] and Lewis [1999, pp. 10–13] call 'natural' (*i.e.*, 'sparse') properties — that is, respects in which things may genuinely resemble each other — on the one hand, and mere shadows of predicates (*i.e.*, 'abundant' properties), on the other hand. (For instance, an arbitrary disjunction of natural properties is not a natural property since, for instance, being five grams or positively electrically charged is not a genuine respect in which objects may resemble each other.) For more on the distinction in mathematics between natural and non-natural properties, see [Corfield, 2003; Lange, unpublished; Tappenden, 2008a,b] – and, of course, [Lakatos, 1976].

[9]'Chemin faisant, d'autres rapports, entre la différentielle, la différence, l'état varié et les nombres, se sont manifestés; il a fallu en rechercher la cause; et tout s'est expliqué fort heureusement, quand, après avoir dépouillé, par une sévère abstraction, ces fonctions de leurs qualités spécifiques, on a eu simplement à considérer les deux propriétés qu'elles possèdent en commun, d'être distributives et commutatives entre elles.'

By 'the reason for so close a resemblance in the solution of two different kinds of equations', Boole means the explanation.

In the context in which I presented them, the salient feature of the binomial theorem conjoined with the differentiation-expansion theorem is their similarity: that they have a certain form in common. On my view of mathematical explanation (see [Lange, forthcoming]), as I mentioned in the previous section, a proof explains (in a given context) why this pair of theorems holds if and only if the proof exploits the same kind of feature in the setup as was (in that context) salient in the result. Since in this context the salient feature of the result (the binomial-expansion theorem conjoined with the differentiation-expansion theorem) is that they have a certain form in common, an explanation of this result is a proof of it that works by appealing to another property common to exponentiation and differentiation and then proceeding in the same way for both operations rather than having to treat the two operations differently. Such an explanation is supplied by the proof of the two theorems that exploits their both obeying the three 'laws of combination'.

The 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. That greater depth presumably arises at least partly from the fact that the proof exploits a structure that also explains the similarity between the differentiation and exponentiation expansion theorems. This deeper proof is able to answer a why question that a proof from premises concerning only differentiation cannot answer: why is there such an analogy between the differentiation and exponentiation expansion theorems?

Again, there are cases of *scientific* depth that work in much the same way — that is, where a common abstract structure explains a similarity between two otherwise unrelated phenomena (making their similarity no coincidence) and so purchases depth by answering a why question that separate dissimilar explanations of the two phenomena cannot correctly answer. For example, the analogies between certain derivative laws in electrostatics, hydrodynamics, and thermodynamics — what James Clerk Maxwell [1890, p. 156] called 'physical analogies' — can be explained by the mathematical isomorphism in the more fundamental equations of these domains, in which (for example) electrical potential, pressure and temperature play analogous roles. The common underlying mathematical architecture of these cases is responsible for the derivative laws' taking the same form in each of them. In view of that common underlying structure, it is no mathematical coincidence that the derivative laws are analogous.[10]

[10] (i) Although it is no *mathematical* coincidence that the derivative laws are analogous, this analogy remains a *physical* coincidence in that there is no significant common explainer for the various derivative laws. For instance, electrical potential difference is not ultimately the same as fluid pressure difference. Therefore, we could still ask why the more fundamental laws of electrostatics, hydrodynamics, and thermodynamics are analogous. (ii) For a specific example of such an analogy among derivative laws, see [Lange, 2010, pp. 332–335]. (iii) For another kind of example where an abstract structure common to many physically disparate phenomena explains various similarities in the derivative laws governing them, making that similarity no coincidence, see the 'dimensional explanations' in [Lange, 2009].

There are many other mathematical examples where depth seems to come from the same factors as it does when the laws of combination unify the two expansion theorems. For instance, projective geometry uses points and lines at infinity to unify theorems that in Euclidean geometry have no common proof (see [Lange, unpublished]). Likewise, complex numbers allow apparently very dissimilar sequences of real numbers to have something in common that results in their having similar convergence behavior. Complex arithmetic thus provides a deeper understanding of the similarity in the two sequences' convergence behavior than can be supplied by separate derivations — purely in terms of real numbers — of the two sequences' convergence behavior. The similarity in their convergence behavior is thereby revealed to be no coincidence, but rather to result from an underlying similarity visible only on the complex plane.[11]

4. DEEP DEPTH AND SHALLOW DEPTH

The mathemagical theorem examined in Section 2 strikes me as pretty shallow, whereas the general expansion theorem examined in Section 3 strikes me as pretty deep. With these examples, I have suggested that there can be deeper and shallower proofs both of deep results and of shallow results. Do deep theorems have at least one deep proof? Is one of these notions of depth parasitic on the other? I do not know. If a deepish result lacks a deep proof, perhaps it must figure in many deep proofs of other results or be a special case of a result that figures in many deep proofs of other results.

The proof of the differentiation expansion theorem from the three laws of combination seems to me deeper than a proof of the same theorem that exploits the details of differentiation and so cannot be generalized. The depth of this proof seems related to the depth of the theorem that any operation satisfying the three laws of combination has such an expansion theorem. The depth of the theorem, in turn, may have something to do with its generality in exploiting an important abstract structure common to many otherwise disparate operations. In contrast, the depth of the proof of the mathemagical theorem that works by exploiting the doubly infinite Fibonacci sequence is apparently not as directly connected to any theorem's depth since the theorems in the neighborhood are not deep. I wonder whether there is a distinction to be drawn along these lines between *deep* depth and *shallow* depth. Where there is deep depth, there is a theorem that would repay further study or an underlying structure that should itself become an object of mathematical investigation. Where there is shallow depth, there is an explanatory proof that answers many why questions (especially why questions raised but left unanswered by other proofs) but not a theorem that generalizes the result being proved or suggests a richer context in which to place it that would repay further exploration.

[11]For further discussion and an example, see [Lange, 2010, pp. 328–332]. As another example along the same lines, consider how the mathematician Edward Frenkel [2013, p. 97] characterizes the Langlands Program: 'It points to deep and fundamental connections between different areas of mathematics. So naturally, we want to know what is really going on here: why might these hidden connections exist? And we still don't fully understand it.'

Depth, generality, and explanatory priority often appear to be linked in science as well.[12] For instance, symmetry principles and conservation laws are commonly identified as both extremely general (covering a wide range of physically disparate kinds of fundamental interactions) and extremely deep (as when a symmetry principle or conservation law gives a more fundamental explanation of some phenomenon than the force laws and other dynamical laws do). Steven Weinberg [1999, p. 73] has remarked that 'the symmetry group of nature is the deepest thing that we can understand about nature today'. Conservation laws seem to have been recognized as having considerable depth just when they were discovered to be associated with spacetime symmetries and hence to be sufficiently general that they were independent of the original dynamical theory in which they were first found (namely, classical physics) and so could survive that theory's demise.[13]

The idea of a mathematical theorem being deep seems related to the idea of a theorem's being mathematically powerful or important. Here I do *not* mean that the theorem is useful for many practical or scientific applications, but rather that the theorem is *mathematically* useful — roughly speaking, that it is useful in proving or perhaps in explaining a wide variety of other results or that it is useful in having a wide variety of extensions, generalizations, analogs, abstractions, and so forth regarding a wide variety of mathematical domains. Obviously, I am not entirely sure what mathematical 'power' or 'importance' amounts to precisely! But I think that mathematicians care about revealing whether or not various results are important. Furthermore, an important fact about mathematical importance is that when mathematicians discover that some result is important, its importance can *itself* become a fact demanding and

---

[12]The relation between generality and explanatory priority in science is not straightforward. For instance, it is not the case in science that one fact F is explanatorily prior to another fact G if F entails G but not *vice versa* — nor is it the case that F is explanatorily prior to G if F is more general than G in the sense that F is 'All Ps are Q', G is 'All Rs are Q', and it is a broadly logical truth that all Ps are R whereas it is not a broadly logical truth that all Rs are P. For instance, that all emerubies (= things that are emeralds or rubies) are gred (= green if emeralds and red if rubies) does not explain why all emeralds are gred.

One attempt to understand the relation in science between generality and explanatory depth is made by Woodward and Hitchcock [2003]. I briefly critique this approach in [Lange, 2011b]. Another attempt is made by Strevens [2010, pp. 136–137]. He identifies 'two dimensions' of depth in scientific explanations: (i) 'an explanation is deep when it drills down to the explanatorily foundational level, to the ultimate explanatory basis' in 'the web of relations of causal influence orchestrated by the fundamental physical laws', and (ii) explanations that are deep show 'that the phenomenon to be explained depends on only a kind of "deep causal structure" of the system in question . . . the more abstract — that is, the more general — properties of the system'. I agree that these ideas characterize many instances of scientific explanatory depth. But in [Lange, 2009; 2011a; 2012; 2013a,b] I argue in various ways that not all scientific explanation derive their explanatory power by virtue of describing the world's *causal* structure. Symmetry principles and conservation laws, for example, supply non-causal scientific explanations.

[13]For more on the way that conservation laws are explained by spacetime symmetries and so are independent of any particular dynamics, see [Lange, 2011a; 2012; 2013a]. There I also discuss examples where symmetries and conservation laws explain facts in a more fundamental way than dynamical laws entailing them do.

receiving an explanation. Mathematicians want to know *why* it turns out to be so important; they want to understand *where* its power comes from. In some cases, at least, an account of why the result is so powerful amounts to a proof of the result that reveals how it follows from premises that are themselves powerful independently of this particular result.[14]

A result's importance or depth or power can be especially mysterious when the result is shown to follow from mathematical premises that seem to hold no particular importance or depth or power.[15] As an example, consider Cauchy's inequality: that if n is a positive integer and $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are non-negative real numbers, then $(a_1^2 + \cdots + a_n^2)(b_1^2 + \cdots + b_n^2) \geq (a_1 b_1 + \cdots + a_n b_n)^2$. The mathematician Michael Steele [2004, p. 1] says about it:

> [T]here is no doubt that this is one of the most widely used and most important inequalities in all of mathematics. A central aim of this [book] is to suggest a path to mastery of this inequality, its many extensions, and its many applications — from the most basic to the most sublime.

(Hardy, Littlewood, and Pólya [1952, p. 16] likewise characterize Cauchy's inequality as 'very important'.) Steele uses mathematical induction to prove Cauchy's inequality. The inductive argument turns on the fact that $x^2/2 + y^2/2 \geq xy$, which follows from $x^2 + y^2 - 2xy \geq 0$, which follows from $(x - y)^2 \geq 0$. This provokes Steele [2004, p. 19] to remark: 'one might rightly wonder how so much value can be drawn from a bound which comes from the trivial observation that $(x - y)^2 \geq 0$.' In other words, Steele is asking why Cauchy's inequality is so powerful, so deep — that is, for a proof of Cauchy's inequality that reveals where its power comes from. (He is asking not for what *constitutes* its importance, but rather for what *accounts for* its importance.)

Steele answers his why question by pointing out that if $a = x^2$ and $b = y^2$, then $x^2/2 + y^2/2 \geq xy$ is equivalent to $a/2 + b/2 \geq \sqrt{ab}$, *i.e.*, $2a + 2b \geq 4\sqrt{ab}$. Now $2a + 2b$ is the perimeter of a rectangle with sides a and b, whereas $4\sqrt{ab}$ is the perimeter of a square with side $\sqrt{ab}$, *i.e.*, a square of the same area as the rectangle with sides a and b. So the inequality used to prove Cauchy's inequality says that among all rectangles of a given area, the square has the smallest perimeter. This is a powerful, deep result and Steele points out that it derives its depth independently of Cauchy's inequality. Steele says that it is the 'rectangular version' of the fact that of all planar regions with a given area, the circle has the smallest perimeter. As Steele [2004, pp. 19–20] concludes, 'we

---

[14](i) Without the requirement that the premises be powerful *independent of this particular result*, any proof at all of the result would do to explain its power (contrary to mathematical practice) since presumably the premises from which the result follows thereby inherit the result's power. (ii) I have just suggested how (in some cases, at least) mathematicians explain why a result is so powerful. I am not intending that this suggestion be subsumed under my suggestion in Section 2 that a proof explains why that result holds if and only if the proof exploits the same kind of feature in the setup (*e.g.*, the same symmetry) as is salient in the result. I do not know whether explaining why T is so powerful is like explaining why T holds (with T's power as its salient feature).

[15]To reveal that the result follows from the axioms, then, does not suffice to make its power the least bit salient or mysterious, since the axioms are presumably well-known to be powerful.

now see more clearly why $x^2/2 + y^2/2 \geq xy$ might be powerful: it is part of that great stream of results that links symmetry and optimality.' We have here an account of why one result turns out to be so deep in terms of the depth (independently acquired) of another result from which it follows.

I will conclude with one final comment about 'shallow depth' — that is, where there is no deep theorem but there are deeper and shallower proofs or methods or strategies for proving a given result, where the deeper way answers why questions raised but left unanswered by the shallower way. Inspired by Martin Gardner's mathematical trick, I wonder whether this sort of contrast between deeper and shallower proofs arises often when (what initially seem to be) mathematical 'tricks' are used to solve various mathematical problems. A shallower solution merely exploits the trick, leaving us wondering why the trick works. By contrast, a deeper solution gives us some motivation behind the trick — not necessarily an account of how the mathematician thought up the trick (the 'context of discovery'), but an explanation for the trick's success (*i.e.*, an account of why it works). A deeper solution reveals that the trick did not just happen accidentally to pay off; there is a reason why it succeeded (*i.e.*, a mathematical explanation). It is therefore not really a 'trick' (in the pejorative sense) at all.[16] The problem is not deep (this is 'shallow depth'), but one solution of it can be deeper than another.

For instance, sometimes a pesky integral succumbs to a clever substitution. Here is an example where a change of variables allows a stubborn integral to be solved.

---

[16]Its success was therefore 'inevitable' and perhaps in this way could have justly been foreseen in advance of its having been tried. Of course, all mathematical results (even 'accidental' ones) are inevitable in that they are mathematically necessary; none is a product of chance or of some contingent fact. Nevertheless, mathematicians do speak non-trivially of certain results as 'inevitable' and of certain proofs as revealing the result's inevitability. I believe 'inevitability' in this sense to be closely tied to mathematical explanation. As Sawyer [1955, p. 26] says: 'In . . . an illuminating proof, the result does not appear as a surprise in the last line; you can see it coming all the way.' On my view of mathematical explanation [Lange, forthcoming], when a mathematical result exhibits some salient feature (such as a symmetry or unity), a proof explains why that result holds if and only if the proof exploits the same kind of feature in the setup as was salient in the result. Thus, an explanation of some result traces its characteristic feature to a similar feature of the setup. By noticing that the setup possesses this feature, one could perhaps have justly foreseen that the result would possess this feature — perhaps even before having discovered the details of the result or the proof that accounts for it. (That the setup exhibits some symmetry anticipates that the result will do so.) Being able to foresee a result's characteristic features seems to me the same as appreciating a result's 'inevitability.' I take 'being able to foresee' from Henri Poincaré:

> [W]hen a rather long calculation has led to some simple and striking result, we are not satisfied until we have shown that we should have been *able to foresee*, if not this entire result, at least its most characteristic traits. . . . To obtain a result of real value, it is not enough to grind out calculations, or to have a machine to put things in order . . . . The machine may gnaw on the crude fact, the soul of the fact will always escape it. [1913, pp. 373–374]

P.A.M. Dirac (as quoted in [Wilczek and Devine, 1987, p. 102] makes a similar remark: 'I consider that I understand an equation when I can predict the properties of its solutions, without actually solving it.' One way to prove that the solutions have a certain property is actually to solve the equation and then to infer the property from the exact solutions. But (on my view) such a proof may be unable to explain why the solutions have this property, whereas a proof that explains why the solutions have this property is a proof that exploits a similar feature of the equation.

The solution is

$$\int_0^{\pi/2} \frac{dx}{1+\tan^m x} = \pi/4$$

This integral is not easily dealt with by any of the standard methods. But here is a nifty maneuver:

Let $t = \pi/2 - x$, so $x = \pi/2 - t$ and $dx = -dt$. As for the bounds of integration, $x = 0$ becomes $t = \pi/2$ and $x = \pi/2$ becomes $t = 0$. So the desired integral I becomes

$$\begin{aligned}
\int_{\pi/2}^{0} \frac{-dt}{1+\tan^m(\frac{\pi}{2}-t)} &= \int_{\pi/2}^{0} \frac{-dt}{1+\cot^m t} \\
&= \int_0^{\pi/2} \frac{dt}{1+\cot^m t} \\
&= \int_0^{\pi/2} \frac{\tan^m t\, dt}{1+\tan^m t} \\
&= \int_0^{\pi/2} \frac{[1+\tan^m t - 1]dt}{1+\tan^m t} \\
&= \int_0^{\pi/2} dt - \int_0^{\pi/2} \frac{dt}{1+\tan^m t} \\
&= \frac{\pi}{2} - I.
\end{aligned}$$

We have thus regenerated the original integral. So $I = \pi/2 - I$, and so $I = \pi/4$.

Such a clever change of variables might well leave us wondering why in the world anyone would have thought that this maneuver would turn out to work — and, relatedly, why it turns out to work.[17] If we had had the explanation of why it works before we tried out the maneuver, then we would have been able to predict that it would work before seeing that it does.[18]

In fact, the reason why it works is that for any value of m, over the given domain of integration the function being integrated is symmetric about the point in the middle of Figure 4 — that is, the point $(\pi/4, 1/2)$. Figure 4 illustrates this symmetry for two values of m. The 'clever' change of variables flips the function left-to-right within the rectangle marked out by the function over the domain of integration and so allows the curve's symmetry over the rectangle's midpoint to be exploited. That is a deeper way of looking at the change of variables: that it exploits the function's symmetry.

But this is not *deep* depth; it does *not* leave us with a theorem that identifies the result being proved as a special case of some broader theorem, creating unity among apparent diversity by revealing some abstract explanatory structure common to many cases beyond the one that is the subject of the proof. Unlike the expansion example in

[17] Tappenden [2005, pp. 171, 200] mentions another example where a clever substitution transforms an integral from 'nasty' to 'nice'; Tappenden calls this transformation 'miraculous', a term suggesting that it might become the proper subject of a mathematical explanation.

[18] See fn 16. I am not endorsing the 'explanation-prediction symmetry thesis' generally.

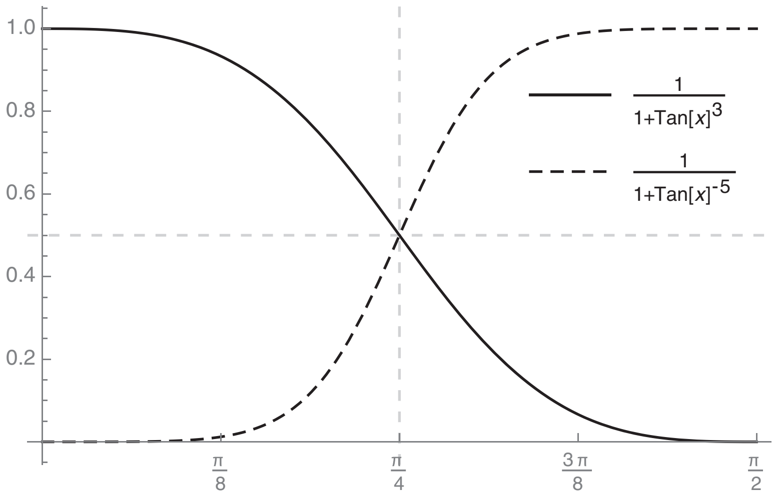

Fig. 4. After [Nahin, 2009, p. 319].

Section 3, this example does not leave us with an abstract structure that is itself worthy of further mathematical investigation.

REFERENCES

Armstrong, David [1978]: *Universals and Scientific Realism*, Vol. 1. Cambridge: Cambridge University Press.

Benjamin, Arthur T., and Jennifer J. Quinn [2003]: *Proofs That Really Count*. Washington, D.C.: The Mathematical Association of America.

Boole, George [1841]: 'On the integration of linear differential equations with constant coefficients', *Cambridge Mathematical Journal* **2**, 114–119.

Corfield, David [2003]: *Towards a Philosophy of Real Mathematics*. Cambridge: Cambridge University Press.

Frenkel, Edward [2013]: *Love and Math: The Heart of Hidden Reality*. New York: Basic Books.

Gardner, Martin [1979]: *Mathematical Circus*. New York: Knopf.

Gregory, Duncan [1837]: 'On the solutions of linear differential equations with constant coefficients', *Cambridge Mathematical Journal* **1**, 22–32.

——— [1841]: *Examples of the Processes of the Differential and Integral Calculus*. London: Deighton.

Hardy, G.H. [1967]: *A Mathematician's Apology*. Cambridge: Cambridge University Press.

Hardy, G.H., J.E. Littlewood, and G. Pólya [1952]: *Inequalities*, 2nd ed. Cambridge: Cambridge University Press.

Hunt, Bruce J. [2010]: *Pursuing Power and Light*. Baltimore: Johns Hopkins University Press.

Koppelman, Elaine [1971]: 'The calculus of operations and the rise of abstract algebra', *Archive for History of Exact Sciences* **8**, 155–242.

Lakatos, Imre [1976]: *Proofs and Refutations*. J. Worrall and E. Zahar, eds. Cambridge: Cambridge University Press.

Lange, Marc [2009]: 'Dimensional explanations', *Noûs* **43**, 742–775.

——— [2010]: 'What are mathematical coincidences (and why does it matter)?', *Mind* **119**, 307–340.

——— [2011a]: 'Conservation laws in scientific explanations: Constraints or coincidences?', *Philosophy of Science* **78**, 333–352.

——— [2011b]: 'Replies to my critics', *Metascience* **20**, 45–52.

——— [2012]: ' "There sweep great general principles which all the laws seem to follow" ', in Karen Bennett and Dean Zimmerman, eds, *Oxford Studies in Metaphysics*, Volume 7, pp. 154–185. Oxford: Oxford University Press.

——— [2013a]: 'How to explain the Lorentz transformations', in Stephen Mumford and Matthew Tugby, eds, *Metaphysics and Science*, pp. 73–98. Mind Association Occasional Series. Oxford: Oxford University Press.

——— [2013b]: 'What makes a scientific explanation distinctively mathematical?', *British Journal for the Philosophy of Science* **64**, 485–511.

——— [forthcoming]: 'Aspects of mathematical explanation: Symmetry, unity, and salience', *Philosophical Review*.

——— [unpublished]: 'Explanation, existence, and natural properties in mathematics — A case study: Desargues' Theorem.'

Leibniz, Gottfried Wilhelm [2004]: *Sämtliche Schriften und Briefe*. Deutsche Akademie der Wissenschaften, ed. Reihe III, Band VI, Part A. Darmstadt, Leipzig, and Berlin: Akademie-Verlag.

Lewis, David [1999]: 'New work for a theory of universals', in D. Lewis, *Papers in Metaphysics and Epistemology*, pp. 8–55. Cambridge: Cambridge University Press.

Maxwell, James Clerk [1890]: *The Scientific Papers of James Clerk Maxwell*, Vol. 1. W.D. Niven, ed. Cambridge: Cambridge University Press.

Nahin, Paul J. [2009]: *Mrs. Perkins's Electric Quilt*. Princeton: Princeton University Press.

Poincaré, Henri [1946]: 'The future of mathematics', in H. Poincaré, *The Foundations of Science*, pp. 369–383. George Bruce Halsted, trans. Lancaster, Penn.: The Science Press.

Sawyer, W.W. [1955]: *Prelude to Mathematics*. Harmondsworth: Penguin.

Servois, François-Joseph [1814–15]: 'Réflexions sur les divers systèmes d'exposition des principes du calcul différentiel, et en particulier, sur la doctrine des infiniment petits', *Annales des Mathématiques Pures et Appliquées* **5**, 141–170.

Steele, J. Michael [2004]: *Cauchy–Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities.* Cambridge: Cambridge University Press.

Strevens, Michael [2010]: *Depth: An Account of Scientific Explanation*. Cambridge, Mass.: Harvard University Press.

Tappenden, Jamie [2005]: 'Proof style and understanding in mathematics – I: Visualization, unification, and axiom choice', in P. Mancosu, K.F. Jørgensen, S.A. Pedersen, eds, *Visualization, Explanation, and Reasoning Styles in Mathematics*, pp. 147–214. Dordrecht: Springer.

——— [2008a]: 'Mathematical concepts and definitions', in P. Mancosu, ed., *The Philosophy of Mathematical Practice*, pp. 256–275. Oxford: Oxford University Press.

——— [2008b]: 'Mathematical concepts: Fruitfulness and naturalness', in P. Mancosu, ed., *The Philosophy of Mathematical Practice*, pp. 276–301. Oxford: Oxford University Press.

Weinberg, Steven [1999]: 'Towards the final laws of physics', in R.P. Feynman and S. Weinberg, eds, *Elementary Particles and the Laws of Physics*, pp. 61–110. Cambridge: Cambridge University Press.

Wilczek, Frank, and Betsy Devine [1987]: *Longing for the Harmonies: Themes and Variations from Modern Physics*. New York: Norton.

Woodward, James, and Christopher Hitchcock [2003]: 'Explanatory generalizations, Part II: Plumbing explanatory depth', *Noûs* **37**, 181–199.