Journal of Humanistic Mathematics

Volume 6 | Issue 1 January 2016

Explanatory Proofs and Beautiful Proofs

Marc Lange
University of North Carolina at Chapel Hill

Recommended Citation

Marc Lange, "Explanatory Proofs and Beautiful Proofs," *Journal of Humanistic Mathematics*, Volume 6 Issue 1 (January 2016), pages 8-51. DOI: 10.5642/jhummath.201601.04. Available at: https://scholarship.claremont.edu/jhm/vol6/iss1/4

Explanatory Proofs and Beautiful Proofs

Cover Page Footnote

My thanks to the organizers of and the participants at the Conference on Beauty and Explanation in Mathematics, University of Umea, Sweden, in March 2014, and to two referees for this journal.

Explanatory Proofs and Beautiful Proofs[1]

Marc Lange

Department of Philosophy, University of North Carolina at Chapel Hill, USA
mlange@email.unc.edu

Abstract

This paper concerns the relation between a proof's beauty and its explanatory power — that is, its capacity to go beyond proving a given theorem to explaining why that theorem holds. Explanatory power and beauty are among the many virtues that mathematicians value and seek in various proofs, and it is important to come to a better understanding of the relations among these virtues. Mathematical practice has long recognized that certain proofs but not others have explanatory power, and this paper offers an account of what makes a proof explanatory. This account is motivated by a wide range of examples drawn from mathematical practice, and the account proposed here is compared to other accounts in the literature. The concept of a proof that explains is closely intertwined with other important concepts, such as a brute force proof, a mathematical coincidence, unification in mathematics, and natural properties. Ultimately, this paper concludes that the features of a proof that would contribute to its explanatory power would also contribute to its beauty, but that these two virtues are not the same; a beautiful proof need not be explanatory.

Keywords: *proof; explanation; beauty; unification; symmetry; coincidence*

1. Introduction

There are many virtues that a mathematical proof may exhibit. These virtues include accessibility to a given audience, beauty, brevity, depth, el-

[1]A portion of this paper is from my "Aspects of Mathematical Explanation: Symmetry, Unity, and Salience", originally published in *The Philosophical Review*, Volume **123** Number 4 (October 2014), pages 485–531.

egance, explanatory power, fruitfulness, generalizability, purity, and visualizability. It is not easy to say precisely what each of these virtues amounts to — or whether mathematicians are appealing to the same virtue each time they invoke "beauty" (for instance). Nevertheless, these virtues clearly figure in mathematical practice. Mathematicians praise a proof for exhibiting one of these virtues, though there need not be any sense in which a proof that exhibits one of them is "better, all things considered" than a proof that lacks any of these virtues. Furthermore, even though a proof that lacks any of these virtues remains just as valid as a virtuous proof, mathematicians who already possess a proof of a given theorem nevertheless seek additional proofs of it that exhibit virtues absent from all previous proofs. A proof that exhibits one of these virtues is often a valuable discovery even if it fails to exhibit any of the other virtues.

Some of these virtues (such as accessibility to a given audience, brevity, and fruitfulness) presumably qualify as virtues because their possession by a proof makes that proof valuable as a means to other goals (such as communicating with an audience or discovering new theorems or proofs). Others of these virtues (such as explanatory power and purity, perhaps) may be valuable in themselves; mathematics may seek them not as a means to some end, but as ends in themselves. Some of these virtues may be entirely independent of others. On the other hand, some may be reducible to others, or they may stand in some more complicated relation.

This paper investigates the relation between two of these virtues: explanatory power and beauty. I will spend much of the paper investigating what it would be for a given proof to succeed not merely in proving its theorem, but also in explaining why its theorem holds. I will offer an account of what makes certain proofs but not others explanatory, and I will compare my account with others that have been proposed. At the paper's close, I will tentatively suggest that the features of a proof that would contribute to its explanatory power would also contribute to its beauty, but that these two virtues are not the same; a beautiful proof need not be explanatory.

2. Explanatory proofs in mathematics

Two mathematical proofs may prove the same theorem from the same axioms, though only one of these proofs explains why that theorem is true. One of my goals in this paper will be to identify the ground of this distinction.

Accordingly, my focus will be on the course that a given proof takes between its premises and its conclusion. The distinction between explanatory and non-explanatory proofs from the same premises must rest on differences in the way they extract the theorem from the axioms.

To clarify this idea, let's look briefly at an example (to which I will return in Section 7). Take an ordinary calculator keyboard, though without the zero (Figure 1):

7	8	9
4	5	6
1	2	3

Figure 1: Calculator keyboard.

We can form a six-digit number by taking the three digits on any row, column, or main diagonal on the keyboard in forward and then in reverse order. For instance, the bottom row taken from left to right, and then right to left, yields 123321. There are sixteen such "calculator numbers" (321123, 741147, 951159 ...). As you can easily verify (with a calculator!), every calculator number is divisible by 37. But a proof that checks each of the calculator numbers separately does not explain why every calculator number is divisible by 37. Compare this case-by-case proof to the following proof:

> The three digits from which a calculator number is formed are three integers a, $a+d$, and $a+2d$ in arithmetic progression. Take any number formed from three such integers in the manner of a calculator number — that is, any number of the form $10^5a + 10^4(a+d) + 10^3(a+2d) + 10^2(a+2d) + 10(a+d) + a$. Regrouping, we find this equals to $a(10^5 + 10^4 + 10^3 + 10^2 + 10 + 1) + d(10^4 + 2 \times 10^3 + 2 \times 10^2 + 10) = 111111a + 12210d = 1221(91a + 10d) = (3 \times 11 \times 37)(91a + 10d)$.[2]

This proof explains why all of the calculator numbers are divisible by 37; as a mathematician says, this proof (unlike the case-by-case proof) reveals the result to be "no coincidence" [36]. Later I will propose an account of what makes this proof but not the case-by-case proof explanatory.

[2] The example appears in an unsigned "gleaning" on page 283 of the December 1986 issue of *The Mathematical Gazette*. Roy Sorensen [43] called this lovely example to my attention. He also cited [36], from which this explanatory proof comes.

In Section 3, I will present another example in which two proofs of the same theorem differ (according to a mathematician) in explanatory power. In Section 4, I will try to spell out the difference between these proofs that is responsible for their difference in explanatory power. Roughly speaking, I will suggest that one of these proofs is explanatory because it exploits a symmetry in the problem — a symmetry of the same kind as the symmetry that initially struck us in the result being explained. In Sections 5 and 6, I will present several other, diverse examples where different proofs of the same theorem have been recognized as differing in explanatory power. In each example, I will suggest that this difference arises from a difference in whether the proofs exploit a symmetry in the problem that is like a striking symmetry in the theorem. In addition, these cases will illustrate how "brute force" proofs fail to explain when the theorem exhibits a striking symmetry and also why some auxiliary constructions but not others are "artificial". In Section 7, I will generalize my proposal to explanations that do not exploit symmetries (as in the calculator-number example). I will argue that in many cases, at least, what it means to ask for a proof that explains is to ask for a proof that exploits a certain kind of feature in the problem: the same kind of feature that is outstanding in the result being explained. The distinction between proofs explaining why some theorem holds and proofs merely establishing that it holds arises only when some feature of the result is salient. In Section 8, I will briefly contrast my account of mathematical explanation with those of Steiner [47], Kitcher [22, 23] and Resnik and Kushner [40]. Finally, in Section 9, I will offer a conjecture regarding the relation between a proof's explanatory power and its beauty.

For at least six decades, philosophy of science has been concerned with understanding what it is to give a scientific explanation (such as an explanation of why the dinosaurs became extinct or an explanation of why the pressure of a gas rises with its temperature under constant volume). By contrast, explanation in mathematics has been little explored by philosophers.[3] Its neglect is remarkable. Mathematical proofs that explain why some theorem holds

[3] In the philosophical literature on explanation, the existence of explanations in mathematics is often acknowledged; for one example per decade, see [33, page 16], [39, page 4], [42, page 96], [41, page 2], and [38, page 2]. But explanation in mathematics is typically not examined at length; in the five works just cited, either it is set aside as not germane to the kind of explanation being examined, or it is given cursory treatment, or it is just ignored after being acknowledged once.

were distinguished by ancient Greek mathematicians from proofs that merely establish that some theorem holds [17], and this distinction has been invoked in various ways throughout the history of mathematics.[4] Fortunately, mathematical explanation has now begun to receive greater philosophical attention. As Mancosu remarks, the topic's "recent revival in the analytic literature is a welcome addition to the philosophy of mathematics" [31, page 134].

A key issue in the philosophical study of scientific explanation is the source of explanatory asymmetry: why does one fact explain another rather than vice versa? Some philosophers have argued that explanatory priority is generally grounded in causal priority: causes explain their effects, not the reverse. By contrast, in mathematical explanations consisting of proofs, the source of explanatory priority cannot be causal (or temporal) priority. Rather, at least part of its source would seem to be that axioms explain theorems, not vice versa. Of course, there may be several different ways to axiomatize a given branch of mathematics. Perhaps only some of these axiomatizations are correct for explanatory purposes. Or perhaps any axiomatization is equally good for explanatory purposes, but a proof's status as explanatory is relative to a given axiomatization.

In any case, I will be concerned with a logically prior issue: the distinction between explanatory and non-explanatory proofs of the same theorem from the same axioms.[5] Nevertheless, I will not contend that all mathe-

[4] "Explaining why the theorem holds" is just explaining the theorem. It is distinct from explaining why we should (or do) believe the theorem. This distinction is familiar from scientific explanation.

[5] Of course, mathematical facts are often used to explain why certain contingent facts hold. Sometimes mathematical facts may even play the central role in explaining some physical fact. (For recent discussions, see [3] and [27]; see also [48].) But these are not "mathematical explanations" of the kind that I will be discussing, which have mathematical theorems rather than physical facts as their target. In a conversation, we might "explain" why (or how) some mathematical proof works (e.g., by making more explicit the transitions between steps). A textbook might "explain" how to multiply matrices. A mathematics popularizer might "explain" an obscure theorem by unpacking it. However, none of these is the kind of "mathematical explanation" with which I will be concerned. None involves explaining why some result holds — just as Hempel [20, page 80] pointed out that an account of scientific explanation does not aim to account for what I do when I use gestures to "explain" to a Yugoslav garage mechanic how my car has been misbehaving. I am also not concerned with historical or psychological explanations of why mathematicians held various beliefs or how a given mathematician managed to make a certain discovery.

matical explanations consist of proofs. Indeed, I will give some examples of mathematical explanations that are not proofs.

We will see several examples where mathematicians have distinguished proofs that explain why some theorem holds from proofs that merely establish that it holds.[6] For instance, in the *Port-Royal Logic* of 1662, Pierre Nicole and Antoine Arnauld characterized indirect proof (that is, proof of p by showing that $\neg p$ implies a contradiction) as "useful" but non-explanatory:

> ... such Demonstrations constrain us indeed to give our Consent, but no way clear our Understandings, which ought to be the principal End of Sciences: for our Understanding is not satisfied if it does not know not only that a thing is, but why it is? which cannot be obtain'd by a Demonstration reducing to Impossibility. [35, page 422 (Part IV, chapter *ix*)]

Nicole and Arnauld obviously took explanation to be as important in mathematics as it is in science. More recently, the mathematician William Byers has characterized a "good" proof as "one that brings out clearly the reason why the result is valid" [9, page 337]. Likewise, researchers on mathematics education have recently argued empirically that students who have proved and are convinced of a mathematical result often still want to know why the result is true [32], that students assess alternative proofs for their "explanatory power" [19, page 399], and that students expect a "good" proof "to convey an insight into why the proposition is true" even though explanatory power "does not affect the validity of a proof" [5, page 24]. However, none

[6] To avoid the corrupting influence of philosophical intuitions, I have also tried to use examples from workaday mathematics rather than from logic, set theory, and other parts of mathematics that have important philosophical connections. But by focusing on proofs that mathematicians themselves recognize as explanatory, I do not mean to suggest that philosophers must unquestioningly accept the verdicts of mathematicians. Indeed, some mathematicians, such as Gale [11], deny that there is any distinction between explanatory and non-explanatory proofs. (Some philosophers agree; see, for example, [16, page 81]. Gale later changed his mind [12, page 41].) But just as an explication of scientific explanation should do justice to scientific practice (without having to fit every judgment of explanatory power made by every scientist), so an explication of mathematical explanation should do justice to mathematical practice. Regarding the examples I will discuss, I have found the judgments made by working mathematicians of which proofs do (and do not) explain to be widely shared and easily appreciated by non-mathematicians. It is especially important that an account of mathematical explanation fit such cases.

of this work investigates what it is that makes certain proofs but not others explanatory. This question will be my focus in the next few sections.

3. Zeitz's biased coin: A suggestive example of mathematical explanation

Consider this problem from the Bay Area Math Meet, San Francisco, April 29, 2000:

> A number p between 0 and 1 is generated randomly so that there is an equal chance of the generated number's falling within any two intervals of the same size inside $[0, 1]$. Next a biased coin is built so that p is its chance of landing heads. The coin is then flipped 2000 times. What is the chance of getting exactly 1000 heads?

The mathematician Paul Zeitz gives the answer:

> The amazing answer is that the probability is 1/2001. Indeed, it doesn't matter how many heads we wish to see — for any [integer r] between 0 and 2000, the probability that r heads occur is 1/2001. [53]

That is, each of the 2001 possible outcomes (from 0 heads to 2000 heads) has the same likelihood. That is remarkable. It prompts us to ask, "Why is that?" (There is nothing special about 2000 tosses; the analogous result holds for any number n of tosses.)

Here is an elaboration of one proof that Zeitz sketches. (I give all of the gory details, but you may safely skim over them, if you wish.)

> If you flip a coin n times, where p is the chance of getting a head on any single flip, then the chance of getting exactly r heads is
>
> $$\binom{n}{r} p^r (1-p)^{n-r}.$$
>
> (The chance of getting a particular sequence of r heads and $(n-r)$ tails is $p^r(1-p)^{n-r}$, and there are
>
> $$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

different ways of arranging r heads and $(n-r)$ tails.) In our problem (where $n = 2000$), the total chance of getting exactly r heads is the sum, taken over all possible p's, of the chance of getting r heads if the coin's bias is p, multiplied by the chance dp that p is the coin's bias. This sum is an integral:

$$\int_0^1 \binom{n}{r} p^r (1-p)^{n-r} dp = \binom{n}{r} \int_0^1 p^r (1-p)^{n-r} dp$$

Let's integrate "by parts" using $\int u \cdot dv = uv - \int v \cdot du$. Let $u = (1-p)^{n-r}$ and $dv = p^r dp$. Then $du = -(n-r)(1-p)^{n-r-1} dp$ and $v = p^{r+1}/r+1$. So

$$\int_0^1 p^r (1-p)^{n-r} dp = (1-p)^{n-r} \cdot \left.\frac{p^{r+1}}{r+1}\right|_0^1 + \frac{n-r}{r+1} \int_0^1 p^{r+1} (1-p)^{n-r-1} dp.$$

The first term on the right side equals zero at both $p = 0$ and $p = 1$. The integral in the second term on the right side takes the same form as the integral on the left side, so another integration by parts yields

$$\int_0^1 p^{r+1} (1-p)^{n-r-1} dp = \frac{n-r-1}{r+2} \int_0^1 p^{r+2} (1-p)^{n-r-2} dp.$$

Repeatedly integrate by parts until $(1-p)$'s exponent has decreased to 0. The remaining integral is

$$\int_0^1 p^n (1-p)^0 dp = \int_0^1 p^n dp = \left.\frac{p^{n+1}}{n+1}\right|_0^1 = \frac{1}{n+1}.$$

So all together

$$\begin{aligned}
\int_0^1 p^r (1-p)^{n-r} dp &= \left(\frac{n-r}{r+1}\right)\left(\frac{n-r-1}{r+2}\right) \cdots \left(\frac{1}{n}\right)\left(\frac{1}{n+1}\right) \\
&= \left(\frac{1}{n+1}\right) \frac{1}{\frac{n(n-1)(n-2)\cdots(r+2)(r+1)}{(n-r)(n-r-1)\cdots(2)}} \\
&= \left(\frac{1}{n+1}\right) \frac{1}{\binom{n}{n-r}}.
\end{aligned}$$

> This result times $\binom{n}{n-r}$ is the total chance of getting exactly r heads in our problem. But $\binom{n}{n-r} = \binom{n}{r}$ since the number of different arrangements of (say) exactly $n-r$ tails in n tosses equals the number of different arrangements of exactly r heads in n tosses. So the total chance of getting exactly r heads in our problem is
> $$\binom{n}{r}\left(\frac{1}{n+1}\right)\frac{1}{\binom{n}{n-r}} = \frac{1}{n+1}$$
> the result we sought.

Although this proof succeeds, Zeitz says that it "shed[s] no real light on why the answer is what it is. ... [It] magically produced the value $\frac{1}{n+1}$" [53]. Although Zeitz does not spell out this reaction any further, I think we can readily sympathize with it. This proof makes it seem like an accident of algebra, as it were, that everything cancels out so nicely, leaving us with just $\frac{1}{n+1}$.

Of course, nothing in math is genuinely accidental; the result is mathematically necessary. Nevertheless, until the very end, nothing in the proof suggested that every possible outcome (for a given n) would receive the same chance.[7] The result simply turns out to be independent of r, and this fact remains at least as remarkable after we have seen the above proof as it was before. I think that many of us would be inclined to suspect that there is some reason why the chance is the same for every possible outcome (given n) — a reason that eludes the above proof. (Notice how natural it becomes in this context to talk of "reasons why" the result holds.)

Zeitz says that in contrast to the foregoing argument, the following argument allows us to "understand *why* the coin problem had the answer that it did" [53, his emphasis]:

> Think of the outcomes of the n coin tosses as dictated by n further numbers generated by the *same* random-number generator that generated the coin's chance p of landing heads: a number less than

[7] Regarding another method of tackling this integral (using generating functions), Zeitz says: "The magical nature of the argument is also its shortcoming. Its punchline creeps up without warning. Very entertaining, and very instructive in a general sense, but it doesn't shed quite enough light on this particular problem. It shows us how these $n+1$ probabilities were uniformly distributed. But we still don't know why" [55, page 352].

> p corresponds to a head, and a number greater than p corresponds to a tail. (The chance that the number will be less than p is obviously p – the chance of a head.) Thus, the same generator generates $n+1$ numbers in total. The outcomes are all heads if the first number generated (p) is larger than each of the n subsequent numbers, all but one of the outcomes are heads if the first number generated is larger than all but one of the n subsequent numbers, and so forth. Obviously, if we were to rank the $n+1$ generated numbers from smallest to largest, then the first generated number (p) has the same chance of being ranked first as it has of being ranked second, and likewise for any other position. Hence, every possibility (from 0 heads to n heads) is equally likely, so each has chance $\frac{1}{n+1}$.

In light of this proof, Zeitz concludes that the result "is not unexpected, magical algebra. It is just simple, almost inevitable symmetry" [53]. I agree: this proof explains why every possible outcome has the same chance (and, therefore, why each possible outcome's chance is $\frac{1}{n+1}$). The proof explains the symmetry in the chances by showing how it arises from a symmetry in the setup (rather than as an algebraic miracle): in effect, the same random-number generator is used for generating p as for generating each of the n coin-toss outcomes, and when $n+1$ numbers are so generated and listed from smallest to largest, every position on the list is equally likely to end up being occupied by the first number generated.[8] A symmetry in the setup accounts for the same symmetry in the chances of the possible outcomes.

In short, our curiosity was initially aroused by the symmetry of the result: that, remarkably, every possible outcome has the same chance. The first proof did not satisfy us because it failed to exploit any such symmetry in the setup. We suspected that there was a reason for the result — a hidden "evenness" in the setup that is responsible for the same "evenness" in the result. The second proof revealed the setup's hidden symmetry and thereby explained the result.

[8] Zeitz: "The probabilities were uniform because the numbers [generated randomly] were uniform, and thus their rankings [that is, the place of the first generated number among the others, as ranked from smallest to largest] were uniform. The underlying principle, the 'why' that explains this problem, is ... Symmetry" [55, page 353].

4. Explanation by symmetry

The example of Zeitz's coin suggests the following proposal. Often a mathematical result that exhibits symmetry of a certain kind is explained by a proof showing how it follows from a similar symmetry in the problem. Each of these symmetries consists of some sort of invariance under a given transformation; the same transformation is involved in both symmetries. For instance, in the example of Zeitz's coin, both symmetries involve invariance under a switch from one possible outcome (e.g., 1000 heads and 1000 tails) to any other (e.g., 999 heads and 1001 tails). In such a case, what makes a proof that appeals to an underlying symmetry in the setup count as "explanatory" in contrast to other proofs of the same result? Nothing beyond the fact that the symmetry of the result was what drew our attention in the first place.

For instance, in the example of Zeitz's coin, had the result been some complicated, unremarkable function of n and r, then the question "Why is *that* the chance?" would probably have amounted to nothing more than a request for a proof. One proof might have been shorter, less technical, more pleasing or accessible to some audience, more elegant in some respect, or more fully spelled out than some other proof. But any proof would have counted as answering the question. There would have been no distinction between a proof that explains the result and a proof that merely proves it.

However, the symmetry of the result immediately struck us, and it was made further salient by the first proof we saw, since in that proof, the symmetry of the solution emerged "magically" from out of the fog of algebra. Its origin was now especially puzzling. The symmetry, once having become salient, prompts the demand for an explanation: a proof that traces the result back to a similar symmetry in the problem. (There need not be any such proof; a mathematical fact may have no explanation.[9]) In light of the salience of the symmetry, there is a point in asking for an explanation over and above a proof. A proof that exploits the symmetry of the setup is privileged as explanatory because the symmetry of the result is especially striking.

My proposal predicts that mathematical practice contains many other examples where an explanation of some result is distinguished from a mere

[9]For example, after asking *why* a given Taylor series fails to converge, Spivak [44, page 482] says, "Asking this sort of question is always dangerous, since we may have to settle for an unsympathetic answer: it happens because it happens — that's the way things are!"

proof of it only in view of the result's exhibiting a puzzling symmetry — and where only a proof exploiting such a symmetry in the problem is recognized as explaining why the solution holds. I will now present several examples of this phenomenon.

5. A theorem explained by a symmetry in the unit imaginary number

Consider this theorem (first proved by d'Alembert in 1746):

> If the complex number $z = a + bi$ (where a and b are real) is a solution to $z^n + a_{n-1}z^{n-1} + \cdots + a_0 = 0$ (where the a_i are real), then z's complex conjugate $\overline{z} = a - bi$ is also a solution.

Why is this true?

We can prove this theorem directly by evaluating $\overline{z}^n + a_{n-1}\overline{z}^{n-1} + \cdots + a_0$. First, we show by calculation that $\overline{z}\,\overline{w} = \overline{zw}$:

> Let $z = a + bi$ and $w = c + di$. Then $\overline{z}\,\overline{w} = (a - bi)(c - di) = ac - bd + i(-bc - ad)$. But $zw = (a+bi)(c+di) = ac - bd + i(bc + ad)$ and so $\overline{zw} = ac - bd - i(bc + ad) = \overline{z}\,\overline{w}$.

Hence, $\overline{z}^2 = \overline{z}\,\overline{z} = \overline{zz} = \overline{z^2}$, and likewise for all other powers. Therefore, $\overline{z}^n + a_{n-1}\overline{z}^{n-1} + \cdots + a_0 = \overline{z^n} + a_{n-1}\overline{z^{n-1}} + \cdots + a_0$. Now we show by calculation that $\overline{z} + \overline{w} = \overline{z + w}$:

> Let $z = a + bi$ and $w = c + di$. Then $\overline{z} + \overline{w} = (a - bi) + (c - di) = a + c + i(-b - d)$ and $\overline{z + w} = \overline{a + bi + c + di} = a + c - i(b + d) = \overline{z} + \overline{w}$.

Thus, $\overline{z^n} + a_{n-1}\overline{z^{n-1}} + \cdots + a_0 = \overline{z^n + a_{n-1}z^{n-1} + \cdots + a_0}$, which equals $\overline{0}$ and hence 0 if z is a solution to the original equation.

Although this proof shows d'Alembert's theorem to be true, it pursues what mathematicians call a "brute force" approach. That is, it simply calculates everything directly, plugging in everything we know and grinding out the result. The striking feature of d'Alembert's theorem is that the equation's nonreal solutions all come in pairs where one member of the pair can be transformed into the other by the replacement of i with $-i$. Why does exchanging i for $-i$ in a solution still leave us with a solution? This symmetry just works out that way ("magically") in the above proof. But we are inclined to suspect that there is some reason for it. In other words, the symmetry in d'Alembert's theorem puzzles us, and in asking for the theorem's

"explanation", we are seeking a proof of the theorem from some similar symmetry in the original problem — that is, a proof that exploits the invariance of the setup under the replacement of i with $-i$.

The sought-after explanation is that $-i$ could play exactly the same roles in the axioms of complex arithmetic as i plays. Each has exactly the same definition: each is exhaustively captured as being such that its square equals -1. There is nothing more to i (and to $-i$) than that characterization. Of course, i and $-i$ are not equal; each is the negative of the other. But neither is intrinsically "positive", for instance, since neither is greater than (or less than) zero. They are distinct, but they are no different in their relations to the real numbers. Whatever the axioms of complex arithmetic say about one can also be truly said about the other. Since the axioms remain true under the replacement of i with $-i$, so must the theorems — for example, any fact about the roots of a polynomial with real coefficients. (The coefficients must be real so that the transformation of i into $-i$ leaves the polynomial unchanged.) The symmetry expressed by d'Alembert's theorem is thus grounded in the same symmetry in the axioms.

Here we have another example where a proof is privileged as explanatory because it exploits a symmetry in the problem — a symmetry of the same kind as initially struck us in the fact being explained. Furthermore, this is a good example with which to combat the impression that a proof's being explanatory is no more objective (no less "in the eye of the beholder") than a proof's being understandable, being of interest, or being sufficiently spelled out. Mathematicians largely agree on whether or not a proof is aptly characterized as "brute force", and I suggest that no "brute force" proof is explanatory when the theorem exhibits a striking symmetry. A brute force approach is not selective. It sets aside no features of the problem as irrelevant. Rather, it just "ploughs ahead" like a "bulldozer" [2, page 215], plugging everything in and calculating everything out. (The entire polynomial, not just some piece or feature of it, was used in the first proof above of d'Alembert's theorem.) In contrast, an explanation must be selective. It must pick out a particular feature of the setup and deem it responsible for (and other features irrelevant to) the result being explained. (Shortly we will see another example in which a brute-force proof is explanatorily impotent.) Mathematicians commonly say that a brute-force solution supplies "little understanding" and fails to show "what's going on" [29, pages 29-30].

Suppose that we had begun not with d'Alembert's theorem, but with some particular instances of it (as d'Alembert might have done). The solutions of $z^3 + 6z - 20 = 0$ are 2, $-1 + 3i$, and $-1 - 3i$. The solutions of $z^2 - 2z + 2 = 0$ are $1 - i$ and $1 + i$. In both examples, the solutions that are not real numbers are pairs of complex conjugates. Having found many examples like these, one might ask: Why is it that in all of the cases we have examined of polynomials with real coefficients, their non-real roots all fall into complex-conjugate pairs? Is it a coincidence, or are they all like that? One possible answer to this why question is that they are not all like that; we have simply gotten lucky by having examined an unrepresentative group of examples. Another possible answer is that our examples were unrepresentative in some systematic way: all polynomials of a certain kind (e.g., with powers less than 4) have their non-real solutions coming in complex-conjugate pairs, and all of the polynomials we examined were of that kind. In fact, as we have seen, d'Alembert's theorem is the explanation; any polynomial with exclusively real coefficients has all of its non-real roots coming in complex-conjugate pairs. Here we have a mathematical explanation that consists not of a proof, but merely of a theorem.

However, it is not the case that just any broader mathematical theorem that subsumes the examples to be explained would suffice to account for them. After all, we could have subsumed those two cases under this gerrymandered theorem: *For any equilateral triangle or equation that is either* $z^3 + 6z - 20 = 0$ *or* $z^2 - 2z + 2 = 0$ (the two cases above), *either the triangle is equiangular or the equation's non-real solutions all form complex-conjugate pairs.* This theorem does not explain why the two equations have the given feature. (Neither does a theorem covering just these two cases.) Plausibly, whether a theorem can be used to explain its instances depends on whether that theorem has a certain kind of explanation. In any case, my concern in this paper is with the way that a certain proof of some theorem can explain why that theorem holds rather than with mathematical explanations where a theorem explains why one or more of its instances hold.

Here is another example (also discussed by Kitcher [23, pages 425-426]) of a proof that is widely respected as possessing explanatory power because it derives a result exhibiting a salient symmetry from a similar symmetry in the setup. It had been well known before Lagrange that a cubic equation of the form $x^3+nx+p = 0$, once transformed by $x = y-n/3y$, becomes a sixth-degree equation $y^6 + py^3 + n^3/27 = 0$ (the "resolvent") that, miraculously,

is quadratic in y^3. Lagrange aimed to determine why: "I gave reasons why [*raison pourquoi*] this equation, which is always of a degree greater than that of the given equation, can be reduced..." [25, page 242]. Lagrange showed that exactly the resolvent's solutions y can be generated by taking $1/3(a_1 + \omega a_2 + \omega^2 a_3)$ and replacing a_1, a_2, and a_3 with the cubic's three solutions x_1, x_2, and x_3 in every possible order — where $\omega = (-1+\sqrt{3}i)/2$, one of the cube roots of unity. But the three solutions generated by even permutations of x_1, x_2, and x_3 — namely $1/3(x_1+\omega x_2+\omega^2 x_3)$, $1/3(x_2+\omega x_3+\omega^2 x_1)$, and $1/3(x_3+\omega x_1+\omega^2 x_2)$ — all have the same cubes (since $1 = \omega^3 = (\omega^2)^3$) — and likewise for the three solutions generated by odd permutations. Since y^3 takes on only two values, y^3 must satisfy a quadratic equation. So "this is why the equation that y satisfies proves to be a quadratic in y^3" [24, page 602]. The symmetry of $1/3(a_1+\omega a_2+\omega^2 a_3)$ under permutations of the three x_i explains the symmetry that initially strikes us regarding the sixth-degree equation (that three of its six roots are the same, and the remaining three are, too). As mathematicians commonly remark, y^3 "assumes two values under the six permutations of the x. It is for this reason that the equation of degree six which $[y]$ satisfies is in fact a quadratic in $[y]^3$" [21, page 51].

6. Two geometric explanations that exploit symmetry

Proofs in geometry can also explain by exploiting symmetries. Consider the theorem:

> If $ABCD$ is an isosceles trapezoid as shown in Figure 2 (AB parallel to CD, $AD = BC$) such that $AM = BK$ and $ND = LC$, then $ML = KN$.

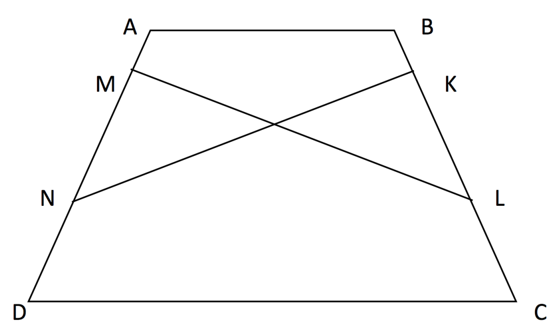

Figure 2: An isosceles trapezoid.

A proof could proceed by brute-force coordinate geometry: *first let D's coordinates be $(0,0)$, C's be $(0,c)$, A's be (a,s), and B's be (b,s), and then solve algebraically for the two distances ML and KN, showing that they are equal.* A more inventive, Euclid-style option would be to draw some auxiliary lines and to exploit the properties of triangles:

> Draw the line from N perpendicular to CD; call their intersection P (see Figure 3 below); likewise draw line LS.

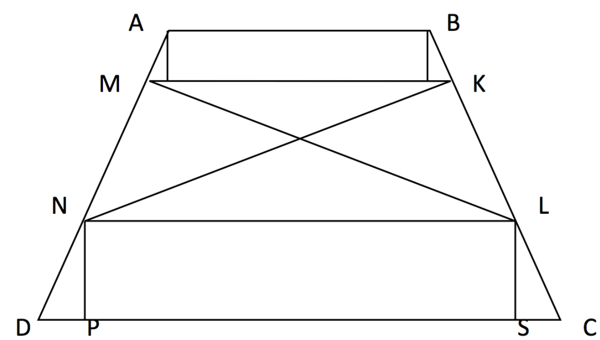

Figure 3: A Euclid-style proof of isosceles trapezoid theorem.

> Consider triangles DNP and CLS: angles D and C are congruent (since the trapezoid is isosceles), $ND = LC$ (given), and the two right angles are congruent. Having two angles and the non-included side congruent, triangles DNP and CLS are congruent, so their corresponding sides NP and LS are congruent. They are also parallel (being perpendicular to the same line). That these two opposite sides are both congruent and parallel shows $PNLS$ to be a parallelogram. Hence, NL is parallel to DC. By the same argument with two new auxiliary lines, AB is parallel to MK. Therefore, MK and NL are parallel (since they are parallel to lines that are parallel to each other), so $MKLN$ is a trapezoid. Since $MN = AD - AM - ND$, $KL = BC - BK - LC$, $AM = BK$, $AD = BC$, and $NK = LC$, it follows that $MN = KL$. As corresponding angles, $\angle KLN = \angle LCS$; since triangle CLS is congruent to triangle DNP, $\angle LCS = \angle NDP$; as corresponding angles, $\angle NDP = \angle MNL$. Therefore, $\angle KLN = \angle MNL$. From this last identity (and that $NL = NL$, $MN = KL$), it follows (by having two sides and their included angle

congruent) that triangles MNL and KLN are congruent, and so their corresponding sides ML and KN are the same length.

This proof succeeds, but only by using a construction that many mathematicians would regard as artificial or "clever." (See [52], for instance, from where I have taken this example.) The construction is artificial because the proof using it seems forced to go to elaborate lengths — all because it fails to exploit the feature of the figure that most forcibly strikes us: its symmetry with respect to the line between the midpoints of the bases (Figure 4).

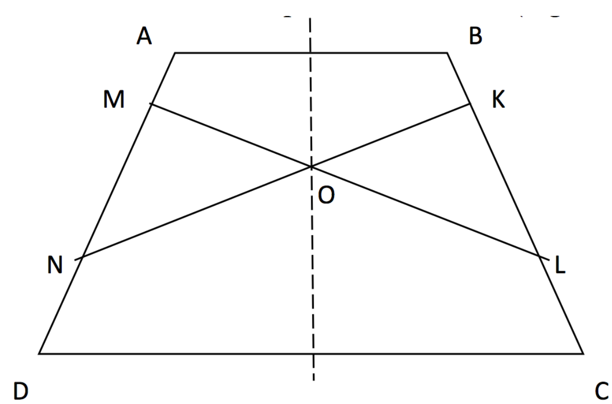

Figure 4: Symmetry in the trapezoid figure.

The theorem (that $ML = KN$) "makes sense" in view of the overall symmetry of the figure. Intuitively, a proof that fails to proceed from this symmetry strikes us as failing to focus on "what is really going on": that we have here the same figure twice, once on each side of the line of symmetry. Folding the figure along the line of symmetry, we find that NO coincides with LO and that MO coincides with KO, so that $MO + OL = KO + ON$, and hence $ML = KN$. Of course, to make this proof complete, we must first show that the point at which ML intersects KN lies on the line of symmetry. But that is also required by the overall symmetry of the figure: if they intersect off of the line of symmetry, then the setup will be symmetrical only if there is another point of intersection at the mirror-image location on the other side of the line of symmetry, but two lines (ML and KN) cannot intersect at more than one point.

Of course, this proof exploits a very simple symmetry: mirror reflection across a line. A proof in geometry can explain by virtue of exploiting a

more intricate symmetry in the setup. For instance, consider one direction of Menelaus' theorem:

> If the three sides of triangle ABC are intersected by a line l_1 (see Figure 5), where C' (A', B') is the point where l_1 intersects line AB (BC, CA, respectively), then
> $$\frac{AC'}{BC'}\,\frac{BA'}{CA'}\,\frac{CB'}{AB'} = 1.$$

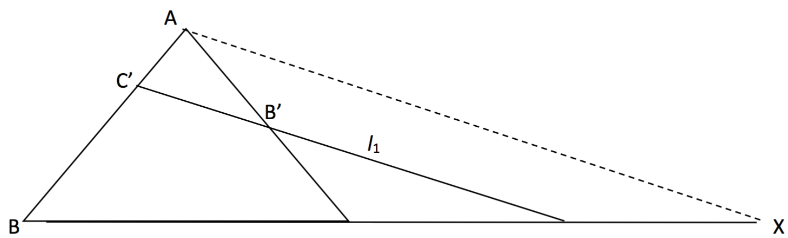

Figure 5: Menelaus' theorem.

We are immediately struck by the symmetry on the left side of this equation. (Indeed, we inevitably use the symmetry to get a grip on what the left side is all about.) The left side consists of a framework of primed points $\frac{\cdot C'}{\cdot C'}\,\frac{\cdot A'}{\cdot A'}\,\frac{\cdot B'}{\cdot B'}$ exhibiting an obvious symmetry: "A", "B", and "C" all play the same role around the primes (modulo the order from left to right, which makes no difference to the product of the three terms). Within this framework, the three unprimed points are arranged so that "A", "B", and "C" all again play the same role: each appears once on the top and once on the bottom, and each is paired once with each of the other two letters primed. These constraints suffice to fix the expression modulo the left-right order, which does not matter to their product, and modulo the inversion of top and bottom, which does not matter since the equation sets the top and bottom equal. In short, the left-hand expression is invariant (modulo features irrelevant to the equation) under any systematic interchange of "A", "B", and "C" around the other symbols. In the literature, the primed points are almost always named as I have named them here (e.g., with C' as the point where l_1 intersects line AB) in order to better display the symmetry of the expression.

Having recognized this symmetry in the theorem, we regard any proof of the theorem that ignores it as failing to explain why the theorem holds. For instance, consider this proof:

> Draw the line through A parallel to l_1 (dotted line in Figure 5); let X be its point of intersection with line BC. As corresponding angles, $\angle BA'C' = \angle BXA$. Therefore (since they also share $\angle B$), triangles $BC'A'$ and BAX are similar, so their corresponding lengths are in a constant proportion. In particular, $AC'/BC' = XA'/BA'$. Likewise, as corresponding angles, $\angle CB'A' = \angle CAX$. Therefore (since they also share $\angle B'CA'$), triangles ACX and $B'CA'$ are similar, so their corresponding lengths are in a constant proportion. In particular, $AB'/CB' = XA'/CA'$. Solving this for $XA' = (AB')(CA')/CB'$ and substituting the resulting expression for XA' into the earlier equation yields $AC'/BC' = AB'/BA'\, CA'/CB'$. The theorem follows by algebra.

Einstein says that this proof is "not satisfying" [30] and Bogomolny agrees [6]. Both cite the fact that (in Einstein's words) "the proof favors, for no reason, the vertex A [since the auxiliary line is drawn from that vertex], although the proposition [to be proved] is symmetrical in relation to A, B, and C" [30]. I agree: this argument depicts the symmetric result as arising "magically", whereas to explain why the theorem holds, we must proceed entirely from the symmetries of the figure over A-B-C.

Following Brunhes [8, page 84], Bogomolny [6] offers such a proof, which I now elaborate (see Figure 6):

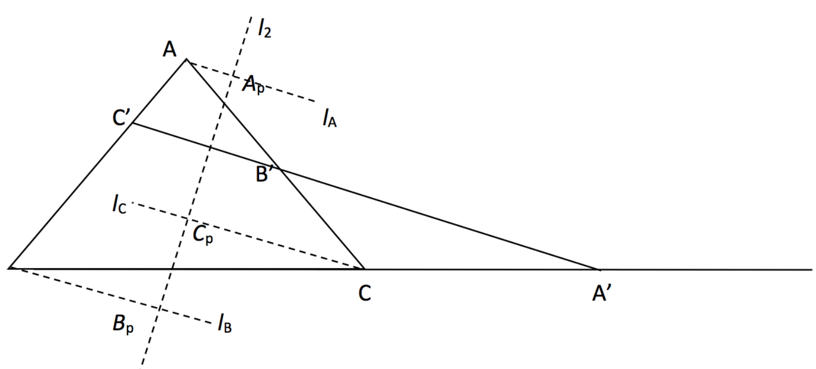

Figure 6: Proof of Menelaus' theorem.

> Add a line l_2 perpendicular to l_1. Project A onto l_2 by a line l_A from A parallel to l_1; let A_p be the point on l_2 to which A

is projected. Perform the same operation on B and C, adding lines l_B and l_C, points B_p and C_p. Of course, A', B', and C' all project to the same point on l_2 (since l_1 is their common line of projection), which can equally well be called "A'_p", "B'_p", or "C'_p". The following equation exhibits the same symmetry as the theorem:

$$\frac{A_pC'_p}{B_pC'_p}\,\frac{B_pA'_p}{C_pA'_p}\,\frac{C_pB'_p}{A_pB'_p} = 1.$$

This equation is true (considering that A'_p, B'_p, and C'_p are the same point, allowing a massive cancellation) and is strikingly invariant (modulo features irrelevant to the equation: left-right order and inversion of top and bottom) under systematic interchange of "A", "B", and "C" around the primes and subscript "p"s — the same symmetry that the theorem possesses. To arrive at the theorem, all we need to do is to find a way to remove the subscript "p"s from this equation, which is easily done. For any side of the triangle, l_1 and the two lines projecting its endpoints onto l_2 constitute three parallel lines, and all three are crossed by the side and by l_2. Now we use the lemma: *Whenever two transversals cross three parallel lines, the two segments into which the three parallels cut one transversal stand in the same ratio as the two segments into which the three parallels cut the other transversal.* That is, the ratio of one transversal's segments is preserved in the ratio of their projections onto the other transversal. By projecting each side onto l_2, we find

$$\frac{A_pC'_p}{B_pC'_p} = \frac{AC'}{BC'} \qquad \frac{B_pA'_p}{C_pA'_p} = \frac{BA'}{CA'} \qquad \frac{C_pB'_p}{A_pB'_p} = \frac{CB'}{AB'}.$$

Thus the theorem is proved by an argument that begins with an equation that treats A, B, and C identically and in each further step treats them identically. This proof reveals how features of the setup that are A-B-C symmetric are responsible for the symmetry of the theorem. The symmetry of the result does not just come out of nowhere. The general strategy of the proof is to project A', B', and C' onto the very same point (thereby treating them identically) by projecting the triangle's three sides onto the same line.

This explanation also shows that the auxiliary lines of a proof need not be "artificial"; that is, the use of auxiliary lines does not suffice to make the

proof non-explanatory. Although the auxiliary lines in the Euclidean proof of the trapezoid theorem were mere devices to prove the theorem, and likewise for line AX in the first proof of Menelaus' theorem, the scheme of auxiliary lines in the second proof of Menelaus' theorem is A-B-C symmetric [6].

7. Generalizing the proposal: mathematical explanations that do not exploit symmetries

I have now given several examples of mathematical explanations consisting of proofs that exploit symmetries. However, I do not mean to suggest that only a proof that appeals to some symmetry can explain why a mathematical theorem holds. Rather, I am using proofs by symmetry to illustrate the way in which certain proofs manage to become privileged as explanatory. Symmetries are not somehow intrinsically explanatory in mathematics. Rather, some symmetry in a mathematical result is often salient to us, and consequently, in those cases, a proof that traces the result back to a similar symmetry in the problem counts as explaining why the result holds. Some feature of a mathematical result other than its symmetry could likewise be salient, prompting a why question answerable by a proof deriving the result from a similar feature of the given. What it means to ask for a proof that explains is to ask for a proof that exploits a certain kind of feature in the setup — the same kind of feature that is outstanding in the result. The distinction between proofs that explain why some theorem holds and proofs that merely establish that it holds exists only when some feature of the result being proved is salient. That feature's salience makes certain proofs explanatory. A proof is accurately characterized as an explanation (or not) only in a context where some feature of the result being proved is salient.

My proposal predicts that if the result exhibits no noteworthy feature, then to demand an explanation of why it holds, not merely a proof that it holds, makes no sense. There is nothing that its explanation over and above its proof would amount to until some feature of the result becomes salient.[10]

[10] There may also be cases where the result exhibits a feature that is only slightly salient. If some proofs but not others exploit a similar feature in the problem, this difference would ground only a slight distinction between proofs that explain why and proofs that merely establish that some theorem holds. Another way for intermediate cases to arise is for a certain feature to be salient in the result, but for proofs to exploit to varying degrees a similar feature in the set up — rather than for any proof to proceed entirely from such a feature. See also footnote 15.

This prediction is borne out. For example, there is nothing that it would be for some proof to explain why and not merely to prove that

$$\int_1^3 (x^3 - 5x + 2)dx = 4.$$

Nothing about this result calls for explanation.

My proposal also predicts that if a result exhibits some noteworthy feature, but no proof traces that result to a similar feature in the setup, then the result has no explanation. This prediction is also borne out. Take the following example of a "mathematical coincidence" given by the mathematician Timothy Gowers:

> [C]onsider the decimal expansion of e, which begins $2.718281828\ldots$. It is quite striking that a pattern of four digits should repeat itself so soon — if you choose a random sequence of digits then the chances of such a pattern appearing would be one in several thousand — and yet this phenomenon is universally regarded as an amusing coincidence, a fact that does not demand an explanation [15, page 34]; cf. [4, page 140].

I take it, then, that mathematicians regard this fact as having no explanation. Of course, there are many ways to derive e's value, and thus to derive that the third-through-sixth digits of its base-ten representation are repeated in the seventh-through-tenth digits. For example, we could derive this result from the fact that e equals the sum of $1/n!$ for $n = 0, 1, 2, 3\ldots$. However, such a proof does not explain *why* the seventh-through-tenth digits repeat the third-through-sixth digits. It merely proves that they do. On my view, that is because the expression $1/0! + 1/1! + 1/2! + \cdots$ from which the proof begins does not on its face exhibit any feature similar to the repeated sequence of digits in e's decimal expansion. (None of the familiar expressions for calculating e makes any particular reference to base 10.) There is, I suggest, no reason why that pattern of digits repeats. It just does.

Let's now return to the example from Section 2: that every "calculator number" is divisible by 37. Is this fact a coincidence? (This is the question asked by the title of the *Mathematical Gazette* article that contributed this example to the mathematical literature.)

The striking thing about this result is that it applies to every single calculator number. In other words, the result's "unity" is salient. A proof

that simply takes each calculator number in turn, separately showing each to be divisible by 37, treats the result as if it were a coincidence. That is, it fails to explain why all of the calculator numbers are divisible by 37. Indeed, a case-by-case proof merely serves to highlight the fact that the result applies to every single calculator number. Especially in light of this proof, an explanation would be a proof that proceeds from a property common to each of these numbers (where this property is a genuine respect in which these numbers are similar, i.e., a mathematically "natural" property — unlike the property of being either 123321 or 321123 or ...) and that is common to them precisely because they are calculator numbers.[11] I gave such a proof in Section 2. I took it from a later *Mathematical Gazette* article entitled "No Coincidence" [36]. That proof exploits the fact that every calculator number can be expressed as $10^5a+10^4(a+d)+10^3(a+2d)+10^2(a+2d)+10(a+d)+a$ where a, $a+d$, $a+2d$ are three integers in arithmetic progression. These three integers, of course, are the three digits on the calculator keypads that, taken forwards and backwards, generate the given calculator number. Hence, this proof traces the fact that every calculator number is divisible by 37 to a property that they have in common by virtue of being calculator numbers.

In short, an explanation of this result consists of a proof that treats every calculator number in the same way.[12] This unified treatment makes the proof explanatory because what strikes us as remarkable about the result, especially in light of the case-by-case proof, is its unity: that it identifies a property common to every calculator number. The point of asking for an explanation is then to ask for a proof that exploits some other feature common to them because they are calculator numbers.[13]

[11]In [28], I am more explicit about what it takes for a mathematical predicate to denote a mathematically natural property — a genuine respect of similarity. Strictly speaking, that a given result identifies a property common to every single case of a certain sort is just a symmetry in the result — for example, that under a switch from one calculator number to any other, divisibility by 37 is invariant. Accordingly, I have already argued that a proof that works by exploiting the same sort of symmetry in the setup counts as explanatory. My view does not require that a theorem's displaying a striking symmetry be sharply distinguished from a theorem's being striking for its treating various cases alike.

[12]In [28], I am more explicit about what it takes to treat them all "in the same way". I also explore more general philosophical issues concerning explanation, especially the connections between explanation in mathematics and scientific explanation.

[13]If a result is remarkable for identifying something common to each case of an apparently diverse lot, then those "cases" may themselves be general results — as when each

A proof does not have to treat each instance separately in order to fail to treat them all together. Consider this proof by mathematical induction of the fact that the product of any three consecutive natural numbers is divisible by 6:

> The product of 1, 2, and 3 is 6, which is divisible by 6.
>
> Suppose that the product of $(n-1)$, n, and $(n+1)$ is divisible by 6. Let's show that the product of n, $(n+1)$, and $(n+2)$ is divisible by 6. By algebra, that product equals $n^3 + 3n^2 + 2n = (n^3 - n) + 3n(n+1)$. Now $(n^3 - n) = (n-1)n(n+1)$, so by

---

case is a theorem and the result identifies something common to each of them. Entirely dissimilar proofs of two theorems would fail to explain why those theorems involve a common element. On the other hand, proofs of each theorem may explain this pattern if the proofs themselves exploit a common element. Note the explanatory language in this remark from the mathematician Terry Gannon:

> There are lots of 'meta-patterns' in mathematics, i.e., collections of seemingly different problems that have similar answers, or structures that appear more often than we would have expected. Once one of these meta-patterns is identified it is always helpful to understand what is responsible for it . . . To give a trivial example, years ago while the author was writing up his PhD thesis he noticed in several places the numbers 1, 2, 3, 4, and 6. For instance, $\cos(2\pi r) \in \mathbb{Q}$ for $r \in \mathbb{Q}$ iff the denominator of r is 1, 2, 3, 4, or 6. Likewise, the theta function $\Theta[\mathbb{Z} + r](\tau)$ for $r \in \mathbb{Q}$ can be written as $\sum a_i \theta_3(b_i \tau)$ for some a_i, $b_i r \in \mathbb{R}$ iff the denominator of r is 1, 2, 3, 4, or 6. This pattern is easy to explain: they are precisely those positive integers n with Euler totient $\varphi(n) \leq 2$, that is, there are at most two positive numbers less than n coprime to n. The various incidences of these numbers can usually be reduced to this $\varphi(n) \leq 2$ property [13, page 168].

We can understand the general idea here even if we do not know what Jacobi's theta function is or how $\varphi(n) \leq 2$ is connected to these two theorems. The general idea is that proofs of the two theorems (one about cos, the other about $\Theta[\mathbb{Z} + r]$) can explain why $\{1, 2, 3, 4, 6\}$ figures in both if each of those proofs exploits exactly the same feature of $\{1, 2, 3, 4, 6\}$ — e.g., nothing about $\{1, 2, 3, 4, 6\}$ save that this set contains all and only the positive integers where $\varphi(n) \leq 2$. (A proof exploits something other than a single feature of $\{1, 2, 3, 4, 6\}$ if, for instance, it determines which numbers n are such that $\varphi(n) \leq 2$, and then proceed case-wise from there.) Since the striking feature of the two theorems (when presented as Gannon does) is that $\{1, 2, 3, 4, 6\}$ figures in both, the point of asking why the two theorems hold is plainly to ask for proofs of the two theorems where (non-trivially) both proofs exploit exactly the same feature of $\{1, 2, 3, 4, 6\}$. That both theorems involve exactly the positive integers n where $\varphi(n) \leq 2$ explains why both theorems involve $\{1, 2, 3, 4, 6\}$, and hence involve exactly the same integers.

> hypothesis, it is divisible by 6. And $n(n+1)$ is even, so $3n(n+1)$ is divisible by 3 and by 2, and therefore by 6. Hence, the original product is the sum of two terms, each divisible by 6. Hence, that product is divisible by 6.

This proof fails (by a whisker — see footnote 15) to treat every product of three consecutive natural numbers alike. Instead, it divides them into two classes: the product of 1, 2, and 3; and all of the others. This proof, then, does not supply a common reason why all of the products of three consecutive natural numbers are divisible by 6. Rather, it treats the first as a special case. Insofar as we found the theorem remarkable for identifying a property common to every triple of consecutive natural numbers, our point in asking for an explanation was to ask for a proof that treats all of the triples alike. This feature of the theorem is made especially salient by a proof that does treat all of the triples alike:

> Of any three consecutive natural numbers, at least one is even (i.e., divisible by 2) and exactly one is divisible by 3. Therefore, their product is divisible by $3 \times 2 = 6$.

This proof proceeds entirely from a property possessed by every triple.[14] Like the explanation of the fact that every calculator number is divisible by 37, this proof traces the result to a property common to every instance and so (when the unity of the result is salient) explains why the result holds.[15]

---

[14] Zeitz [54] mentions this theorem in passing but neither discusses it nor gives any proof of it. I wonder if he had in mind either of the proofs I discuss.

[15] It follows that in a context where the result's unity is salient, a proof by mathematical induction cannot explain the result, since a proof by mathematical induction always treats the first instance as a special case. However, the inductive proof of the triplet theorem is a far cry from the proof of the calculator-number theorem that treats each of the sixteen calculator numbers separately. The inductive proof nearly treats every triplet alike. It gives special treatment only to the base case; all of the others receive the same treatment. Therefore, although this inductive proof is not an explanation (when the result's unity is salient), it falls somewhere between an explanation and a proof utterly lacking in explanatory power. A result (displaying unity as its striking feature) having such a proof by induction, but (unlike the triplet theorem) having no proof that treats every case alike, has no fully qualified explanation but is not an utter mathematical coincidence either, since the inductive proof ties all but one of its cases together. On my account,

A proof may focus our attention on a particular feature of the result that would not otherwise have been salient. The proof may even call our attention to this feature because the proof conspicuously fails to exploit it. When this happens, the proof fails to qualify as explanatory. For instance, take one standard proof of the formula for the sum S of the first n natural numbers $1 + 2 + \cdots + (n-1) + n$.

> There are two cases.
>
> When n is even, we can pair the first and last numbers in the sequence, the second and second-to-last, and so forth. The members of each pair sum to $n+1$. No number is left unpaired, since n is even. The number of pairs is $n/2$ (which is an integer, since n is even). Hence, $S = (n+1)n/2$.
>
> When n is odd, we can pair the numbers as before, except now the middle number in the sequence is left unpaired. Again, the members of each pair sum to $n+1$. This time there are $(n-1)/2$ pairs, since the middle number $(n+1)/2$ is unpaired. The total sum is the sum of the paired numbers plus the middle number: $S = (n+1)(n-1)/2 + (n+1)/2$. This simplifies to $(n+1)n/2$ — remarkably, the same as the expression we just derived for even n.

Before having seen this proof, we would not have found it remarkable that the theorem finds that the same formula applies to both even n and odd n. However, this feature of the result strikes us forcibly in light of this proof. We might then well wonder: Is it a coincidence that the same formula emerges in both cases? This proof depicts it as an algebraic miracle. Accordingly, in this context, to ask for the reason why the formula holds, not merely a proof that it holds, is to ask for the feature (if any) common to both of these kinds of cases from which the common result follows.

Indeed there is such a feature; the result is no coincidence. Whether n is even or odd, the sequence's midpoint is half of the sum of the first and last numbers: $(1+n)/2$. Furthermore, all sequences of both kinds consist of numbers balanced evenly about that midpoint. In other words, for every number in the sequence exceeding the midpoint by some amount, the

---

the triplet theorem's inductive proof is inferior in explanatory power to the fully unifying proof, but nevertheless retains some measure of explanatory significance.

sequence contains a number less than the midpoint by the same amount. Allowing the excesses to cancel the deficiencies, we have each sequence containing n numbers of $(1+n)/2$ each, yielding our formula. This is essentially what happens in the standard proof of the formula, where each member in the sequence having an excess is paired with a member having an equal deficiency:

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

> If we pair the first terms, the second terms, and so forth in each sum, then each pair adds to $(n+1)$, and there are n pairs. So $2S = n(n+1)$, and hence $S = n(n+1)/2$.

This proof is just slightly different from the proof that deals separately with even n and odd n. The use of two sequences could be considered nothing but a trick for collapsing the two cases. Yet it is more than that. It brings out something that the earlier proof obscures: that the common result arises from a common feature of the two cases. For this reason, the earlier proof fails to reveal that the formula's success for both even n and odd n is no coincidence. What allows the second proof to show that this is no coincidence? It traces the result to a property common to the two cases: that the terms are balanced around $(1+n)/2$. (Whether or not a term in the sequence actually occupies that midpoint is irrelevant to this balancing.) The earlier proof does not exploit this common feature. Rather, it simply works its way through the two cases and magically finds itself with the same result in both.[16]

Similar phenomena arise in some explanations that are not proofs. For example, mathematicians such as Cardano and Euler had developed various tricks for solving cubic and quartic equations. One of Lagrange's tasks in his monumental 1770-1 memoir "Reflections on the Solution of Algebraic Equations" was to explain why his predecessors' various methods all worked ("*pourquoi ces méthods réussissent*" [26, page 206]; cf. [24, page 601]). By showing that these different procedures all amount fundamentally to the same method, Lagrange showed that it was no coincidence that they all worked. In other words, faced with the fact that Cardano's method works, Euler's

[16] Steiner deems the second proof "more illuminating" than an inductive proof [47, page 136]. He does not discuss mathematical coincidences or contrast the second proof with separate proofs for the even and odd cases.

method works, Tschirnhaus's method works, and so forth, one feature of this fact is obviously salient: that it identifies something common to each of these methods (namely, that each works). Lagrange's explanation succeeds by tracing this common feature to other features that these methods have in common — in sum, that they are all fundamentally the same method: "These methods all come down to the same general principle..." [26, page 355], quoted in [21, page 45]. Part of the point in asking why these methods all work is to ask whether their success can be traced to a feature common to them all or is merely a coincidence. (Lagrange then asked why this common method works; earlier we saw his explanation of the "resolvent".)

Simplicity is another feature that generally stands out when a mathematical result possesses it.[17] An especially simple result typically cries out for a proof that exploits some similar, simple feature of the setup. In contrast, a proof where the result, in all of its simplicity, appears suddenly out of a welter of complexity — through some fortuitous cancellation or clever manipulation — tends merely to heighten our curiosity about why the result holds. Such a proof leaves us wanting to know where such a simple result came from.

[17]Poincaré says that a brute-force proof is unsatisfying when the result exhibits some noteworthy feature such a symmetry or simplicity: "[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" [37, pages 373-374].

Is it unilluminating to characterize a "salient feature" as a feature that gives content (in the manner I have described) to a demand for mathematical explanation, while in turn characterizing "mathematical explanation" in terms of a salient feature? I do not think that my proposal is thereby rendered trivial or unilluminating. We can say plenty about salience other than through its role in mathematical explanation (by way of paradigm cases such as those given in this paper, the kinds of features that are typically salient, how features become salient, and so forth) and we can recognize a feature as salient apart from identifying a proof as explanatory. We can likewise say plenty about mathematical explanation apart from its connection to salience. The connection my account alleges between salience and mathematical explanation does no more to trivialize my account than van Fraassen's pragmatic approach to scientific explanation is trivialized by the fact that it characterizes a "scientific explanation" as an answer to a question defined in terms of a contrast class, while it characterizes the "contrast class" in a given case as consisting of the possible occurrences that are understood to be playing a certain role in a given question demanding a scientific explanation [51, page 127].

After such a proof is given, there often appears a sentence like the following: "The resulting answer is extremely simple despite the contortions involved to obtain it, and it cries out for a better understanding" [45, page 14]. Then another proof that traces the simple result to a similar, simple feature of the problem counts as explaining why the result holds.

Examples of this kind often arise in proofs of "partition identities." The number $p(n)$ of "partitions of n" is the number of ways that the non-negative integer n can be expressed as the sum of one or more positive integers (irrespective of their order in the sum). For instance, $p(5) = 7$, since 5 can be expressed in 7 ways: as 5, $4+1$, $3+2$, $3+1+1$, $2+2+1$, $2+1+1+1$, and $1+1+1+1+1$. (By convention, $p(0) = 1$.) Here is a "partition identity" (proved by Euler in 1748): *The number $O(n)$ of partitions of n into exclusively odd numbers ("O-partitions") equals the number $D(n)$ of partitions of n into parts that are distinct, i.e., that are all unequal ("D-partitions").* For instance, $O(5) = 3$ since 5, $3+1+1$, and $1+1+1+1+1$ are the O-partitions of 5, and $D(5) = 3$ since 5, $4+1$, and $3+2$ are the D-partitions of 5.

There are two standard ways of proving partition identities: either with "generating functions" or with "bijections". The generating function $f(q)$ for a sequence $a_0, a_1, a_2, \ldots$ is $a_0q^0+a_1q^1+a_2q^2+\cdots = a_0+a_1q+a_2q^2+\cdots$. It does not matter whether the sum in the generating function converges because it is merely a device for putting the sequence on display; "q^n" does not stand for some unknown quantity, but just marks the place where a_n appears. Such a "formal power series" can generally be manipulated in precisely the same manner as a genuine power series. For example, "Consider (or as that word often implies, 'look out, here comes something from left field')" [53, page 6]

$$(1+q+q^2+q^3\cdots)(1+q^2+q^4+q^6\cdots)(1+q^3+q^6+q^9\cdots)(1+q^4+q^8+q^{12}\cdots)\cdots$$

Multiplication of q^3 from the first factor with 1's from each other factor will contribute q^3 to the product. Another q^3 is contributed by q from the first factor multiplied by q^2 from the second and 1's from each of the rest. Any combination of exponents adding to 3 contributes a q^3. Each way of contributing q^3 corresponds to a partition of 3; for the mth factor, its representative's exponent divided by m equals the number of times that m appears as a part of that partition. For example, q^3 from the first factor and 1's from every other factor corresponds to the partition with three 1's and no other parts, whereas q from the first factor, q^2 from the second, and 1's from the

rest corresponds to the partition with one 1, one 2, and no other parts. Thus, the above product is $p(n)$'s generating function: $p(0) + p(1)q + p(2)q^2 + \cdots$. Since formally $(1 + q + q^2 + q^3 + \cdots)(1 - q) = 1$, the above product is

$$\left(\frac{1}{1-q}\right)\left(\frac{1}{1-q^2}\right)\left(\frac{1}{1-q^3}\right)\cdots$$

By including only the factors corresponding to the number of 1's, number of 3's, number of 5's, and so on:

$$(1 + q + q^2 + q^3 \cdots)(1 + q^3 + q^6 + q^9 \cdots)(1 + q^5 + q^{10} + q^{15} \cdots)$$

we produce the generating function for $O(n)$:

$$\left(\frac{1}{1-q}\right)\left(\frac{1}{1-q^3}\right)\left(\frac{1}{1-q^5}\right)\cdots$$

Returning to the mth factor $(1 + q^m + q^{2m} + q^{3m} \cdots)$ of $p(n)$'s generating function, we see that the terms beyond q^m allow m to appear two or more times in the partition, so their removal yields $D(n)$'s generating function

$$(1 + q)(1 + q^2)(1 + q^3)\cdots$$

By manipulating the generating functions in various ways (justified for infinite products by taking to the limit various manipulations for finite products), we can show (as Euler first did) that the two generating functions are the same, and hence that $O(n) = D(n)$:

$$\begin{aligned}
\left(\frac{1}{1-q}\right)\left(\frac{1}{1-q^3}\right)\left(\frac{1}{1-q^5}\right)\cdots &= \left(\frac{1}{1-q}\right)\left(\frac{1-q^2}{1-q^2}\right)\left(\frac{1}{1-q^3}\right)\left(\frac{1-q^4}{1-q^4}\right)\cdots \\
&= \left(\frac{1-q^2}{1-q}\right)\left(\frac{1-q^4}{1-q^2}\right)\left(\frac{1-q^6}{1-q^3}\right)\left(\frac{1-q^8}{1-q^4}\right)\cdots \\
&= \left(\frac{(1-q)(1+q)}{1-q}\right)\left(\frac{(1-q^2)(1+q^2)}{1-q^2}\right)\cdots \\
&= (1+q)(1+q^2)(1+q^3)\cdots.
\end{aligned}$$

Wilf terms this "a very slick proof" [53, page 10], which is to say that it involves not only an initial generating function "from left field", but also a sequence of substitutions, manipulations, and cancellations having no motivation other than that, miraculously, it works out to produce the simple result in the end. Proofs of

partition identities by generating functions, although sound and useful, in many cases "begin to obscure the simple patterns and relationships that the proof is intended to illuminate" [7, page 46].

In contrast, "[a] common feeling among combinatorial mathematicians is that a simple bijective proof of an identity conveys the deepest understanding of why it is true [1, page 9, italics in the original]. A "bijection" is a 1-1 correspondence; a bijective proof that $O(n) = D(n)$ finds a way to pair each O-partition with one and only one D-partition. Let us look at a bijective proof from Sylvester [49] that $O(n) = D(n)$. Display each O-partition as an array of dots, as in this representation (Figure 7) of the partition $7 + 7 + 5 + 5 + 3 + 1 + 1 + 1$ of 30:

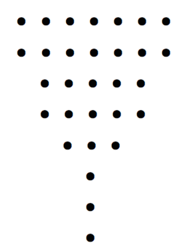

Figure 7: A partition of 30.

Each row has the number of dots in a part of the partition, with the rows weakly decreasing in length and their centers aligned. (Each row has a center dot since each part is odd.) Here is a simple way to transform this O-partition into a D-partition. The first part of the new partition is given by the dots on a line running from the bottom up along the center column and turning right at the top — 11 dots. The next part is given by the dots on a line running from the bottom, up along a column one dot left of center, turning left at the top — 7 dots. The next part runs from the bottom upward along a column one dot right of center, turning right at last available row (the second row from the top) — 6 dots. This pattern leaves us with the fish-hook diagram (Figure 8):

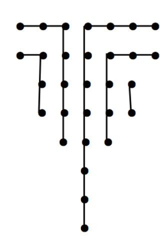

Figure 8: Fish-hook diagram of partition.

The result is a D-partition $(11 + 7 + 6 + 4 + 2)$, and the reverse procedure on that partition returns the original O-partition. With this bijection between O-partitions and D-partitions, there must be the same number of each. The key to the proof is that by "straightening the fish hooks", we can see the same diagram as depicting both an O-partition and a D-partition.

Because the bijection is so simple, this proof traces the simple relation between $O(n)$ and $D(n)$ to a simple relation between the O-partitions and the D-partitions. Moreover, the simple feature of the setup that the proof exploits is similar to the result's strikingly simple feature: the result is that $O(n)$ and $D(n)$ are the same, and the simple bijection reveals that n's O-partitions are essentially the same objects as n's D-partitions, since one can easily be transformed into the other. It is then no wonder that $O(n) = D(n)$, since effectively the same objects are being counted twice. This is the source of the explanatory power of a simple bijective proof. As Wilf (personal communication) puts it, a simple bijective proof reveals that "in a sense the elements [of the two sets of partitions] are the same, but have simply been encoded differently."[18]

That is, the bijective proof shows that in counting O-partitions or D-partitions, we are effectively counting the same set of abstract objects. Each of these objects can be represented by a dot diagram; the same diagram can be viewed as representing an O-partition and a D-partition. To emphasize this point [1, pages 16-17], consider this partition identity: *The number of partitions of n with exactly m parts equals the number of partitions of n having m as their largest part.* For instance, Figure 9 shows the partitions of 7 with exactly three parts.[19] Figure 10 shows each of these seen from a different vantage point — namely, after being rotated one-quarter turn counterclockwise and then reflected across a horizontal line above it. These are the partitions of 7 that have 3 as their largest part.

[18] Perhaps (contrary to the passages I have quoted) some combinatorial mathematicians regard the proof from generating functions as explanatory just like the simple bijective proof. Perhaps, then, I should restrict myself to identifying what it is that makes the bijective proof explanatory and not argue that the generating-function proof lacks this feature, but merely that it is perceived as lacking this feature by those who regard it as less explanatorily powerful than the bijective proof. (See the remark quoted in footnote 20.)

[19] The dot diagrams given here all have their rows weakly decreasing in length (just like the earlier diagram), but (unlike the earlier diagram) their rows do not have their centers aligned. (Since the parts are not all odd, some rows do not have center dots.) Rather, each row is aligned on the left. Clearly, the same partition may be displayed in various dot diagrams.

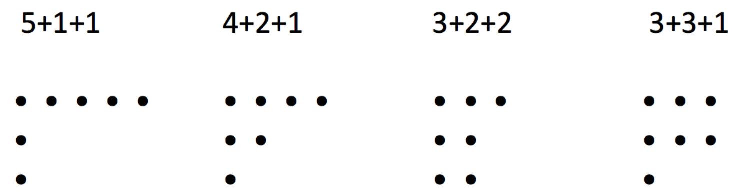

Figure 9: The partitions of 7 with exactly three parts.

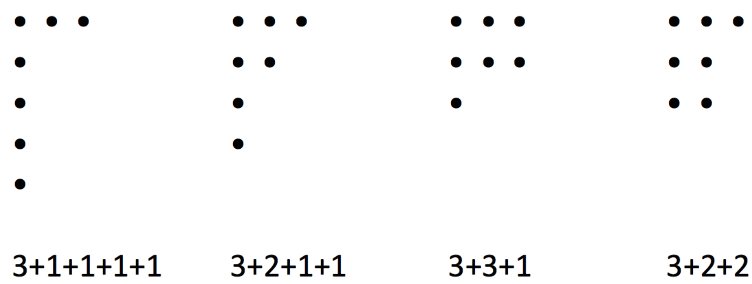

Figure 10: The partitions of 7 that have 3 as their largest part.

Clearly, then, both sets of partitions are represented by the same set of four arrays. A partition of one kind, seen from another vantage point, is a partition of the other kind. The number of partitions of one kind is the same as the number of partitions of the other kind because the same abstract object can be represented as either kind of partition, and the number of those abstract objects is the same no matter how we represent them.

Whereas the generatingfunctionology "seems like something external to the combinatorics" (Andrews, personal communication) that just miraculously manages to yield the simple result, a simple bijective proof shows that there is the same number of partitions of two kinds because partitions of those two kinds, looked at abstractly, are the same things seen from different perspectives. Such a proof "makes the reason for the simple answer completely transparent" [45, page 15]; "[i]t provides a natural' explanation . . . unlike the generating function proof which depended on a miraculous trick" [46, page 24].[20]

[20] I do not contend that any bijective proof (regardless of the bijection's complexity or artificiality) is explanatory, that no generating function proof is explanatory, or that there is never a good reason to seek a generating function proof once a simple, explanatory bijective proof has been found. Each kind of proof may have value; for instance, a gen-

8. Comparison to other proposals

Now I will briefly contrast my approach to proofs that explain with several others in the recent literature.

According to Steiner [47], a proof explains why all S_1's are P_1 if and only if it reveals how this theorem depends on S_1's "characterizing property" — that is, on the property essential to being S_1 that is just sufficient to distinguish S_1's from other entities in the same "family" (for example, to distinguish triangles from other kinds of polygons). To reveal the theorem's dependence on the characterizing property, the proof must be "generalizable". That is, if S_1's characterizing property is replaced in the proof by the characterizing property for another kind S_2 in the same family (but the original "proof idea" is maintained), then the resulting "deformation" of the original proof proves that all S_2's are P_2 for some property P_2 incompatible with P_1. Thus, the original explanation helps to show that there are different but analogous theorems for different classes in the same family.

Steiner's proposal nicely accommodates some of the explanatory proofs I have examined (at least under a natural reading of the relevant "family" and "proof idea"). For instance, the proof I presented as explaining why the product of any three consecutive natural numbers is divisible by $6(= 1 \times 2 \times 3)$ could be deformed to prove that the product of any four consecutive natural numbers is divisible by $24(= 1 \times 2 \times 3 \times 4)$. However, this four-number result could also be proved by mathematical induction — a proof that is a "deformation" of the proof by induction of the three-number result. Yet (I have argued) the inductive proof of the three-number result is not explanatory (in a context where the result's treating every case alike is salient), and its "generalizability" in Steiner's sense does not at all incline me to reconsider that verdict.[21] The proof treats the first triplet

erating function proof may be much shorter or help us to find proofs of new theorems. Also bear in mind that "[t]he precise border between combinatorial [i.e., bijective] and non-combinatorial proofs is rather hazy, and certain arguments that to an inexperienced enumerator will appear non-combinatorial will be recognized by a more facile counter as combinatorial, primarily because he or she is aware of certain standard techniques for converting apparently non-combinatorial arguments into combinatorial ones" [45, page 13].

[21]Steiner says that "inductive proofs usually do not allow deformation" and hence are not explanatory, because by replacing S_1s characterizing property with S_2s in the original inductive proof, we do not automatically replace the original theorem at the start of the inductive step with the new theorem to be proved — and in an inductive argument, the theorem must be introduced at the start of the inductive step [47, page 151]. It seems to me, however, that if we take the inductive proof that the product of any three consecutive natural numbers is divisible by $6(= 1 \times 2 \times 3)$ and replace the initial reference to three

of natural numbers differently from all of the others, rather than identifying a property common to every triplet that makes each divisible by 6.

Furthermore, some of the explanatory proofs I have identified simply collapse rather than yield new theorems when they are deformed to fit a different class in what is presumably the same "family". For instance, the proof from the isosceles trapezoid's symmetry does not go anywhere when we shift to a non-isosceles trapezoid, since the symmetry then vanishes.[22] For some of the other explanatory proofs I have presented, it is unclear what the relevant "family" includes. What, for instance, are the other classes in the family associated with d'Alembert's theorem that roots come in complex-conjugate pairs? Even if we could ultimately discover such a family, we do not need to find it in order to recognize the explanatory power of the proof exploiting the setup's invariance under the replacement of i with $-i$. (I will return to this point momentarily.)

I turn now to Kitcher [22, pages 208-9, 227]; [23, pages 423-6, 437], who offers a unified account of mathematical and scientific explanation; in fact, he sees explanation of all kinds as involving unification. Roughly speaking, Kitcher says that an explanation unifies the fact being explained with other facts by virtue of their all being derivable by arguments of the same form. Explanations instantiate argument schemes in the optimal collection ("the explanatory store") — optimal in that arguments instantiating these schemes manage to cover the most facts with the fewest different argument schemes placing the most stringent constraints upon arguments. An argument instantiating an argument scheme excluded from the explanatory store fails to explain.

Thus, Kitcher sees a given mathematical proof's explanatory power as arising from the proof's relation to other proofs (such as their all instantiating the same scheme or their covering different facts). My account contrasts with Kitcher's (and with Steiner's) in doing justice to the fact that (as we have seen in various examples) we can appreciate a proof's explanatory power (or impotence) just from

consecutive natural numbers with a reference to four consecutive natural numbers, then the first step of the inductive proof is automatically that the first case of four consecutive natural numbers is obviously divisible by their product ($1 \times 2 \times 3 \times 4 = 24$), and this gives us immediately the new theorem to be proved (namely, that the product of any four consecutive natural numbers is divisible by 24) for use at the start of the second step. So in this case, the inductive proof permits deformation.

[22] Resnik and Kushner [40, pages 147-152] likewise argue that the standard proof of the intermediate value theorem (which, they say, is explanatory if any proofs are) simply collapses if any attempt is made to shift it to cover anything besides an interval over the reals (such as over the rationals or a disjoint pair of intervals over the reals).

examining the details of that proof itself, without considering what else could be proved by instantiating the same scheme (or "proof idea") or how much coverage the given proof adds to whatever is covered by proofs instantiating other schemes. In addition, Kitcher regards all mathematical explanations as deriving their explanatory power from possessing the same virtue: the scheme's membership in the "explanatory store". It seems to me more plausible (especially given the diversity of our examples) to expect different mathematical explanations to derive their explanatory power by virtue of displaying different traits. On my approach, different traits are called for when the result being explained has different salient features.

A typical proof by "brute force" uses a "plug and chug" technique that is perforce applicable to a very wide range of problems. Presumably, then, its proof scheme is likely to belong to Kitcher's "explanatory store". (Not every brute-force proof instantiates the same scheme, but a given brute-force proof instantiates a widely applicable scheme.) For example, as I mentioned in Section 6, we could prove the theorem regarding isosceles trapezoids by expressing the setup in terms of coordinate geometry and then algebraically grinding out the result. The same strategy could be used to prove many other geometric theorems. Nevertheless, these proofs lack explanatory power. This brute-force proof of the trapezoid theorem is unilluminating because it begins by expressing the entire setup in terms of coordinate geometry and never goes on to characterize various particular features of the setup as irrelevant. Consequently, it fails to pick out any particular feature of isosceles trapezoids (such as their symmetry) as the feature responsible for the theorem. On my view, a brute-force proof is never explanatory when the salient feature of the theorem being explained is its symmetry or simplicity or some other such feature, since a brute-force proof does not exploit any such feature. (However, if the theorem's salient feature is its unity, then a brute-force proof can explain, since it may well treat all cases alike.)

A new proof technique can explain why some theorem holds even if that technique allows no new theorems to be proved. My approach can account for this feature of mathematical explanation. It is more difficult to accommodate on Kitcher's proposal, since any explanatory argument scheme must earn its way into the "explanatory store" by adding coverage (without unduly increasing the number of schemes or decreasing the stringency of their constraints).[23]

I turn finally to Resnik and Kushner [40], who doubt that any proofs explain simpliciter. They contend that a proof's being "explanatory" to a given audience

[23] Tappenden [50] offers a similar objection to this "winner take all" feature of Kitcher's account.

is nothing more than its being the kind of proof that the audience wants — perhaps in view of its premises, its strategy, its perspicuity, or the collateral information it supplies (or perhaps any proof whatsoever of the theorem would do). Contrary to Resnik and Kushner, I do not think that whenever someone wants a certain kind of proof, for whatever reason, then such a proof qualifies for them as explaining why the theorem being proved holds. Rather, mathematical practice shows that an explanatory proof requires some feature of the result to be salient and requires the proof to exploit a similar noteworthy feature in the problem. Thus, the demand for an explanation is not simply the demand for a certain kind of proof; the demand arises from a certain feature of the result and is satisfied only by a proof that involves such a feature from the outset. For example, we may want to see a proof of the "calculator number" theorem that proceeds by checking each of the sixteen calculator numbers individually. But this proof merely heightens our curiosity, motivating us to seek the reason why all of the calculator numbers are divisible by 37. It is not the case that any kind of proof we happen to want counts as an explanation when we want it.

9. Conclusion

I have tried to identify the basis on which certain proofs but not others are explanatory. Of course, I have not shown that all explanatory proofs work in the manner I have identified. I claim only to have sketched one very common way in which various proofs manage to explain. Admittedly, several fairly elastic notions figure in my idea of a proof's exploiting the same kind of feature in the problem as was salient in the result. This elasticity allows my proposal to encompass a wide range of cases (as I have shown). Insofar as the notions figuring in my proposal have borderline cases, there will correspondingly be room for mathematical proofs that are borderline explanatory.[24] However, the existence of such cases would not make a proof's explanatory power rest merely "in the eye of the beholder."

As I noted at the outset, it is challenging to find a source of explanatory asymmetry for mathematical explanations, since the usual suspects in scientific explanation (such as the priority of causes over their effects, and the priority of more fundamental laws of nature over more derivative laws) are unavailable. In response, I have gestured toward the priority that axioms in mathematics have over theorems, but I have also emphasized mathematical explanations that operate in connection with "problems", each of which is characterized by a "setup" and a "result". This structure of setup and result adds an asymmetry that enables

[24] See footnotes 10 and 15.

mathematical explanation to get started by allowing why questions to be posed (as in our first example, where Zeitz declares that he wants to "understand why the coin problem [i.e., the setup] had the answer [i.e., the result] that it did"). Of course, we bring this structure of "setup" and "result" to the mathematics, just as we bring the sensibility that privileges certain features as salient. But although the distinction between explanatory and non-explanatory proofs in the cases I have examined arises only when these cases are understood in terms of setups and results exhibiting salient features, it does not follow that a proof is explanatory merely by virtue of striking its audience as explanatory.

Of course, if some extraterrestrials differ from us in which features of a given theorem they find salient, then it follows from my account that those extraterrestrials will also differ from us in which proofs they ought to regard as explanatory. I embrace this conclusion. In different contexts, we properly regard different proofs of the same theorem to be explanatory — namely, in contexts where different features of the theorem are salient. Furthermore, if some extraterrestrials differed from us so much that they never regarded symmetry, unity, and simplicity as salient, then even if we and they agreed on the truth of various mathematical theorems, their practices in seeking and refining proofs of these theorems would differ so greatly from ours that it would be a strain to characterize them as doing mathematics. As cases such as the embrace of imaginary numbers illustrate, the search for mathematical explanations often drives mathematical discovery and innovation. Such extraterrestrials would not be seeking the same things as we do in doing mathematics.

Let me conclude by offering a conjecture regarding the relation between a mathematical proof's explanatory power and its beauty. I suggest that all explanatory proofs are beautiful (or, at least, not ugly). They derive their beauty from exactly what gives them their explanatory power, namely, from their exploiting precisely the feature in the setup that is salient in the theorem (such as its symmetry, unity, or simplicity). Moreover, the possession of such a feature renders a theorem beautiful (or, at least, not ugly).

This proposal accounts for many aspects of beauty in mathematics. For instance, it accounts for the fact that brute-force proofs are not reckoned to be beautiful when the theorem being explained is strikingly symmetrical or simple. As Hardy pungently comments: "'enumeration by cases' ... is one of the duller forms of mathematical argument" [18, page 113]. Likewise, Bogolmy characterizes the explanatory and non-explanatory proofs of Menelaus' theorem as "elegant" and "ugly", respectively [6]. Similarly, consider Morley's theorem (Figure 11): that for any Euclidean triangle, an equilateral triangle is formed by the three intersection points of adjacent interior angle trisectors.

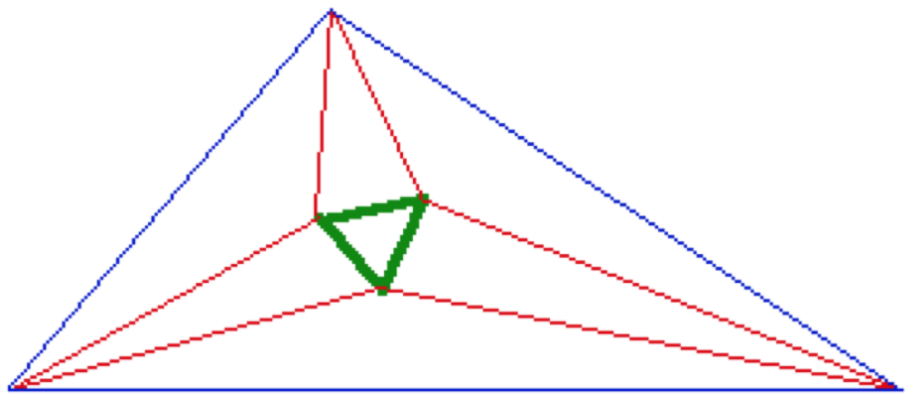

Figure 11: Morley's theorem.

This result is widely deemed to be beautiful in virtue of the striking symmetry that it reveals to be buried within even the most ungainly triangle. However, no proof of the theorem has yet been found that is beautiful or that explains why the theorem holds; none exploits some hidden symmetry in the setup. The proofs it has received are characterized as not only ugly — as involving "elephantine" geometrical constuctions [10] — but also explanatorily impotent:

> [E]ven when I read the much simpler proof based on trigonometry, or the fairly simple geometric proof due to Navansiengar, there was still too much complexity and lack of motivation. (A series of lucky breaks!) Were we to give up, forever, understanding the Morley Miracle? [34, page 31]

Perhaps Morley's "miracle" has no explanation and no beautiful proof, despite being a beautiful theorem.

This proposal leaves room for beautiful proofs of theorems that have no salient features and hence no explanations. It also does not rule out beautiful, non-explanatory proofs of theorems that have explanations. There may be many different ways in which proofs come to qualify as "beautiful". This proposal also suggests that like its explanatory power, a proof's beauty may be context sensitive since it may depend on a certain feature of the theorem being salient.

Mathematicians do occasionally reflect upon explanation in mathematics. For instance, Timothy Gowers writes:

> [Some] branches of mathematics derive their appeal from an abundance of mysterious phenomena that demand explanation. These might be striking numerical coincidences suggesting a deep relationship between areas that appear on the surface to have nothing to do with each other, arguments which prove interesting results by brute force and therefore do not satisfactorily explain them, proofs that apparently depend on a series of happy accidents . . . [14, page 73].

I hope that this paper has managed to unpack some of these provocative remarks.

References

[1] George E. Andrews and Kimmo Eriksson, *Integer Partitions*, Cambridge University Press, Cambridge, 2004.

[2] Michael Atiyah, "How research is carried out," in *Collected Works*, vol. 1, Clarendon, Oxford, 1988, pages 211-216.

[3] Alan Baker, "Are there genuine mathematical explanations of physical phenomena?," *Mind*, Volume **114** (2005), pages 223-238.

[4] Alan Baker, "Mathematical accidents and the end of explanation," in *New Waves in Philosophy of Mathematics*, Otávio Bueno and Øystein Linnebo, eds., Palgrave Macmillan, Houndmills, 2009, pages 137-159.

[5] A. W. Bell, "A study of pupils' proof-explanations in mathematical situations," *Educational Studies in Mathematics*, Volume **7** (1976), pages 23-40.

[6] A. Bogomolny, Menelaus theorem: proofs ugly and elegant - A. Einstein's view — *From Interactive Mathematics Miscellany and Puzzles*; available at `http://www.cut-the-knot.org/Generalization/MenelausByEinstein.shtml#F`, accessed on April 12, 2009.

[7] David M. Bressoud, *Proofs and Confirmations: The Story of the Alternating Sign Matrix Conjecture*, Cambridge University Press, Cambridge, 1999.

[8] Edmond Brunhes [Frère Gabriel-Marie], *Exercices de Géometrié*, Editions Jacques Gabay, Paris, 1991, reprint of the 6$^{\text{th}}$ ed., A. Mame, Tours, 1920.

[9] William Byers, *How Mathematicians Think*, Princeton University Press, Princeton, 2007.

[10] Richard L. Francis, "Modern mathematical milestones: Morley's mystery," *Missouri Journal of Mathematical Sciences*, Volume **14** Number 1 (2002). [no page number]

[11] David Gale, "Proof as explanation," *The Mathematical Intelligencer*, Volume **12** Number 1 (1990), page 4.

[12] David Gale, "Mathematical entertainments," *The Mathematical Intelligencer*, Volume **13** Number 1 (1991), pages 40-42.

[13] Terry Gannon, *Moonshine Beyond the Monster*, Cambridge University Press, Cambridge, 2006.

[14] W.T. [Timothy] Gowers, "The two cultures of mathematics," in *Mathematics: Frontiers and Perspectives*, V. Arnold, M. Atiyah, P. Lax, and B. Mazur, eds., American Mathematical Society, Providence, 2000, pages 65-78.

[15] W.T. [Timothy] Gowers, "Mathematics, memory, and mental arithmetic", in *Mathematical Knowledge*, Mary Leng, Alexander Paseau, and Michael Potter, eds., Oxford University Press, Oxford, 2007, pages 33-58.

[16] Emily R. Grosholz, "The partial unification of domains, hybrids, and the growth of mathematical knowledge," in *The Growth of Mathematical Knowledge*, Emily Grosholz and Herbert Breger, eds., Kluwer, Dordrecht, 2000, pages 81-91.

[17] Orna Harari, "Procus' account of explanatory demonstrations in mathematics and its context," *Archiv für Geschichte der Philosophie*, Volume **90** Number 2 (2008), pages 137-164.

[18] G. H. Hardy, *A Mathematician's Apology*, Cambridge University Press, Cambridge, 1940.

[19] Lulu Healy and Celia Hoyles, "A study of proof conceptions in algebra," *Journal for Research in Mathematics Education*, Volume **31** (2000), pages 396-428.

[20] Carl G. Hempel, "Explanation and prediction by covering laws," in *The Philosophy of Carl G. Hempel*, James Fetzer, ed., Oxford University Press, New York, 2001, pages 69-86.

[21] B. Melvin Kiernan, "The development of Galois theory from Lagrange to Artin," *Archive for History of Exact Sciences*, Volume **8** (1971), pages 40-154.

[22] Philip Kitcher, *The Nature of Mathematical Knowledge*, Oxford University Press, Oxford, 1984.

[23] Philip Kitcher, "Explanatory unification and the causal structure of the world", in *Scientific Explanation*, Philip Kitcher and Wesley Salmon, eds., Minnesota Studies in the Philosophy of Science, vol. 13, University of Minnesota Press, Minneapolis, 1989, pages 410-505.

[24] Morris Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York, 1972.

[25] Joseph-Louis Lagrange, "Note XIII: Sur la Résolution des Équations Algébriques," in *Traité de la Résolution des Équations Numériques de Tous les Degrés*, 3rd ed., Bachelier, Paris, 1826.

[26] Joseph-Louis Lagrange, "Refléxions sur la Résolution Algébrique des Équations," in *Œuvres de Lagrange*, vol. 3, J.-A. Serret, ed., Gauthier-Villars, Paris, 1869, pages 205-421.

[27] Marc Lange, "What makes a scientific explanation distinctively mathematical?," *British Journal for the Philosophy of Science*, Volume **64** (2013), pages 485-511.

[28] Marc Lange, *Because Without Cause*, Oxford University Press, New York (to appear).

[29] Mark Levi, *The Mathematical Mechanic*, Princeton University Press, Princeton, 2009.

[30] Abraham S. Luchins and Edith H. Luchins, "The Einstein-Wertheimer correspondence on geometric proofs and mathematical puzzles," *Mathematical Intelligencer*, Volume **12** Number 2 (1990), pages 35-43.

[31] Paolo Mancosu, "Mathematical explanation: Why it matters," in *The Philosophy of Mathematical Practice*, Paolo Mancosu, ed., Oxford University Press, Oxford, 2008, pages 134-150.

[32] Vimolan Mudaly and Michael de Villiers, "Learners' needs for conviction and explanation within the context of dynamic geometry," *Pythagoras*, Volume **52** Number 8 (2000), pages 20-23.

[33] Ernest Nagel, *The Structure of Science*, Harcourt, Brace, New York, 1961.

[34] Donald J. Newman, "The Morley miracle," *The Mathematical Intelligencer*, Volume **18** Number 1 (1996), pages 31-32.

[35] Pierre Nicole and Antoine Arnauld, *Logic; or the Art of Thinking*, William Taylor, London, 1717.

[36] Eric Nummela, "No coincidence," *The Mathematical Gazette*, Volume 71 Number 456 (1987), page 147.

[37] Henri Poincaré, "The future of mathematics," in *Science and Method* (trans. George Bruce Halsted), *The Foundations of Science*, The Science Press, New York, 1913, pages 369-383.

[38] Stathis Psillos, *Causation & Explanation*, Acumen, Chesham, 2002.

[39] Nicholas Rescher, *Scientific Explanation*, Free Press, New York, 1970.

[40] Michael Resnik and David Kushner, "Explanation, independence and realism in mathematics," *British Journal for the Philosophy of Science*, Volume **38** (1987), pages 141-158.

[41] J. J. C. Smart, "Explanation — Opening address," in *Explanation and Its Limits*, Dudley Knowles, ed., Cambridge University Press, Cambridge, 1990, pages 1-20.

[42] Elliott Sober, *The Nature of Selection*, University of Chicago Press, Chicago, 1984.

[43] Roy Sorensen, *Mathematical coincidences* (to appear).

[44] Michael Spivak, *Calculus*, 2nd ed., Publish or Perish, Berkeley, 1980.

[45] Richard P. Stanley, *Enumerative Combinatorics*, 2nd ed., vol. 1, Cambridge University Press, Cambridge, 2012.

[46] Richard P. Stanley and Anders Björner, *A Combinatorial Miscellany*, L'Enseignement Mathématique, Geneva, 2010.

[47] Mark Steiner, "Mathematical explanation," *Philosophical Studies*, Volume **34** (1978), pages 135-151.

[48] Mark Steiner, "Mathematics, explanation, and scientific knowledge," *Noûs*, Volume **12** (1978), pages 17-28.

[49] J. J. Sylvester, "A constructive theory of partitions, arranged in three acts, an interact and an exodion," *American Journal of Mathematics* **5** (1882), pages 251-330.

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

[51] Bas Van Fraassen, *The Scientific Image*, Clarendon, Oxford, 1980.

[52] Shlomo Vinner and Evgeny Kopelman, "Is symmetry an intuitive basis for proof in Euclidean geometry?," *Focus on Learning Problems in Mathematics*, Volume **20** Numbers 2-3 (1998), pages 14-26.

[53] Herbert Wilf, Lectures on Integer Partitions at University of Victoria (2000); available at http://www.math.upenn.edu/~wilf/PIMS/PIMSLectures.pdf, accessed on September 24, 2009.

[54] Paul Zeitz, "Graph Theory," handout for the Berkeley Math Circle on October 8, 2000.

[55] Paul Zeitz, *The Art and Craft of Problem Solving*, 2nd ed., Wiley, Hoboken, 2007.