Philosophical Studies
https://doi.org/10.1007/s11098-026-02651-6

Perfecting the symmetry: a variety of inference to the best explanation in mathematics

Marc Lange[1]

Received: 16 September 2025 / Accepted: 10 October 2026
© The Author(s) 2026

Abstract
Although a given mathematical theorem can typically be proved in many ways, only some of those proofs are regarded by mathematicians as *explaining why* that theorem holds. Furthermore, mathematicians routinely use non-deductive arguments to *confirm* conjectures before they have found proofs (or disproofs) of them. These two features of mathematical practice suggest that some confirmation in mathematics proceeds by "inference to the best explanation" (IBE). Furthermore, a mathematical theorem may exhibit a striking symmetry, and a proof of such a theorem would have to exploit a similar symmetry in order to explain why that theorem holds. This aspect of mathematical explanation reveals an important, heretofore unrecognized form of IBE in mathematics. In such a case, a conjecture is confirmed by some mathematical discovery partly because the conjecture "perfects the symmetry" in that discovery. This paper examines one historical example of such an IBE: Brahmagupta and Snell (independently) regarded Heron's formula for the area of triangles as confirming a conjecture not merely because that conjecture generalizes Heron's formula, but also because that conjecture perfects the formula's symmetry.

Keywords Brahmagupta · Confirmation · Conjectures · Explanation · Inference to the best explanation · Heron's theorem · Mathematics · Snell

1 The confirmation of mathematical conjectures

Although proof occupies a privileged place in mathematics, mathematicians frequently possess evidence that confirms a conjecture but falls short of proving it. This evidence may justify mathematicians in regarding the conjecture as promising and

✉ Marc Lange
marc.lange@rutgers.edu

[1] Department of Philosophy, Rutgers University–New Brunswick, New Brunswick, USA

Published online: 06 November 2026

in expending considerable effort trying to prove it.[1] For example, Gauss recorded in his mathematical diary for May 30, 1799 that he had found (by laborious computation) that the reciprocal of the arithmetic–geometric mean iteration (agm) of 1 and $\sqrt{2}$ agrees to at least 11 decimal places with $\frac{2}{\pi}\int_0^1 \frac{dx}{\sqrt{1-x^4}}$, which had been calculated to 17 decimal places by Stirling in 1730.[2] Prompted by this discovery, Gauss worked to prove (and eventually did prove) this equality.

This paper identifies one important variety of non-deductive reasoning by which mathematical conjectures have sometimes been confirmed by mathematical evidence. This variety of reasoning has not previously been discussed. I will work my way towards it in the first three sections of this paper.

What does it take for a piece of mathematical evidence to confirm (without proving) a given conjecture? George Pólya, in his classic 1954 book *Mathematics and Plausible Reasoning*, builds his account of the confirmation of mathematical conjectures around "the fundamental inductive pattern"[3] that "theories are confirmed by their [non-trivial] consequences or predictions" (Pólya, 1954:II, p. 4). On Pólya's picture, a confirmed conjecture typically generalizes the pattern exhibited by the data confirming it. This hypothetico-deductive picture of confirmation seems to fit many mathematical examples, including Gauss's taking the match between agm(1,$\sqrt{2}$)'s first 11 decimal places and Stirling's integral as confirming a generalization of that match: that the two quantities agree to all decimal places.

This picture of a mathematical conjecture as confirmed by its successful predictions (where the conjecture typically generalizes the evidence confirming it) has a great deal of common-sense plausibility. But it fails to address a crucial aspect of confirmation familiar from many philosophical discussions of confirmation in science: that confirmation is selective. A finite body of evidence fits *many* hypothetical, more general patterns, but it is not the case that the evidence confirms, of *each* of those patterns, that it will continue. Only some of these hypotheses are confirmed.

For instance, a mathematical conjecture can be an *arbitrary* conjunction, such as Goldbach's conjecture (that every even integer greater than 2 can be expressed as the sum of two primes) conjoined with some presumably unrelated conjecture, such as the Twin Prime conjecture (that there are infinitely primes p where p+2 is also prime). Mathematicians might well regard the verification of Goldbach's conjecture for all even numbers below 10^{18} as confirming another of that conjecture's predictions (e.g., for some very large, even number N that has not yet been demonstrated to be the

[1] Brown (2008: p. 164; 2017: p. 48) and Franklin (2013: p. 12) have pointed this out.

[2] The agm(x,y) is the limit of both of the linked sequences a_i and b_i where $a_{n+1} = (a_n + b_n)/2$, $b_{n+1} = \sqrt{a_n b_n}$, $a_0 = x$, $b_0 = y$. 1/agm(1,$\sqrt{2}$) = 0.8346268… Borwein and Devlin (2009: pp. 4–5) and Rodriguez Villegas (2007: ix) portray Gauss as justified in regarding his evidence as strongly confirmatory. (My thanks to James Robert Brown for suggesting the former book.) This example of Gauss's approach is not unusual. In 1817, Gauss wrote (quoted by Gray, 2018: ix): "It is characteristic of higher arithmetic that many of its most beautiful theorems can be discovered by induction with the greatest of ease but have proofs that lie anywhere but near at hand…" (Regarding "induction", see note 3.)

[3] By "inductive", Pólya is not referring to mathematical induction, which is a form of proof (i.e., deductive reasoning). Rather, Pólya means the kind of reasoning by which hypotheses are confirmed but not proved.

sum of two primes, that N accords with Goldbach's conjecture). But mathematicians would not —and *should* not—regard the many observed instances of Goldbach's conjecture (without any exceptions having been found) as confirming the Twin Prime conjecture, even though the evidence is entailed by a conjecture (the conjunction of Goldbach's conjecture and the Twin Prime conjecture) that also entails the Twin Prime conjecture. Evidently, some considerations beyond mere relations of logical entailment (which is all that Pólya's hypothetico-deductive "fundamental inductive pattern" takes into account) must step in to determine which conjectures have their unverified predictions confirmed by the verification of their other predictions.[4]

2 IBE in science

Pólya clearly intends to model the confirmation of *mathematical* conjectures on the confirmation of *scientific* hypotheses. In science as in mathematics, relations of logical consequence must be supplemented by other considerations to determine whether the hypothesis's success in one prediction confirms another of its predictions. One consideration that has often been thought (e.g., by Lipton, 2004: p. 66) to play this important role in science (at least in some cases) is *explanation*. Oftentimes a key part of the reason that we are justified in taking F as confirming G is that we have some justified confidence that F and G (should they both turn out to be true) would have a common explanation (i.e., would hold for the same reason).[5] For instance, our discovery that one stone is green confirms the hypothesis that another stone (whose color we have not yet checked) is green insofar as we are justly confident that the two stones consist of the same mineral and so have their colors (whatever they turn out to be) for the same reason (presumably having to do with their common molecular constitution). On the other hand, suppose we justly believe that it would be a coincidence for each of the families on my block to have two children; in other words, we justly believe that if this pattern holds, then it is not the case that each family has two children for the same reason. Rather, we justly believe that different causes are then at work for each family. In that case, our discovering that the Jones and Brown families on my block have two children would not (without some special background information) justly raise our confidence that the Smith family on the block also has two children. In this way, our prior just opinions about the explanation of our evidence can sometimes help to determine which hypotheses we can justly regard as having their predictive accuracy confirmed by that evidence.

If the degree of confirmation is strongly influenced by prior opinions expressly concerning the sort of explanation that the confirming evidence is liable to have, then

[4] That a generalization's past predictive success may fail to confirm some of the generalization's other consequences, even if it confirms the generalization, is very familiar from the literature on confirmation in science. Two classic discussions are Goodman, (1983: pp. 66-83) and Salmon, (1975). See note 5.

[5] That explanatory considerations (e.g., our justified confidence that there is a live hypothesis H that would explain both F and G) sometimes enable F to justly confirm G does not entail any simple general rule along these lines. For instance, it is not a general rule that if F justly confirms H and if H entails G, then F justly confirms G. That rule (what Hempel (1965) calls the "special consequence condition") has long been known to have counterexamples for incremental confirmation (Salmon, 1975).

that confirmation operates by a form of reasoning often called (somewhat misleadingly) "Inference to the Best Explanation" (IBE).[6] One reason that this terminology is misleading is that IBE need not involve any *inference* to a hypothesis. When the evidence is not strong enough to warrant inferring any of the competing hypotheses to be true, IBE uses explanatory considerations as a guide to the degrees of confidence that we should attribute to the rival hypotheses. Furthermore, according to IBE, explanatory considerations are only one kind of factor that we should use in judging a hypothesis's plausibility; explanatory considerations can be outweighed by other factors.

Perhaps the most obvious reason why the label "inference to the *best* explanation" is unhelpful is that it leaves unspecified what makes one candidate explanation *better* than another. When employing IBE, a scientist relies upon her tentative, defeasible, justified expectations regarding the kind of explanation the evidence has. For example, Darwinians in the nineteenth century argued that the anatomical similarities among a human's arm, cat's leg, whale's flipper, and bat's wing would receive a better explanation by their having a common cause in the anatomy of their common ancestor than by their being caused by separate, independent acts of divine creation. Darwinians justly regarded the anatomical similarities as unlikely to be a coincidence, but rather as liable to have a common explanation. The same sort of justified explanatory expectations underwrote the confirmatory power of many other pieces of evidence for evolution from common ancestors. As Darwin (1889: p. 421) wrote: "It can hardly be supposed that a false theory would explain, *in so satisfactory a manner* as does the theory of natural selection, the several large classes of facts above specified" (where I have emphasized Darwin's reference to the *quality* of the potential explanation). Darwin added that the same "method of arguing" not only is "used in judging of the common events of life", but also "has often been used by the greatest natural philosophers." He mentioned that the available evidence had once fit both the Copernican and the Ptolemaic models of the heavens, but Copernicans were entitled to judge their model to be more plausible because it supplies better explanations of various observed relationships between the Sun and the planets—relationships that the Ptolemaic model can reproduce but must regard as mere coincidences.[7]

Why were astronomers justified in regarding the Copernican model as deriving some epistemic support from the *quality* of the explanations that it would supply—from what Lipton (2004) calls the model's "loveliness"? More generally, why should a hypothesis's loveliness (over and above its logical and statistical relations to the evidence, including to the known facts it would explain) make any difference to its likeliness? Lipton (2004: pp. 70,144) calls this "Voltaire's objection" to IBE. It would take more space than I have available here to properly address this challenge. But I will briefly say how I would reply to it, since I want to avoid giving the impres-

[6] For the tradition of characterizing some confirmation as operating by IBE, see the surveys and references in Lipton, (2004) and Lycan, (2002). (IBE has also had its share of critics, notably van Fraassen, 1980, 1989 and Salmon, 2001.)

[7] Darwin's historical claim is correct. For instance, Kepler wrote in 1596 that the most "important" consideration favoring Copernicus's model over Ptolemy's is that in Copernicus's model, "reasons are supplied for a great many … matters for which Ptolemy for all his many motions could give no reason." (Kepler, 1981: pp. 75–6).

sion that IBE must rest on some a priori belief that a possible world where there are lovelier explanations is more likely (which is the Panglossian "best of all possible worlds" idea that motivated Lipton to name this challenge after Voltaire). I also want to avoid leaving the impression that (as Lycan (2002), among others, has held) no deeper justification can be offered for appealing to a hypothesis's loveliness to support its likeliness. I also do not want to give the impression that "loveliness" has no epistemic significance, but instead reflects merely an aesthetic preference for certain theories (as Lipton's term unwittingly suggests)—or that IBE reflects (as van Fraassen (1980: p. 87) maintains) "specifically human concerns, a function of our interests and pleasures, which make some theories more valuable and appealing to us than others".

On my view (following Lange, 2022, 2024), it is our prior knowledge of various other explanations that determine in a given case which features of a potential explanation would enhance its "loveliness". Investigators are sometimes epistemically justified, by the various explanations that they have already discovered, in expecting that certain other, as yet unexplained facts will turn out to have the same sorts of explanations. A hypothesis that would supply some fact with the expected sort of explanation is thereby supported (to some degree and defeasibly). In supplying explanations having the expected features, a hypothesis is "lovelier"; this is (on my view) what loveliness ultimately consists in.

In the Copernican case, for instance, the celestial phenomena that were to be explained included various "suspicious" correlations among the positions and motions of the celestial bodies in Earth's sky. These correlations are "suspicious" in that they suggest a certain sort of explanation: that the Sun's and planets' positions and motions in Earth's sky are not independent, but have important common causes. These correlations suggest such common causes because these correlations resemble various other, familiar correlations that are explained (we have already discovered) by the correlated events having common causes. (For instance, we know that the reason why certain people all developed the same illness at the same time was that they all ate the same tainted food—a common cause explaining the correlation among the illnesses.) The Copernican model's "loveliness" is that it would provide these "suspicious" correlations with the sorts of explanations that our knowledge of other explanations justly leads us to expect. (According to Copernicanism, the key common cause of the solar and planetary positions and motions in the Earth's sky is the Earth's own orbital motion around the Sun.)

The Copernican model's loveliness thereby contributes to its likeliness. In other cases where IBE operates, other features of various putative explanations make the hypotheses that would supply them more or less "lovely". (Supplying a common cause makes Copernicanism lovelier but may make a putative explanation of some other phenomenon simplistic and hence less lovely). Moreover, in different cases of IBE, investigators have different particular empirical reasons for expecting various facts to have certain sorts of explanations. But in all of these IBEs, a hypothesis's loveliness justly influences its likeliness because we should judge the quality of the explanations it would supply in light of the other explanations that we have already discovered. Loveliness, on this view, is an epistemic rather than an aesthetic consideration. The knowledge underwriting an instance of IBE concerns other explanations

per se, over and above knowledge of the logical and statistical relations between the hypothesis being confirmed and our evidence.

Beginning in the next section, I will argue that IBE sometimes operates in the justified confirmation of mathematical conjectures much like IBE sometimes operates in the justified confirmation of scientific hypotheses.[8] But nothing that I will say below will depend on the details of the account I have just sketched of why loveliness justly bears on likeliness. In particular, my argument (beginning in Sect. 5) for the importance of a previously philosophically underappreciated kind of IBE in math (that involves "perfecting the symmetry") will not presuppose this account of IBE's basis (beyond that loveliness does justly bear on likeliness).

[8] I do not regard IBE as operating in *every* case where a mathematical conjecture is justly confirmed but not proved (any more than IBE operates whenever a scientific hypothesis is justly confirmed). Many other kinds of justified prior opinions besides opinions about explanations can underwrite confirmation in mathematics (just as in science). Although Pólya (1954) officially regards hypothetico-deductivism as the fundamental type of all good non-deductive reasoning, his book's many examples actually illustrate the tremendous diversity of the justified prior opinions that in mathematical practice underwrite the justified confirmation of conjectures. Sometimes these opinions concern various different sorts of analogies between the theorem being confirmed and other theorems that have already been proved (or confirmed). In some other cases, these opinions concern the way that a familiar proof strategy (and its track record of applicability) justly appears to hold promise for proving the given conjecture.

As I emphasized in Sect. 1, hypothetico-deductivism (that a live hypothesis is justly confirmed by the discovery that one of its predictions succeeded) is insufficient to account for the cases where a hypothesis's *predictions* are justly confirmed (to some degree) by the hypothesis's having made a successful prediction (such as the confirmation that Goldbach's conjecture holds in unexamined cases by its holding in all cases that have been checked). That is because typically, the entire hypothesis (not just the particular prediction that was discovered to succeed) is needed to entail the prediction to be confirmed. So even if hypothetico-deductivism accounted for the justified confirmation of Goldbach's conjecture by its known successful predictions, hypothetico-deductivism cannot suffice to account for the justified confirmation of the conjecture's holding in unexamined cases by its known successful predictions. (Recall note 4.) Confirmation of a conjecture's holding in unexamined cases depends on justified prior opinions over and above the logical and statistical relations between the evidence and what is being confirmed.

Here is another illustration of the diversity of the justified prior opinions that sometimes underwrite the confirmation of mathematical conjectures. As Franklin (2013: p. 13) emphasizes, justified beliefs regarding an expert mathematician's track record of making conjectures that turned out to be true can justly confirm (to some degree) the mathematician's next conjecture. In view of the diversity of the justified prior opinions that underwrite justified confirmation in mathematics, I am not maintaining that IBE (where confirmation is underwritten by justified prior opinions regarding explanatory quality per se) operates in *every* case where mathematical conjectures are justly confirmed (without being proved). But (I maintain) IBE operates in mathematics much as it does in science in constituting an important kind of plausibility argument.

3 IBE in math

I maintain that IBE operates in the justified confirmation of some mathematical conjectures much as it does in the justified confirmation of some scientific hypotheses.[9] Let's look at a workaday mathematical example. Take an ordinary calculator keypad (without the zero key):

7	8	9
4	5	6
1	2	3

We can form a six-digit number by taking the three digits on any row, column, or main diagonal on the keypad in one direction and then in the opposite direction. 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,…). Suppose we took the first two calculator numbers in my list and verified (using a calculator!) that both of them are divisible by 37. Should this evidence raise your confidence, of the third calculator number, that it is divisible by 37?

That depends on whether you justly believe that if every calculator number is in fact divisible by 37, then this pattern is no coincidence. That is, whether you should take the discovery that the first two calculator numbers are divisible by 37 as confirming that this pattern will continue depends on what you justly believe the explanation of this pattern would be, if it did continue—just as whether you should take the discovery that the Smith and Brown families on my block each have two children as confirming that this pattern will continue typically depends on whether or not you justly believe that this pattern, if it did continue, would be coincidental.

For it to be *no* coincidence that all of the families on my block have two children would be for there to exist a *common cause* operating in the case of each family (perhaps a chemical in the soil!) that explains why each of them has two children. What would it be for all of the calculator numbers' divisibility by 37 to be no coincidence? There could be a *common reason* operating in the case of each calculator number that explains why all of them are divisible by 37. The pattern is no coincidence exactly when there is such a common reason.

The initial discovery that every calculator number is divisible by 37 was made by checking each of the sixteen calculator numbers separately. Although this proof was utterly convincing, it failed to reveal whether or not this pattern is a coincidence. That is why when this discovery was published (as an unsigned "gleaning" in the Decem-

[9] Of course, IBE is used in certain "indispensability arguments" for mathematical realism (Colyvan, 2012: pp. 41–51, pp. 90–95); these philosophical arguments aim to use IBE to argue for the reality of platonistic mathematical entities from their alleged role in certain explanations in *science*, in the same way as IBE is allegedly used by scientists to argue for the reality of various unobservable physical entities. My concern, by contrast, is with IBE as applied to certain explanations in *mathematics*. The arguments that I am interpreting as operating by IBE are given by mathematicians, not by philosophers, and purport to support various ordinary mathematical claims, not any particular (e.g., platonistic) philosophical interpretation of mathematics.

ber 1986 issue of *The Mathematical Gazette*), the piece was entitled with the question "A Calculator Coincidence?" asking why this pattern holds.[10] This question was soon answered by Nummela (1987) in a response simply entitled "No Coincidence," where he gave the following proof that all calculator numbers are divisible by 37:

> 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 equal to $a(10^5 + 10^4 + 10^3 + 10^2 + 10 + 1) + d(10^4 + 2.10^3 + 2.10^2 + 10)$ $= 111111a + 12210d = 1221(91a + 10d) = (3 \text{ x } 11 \text{ x } 37)\ (91a + 10d)$.

This proof, unlike the case-by-case proof, *explains why* all of the calculator numbers are divisible by 37; as Nummela says, it reveals the result to be no coincidence. The explanation works by identifying a form that all calculator numbers share by virtue of being calculator numbers. This common form constitutes the common reason why all of them are divisible by 37.

My point is that when a mathematician ascertains that the first two calculator numbers are both divisible by 37, the mathematician should typically regard this discovery as confirming that the pattern will continue only insofar as the mathematician is justly confident that it would be no coincidence for the pattern to hold of all calculator numbers—that is, insofar as the mathematician is justly confident that if the pattern holds of all calculator numbers, then there exists a common reason for each calculator number's divisibility by 37. In this example of confirmation in mathematics, IBE operates much as it does in science: Explanatory considerations justly influence whether a conjecture's making some successful predictions should be taken as confirming the rest of its predictions.

There are many examples like the calculator-number case—that is, where mathematicians describe a case-by-case proof of some general theorem as failing to explain why the theorem holds; an explanatory proof would have to give a unified, uniform treatment of the various cases falling under the theorem. If the theorem has no such proof, then it is a coincidence. For example, Baker (2009: p. 148) characterizes the notoriously disunified proof of the four-color theorem as failing to explain why the theorem holds: "There are 1476 different sub-cases that are individually considered. Thus, the proof is very unexplanatory." The graph-theorist Paul Seymour (2016: p. 417) agrees, strongly suspecting that the four-color theorem is no coincidence—that it has a unified, uniform proof (as yet unknown) that explains why four colors suffice for any planar map. Seymour says that "the massive case analysis of the computer proof" fails to identify "the 'real' reason the [theorem] is true; what exactly is it about planarity that implies that four colours suffice?".

In examples where an explanatory proof of some mathematical claim would have to give a unified, uniform treatment of all of the cases within the claim's scope, it is

[10] The gleaning appeared on page 283 of volume 70, number 454. I am grateful to Roy Sorensen for calling my attention to this example in a lovely unpublished essay. This example of an explanation in mathematics is discussed further in Lange, (2017).

easy to imagine IBE operating when a mathematician properly judges whether the claim's holding in all of the already examined cases confirms that this pattern will continue. For instance, a mathematics student might have learned that notoriously, the prime numbers fall into no simple general patterns. Having learned (perhaps to her surprise) that the six numbers 31, 331, 3331, 33331, 333331, and 3333331 are all prime, the student is justly confident that even if the next number in this sequence turns out to be prime, that seven-number pattern is coincidental; there would be a case-by-case proof that all seven numbers are prime, but there would be no unified, uniform proof—that is, no proof proceeding from the form that they all share (in base 10). Therefore, the student (using IBE) justly fails to regard the primality of the first six numbers in this sequence as confirming the seventh's primality. (Although it is prime, the eighth member of the sequence is not; that the first seven are prime is commonly termed a "coincidence."[11]).

4 A new variety of IBE in math

We have now arrived at this paper's main subject: an important but neglected variety of non-deductive reasoning by which mathematical conjectures are sometimes confirmed by mathematical evidence. The sketch I have given so far of mathematical IBE is seriously incomplete because (as we will see in Sect. 5) not every explanatory proof in mathematics derives its explanatory power from giving a unified, uniform treatment of all of the cases in the scope of the theorem being proved (and thereby explained), making that theorem no coincidence. Instead, some prominent explanatory proofs derive their explanatory power by virtue of their exploiting a *symmetry* in the theorem's setup—in particular, a symmetry that is similar to a striking symmetry in the theorem being proved (and thereby explained). The remainder of this paper will argue that in such cases, IBE can enhance a mathematical discovery's confirmatory impact on some conjecture. I will elaborate how a conjecture can "perfect the symmetry" in some discovery. I will argue that if such a conjecture could be proved by making use of the same symmetry as the discovery (imperfectly) possesses, then that proof would nicely explain why that discovery holds. By IBE, the potential for such a good explanation strengthens the discovery's confirmation of the conjecture.

In the following sections, I will present the case for this form of IBE in mathematics. In Sect. 5, I will argue that some explanatory proofs derive their explanatory power by taking the symmetry that is striking in the theorem being explained and exploiting the same symmetry in the theorem's setup. In Sect. 6, I will bring this kind of explanatory proof together with IBE. I will argue that in mathematical practice, a known theorem sometimes confirms a conjecture that entails it partly because that

[11] In, e.g., Leavitt, (2007: p. 182). (This example is also mentioned by Sorensen in the essay to which I referred in the previous note.) I will have no need here to spell out what it would take for a proof to give a unified, uniform treatment of the various cases falling under the theorem being proved. To do so would require an account of *natural* mathematical properties and kinds (as Baker (2009: pp. 144) puts it) so that an arbitrary pair of proofs cobbled together does not count as a unified uniform proof of the conjunction of the two theorems proved. All of these topics (mathematical coincidence, natural mathematical properties and kinds, explanation in mathematics, unification,…) are discussed in greater detail in Lange (2017).

conjecture perfects the theorem's striking near-symmetry. If there were a proof of the conjecture that exploited the same symmetry as the conjecture strikingly possesses (and as the theorem nearly possesses), then that proof would be a very good explanation of the theorem. The conjecture's truth would thus help the theorem to have a nice explanation. The existence of such an explanatory proof and the truth of the conjecture that it would prove are justly confirmed by the known theorem partly because of how nicely the theorem would be explained by such a proof. This confirmation operates by IBE; a potential explanation's quality (what Lipton (2004) calls its "loveliness") is contributing to its likeliness.[12] (The existence of such an unknown proof could thus be justly confirmed before such a proof has been discovered. Mathematicians could in this way be justified in putting in the considerable effort required to find such a proof, as I mentioned at the start of Sect. 1.)

In short, the known theorem justly confirms (by IBE) that the conjecture can be proved in this way. The theorem thereby justly confirms the conjecture itself much more strongly than the conjecture would have been justly confirmed merely by virtue of its entailing the known theorem.

In Sects. 7–10, I will support this argument by presenting a historical example that plausibly involved this sort of IBE. Many mathematicians have maintained that Brahmagupta and Snell each used this sort of argument (which I understand as a form of IBE) to underwrite their confidence in a conjectured generalization of Heron's formula for the area of triangles. Mathematicians now widely regard Heron's formula as exhibiting a striking near-symmetry that is perfected by "Brahmagupta's formula" (as it is now generally called). Before it had been proved, Brahmagupta's formula derived additional support (that is, support beyond what it received merely by entailing Heron's formula) from its perfecting the symmetry exhibited by Heron's formula. This additional confirmation through IBE justly helped to make Brahmagupta's formula sufficiently plausible to be worthy of the considerable effort that Brahmagupta and Snell had to expend in trying to prove it.

5 Explanations in math that exploit symmetry

If a mathematical result exhibits a striking symmetry, then the difference between proofs of it that explain why it holds and proofs that are not explanatory is that all and only the explanatory proofs exploit the same kind of symmetry in the result's "given" (its "setup") as the result itself exhibits. Let's look at one example.[13] 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 come in pairs of the form $a + bi$ and $a - bi$. Here is the general theorem (proved by d'Alembert in 1746):

[12] At the end of Sect. 2, I sketched how I would account for loveliness's contribution to likeliness. But I am not depending on this account of IBE here.

[13] Lange (2017) has argued for this at greater length and given many examples, including this one.

> 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 that? Why do the solutions come in complex-conjugate pairs?

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

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

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

Thus, $\overline{z^n} + a_{n-1}\,\overline{z^{n-1}} + \ldots + a_0 = \overline{z^n + a_{n-1}z^{n-1} + \cdots + a_0}$, 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 is merely the application of unrelenting algebraic force, doggedly plugging everything in and grinding out the result. It entirely disregards the outstanding feature of d'Alembert's theorem: the symmetry that it reveals between *i* and *–i*. The above proof depicts this symmetry as just falling out of the algebra. This is no explanation. In asking for an explanation, mathematicians are seeking a proof of the theorem from some similar symmetry in the original problem—that is, a proof that somehow exploits the setup's invariance 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 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 unequal. They are distinct, but they are no different in their relations to the real numbers. Whatever the axioms of complex arithmetic say about one of them is also true about the other. Since the axioms remain true under the replacement of *i* with *–i*, so must any 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 accounted for by the same symmetry in the axioms of complex arithmetic.[14]

Notice that these two proofs of d'Alembert's theorem are alike in the degree to which they unify all of the theorem's cases (such as the solutions of of $z^3 + 6z - 20 = 0$ and the solutions of $z^2 - 2z + 2 = 0$ we saw earlier). Both proofs are unified, uniform proofs; neither treats some of the theorem's cases separately from others. Yet mathematicians regard the proofs as differing in their explanatory power. That is because what makes a proof of d'Alembert's theorem explanatory is not its unification, but rather its exploiting a symmetry in the problem—a symmetry of the same kind (in particular, invariance under the exchange of i with $-i$) as the salient symmetry in the result being explained.

We must now investigate how IBE operates in mathematics where symmetry, not unification, is what differentiates explanatory from non-explanatory proofs.

6 IBE where a conjecture is confirmed by its perfecting a theorem's symmetry

In Sect. 3, I argued that mathematicians sometimes justly confirm a conjecture through IBE when they ascertain various cases exhibiting a striking pattern that the conjecture generalizes so that a proof *explaining why* this pattern holds would have to treat all of its cases in a unified, uniform way. Mathematicians can thereby justly regard the fact that the observed cases fit the pattern as confirming that the other cases will also turn out to fit the pattern. Crucially, mathematicians can justly take this confirmation as stronger insofar as mathematicians are justly confident that there is some (as yet unknown) explanation for the observed cases' fitting the pattern—that the observed cases fit the pattern for some common reason, not coincidentally. The observed cases' fitting the pattern justly confirms more strongly that the other cases also fit the pattern because of the other cases' potential involvement in the explanation of the observed cases, i.e., because the conjecture may be proved by a unified, uniform proof of the observed cases' fitting the pattern. Without these explanatory considerations bringing them together, the observed cases' fitting the pattern and the conjecture that the other cases will also turn out to fit the pattern would typically be an arbitrary conjunction (like Goldbach's conjecture conjoined with the Twin Prime conjecture) so that mathematicians could not justly take one as confirming the other.

As an example, I argued that when a mathematician ascertains that the first two "calculator numbers" are both divisible by 37, typically the mathematician should regard this discovery as confirming that the pattern will continue only insofar as the mathematician is justly confident that it would be no coincidence for the pattern to hold of all calculator numbers—that is, insofar as the mathematician is justly confident that if the pattern turns out to hold of all calculator numbers, then this theorem has an (as yet unknown) explanation (i.e., a unified, uniform proof). The discovery that the first two calculator numbers are both divisible by 37 is made to bear more

[14] For a pungent characterization of this argument as an explanation, see Feynman et al., (1963, pp. 22–7). The mathematician Eugenia Cheng (2004: p. 12) also characterizes it as explanatory.

strongly on the conjecture that the other calculator numbers are divisible by 37 by the conjecture's potential involvement in the discovery's explanation. That involvement arises from the conjecture's potential to have a unified, uniform proof that also covers the first two calculator numbers. The conjecture is justly deemed by mathematicians to be likely (if it is true) to be proved by a proof explaining the discovery that the first two calculator numbers are both divisible by 37 because the same unified, uniform treatment given to the first two calculator numbers can presumably (if the conjecture is true) be extended to all of the other calculator numbers.

Let's now maintain a strict analogy with the previous two paragraphs while changing the salient feature of the theorem being explained. That feature in the previous paragraphs was the theorem's revealing a striking similarity among various cases—that is, revealing a striking pattern running through them and thereby unifying them. Now let's see what happens in an example where the salient feature of the theorem being explained is instead the theorem's exhibiting a striking symmetry (as we saw in Sect. 5).

A mathematician can justly confirm a conjecture through IBE when the mathematician ascertains a theorem exhibiting a striking near-symmetry that the conjecture perfects so that a proof *explaining why* this theorem holds would have to exploit the same kind of symmetry in the setup. A mathematician can then justly regard the theorem as confirming the conjecture. Crucially, a mathematician can justly take this confirmation as stronger insofar as the mathematician is justly confident that the theorem has an (as yet unknown) explanation. The discovered theorem justly confirms the conjecture more strongly because of the conjecture's potential involvement in the theorem's explanation. That involvement arises from the conjecture's potential to be proved by a proof of the theorem that works by exploiting a symmetry of the same kind as the theorem outstandingly exhibits nearly. The conjecture is justly deemed by mathematicians to be likely (if it is true) to be proved by a proof explaining the theorem because the conjecture "perfects the symmetry" that the theorem strikingly exhibits nearly.[15]

In Sects. 7–10, we will examine a historical example where mathematicians plausibly used IBE reasoning of this kind. Heron's theorem was powerful evidence for a conjecture (Brahmagupta's formula) because that conjecture perfects the near-symmetry of Heron's theorem. To grasp this IBE, we will first need to identify the theorem's near-symmetry, to appreciate what makes it imperfect, and to understand what a proof of Heron's theorem would have to be like in order for it to nicely explain why that theorem holds. That will be the agenda of Sects. 7–9. Section 10 will then present the IBE argument by which Heron's theorem strongly confirmed Brahmagupta's formula.

[15] On the view that I sketched at the end of Sect. 2, this confirmation does not reflect a merely aesthetic preference for symmetric mathematical results. Rather, this confirmation would be justified by mathematicians' prior discoveries of proofs explaining various symmetric mathematical results.

7 Heron's theorem

The most familiar formula for the area Δ of a triangle is $\Delta = \frac{1}{2}bh$, where b is the (length of the) triangle's base and h is its height. A formula "that rarely finds a place in elementary courses" (Simmons, 1992: p. 222), generally termed "Heron's theorem" (because Heron's proof is the earliest that survives, though perhaps the theorem was first discovered by Archimedes), expresses Δ in terms of the triangle's three sides a, b, and c:

$$\Delta = \frac{1}{4}\sqrt{(a+b+c)(-a+b+c)(a-b+c)(a+b-c)}. \tag{1}$$

The other common expression for Heron's formula uses the triangle's semiperimeter $s = \frac{a+b+c}{2}$:

$$\Delta = \sqrt{s(s-a)(s-b)(s-c)}. \tag{2}$$

These expressions exhibit two striking symmetries.[16] First, these expressions treat every side in the same way; these expressions are invariant under systematic interchange of a, b, and c. But that a, b, and c all play the same role in (1) and likewise in (2) is not the only symmetry in Heron's formula. After all, another, rarely seen expression of Heron's theorem

$$\Delta = \frac{1}{4}\sqrt{2(a^2b^2 + a^2c^2 + b^2c^2) - (a^4 + b^4 + c^4)} \tag{3}$$

exhibits the symmetry over a, b, and c but obviously is not as symmetric as (1) and (2). Expression (3) loses the second striking symmetry exhibited by (1) and (2): that the various terms in a given expression all take the same form. (Well, this is nearly true: *three* of the four terms do in (1)—and the fourth term almost does. The same holds for the four terms in (2). The symmetry is "imperfect". I will shortly return to this key point.) In (1), three of the four terms are strikingly alike in adding two of the sides and subtracting the third. The three terms represent all three of these possibilities. In (2), three of the four terms are strikingly alike in subtracting one side from s. Again, the three terms represent all three of these possibilities. This striking symmetry over the terms is lost in (3), which is why mathematicians rarely express Heron's formula in this way. That mathematicians are keen to expose the symmetry among the terms, not just the symmetry over the sides, explains why mathematicians use s to express Heron's theorem. Although s is superfluous once a, b, and c have been used, s helps to make the terms more obviously (nearly) symmetrical.

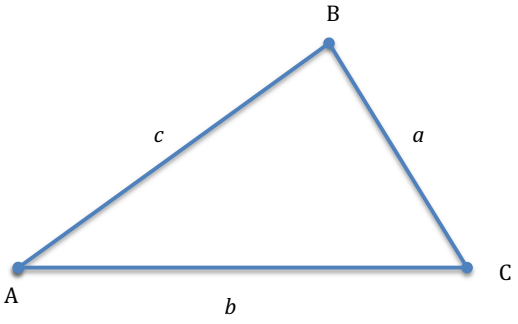
[16] Mathematicians often emphasize the symmetries of Heron's theorem. See (among many examples) Pratt (2009: p. 16) on "the pleasing symmetry of Heron's Formula" and Klain (2004: p. 709) on its "symmetrical and intuitive factorization" making "Heron's formula more memorable." Notably, Conway (1999) wrote that he "would absolutely LOVE to get a 'symmetrical' proof of Heron's formula."

In Sect. 5, I argued that when a mathematical theorem is strikingly symmetric, the difference between proofs that explain why it holds and proofs that do not explain it is that explanatory proofs exploit the same kind of symmetry in the theorem's setup as the theorem strikingly displays. Actually, the distinction in mathematical practice between explanatory and non-explanatory proofs is not sharp; a proof can be explanatory *to some degree*. For a theorem that is strikingly symmetric, a proof is explanatory *insofar* as it exploits the same symmetry in the setup.[17] As we will see, this account fits various proofs of Heron's theorem. A proof may treat the triangle's sides differently by privileging certain sides over others—for instance, by drawing auxiliary lines to one of the sides but no analogous lines to the others, or by treating the longest side (or, at least, a side where none is longer) differently from the others. Insofar as a proof does so, then mathematicians regard the proof to that extent as not explanatory. A non-explanatory proof depicts the formula's symmetry as emerging "magically" out of a series of apparently fortuitous accidents in the course of the proof. By contrast, if a proof treats each of the sides in the same way throughout, then it maintains the setup's symmetry among the sides, inevitably yielding a symmetrical formula. Such a proof is characterized in mathematical practice as more explanatory—though not fully explanatory as long as the theorem's second symmetry (over the terms) is not likewise respected by the proof.[18] (As we saw, that second symmetry is imperfect; I will shortly return to this.)

8 Heron's proof

There are many proofs of Heron's theorem. (For a survey, see Dawson 2015: pp. 187–91.) I will contrast two proofs. They begin in the same way. Draw triangle ABC (Fig. 1) with side a (b,c) opposite angle A (B,C). Draw the three angle bisectors (Fig. 2). A point on angle A's bisector must be equidistant between lines AB and AC, and so the point O (the "incenter") where A's bisector and B's bisector intersect must

Fig. 1 Triangle ABC



[17] Lange (2017: pp. 297–8, 311, 448, 454) recognizes that a proof's explanatoriness is a matter of degree.

[18] Lange (2017: pp. 268–70) gives another example where mathematicians are struck by two salient features of some result (the value of a definite integral) and so to explain this result best, a proof must exploit both sorts of features in the setup.

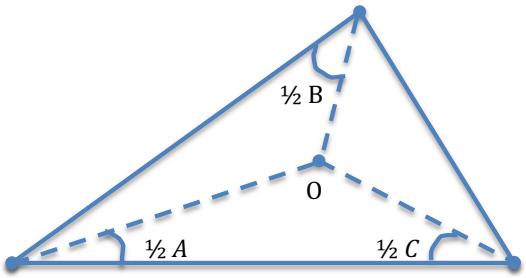


Fig. 2 Angle bisectors decompose triangle ABC into three triangles

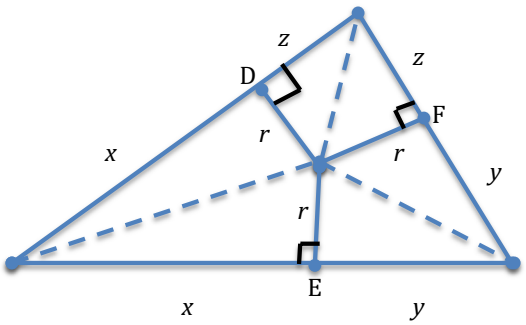


Fig. 3 Triangle ABC decomposed into six triangles

be equidistant from all three sides (and so lie on C's bisector). Draw lines of equal length r from O to each of the sides (Fig. 3). Those lines form right angles with their respective sides (since the shortest segment to a line from a point off that line is perpendicular to that line). Triangle ABC is thus decomposed into six triangles.

Triangles ADO and AEO both have right angles, have congruent angles at A (by virtue of the angle bisector), and have congruent non-included sides (which could be their shared hypotenuse or their sides of length r). So (by the AAS rule), the triangles are congruent. The same argument applies to each of the other two pairs of triangles (BDO and BFO, CEO and CFO). So the six triangles into which ABC has been decomposed form three pairs of congruent triangles. So far, all three vertices A, B, and C have received the same treatment; the proof has been symmetrical in the angles (or, equivalently, in the sides).

Furthermore, triangle ABC's area Δ is the sum of the areas of the three triangles drawn symmetrically over the sides in Fig. 2. Each of those areas is given by ½ (base) (height). So symmetrically over the sides (though no symmetry over the terms in Heron's theorem has been exploited):

$$\tfrac{1}{2}ar + \tfrac{1}{2}br + \tfrac{1}{2}cr = \tfrac{1}{2}(a+b+c)r = sr.$$

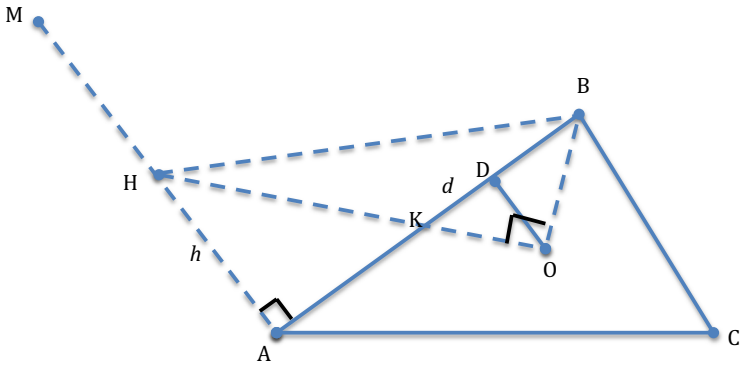


Fig. 4 Heron's proof begins

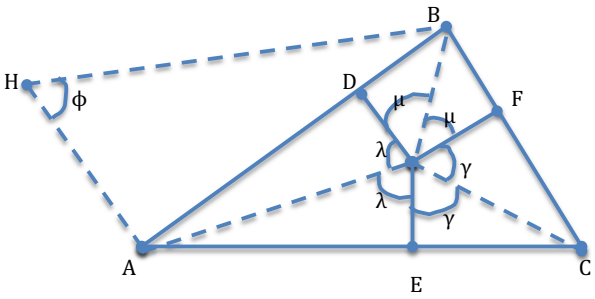


Fig. 5 Heron's proof concludes

Since the triangles in each pair are congruent, let x = AD = AE, y = CE = CF, z = BD = BF (Fig. 3). Therefore, semi-perimeter $s = \frac{a+b+c}{2} = \frac{(y+z)+(x+y)+(x+z)}{2} = \frac{2(x+y+z)}{2} = x + y + z$.

From this point, the two proofs diverge in how they seek r. Heron's proof (see Fig. 4, following Oliver, 1993: pp. 161–2) constructs AM perpendicular to AB, OH perpendicular to OB, and BH. Label point K where OH intersects AB. Let h be the length of AH and d be the length of KD. Angles BOH and BAH are right angles spanned by BH. So B, O, A, and H lie on a semicircle of a circle of diameter BH. Angle AHB (measure ϕ – see Fig. 5) subtends arc AOB, and angle AOB (measure $\lambda + \mu$) subtends the rest of the circle. Therefore, the two angles together subtend the entire circle, so $\phi + \lambda + \mu = \pi$. The six angles around O (Fig. 5) added make $2\lambda + 2\mu + 2\gamma = 2\pi$. Hence, $\gamma = \phi$. Thus, triangles BAH and CFO have two pairs of congruent angles and so are similar. Ratios of their corresponding sides are therefore equal, where AH corresponds to FO and BA corresponds to CF: $\frac{h}{r} = \frac{x+z}{y}$ and so $r = \frac{hy}{x+z}$.

Heron then looks for additional pairs of similar triangles. Triangles HAK and ODK have vertical angles K congruent and right angles D and A congruent, so they are similar. AH (length h) corresponds to DO (length r) and AK (length $x - d$) corresponds to DK (length d): $\frac{h}{r} = \frac{x-d}{d}$, so $hd = rx - rd$, and so $d(h+r) = rx$, and hence $d = rx/(h + r) = \frac{x}{\frac{h}{r}+1}$.

Triangles BOD and OKD have right angles D congruent, and angles OBD and KOD are both complementary to angle DOB and so are congruent. Therefore, these triangles are similar. KD (length d) corresponds to OD (length r) and DO (length r) corresponds to DB (length z): $\frac{d}{r} = \frac{r}{z}$, and so $r^2 = dz$. So now Heron has three pairs of similar triangles. As Dunham (1990: p. 125) says: "At this point, Heron's readers may be forgiven for suspecting that the mathematician was adrift, wandering aimlessly through an unending series of similar triangles." (No symmetry has been preserved in these steps!) But now he puts together all of these results of similar triangles. Substituting the previous paragraph's expression for d into the above equation for r^2 yields $r^2 = \frac{xz}{\frac{h}{r}+1}$. Using $r = \frac{hy}{x+z}$ from the first pair of similar triangles, we have

$$r^2 = \frac{xz}{\frac{h}{(\frac{hy}{x+z})} + 1} = \frac{xz}{\frac{h(x+z)}{hy} + 1} = \frac{xz}{\frac{x+z}{y} + 1} = \frac{xzy}{x+z+y}$$

An expression symmetric in x, y, and z has emerged miraculously from out of a welter of similar triangles and tedious algebra. Since $x + y + z = s$, we can simplify further to

$$r^2 = \frac{xyz}{s}.$$

From here, the proof is symmetric in the sides: Since $s = x + y + z$ and $a = y + z$, we have $s - a = x$ and symmetrically $y = s - c$ and $z = s - b$. So $r^2 = \frac{(s-a)(s-b)(s-c)}{s}$. Since we found earlier that $\Delta = sr$, we have $\Delta = s\sqrt{\frac{(s-a)(s-b)(s-c)}{s}} = \sqrt{s(s-a)(s-b)(s-c)}$, which is Heron's formula.

Mathematicians commonly remark on the surprise ending of Heron's proof. For instance, Dunham (1990: p. 119) writes that Heron's proof is "at once extremely circuitous, extremely surprising, and extremely ingenious. … [It] boasts one of the best surprise endings in mathematics. As with a good Agatha Christie novel, readers of Heron's proof can be within a few lines of the end and still have no idea how the matter will be resolved." At least part of what makes the windup of Heron's proof so surprising is the contrast between the final result's symmetry and the messiness of the algebraic expressions preceding it. The result's symmetry does not seem inevitable but rather seems to come out of nowhere; the final result would have been much less surprising had the proof somehow managed to maintain any of the setup's symmetries.

Likewise, as I mentioned earlier, mathematicians frequently note the lack of any motivation apparent to the reader for the auxiliary lines that Heron draws: "The pur-

pose of the auxiliary lines in Heron's diagram is not at all apparent at the outset" (Dawson 2015: p. 188). Mathematicians commonly make similar remarks about various other proofs of Heron's theorem. For example, regarding Euler's proof, Dunham (1999: p. 142) comments that the auxiliary lines appear far from inevitable, asking rhetorically: "How…did Euler know to construct ΔNOV in his proof of Heron's formula? How did he know which similar triangles would prove fruitful in leading him to the desired conclusion?" The auxiliary constructions would not have seemed arbitrary had they given equal treatment to all of the parts playing the same role in the setup, maintaining the setup's striking symmetries rather than breaking them.[19]

The sudden symmetry of the outcome, by contrast with the asymmetry of the steps preceding it, serves merely to increase our puzzlement about why everything ends up boiling down so quickly, turning out so nicely in the end. Why does Heron's formula hold? Mathematicians sometimes explicitly note that Heron's proof heightens this why-question rather than answering it—and that Heron's proof is non-explanatory for the reason that I identified earlier: because Heron's proof fails to preserve and exploit the symmetries in the setup that match the formula's striking symmetries. For instance, Oliver (1993: p. 162) explicitly links the proof's failure to provide understanding of Heron's result to the proof's failure to exploit the symmetries: the proof "exhibits none of the symmetry of the problem or of the results. … [W]e seem to gain little or no true understanding. In particular, why must our chain of logic take us way outside the triangle when all the pertinent geometry is contained therein?" Those external constructions break the setup's symmetry. Dawson (2015: p. 189) says that he agrees with Oliver that the proof "serves to *verify* rather than *explain*" the "result."

9 A more explanatory proof

Mathematicians cite another proof (Ballentine, 1954, Dobbs, 1984, Oliver, 1993: pp. 162–3, Nelson, 2001) as more explanatory than Heron's. This proof also begins with the incenter (a symmetrical construction) but then finds r in a different way.

Since the tangent of an angle in a right triangle is the ratio of the sides opposite and adjacent to the angle, in triangle AEO (recalling that $s = x + y + z$) $\tan\frac{1}{2}A = \frac{r}{x} = \frac{r}{s-(y+z)} = \frac{r}{s-a}$. Proceeding symmetrically for B and for C as for A, we arrive at symmetric results:

$$\tan\frac{1}{2}B = \frac{r}{s-b},$$

$$\tan\frac{1}{2}C = \frac{r}{s-c}.$$

[19] Lange (2017: pp. 247-50) discusses a similar example: a result concerning triangles (Menelaus's theorem) that displays a striking symmetry over the triangle's three vertices. Mathematicians regard a proof that draws auxiliary lines breaking that symmetry as failing to explain why the theorem holds. By contrast, a proof that mathematicians regard as explanatory begins with a feature of the setup in which the three vertices play the same role and then treats them identically in each subsequent step (e.g., by drawing the same sorts of auxiliary lines to each vertex).

Since the interior angles of any triangle add to 180°,

$$\tfrac{1}{2}A + \tfrac{1}{2}B + \tfrac{1}{2}C = 90^\circ.$$

Here again we preserve the symmetry among the three angles: the sum of any two is complementary to the third. Let's exploit this symmetry in the setup to arrive at another result that is symmetric across all three angles. Move any one of the three angles to the equation's right side—it makes no difference which since the same (symmetric) formula will result momentarily. I'll move ½ *C*:

$$\tfrac{1}{2}A + \tfrac{1}{2}B = 90^\circ - \tfrac{1}{2}C$$

So

$$\tan\left(\tfrac{1}{2}A + \tfrac{1}{2}B\right) = \tan\left(90^\circ - \tfrac{1}{2}C\right)$$

Using the formula for the tangent of the sum of two angles (on the left side) and the formula $\tan(90^\circ - x) = \cot x$ (on the right side),

$$\frac{\tan\frac{1}{2}A + \tan\frac{1}{2}B}{1 - \tan\frac{1}{2}A\tan\frac{1}{2}B} = \frac{1}{\tan\frac{1}{2}C}$$

Cross-multiplying

$$\tan\frac{1}{2}A\ \tan\frac{1}{2}C + \tan\frac{1}{2}B\ \tan\frac{1}{2}C = 1 - \tan\frac{1}{2}A\ \tan\frac{1}{2}B$$

we arrive at what Oliver (1993:p. 162) calls "the symmetrical key equation"

$$\tan\frac{1}{2}A\ \tan\frac{1}{2}C + \tan\frac{1}{2}B\ \tan\frac{1}{2}C + \tan\frac{1}{2}A\ \tan\frac{1}{2}B = 1.$$

Inserting the values of these tangents found earlier, we now have a symmetrical expression over the three sides and over three of the four terms involving *s* in Heron's formula:

$$\frac{r^2}{(s-a)(s-c)} + \frac{r^2}{(s-b)(s-c)} + \frac{r^2}{(s-a)(s-b)} = 1.$$

Multiplying both sides by the symmetrical $(s-a)(s-b)(s-c)$

$$r^2(s-b+s-a+s-c) = (s-a)(s-b)(s-c).$$

Since $(s-b+s-a+s-c) = 3s-(a+b+c) = 3s-2s = s$,

$$r^2 s = (s-a)(s-b)(s-c)$$

$$r = \sqrt{\frac{(s-a)(s-b)(s-c)}{s}}$$

The proof concludes as Heron's proof did: $\Delta = sr = s\sqrt{\frac{(s-a)(s-b)(s-c)}{s}} = \sqrt{s(s-a)(s-b)(s-c)}$.

This proof avoids the features that mathematicians regard as depriving Heron's proof of explanatory power. In particular, Oliver (1993: p. 162) takes pains to point out that this proof consists in "a chain of reasoning…that continually exhibits the symmetry of the problem." Strogatz (2015) likewise praises the proof as "symmetric".

10 Perfecting the symmetry and Brahmagupta's formula

However, even in the above proof of Heron's theorem, one of the theorem's terms (s) still looks different from the other three ($(s-a), (s-b)$, and $(s-c)$), breaking what is otherwise a striking symmetry. Furthermore, while the three similar terms are all derived in the same way, the s term arises differently: the three similar terms come from a symmetrical proof of an expression for r, whereas the s term comes from the s in $\Delta = sr$. Because this striking symmetry in Heron's theorem remains incomplete and the above proof still treats the s term differently from the rest, the above proof is not fully explanatory.[20] A fully explanatory proof would be uniform over all four terms and replace the near-symmetry in Heron's formula with a complete symmetry. It would perfect the formula's symmetry.[21]

How could a derivation fully explain Heron's theorem by perfecting the symmetry in its terms? Heron's theorem would have to be proved as a special case of a more general result that not only includes Heron's three similar terms (($s-a$), ($s-b$), and ($s-c$)) but also amends Heron's fourth term (s) to be like these three, making the generalization fully symmetric over these four terms. An explanation of this result would then be a proof that treats all four terms alike, taking the striking symmetry in the result and exploiting the same sort of symmetry in the setup.

Now we come to IBE. Already any live conjecture having Heron's theorem as a special case would in mathematical practice receive some support from entailing

[20] Recall (note 16) that Conway said that he was still looking for a fully symmetric proof of Heron's theorem.

[21] The proof in the previous section is not the only derivation of Heron's formula that treats the three similar terms alike but differently from the fourth. Many mathematicians (e.g., Hersh, 2004; Klain, 2004) have noted that any expression for Δ that is a polynomial in the sides would have to have $s-a = \frac{a+b+c}{2} - a = \frac{a+b+c-2a}{2} = \frac{b+c-a}{2}$ as a root because (by the triangle inequality) the sum of any two sides of a triangle must exceed the third side, so the area is zero in the degenerate case where the third side is equal to the sum of the other two (e.g., where $\frac{b+c-a}{2} = 0$). The same reasoning applies to ($s-b$) and ($s-c$), so any such polynomial must have factors of ($s-a$), ($s-b$), and ($s-c$). But the fourth factor requires an entirely different motivation.

Heron's theorem. This is what Pólya (as we saw in Sect. 1) calls "the fundamental inductive pattern" that "theories are confirmed by their [non-trivial] consequences or predictions." But such a live conjecture would in mathematical practice justly receive additional support (in the absence of other considerations) if it perfected the symmetry of Heron's theorem. This is a form of IBE. In other words, I contend that in mathematical practice, a conjecture having Heron's theorem as a special case would justly receive additional support insofar as it is symmetric over its four similar terms (three of which are Heron's terms $(s-a)$, $(s-b)$, and $(s-c)$) and insofar as mathematicians are justly confident that it would (if true) have a proof (as yet unknown) that gives uniform treatment to the four terms and thereby not only proves Heron's theorem, but also *explains why* Heron's theorem holds.[22] This is a form of IBE since the conjecture derives additional support (over and above the support it receives from entailing Heron's theorem) by virtue of explanatory considerations: Heron's theorem confirms such a conjecture more strongly by virtue of the conjecture's potential involvement in the theorem's explanation. The conjecture's potential involvement arises from the conjecture's not only generalizing Heron's theorem, but also perfecting the symmetry that Heron's theorem strikingly nearly exhibits. That is, the conjecture has a symmetry of the same kind as Heron's theorem nearly exhibits, and so it has the potential (if it is true) to have a proof that exploits this symmetry and thereby explains Heron's theorem.

A conjecture that (mathematicians say) might reasonably have been confirmed in this way (and that historians say may well actually have been so confirmed) is Brahmagupta's formula for the area $\boxdot$ of a cyclic quadrilateral (that is, a quadrilateral whose four vertices lie on a circle) in terms of the sides' lengths (a, b, c, and d):

$$\boxdot = \sqrt{(s-a)(s-b)(s-c)(s-d)}\,.$$

Compare it to Heron's formula

$$\Delta = \sqrt{s(s-a)(s-b)(s-c)}$$

where (in both equations) s is half of the perimeter of the figure (quadrilateral and triangle, respectively). Regarding Brahmagupta's formula, mathematicians widely remark its striking symmetries both over the sides and over the terms—three of which are Heron's terms [(s—a), (s—b), and (s—c)], while the fourth takes the same form as those three but involves the quadrilateral's fourth side. The same symmetries are exhibited in the other common expression for Brahmagupta's formula:

$$\boxdot = \frac{1}{4}\sqrt{(a+b+c-d)(a+b-c+d)(a-b+c+d)(-a+b+c+d)}.$$

[22] On my view (sketched at the end of Sect. 2), mathematicians' confidence that the symmetric conjecture would have an (as yet unknown) explanatory proof is justified by mathematicians' experience with the explanations that they have discovered other mathematical results to have (in particular, by the extent to which mathematicians have found that symmetric results tend to have some proofs that operate symmetrically). Mathematicians' IBE is not the product of some merely aesthetic preference on their part for symmetric conjectures.

Again compare it to Heron's formula:

$$\Delta = \frac{1}{4}\sqrt{(a+b+c)(-a+b+c)(a-b+c)(a+b-c)}.$$

Three of the four terms in Heron's formula are alike in adding two of the sides and subtracting the third (the three terms representing all three of these possibilities) but the fourth term is different. In Brahmagupta's formula, all four terms are alike in adding three of the sides and subtracting the fourth (the four terms representing all four of these possibilities).[23]

Heron's formula is entailed by Brahmagupta's; Heron's is the special, degenerate case of Brahmagupta's where $d = 0$. A triangle is a cyclic quadrilateral with one side of length zero. (Any triangle's vertices all lie on a single circle and so any triangle is automatically cyclic.) It is as if that fourth side was in Heron's formula all along, but simply invisible because it was zero.

Heron flourished sometime in the first three centuries CE (scholars disagree about precisely when); Brahmagupta was born in 598 CE (Pottage, 1974: p. 306). There is no good evidence regarding whether Heron's formula was communicated to India from a Greek source before Brahmagupta's time, or whether it was rediscovered independently by Brahmagupta's Indian predecessors, or whether Brahmagupta himself rediscovered it (Pottage, 1974: pp. 341–3; Atzema, 2015: p. 22). Nevertheless, historians generally (though not unanimously) believe that Brahmagupta arrived at his formula by generalizing Heron's formula. For instance, Pottage (1974: p. 342) writes: "[I]t is unlikely that any would wish to suppose that he reached the rule for quadrilateral areas first, and only then saw that triangles could be treated as an extreme case of quadrilaterals in which one of the sides happened to have zero length. … It is hardly to be doubted that Brahmagupta (or any other mathematician for that matter) would have ascended to the quadrilateral only *after* having found the triangle rule." Likewise, Atzema (2015:p. 22) says that despite the absence of evidence regarding the transmission of Heron's formula to India, "It is tempting to assume that Brahmagupta might have conjectured [his] formula as an obvious generalization of Heron's." The reason that this reconstruction is so plausible, despite the absence of documentation for it, is that Heron's formula gives Brahmagupta's significant support.[24]

My point is not the relatively uncontroversial idea that Brahmagupta's formula entails Heron's and so Brahmagupta's formula (before Brahmagupta proved it)

[23] In the context of Heron's and Brahmagupta's formulas, Hersh (2004: p. 114) also gives a formula for a tetrahedron's volume that is likewise striking for being "symmetric in all six edges." He notes, furthermore, that his proof of this formula is not fully symmetric: "You (temporarily) give up symmetry in order to eventually get some back."

[24] How much support? Reasonable people might differ. But some historians and mathematicians have described the support as very strong indeed. Take Heilbron (1998:220): "While playing with the cyclic quadrilateral, the Indian mathematicians found a generalization of Hero's formula that fit it. The generalization is *exactly what you would expect*…" (my emphasis). Similarly, Simmons (1992:223) notes that the s term in Heron's theorem breaks the symmetry among its other terms and that this confirms a theorem that would perfect the symmetry: "The presence of the factor s under the radical in [Heron's formula] suggests that this formula for the area of the triangle … might be a special case of a more general formula $T = \sqrt{(s-a)(s-b)(s-c)(s-d)}$ for the area of a quadrilateral with sides a, b, c, d…"

received some degree of support from Heron's by Pólya's hypothetico-deductive pattern of reasoning (as Pólya (1954:II, p. 172) himself proposed). My point is that beyond that confirmation, Brahmagupta's formula received additional confirmation from *the way* that it generalizes Heron's: it perfects the striking near-symmetry of Heron's formula. Therefore, it could plausibly have a proof that *explains why* Heron's formula holds (by virtue of that proof's according uniform treatment to all four of the terms and thereby explaining why Brahmagupta's formula holds). This additional support would be by IBE.

Mathematicians themselves sometimes point to this as a weighty "plausibility argument" that Brahmagupta would have been entitled to make for his formula before he proved it, presuming that (by whatever means) he had Heron's theorem already in hand. For instance, after pointing out the "greater formal symmetry" achieved by tweaking Heron's formula into Brahmagupta's by replacing s with $s-d$, Brown (n.d.) writes: "[S]o it seems as if the [resulting] quantity *must* represent…something meaningful…and indeed it does." Then, after stating Brahmagupta's discovery, Brown continues: "It is tempting to think that Brahmagupta might have just imagined the equation based on its formal symmetry." Likewise, after giving the more explanatory proof of Heron's theorem that I gave in Sect. 9, Nelson (2001: pp. 291–2) concludes his paper with "a challenge" for his readers: to give a similar proof of Brahmagupta's theorem. Presumably, then, he regards the more symmetric proof of Heron's formula and the way that Brahmagupta's theorem generalizes Heron's (by perfecting its striking symmetry) as some evidence that such a fully symmetric proof of Brahmagupta's exists.

As I have said, it is unknown whether Brahmagupta actually inherited Heron's theorem and regarded it as providing evidence for his generalization of it. However, more important for my purposes than Brahmagupta's actual trajectory is that present-day mathematicians believe that for Brahmagupta to have taken Heron's formula as providing additional support for Brahmagupta's conjecture—because that conjecture does not merely entail Heron's formula, but also perfects its symmetry—would have been a good argument for Brahmagupta to have made.[25]

Historians have made similar points regarding Willebrord Snell's independent rediscovery of Brahmagupta's formula.[26] Heron's formula (but not Brahmagupta's) was definitely transmitted to Snell; as de Wreede (2007: p. 270) documents, Heron's

[25] For another example of a proof that explains a result by first "perfecting" a striking feature of that result and then tracing that result to a similar feature of the setup, see Lange (2017: pp. 267-8). It contrasts two proofs of the fact that the digit 7 appears 50,000 times when you list all of the whole numbers (in base 10) from 1 to 99,999. The result's striking feature is that 50,000 is *almost exactly* one-half of 99,999. Mathematicians regard as unexplanatory a plodding proof that simply lists all of those numbers and then tallies every appearance of 7 in the list. By contrast, a proof that mathematicians describe as explanatory begins by expanding the listed numbers so that it starts with 0 (an expansion that adds no 7's to the list). This expansion makes 50,000 *exactly* half of the number of whole numbers in the list, perfecting the result's striking feature. The explanatory proof then works by revealing where that ½ of 100,000 figures in the setup.

[26] Snell "was the first European who published this theorem" and believed that it originated with him (de Wreede, 2007: p. 286). Both Brahmagupta (Kichenassamy, 2009: p. 41) and Snell (de Wreede, 2007: p. 286) expressed the cyclic quadrilateral area formula in terms of the semiperimeter s, thereby bringing out its symmetries – though in words rather than in symbolic notation.

theorem "and its proof were a well-known geometrical topic by Snellius's time" and Snell was thoroughly familiar with it. As with Brahmagupta, historians and mathematicians find it plausible that Snell regarded Heron's formula as evidence confirming the formula for cyclic quadrilateral areas—not merely from its entailing Heron's formula as a special case, but also from its perfecting the symmetry of Heron's formula. As de Wreede (2007: pp. 285–6) writes, Snell "may have created the formula inductively from Heron's Theorem. … Probably, his reasoning was similar to Brahmagupta's. Snellius may have guessed the expression as an analogue to Heron's Theorem…" If so, Snell's reasoning was a good "plausibility argument" and (I have argued) a form of IBE.

11 Conclusion

I have just argued that an explanation of Brahmagupta's formula would explain Heron's formula. For this to be so, it is not sufficient that Heron's formula is a special case of Brahmagupta's; in mathematics, even if p entails q, q need not be explained by a proof of p that explains p.[27] What allows an explanation of Brahmagupta's formula to explain Heron's is that Brahmagupta's formula not only entails Heron's, but also perfects the symmetry that is strikingly nearly present in Heron's formula. Because an explanation of Brahmagupta's formula would explain Heron's formula, a conjecture of Brahmagupta's formula would typically receive additional support from the already proved Heron's formula beyond what the conjecture would receive merely from entailing a correct expression for the area of a triangle. Mathematicians recognize this additional confirmation.

I have argued that this confirmation of a conjecture—by virtue of the conjecture's "perfecting the symmetry" of some established result (that is not yet fully explained)—is an important, heretofore unrecognized form of IBE in mathematics.

One key lesson of this paper is that in mathematical practice, IBE is not confined to examples (such as the "calculator-number" theorem in Sect. 3) involving the confirmation of a conjecture that would be explained by a proof deriving its explanatory power from its treating all of the conjecture's cases alike and together. In the IBE argument operating in the calculator number example, the discovery that various cases (some calculator numbers) fit some pattern (are divisible by 37) justly confirms more strongly the conjecture that various unexamined cases (the unexamined calculator numbers) fit that pattern because of the conjecture's potential involvement in the explanation of the observed cases' fitting the pattern—that is, because the conjecture (if true) would likely have a proof that is a unified, uniform proof of the known cases fitting the pattern.

However, as we saw in Sect. 5, not all explanatory proofs in mathematics derive their explanatory power from providing this sort of unification. Some explanatory

[27] On my view, that is because p's striking features need not be striking features of q and what it takes for a proof in a given context to explain the theorem it proves depends on what features of that theorem (e.g., its unification, its exhibiting a certain symmetry) are striking in that context. Lange 2017 argues for this view of explanatory proofs.

proofs instead derive their explanatory power by exploiting the same symmetry in the setup that is striking in the result being explained. In Sect. 6, we saw how a theorem is made to bear more strongly on a conjecture insofar as that conjecture perfects the theorem's striking symmetry so that an explanatory proof of the conjecture would nicely explain why the theorem holds. This sort of IBE could (mathematicians say) reasonably have underwritten Brahmagupta's and Snell's confidence in their conjecture, justifying their efforts to prove their symmetric generalization of Heron's formula.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Atzema, E. (2015). From Brahmagupta to Euler: On the formula for the area of a cyclic quadrilateral. *BSHM Bulletin: Journal of the British Society for the History of Mathematics, 30*(1), 20–34.

Baker, A. (2009). Mathematical accidents and the ends of explanation. In O. Bueno & Ø. Linnebo (Eds.), *New waves in the philosophy of mathematics* (pp. 137–159). Palgrave Macmillan.

Ballentine, J. P. (1954). Note on Hero's Formula. *American Mathematical Monthly, 61*(5), 337.

Borwein, J., & Devlin, K. (2009). *The computer as crucible: An introduction to experimental mathematics*. AK Peters.

Brown, J. R. (2008). *Philosophy of mathematics: A contemporary introduction to the world of proofs and pictures* (2nd ed). Routledge.

Brown, K. (n.d.). Heron's Formula and Brahmagupta's Generalization. MathPages, https://mathpages.com/home/kmath196/kmath196.htm (accessed 20 January 2020).

Brown, J. R. (2017). Proofs and guarantees. *Mathematical Intelligencer, 39*(4), 47–50.

Cheng, E. (2004). Mathematics, Morally. Talk given to the Cambridge University Society for the Philosophy of Mathematics. http://eugeniacheng.com/wp-content/uploads/2017/02/cheng-morality.pdf

Colyvan, M. (2012). *An introduction to the philosophy of mathematics*. Cambridge University Press.

Conway, J. H. (1999). Correspondence with Peter Doyle concerning Heron's formula, http://www.math.dartmouth.edu/~doyle/docs/heron/heron.txt (accessed 11/19/2019).

Darwin, C. (1889). *The origin of species* (6th ed). Murray.

Dawson, J. W. Jr. (2015). *Why Prove It Again?* Cham: Springer.

de Wreede, L. (2007). *Willebrord Snellius (1580–1626): A Humanist Reshaping the Mathematical Sciences*. ISBN: 978–90–393–4628–0. PhD, Utrecht.

Dobbs, D. E. (1984). Proving Heron's formula tangentially. *The College Mathematics Journal, 15*(3), 252–253.

Dunham, W. (1990). *Journey through genius*. Wiley.

Dunham, W. (1999). *Euler the master of us all*. Mathematics Association of America.

Feynman, R., Leighton, R., & Sands, M. (1963). *The Feynman lectures on physics* (volume 1). Addison-Wesley.

Franklin, J. (2013). Non-deductive logic in mathematics: The probability of conjectures. In A. Aberdein & I. J. Dove (Eds.), *The argument of mathematics* (pp. 11–29). Springer.

Goodman, N. (1983). *Fact, fiction and forecast* (4th edition). Harvard University Press.

Gray, J. (2018). *A history of abstract algebra*. Springer.

Heilbron, J. L. (1998). *Geometry civilized*. Clarendon.

Hempel, C. G. (1965). *Aspects of scientific explanation*. Free Press.
Hersh, R. (2004). Heron's area formula: What about a tetrahedron? *The College Mathematics Journal, 35*(2), 112–114.
Kepler, J. (1981). *Mysterium Cosmographicum*, trans. A.M. Duncan. New York: Abaris.
Kichenassamy, S. (2010). Brahmagupta's derivation of the area of a cyclic quadrilateral. *Historia Mathematica, 30*, 28–61.
Klain, D. (2004). An intuitive derivation of Heron's formula. *The American Mathematical Monthly, 111*, 709–712.
Lange, M. (2017). *Because without cause: Non-causal explanation in science and mathematics*. Oxford University Press.
Lange, M. (2022). Putting explanation back into 'inference to the best explanation.' *Noûs, 56*, 84–109.
Lange, M. (2024). A false dichotomy in denying explanatoriness any role in confirmation. *Noûs, 58*, 520–535.
Leavitt, D. (2007). *The Indian clerk*. Bloomsbury USA.
Lipton, P. (2004). *Inference to the best explanation* (2nd ed). Routledge.
Lycan, W. G. (2002). Explanation and epistemology. In P. Moser (Ed.), *The Oxford handbook of epistemology* (pp. 408–433). Oxford University Press.
Nelson, R. B. (2001). Heron's formula via proofs without words. *The College Mathematics Journal, 32*(4), 290–292.
Nummela, E. (1987). No coincidence. *The Mathematical Gazette, 71*(456), 147.
Oliver, B. (1993). Heron's remarkable triangle area formula. *The Mathematics Teacher, 86*(2), 161–163.
Pólya, G. (1954). *Mathematics and Plausible Reasoning, 2 vols*. Princeton University Press.
Pottage, J. (1974). The Mensuration of Quadrilaterals and the Generation of Pythagorean Triads. *Archive for History of Exact Sciences, 12*, 299–354.
Pratt, V. (2009). Factoring Heron. *The College Mathematics Journal*, 40(1), 15–16
Rodriguez Villegas, F. (2007). *Experimental Number Theory*. Oxford University Press.
Salmon, W. (1975). Confirmation and relevance. *Induction. Probability, and confirmation, Minnesota studies in the philosophy of science, vol. VI* (pp. 3–36). University of Minnesota Press.
Salmon, W. (2001). Explanation and confirmation: A Bayesian critique of inference to the best explanation. In G. Hon & S. Rakover (Eds.), *Explanation: Theoretical approaches and applications* (pp. 61–92). Kluwer.
Seymour, P. (2016). Hadwiger's conjecture. In J. F. Nash & M. Rassias (Eds.), *Open problems in mathematics* (pp. 417–437). Springer.
Simmons, G. (1992). *Calculus gems*. McGraw-Hill.
Strogatz, S. 2015. Twitter feed 2 November 2015. https://twitter.com/stevenstrogatz/status/661386838482010112
van Fraassen, B. (1980). *The scientific image*. Clarendon.
van Fraassen, B. (1989). *Laws and symmetry*. Clarendon.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.