Mathematical Fit: A Case Study[†]

Manya Raman-Sundström[*] and Lars-Daniel Öhman[**]

[*]Department of Science and Mathematics Education, Umeå University, Umeå, SE-90187, Sweden.
Email: manya.sundstrom@umu.se

[**]Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, SE-90187, Sweden.
Email: lars-daniel.ohman@umu.se

ABSTRACT

Mathematicians routinely pass judgements on mathematical proofs. A proof might be elegant, cumbersome, beautiful, or awkward. Perhaps the highest praise is that a proof is right, that is, that the proof fits the theorem in an optimal way. It is also common to judge that one proof fits better than another, or that a proof does not fit a theorem at all. This paper attempts to clarify the notion of mathematical fit. We suggest six criteria that distinguish proofs as being more or less fitting, and provide examples from several different mathematical fields.

1. INTRODUCTION

Mathematics, as a subject, stands out from other fields in at least one essential way. Even though proofs derive from axioms, claims can be shown to be true or false. This is part of what makes mathematics satisfying. Similar to the feeling one gets from establishing the truth of a claim, one can often experience a feeling in mathematics that a claim is *right*, that a certain proof *fits* a theorem, or that a particular argument is *exactly* the one needed. These are stronger and somewhat more mysterious requirements than those needed to assert that

a particular claim is true. Similarly one can feel that one proof has a better fit for a certain theorem than another, or that another proof, while being technically correct, jars with the statement of a theorem, being somehow the wrong kind of argument for the given claim.

The purpose of this paper is to clarify different aspects of fit and give concrete examples of proofs that fit or do not fit in different ways. While fit has been discussed, in more or less informal ways in the literature (see [Wechsler, 1978; Sinclair, 2002; 2004] for discussions of fit in mathematics and science, and [Beardsley, 1981] for fit in aesthetics more generally), this paper is to our knowledge the first attempt to specify concretely what fit could be in mathematics.

Before embarking on this task, we consider briefly how the term is used outside of mathematics. The word 'fit' and the related word 'fitting' are somewhat vague terms which can mean many things. The connotations we will appeal to in our characterization of fit include (1) 'snug and correctly in place', (2) 'to be the proper size and shape', and (3) 'be in agreement or harmony with' or 'suitable for a specific purpose'. We will not use these characterisations directly, but they give a flavor of the kind of fit we would like to describe. Notice that some of the characterizations of fit are exact (the key fits the lock) and some gradable (this shoe fits better than that one). In our analyses we will see examples of both the exact and gradable types of fit. The intuition that a particular proof is 'right' is an example of the exact use of fit. The idea that one proof 'fits better than another', is an example of the gradable use.

Proofs are not the only mathematical objects that can possess fit. Definitions, diagrams, even theories, might be fitting, but in this paper we will limit the discussion to proofs. And by proof, we mean in this paper the written proof, which may or may not correspond to ideas or pictures or arguments held in the mind. This distinction will be important, for instance, when discussing the amount of detail given in a proof. In particular we will distinguish between three different kinds of fit — *direct fit*, *presentational fit*, and *familial fit* — which, roughly, relate a particular proof to its underlying ideas, to the form in which it is presented, and to other proofs and theorems.

The game plan is as follows: we will first describe the criteria for mathematical fit (Section 2), then apply these criteria to a set of contrasting proofs of four different theorems (Section 3) which will be summarized in a table (Section 4), and finally we will discuss, rather speculatively, the possible connection between the notion of fit, which we believe is somewhat tractable to systematic analysis, and the notions of mathematical explanation and beauty, which appear at first sight to be less amenable (Section 5).

2. PROPOSED CRITERIA FOR MATHEMATICAL FIT

The criteria for fit presented below came about by analysing a set of approximately twenty proofs, eight of which are given in this paper. This list is not meant to be comprehensive, but contains the criteria that were most salient to our group of mathematicians who discussed the different proofs.

Our focus here is on the description of the criteria rather than the relations between them (though a short comment about the interdependence of criteria will be made after the criteria are introduced). We will also deliberately avoid the question of what makes us value one proof more than another, even though fit might play a role in that judgement. To say that a proof *fits* is just to say to what extent and in what ways the proof fulfills the criteria discussed here. We are interested only in the question of how a proof fits, not in passing judgement on other kinds of claims, such as whether a proof is nice or memorable or surprising.[1]

As stated above, we will distinguish three different kinds of fit: *direct* fit, which deals with the relation between the proof and a theorem, *presentational* fit, which deals with the relation between the proof and the reader, and *familial* fit, which deals with the relation between a proof and mathematics as a whole (*e.g.*, other theorems). The rationale for these terms will be indicated below. In the following sections we present examples of proofs that exemplify different combinations of these different aspects of fit. We use contrasting proofs of the same theorems to illustrate how a proof could be a better or worse fit for a particular theorem, in each of these different dimensions.

2.1.   Criteria for Direct Fit

Direct fit refers to the relationship between a theorem and a proof. We call it direct because it does not rely upon any sort of mediation, such as the psychological process of reading nor the existence of other proofs. This is perhaps the first kind of fit that comes to mind, and the one that most directly captures the feeling that a certain proof is right. We identify two aspects of direct fit, the first of which we call coherence, which deals with the concepts used in the proof, and the second of which we call specificity, which deals with the choice of proof technique or method.

D_1:*Coherence.* The proof is stated in the same terms as the theorem.

> This criterion deals with the language and the conceptual apparatus used in the theorem. If a theorem is stated, for example, in terms of areas, then a proof that coheres will also be stated in terms of areas. If the proof is stated in other terms, such as angle measures, it would not cohere. The introduction of a seemingly unrelated conceptual apparatus would impact the coherence negatively.[2]

---

[1] While fit has positive connotations, that is to say that all other things being equal we would prefer a proof that fits to one that does not, there are certainly other features a proof might have that could trump the quality of fit. For instance the novelty of a cumbersome proof might make it preferable in some contexts to a well-known slick or elegant proof. But this discussion already diverges from the simple aim of this paper, which is to provide criteria for identifying the fit of a given proof.

[2] Note that this criterion is similar to 'purity of method', that is the idea, dating back at least as far as Aristotle, that resources used in solving or in proving a theorem should synchronize with those being used in understanding [Detlefsen and Arana, 2011].

D_2:*Specificity.* The proof employs a tool that uses exactly the right level of technical power for the task.

This criterion deals with the method used to prove a theorem. It captures the intuitive feeling that certain methods or tools are 'right' for a particular proof. The tool should be something that goes above and beyond the standard brute-force methods for approaching a proof of a theorem at hand. For a proof to be specific the tool should seem to be appropriate, 'just the right one needed' to get the proof done; it supplies all that is needed and nothing more.

2.2. Criteria for Presentational Fit

Presentational fit refers to the way a proof is communicated and the extent to which the proof write-up makes the underlying ideas accessible to the reader. This type of fit has a psychological component and is dependent in part on the reader's background knowledge. We deliberately choose the term presentational rather than representational to avoid the many varied connotations of the latter term. We will underscore that throughout this paper we are interested in the *write-up* of the proof, not how the proof might be represented in the mind. We distinguish between two aspects of presentational fit, the first of which, level of detail, refers to the appropriateness of detail in the written-up version of the proof, and the second of which, transparency, deals with how well the written proof matches the ideas the proof conveys.

P_1:*Level of detail.* The underlying ideas are presented with the appropriate amount of detail.

This criterion measures the extent to which a given written proof has an appropriate amount of detail, given the background knowledge that can be assumed of the reader. For instance, it seems proper to give the proof of an entry-level theorem in a field with details regarding basic calculations, but for a higher-level theorem this would be superfluous, as any reader who would be prepared to understand even the statement of the theorem must be expected to know certain basics of the theory.
A proof might fail to have the right level of detail in two ways — it might have too much, for instance including computations that readers could easily do on their own, or it might have too little, leaving to readers arguments or steps that are difficult and not obvious. While to some extent this criterion is subjective, there may be norms of adequacy depending on the given population (for instance, among well-trained topologists) and on the theorem. A proof with the appropriate level of detail allows the main ideas to

---

Historically this distinction was used to separate algebraic and geometric methods, and was employed famously by Hilbert in his *Grundlagen der Geometrie* [1899]. Our notion of coherence is more specific than that of purity of method, though we do not make any claims about understanding and focus only on the resources employed in a proof.

be foregrounded and other aspects of the proof, such as tedious calculations, to be backgrounded, and this is what brings about this feeling of fit.[3]

P_2:*Transparency.* The structure of the argument is clear.

In a proof that is strong on this criterion, it is easy see 'what is going on'. In other words, the structure of the proof is natural for the particular argument, and there is no *deus ex machina* component. A certain kind of argument, like proof by contradiction, could be natural in some contexts and not others. For instance, to prove the infiniteness of a set it is natural to set up a proof by contradiction because the infiniteness is formulated as being *not* finite. However, using a proof by contradiction to prove the Fundamental Theorem of Arithmetic would obscure the central ideas. While the ability to grasp ideas depends on the background of the reader, the criterion of transparency deals with a proof's *potential* to make clear the underlying ideas. In other words, if a proof is transparent, a reader with the appropriate background should be in an ideal position to grasp the ideas of the proof.

2.3. Criteria for Familial Fit

Familial fit refers to the relationship between a particular proof and a family of proofs. We identify two basic ways that a family membership can be established. One, which we call generality, connects proofs via generalization. The other, which we call connectedness, connects proofs via similarities between their ideas. This aspect of fit deals less with the feeling that this proof is the 'right' one, and more about the positioning of a particular proof relative to the rest of mathematics.

F_1:*Generality.* The idea of the proof generalizes to a larger class of theorems.

This criterion deals with how well an underlying idea generalizes to prove a class of theorems. The theorem at hand can be seen as a specific instance of a more general claim, which is still provable by means of the same general proof idea. This type of family relation might be thought of as being vertical, like nested cups, because the more general proof subsumes the less general one, without adding additional ideas or techniques.

F_2:*Connectedness.* The proof idea connects to proof ideas of other theorems.

This criterion also deals with family membership but not via generalization. The proofs may be related via ideas or techniques, but one proof does not subsume the other (though a third, more general, proof might subsume both of them.) This type of family relation might be thought of as being horizontal — if generality is a relationship between parent and child, connectedness is a relationship between siblings. Connectedness is a matter of degree — the more proofs a given proof is related to, in this way, the more connected it is.

---

[3] The foregrounding and backgrounding of information might be connected Lange's [2015] notion of salience in the context of mathematical explanation.

2.4. Comment on Interdependence of Criteria

In the course of analyzing the examples below we will see that these criteria distinguish between different proofs. It seems that the criteria are pairwise independent, in the sense that not all proofs scoring high on criterion A always score high on criterion B, for any choice of A and B. However, it may still be the case that scoring high on some subset of the criteria (say, A, B, and C) always implies scoring high on some other criterion (say, D), so that the set of criteria as a whole is not independent. We have not made an effort to investigate these kinds of relationships; we merely note that they might exist.

Another natural question that might arise is whether the different aspects of fit could work against each other; so a few words about that are in order. In particular, two of the criteria that might appear at first glance to be in conflict are generality, F_1, and specificity, D_2. It does seem to be the case that a very specific proof is unlikely to be general, and *vice versa*. However it is also the case that some proofs involve special cases which are both specific and general at the same time. A *generic example*[4] is a specific example which captures the generality of a claim while at the same time grounding the argument in a concrete situation. Some of our examples have this character, and in that case specificity and generality are not at all in conflict.

3. EXAMPLES

We will consider proofs of four theorems in this paper: the square root of 2 is irrational, the Pythagorean theorem, the complex-conjugate-root theorem, and Pick's theorem. We have purposely chosen theorems that are familiar and have a number of well-known proofs so that the focus can be on the analysis of these proofs, in terms of fit, rather than on the technical details of the proofs. In the first three examples we contrast two proofs for each theorem, the first of which will fit the theorem in some way, and the second of which will not fit to the same extent. We will then consider two proofs of Pick's theorem where there is no clear winner regarding the fit. A table is given in Section 4 to summarize the analyses done in this section.

3.1. The Square Root of 2 is Irrational

Below we contrast two proofs of the theorem that the square root of two is irrational. The first is a standard proof, which we claim fits the theorem in all three ways. The second proof, using base 3 representation, is less well-known. We will see that based on several of our criteria, this proof fits less well than the standard proof.

Theorem 3.1. *The square root of 2 is irrational, that is $\sqrt{2}$ cannot be written as $\sqrt{2} = p/q$ where p and q are positive integers.*

---

[4] See [Balacheff, 1988] for further discussion of generic examples.

For the first proof, we will use the result that any positive integer p has a unique factorization into prime numbers — that is, a version of the fundamental theorem of arithmetic — which we phrase as a lemma. The fundamental theorem of arithmetic was known already by Aristotle, but not proved until Gauss had recognized the need to supply a rigorous proof (see [Agargün and Özkan, 2001]). We omit the proof of the lemma.

Lemma 3.2. *Let p be a positive integer. Then p can be written as*

$$p = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot p_3^{\alpha_3} \cdot \ldots,$$

where the α_i are uniquely determined non-negative integers, and $\{p_1, p_2, p_3, \ldots\}$ is the set of prime numbers.

First proof of Theorem 3.1. Assume, for the sake of contradiction, that $\sqrt{2} = p/q$, where p and q are positive integers. By Lemma 3.2, we may then write p and q as

$$p = 2^{\alpha_1} \cdot 3^{\alpha_2} \cdot 5^{\alpha_3} \cdot \ldots \quad \text{and} \quad q = 2^{\beta_1} \cdot 3^{\beta_2} \cdot 5^{\beta_3} \cdot \ldots$$

respectively. Now, since $\sqrt{2} = p/q$, we get $2 = p^2/q^2$, and subsequently $2q^2 = p^2$.

Observe that

$$p^2 = 2^{2\alpha_1} \cdot 3^{2\alpha_2} \cdot 5^{2\alpha_3} \cdot \ldots \quad \text{and} \quad q^2 = 2^{2\beta_1} \cdot 3^{2\beta_2} \cdot 5^{2\beta_3} \cdot \ldots$$

by standard laws of arithmetic. When multiplying q^2 by 2, the exponent of the prime factor 2 goes up by one, so that

$$2q^2 = 2^{2\beta_1+1} \cdot 3^{2\beta_2} \cdot 5^{2\beta_3} \cdot \ldots$$

Now, from $p^2 = 2q^2$, we get that

$$2^{2\alpha_1} \cdot 3^{2\alpha_2} \cdot 5^{2\alpha_3} \cdot \ldots = 2^{2\beta_1+1} \cdot 3^{2\beta_2} \cdot 5^{2\beta_3} \cdot \ldots$$

Since the prime factorization is unique by Lemma 3.2, we can equate exponents on either side to find that $2\alpha_1 = 2\beta_1 + 1, 2\alpha_2 = 2\beta_2, 2\alpha_3 = 2\beta_3, \ldots$. The equation $2\alpha_1 = 2\beta_1 + 1$ amounts to an even number being equal to an odd number, which is clearly absurd. We thus have a contradiction, and the original assumption must be false. The result follows by *reductio ad absurdum.* □

To what extent does this proof exhibit the three different kinds of fit?

D_1:*Coherence.* The proof is coherent, as the theorem concerns the structure of the integers, and the line of argument in the proof uses only these concepts. In particular, the proof idea, *i.e.*, the idea by virtue of which Theorem 3.1 is true, is that equal integers have equal exponents in their prime factorizations.

D_2:*Specificity.* The proof does not seem to fulfill the criterion of specificity. The technical tool at work in this case is prime factorization. This is adequate to prove the theorem, but is overly powerful. In fact, only the number of factors 2 in p and q is needed for the conclusion to be drawn.

P_1:*Level of Detail.* We have deliberately written this proof so that the level of detail criterion is not fulfilled.[5] This proof gives the relevant information simply and concisely, but some unnecessary trivial steps are included, as for example in the move from the prime factorizations of p and q via $\sqrt{2} = p/q$ to the equation of the prime factorizations of p^2 and $2q^2$. Also, some extraneous information is presented, like the details of equating exponents of prime factors other than 2. Note that a less detailed proof is not necessarily the most pedagogical. The steps given in a less detailed proof may be helpful if you do not have the appropriate background to fill in the missing steps, but given the level of the theorem, and the concepts needed to understand the statement of the theorem, these details seem unnecessary.

P_2:*Transparency.* The transparency criterion is fulfilled. The underlying idea, namely that the parities of the exponent of the prime factor 2 in the expressions for p^2 and q^2 will be different, is easily graspable. As regards the structure of the rest of the argument, the moves from $\sqrt{2} = p/q$ to $2 = p^2/q^2$, to $p^2 = 2q^2$ are directly motivated by how fractions and square roots are introduced in terms of the integers: since fractions are introduced as ordered pairs of integers, and square roots are introduced (in the standard way) as those elements x that satisfy an equation $x^2 = n$, the rewriting steps can be read as a very natural unwinding of the definitions in terms of more basic concepts.

F_1:*Generality.* The proof fulfills the criterion of generality. The proof idea can be adapted without effort to prove that the square root of p is irrational for any prime p (by looking at the exponent of the prime factor p in the final step of the proof), and with some minor effort to the case when n is any integer which is not a perfect square (in which case the prime factorization of n, and the exponents of several prime factors must be examined).

F_2:*Connectedness.* The proof is connected. To establish this, we must describe how the proof idea figures in a range of other proofs in such a way that the proof at hand is a particular instance. For example, the basic divisibility theorems $p|n^2 \implies p|n$ for p prime, and $p|ab \implies p|a \lor p|b$ for p prime, could be quoted.

---

[5] Compare this proof with the following sparser one.

Assume, for the sake of contradiction, that $\sqrt{2} = p/q$, where p and q are positive integers. By Lemma 3.2, we may then write p and q as $p = 2^{\alpha_1} \cdot 3^{\alpha_2} \cdot 5^{\alpha_3} \cdot \ldots$ and $q = 2^{\beta_1} \cdot 3^{\beta_2} \cdot 5^{\beta_3} \cdot \ldots$. Now, from $p^2 = 2q^2$, we get that $2^{2\alpha_1} \cdot 3^{2\alpha_2} \cdot 5^{2\alpha_3} \cdot \ldots = 2^{2\beta_1+1} \cdot 3^{2\beta_2} \cdot 5^{2\beta_3} \cdot \ldots$. Since the prime factorization is unique by Lemma 3.2, we can equate exponents on either side to find that $2\alpha_1 = 2\beta_1 + 1$, which amounts to an even number being equal to an odd number, which is clearly absurd. We thus have a contradiction, and the original assumption must be false.

We can see here a clear example of two criteria pulling in different directions: A proof using only the number of factors 2 in p and q would have been more specific, but less general. It should not, however, be concluded from this that specificity and generality are always negatively correlated. This is the one proof in our sample that fails the criterion for level of detail. It is, in general, fairly easy to imagine proofs that vary in level of detail; so we have not bothered to alter the other proofs. The failure to fulfill this criterion comes from the fact that the details are at a fairly low level compared to what is being proven. These details might be appropriate when first introducing a proof but are not necessary for someone with adequate background knowledge.

The second proof, in contrast, exhibits fewer of the criteria of fit than the first proof. This second proof uses base 3 representation to get a contradiction, which proves the specific claim that the square root of 2 is irrational, but the argument does not carry over to similar claims (*i.e.*, square roots of other primes). The proof relies on the following lemma which can be proven independently of the irrationality of $\sqrt{2}$.

Lemma 3.3. *Let p be an integer, represented in base 3. Then the last non-zero digit of p^2 represented in base 3 must be a 1.*

Proof of Lemma 3.3. We write the congruence $a \equiv b \pmod 3$ as $a \equiv_3 b$. First observe that any zero digits at the end of p will only contribute zeros to the end of p^2; so we may assume that p does not end with a zero. Now, suppose the last digits of p are $\ldots, a_4, a_3, a_2, a_1$, where $a_1 \neq 0$, and the last digits of p^2 are $\ldots, b_4, b_3, b_2, b_1$. We have two cases: $a_1 = 1$ and $a_1 = 2$. If $a_1 = 1$ then by standard laws of arithmetic (in base 3), we have

$$b_1 \equiv_3 a_1^2 = 1^2 = 1;$$

so $b_1 = 1$ and we are finished. If $a_1 = 2$, we have

$$b_1 \equiv_3 a_1^2 = 2^2 \equiv_3 1,$$

and again we are finished. □

Second proof of Theorem 3.1. Assume, for the sake of contradiction, that $\sqrt{2} = p/q$, where p and q are integers. By standard arithmetic we get that $p^2 = 2q^2$.

By Lemma 3.3, the last non-zero digit of p^2 represented in base 3 must be 1, but the last non-zero digit in the representation of $2q^2$ must be 2, since by the same lemma, the last non-zero digit in q^2 must be 1. This is a contradiction, and the original assumption must be false; so the result follows by *reductio ad absurdum*. □

To what extent does this proof exhibit the three different kinds of fit?

D_1:*Coherence.* The proof is not coherent, since the argument in terms of base 3 is in different terms than the statement of the theorem. Of course, if $\sqrt{2}$ is

irrational, it is irrational regardless of the base representation. However, the fact that in one of these representations it is possible to get a contradiction is not immediately obvious. Moreover, the statement of the theorem makes no reference to the specific representation of the integers; so choosing a particular base to carry out the argument imposes a restriction that breaks any possible parallel between the theorem and proof.

D_2:*Specificity.* The proof is specific. The technical tool that works in this case is expressing the integers in base 3. This turns out to be just what is needed to draw the desired conclusion (the different last digits), and it is hard to see how one would reduce this idea while still maintaining this possibility. Arguably, one could restrict this proof to looking at p and q modulo 3, considering only the last digits, but this seems very similar to the present proof.

P_1:*Level of detail.* The proof has an appropriate level of detail. It gives only information relevant for the level of the theorem, simply and concisely.

P_2:*Transparency.* This proof is not transparent. It is not clear why rewriting the integers p and q in base 3 should be relevant, and it seems like a trick — it just happens to work.

F_1:*Generality.* The proof is not general. The base 3 representation cannot be used to prove, for instance, that $\sqrt{19}$ is irrational in this way, and neither can it prove irrationality for the square root of any other prime whose base 3 representation ends with a 1: Let a_3 denote the representation of the integer a in base 3, and observe that $19 = 201_3$. Now consider the equation $201_3 \cdot q_3^2 = p_3^2$. The last non-zero digit on each side will be 1, and no conclusion can be drawn.[6]

F_2:*Connectedness.* The proof is not connected, as far as we can tell. We know of no other proof that uses this kind of argument. For any given prime number one could try to use a similar argument, but we do not know of any current set of proofs that have this character.

While this proof satisfies one criterion for presentational fit — relating to the level of detail — and one criterion for direct fit — being specific — it seems fair to say that the proof is not a very good fit in general for the theorem. The proof is not general, as noted above. The first proof, in contrast, uses prime

---

[6] By studying the possible last digits of squares in different bases, the proof can be adapted to show that other square roots of primes are irrational. For example, the only possible last non-zero digits of squares in base 5 are 1 and 4; so the square roots of 2 and 3 can be shown to be irrational using a similar proof in base 5. In fact, it may even be possible to find, for a given prime p, a suitable base b, in which the only possible last non-zero digits of squares do not coincide with the last non-zero digit of p in base b so that the same proof idea would carry over.

However, given a prime p, the proof as it stands gives no indication of how to find such a suitable base b, and therefore the proof can not be said to be general.
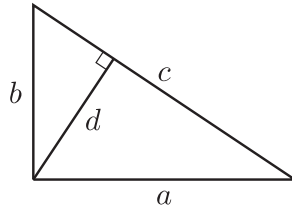
Fig. 1. Dissection of a right triangle

factorization, which allows us to get quickly to the essence of why the theorem must hold. The first proof involves simply writing out what the terms mean, and using the uniqueness of the prime factorization to get a contradiction. This argument can be easily generalized to square roots of any prime numbers, as mentioned above.

3.2. The Pythagorean Theorem

The second example we will consider is the Pythagorean theorem. Again we will contrast a well-known proof, which we claim fits the theorem, with a lesser known one that does not fit. Whereas with the previous example the proof that fits does so mostly based on its algebraic structure, which was presented by prime factorization, in this case the proof that fits does so based on geometric properties. The first proof comes from Euclid's *Elements*, VI. 31, and the second proof is contemporary.

Theorem 3.4. *Let c be the hypotenuse of a right triangle T_0, and let a, b be other two sides. Then the sum of the areas of the squares constructed on sides a and b of T_0 equals the area of the square constructed on the hypotenuse.*

First proof of Theorem 3.4. Consider Figure 1 with line d perpendicular to c. This figure contains three similar triangles, T_1, T_2, and T_0, which lie on sides a, b, and c, respectively. Clearly, the sum of the areas of T_1 and T_2 equals that of T_0. But it is also the case that each of the triangles lies on one of the sides of the original triangle; so the sum of the areas on sides a and b must be the same as the area on side c. Changing the scale factor gives the classic result involving squares on each side. The algebraic details of this argument are given in the Appendix. □

To what extent does this proof exhibit the three different kinds of fit?

D_1:*Coherence.* The proof is coherent. The theorem is stated in terms of areas, as is the proof. We also note that the idea of preserving areas is in line with the more famous proof in Euclid's *Elements*, I. 47, where the areas of the squares constructed on either side are shown to be equal by area-preserving steps, a method that was standard for theorems involving area in ancient Greek mathematics.

D_2:*Specificity.* The proof fulfills the criterion of specificity. The technical tool that works in this case is dividing the original triangle into similar triangles (put more generally, this could be described as dissection). This division allows us to see the crucial relationship, namely that all three triangles are similar and their areas add up. Other dissections could be used, but the present dissection supplies precisely what is needed for the proof.

P_1:*Level of detail.* The proof has an appropriate level of detail. It gives only information relevant for the level of the theorem, simply and concisely.

P_2:*Transparency.* This proof is transparent. The proof consists of two main ideas, presented clearly and in a logical succession, namely the dissection of the triangle into similar triangles, and that the scaling of the areas carries over to arbitrary shapes. The particular choice of the dissecting line may be seen as a trick, but drawing the height in a right-angled triangle and getting similar triangles in this way is a rather standard procedure.

F_1:*Generality.* The proof is general.[7] The generality comes from the fact that the proof works for arbitrary similar shapes constructed on the sides of the triangle.

F_2:*Connectedness.* Whether or not the proof is connected is not as clear as the other criteria, but we are inclined to consider the proof connected. The class of proofs to which the proof can be seen to belong (other classes may be possible) might be taken to be proofs by area preservation, for instance the other ancient Greek proofs that used this method.

The second proof, which uses trigonometry, is in some ways remarkable. It was long thought that the Pythagorean theorem could not be proven using trigonometry, because it would be impossible not to use the fact that $\sin^2 x + \cos^2 x = 1$, which is equivalent to the Pythagorean theorem, and thus make the argument circular. Zimba [2009] found the following proof which avoids circularity, using the subtraction formulas for sine and cosine. We assume that we have the subtraction formulas for sine and cosine,

$$\cos(\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta$$
$$\sin(\alpha - \beta) = \sin \alpha \cos \beta - \cos \alpha \sin \beta.$$

A sketch of how to prove the subtraction formulas for sine and cosine without relying on the Pythagorean theorem is given in the Appendix.

Second proof of Theorem 3.4.. Suppose that α is the angle opposite side a, and β is the angle opposite side b, and without loss of generality that $0 < \beta \leq \alpha < 90°$. We now have

$$\cos \beta = \cos(\alpha - (\alpha - \beta))$$

---

[7] Steiner [1978] gives an account of this generality. He claims that this proof is the most explanatory and most general of all proofs of the Pythagorean theorem.

$$= \cos\alpha\cos(\alpha - \beta) + \sin\alpha\sin(\alpha - \beta)$$
$$= \cos\alpha(\cos\alpha\cos\beta + \sin\alpha\sin\beta) +$$
$$\sin\alpha(\sin\alpha\cos\beta - \cos\alpha\sin\beta)$$
$$= (\cos^2\alpha + \sin^2\alpha)\cos\beta,$$

from which it follows that $\cos^2\alpha + \sin^2\alpha = 1$, since $\cos\beta$ is the ratio between one leg and the hypotenuse of a right triangle, and as such is never zero. The theorem now follows from the definitions of sine and cosine and scaling. □

To what extent does this proof exhibit the three different kinds of fit?

D_1:*Coherence.* The proof is not coherent. Although it may be possible to rephrase the introduction to the proof, at least in its historical context, it seems that the underlying idea was something along the lines of 'is it even possible to give a trigonometric proof of the Pythagorean theorem?', not something more mathematically motivated. The trigonometric language is clearly a different framework from the one in the statement of the theorem.

D_2:*Specificity.* The proof exhibits specificity, in that the tool used (the subtraction formulas) works out to be just what is needed for the conclusion to be drawn. We note that the subtraction formulas can of course be used for proving many other results, but this does not detract much from the specificity. To indicate what would have been a less specific proof, it might have made reference to some abstract generalization of the subtraction formulas to Hilbert spaces.

P_1:*Level of detail.* The proof has an appropriate level of detail. It gives only information relevant for the level of the theorem, simply and concisely.

P_2:*Transparency.* This proof is not transparent. There is no clear sense of direction in the calculations performed. The structure of the proof is clear enough, but it seems that there is little in the way of a natural sequence of ideas, and the introduction of trigonometric quantities seems extraneous. It is hard to see, for instance, why one would want to rewrite $\cos\beta$ as $\cos(\alpha - (\alpha - \beta))$.

F_1:*Generality.* The proof as it stands is not general. It is true that once $\cos^2\alpha + \sin^2\alpha = 1$ is established, one can add the scaling argument to show that the result holds for arbitrary similar shapes, but the scaling argument is not an integral part of the proof.

F_2:*Connectedness.* The proof is not connected. The fact that this proof was only found in 2009 bears witness to its singular nature, and we have not been able to give a family of proofs, of which this is a special case. Arguably, the proof belongs to the loosely defined family of proofs employing trigonometric identities, but it seems this is too general a family to be meaningful.

What is it that makes the first proof seem to fit better than the second? As in the case of the proofs that the square root of two is irrational, the first proof fulfills many more of the criteria for fit, in all three categories, while the second

proof fulfills only a few criteria. The fact that the second proof satisfies the detail and specificity criteria does not seem to compensate for the lack of being transparent and coherent or general and connected. The geometric argument of the first proof gets very quickly to what the Pythagorean theorem is about. The relationship between the areas on the sides of the right triangle is very easy to see once the similarity is established (which is also not very hard to do), and we see that we are proving a more general, deeper claim than what the Pythagorean theorem states. The result is not based on the specific shape of the triangle but on the areas of congruent figures.

3.3. The Complex-Conjugate-Root Theorem

The third theorem that we will consider is the complex conjugate root theorem, which was first proven by d'Alembert [1746] as a corollary of the Fundamental Theorem of Algebra. In this case the two proofs we will consider are not so radically different in terms of fit as the pairs of proofs in the previous examples. Both of the proofs fit to some extent, but one (we claim) fits better. In the two proofs of the conjugate-root theorem, the arguments are very similar, differing mostly in the amount and type of detail provided, and the aspects of the proof that are foregrounded or backgrounded. In the first example, the algebraic details to establish the claim are in the foreground, and in the second example the details are left to the reader and the proof idea is foregrounded.

Theorem 3.5. *Let $f(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0$ be a polynomial in the complex variable z, and suppose all a_i are real numbers. If z_0 is a root of the equation $f(z) = 0$, then so is the complex conjugate $\overline{z_0}$.*

Theorem 3.5, compared with Theorems 3.1 and 3.4, deals with mathematics that is not entirely trivial to all working mathematicians. The first proof, which might be considered brute force is a fairly straightforward algebraic treatment, requiring not much more than high-school mathematics to follow. It evaluates the polynomial at the relevant value. The second proof is a bit more sophisticated, and fleshing out the details requires some abstract algebra. However, one might still expect that any research-level mathematician would feel that the idea given in the second proof would be enough for her to be able to supply the missing details. We will use the following lemma, which can be confirmed by straightforward calculation.

Lemma 3.6. *Let z and w be arbitrary complex numbers. Then $\overline{zw} = \overline{z}\,\overline{w}$ and $\overline{z+w} = \overline{z} + \overline{w}$.*

The proof proceeds by calculation, by showing that $f(z_0) = 0$ implies $f(\overline{z_0}) = 0$.

First proof of Theorem 3.5. Assume that $f(z_0) = 0$. We must show that $f(\overline{z_0}) = 0$, or in other words that $a_n \overline{z_0}^n + a_{n-1} \overline{z_0}^{n-1} + \ldots + a_1 \overline{z_0} + a_0 = 0$.

By applying Lemma 3.6 iteratively, and using the fact that $a_i = \overline{a_i}$ for any real number a_i in the second equality, we get

$$
\begin{aligned}
f(\overline{z_0}) &= a_n \overline{z_0}^n + a_{n-1} \overline{z_0}^{n-1} + \ldots + a_1 \overline{z_0} + a_0 \\
&= \overline{a_n z_0^n} + \overline{a_{n-1} z_0^{n-1}} + \ldots + \overline{a_1 z_0} + \overline{a_0} \\
&= \overline{a_n z_0^n + a_{n-1} z_0^{n-1} + \ldots + a_1 z_0 + a_0} \\
&= \overline{f(z_0)} = \overline{0} = 0. \qquad \qquad \square
\end{aligned}
$$

To what extent does this proof exhibit the three different kinds of fit?

D_1:*Coherence.* The proof is coherent. It proceeds using only concepts introduced in the statement of the theorem.

D_2:*Specificity.* This proof involves only carrying out a simple computation; so it is not specific.

P_1:*Level of detail.* The proof has an appropriate level of detail. It gives only information relevant for the level of the theorem, simply and concisely.

P_2:*Transparency.* This proof is transparent. The direct calculations have a clear beginning and end, and the steps in between follow in a natural order. It is easy to follow the calculations, with the conjugation being undistributed.

F_1:*Generality.* The proof is to some extent general. The proof works for any polynomial with real coefficients; so in this sense it is more general than, for instance, calculating that both $i^2 + 1 = 0$ and $(-i)^2 + 1 = 0$, that is the polynomial $f(z) = z^2 + 1$. However, proofs by direct calculation are particular to the specific claim to be proven.

F_2:*Connectedness.* In a rather weak sense, the proof is connected. It is, at least, a particular instance of the family of proofs weakly held together by being proofs by direct calculation. One can imagine saying something along the lines of 'Aha, so you can prove this by direct calculation'.

The second proof builds on the simple observation that both i and $-i$ when squared equal -1. This symmetry is at the heart of why the algebra above works, and the proof moves forward with reasoning based on this idea.

Second proof of Theorem 3.5. Since the defining property of the imaginary unit i is that $i^2 = -1$, there is nothing that sets it apart from its negative counterpart $-i$ in relation to the real numbers, as we also have $(-i)^2 = -1$. Thus when introducing the imaginary unit, we may take $-i$ for i, and *vice versa*, and the complex conjugation operation only takes us from one version of the complex numbers to the other indistinguishable one.

Therefore any polynomial in a complex variable with real coefficients cannot tell i and $-i$ apart, and so evaluation at z_0 or $\overline{z_0}$ makes no difference for the value of the polynomial. A geometric interpretation is that the complex plane

is reflected over the real axis. This of course leaves the real axis fixed. So the right-hand side of $f(z_0) = 0$ is not changed when f is evaluated at $\overline{z_0}$. ☐

In order to make this second proof more formal, one would introduce the mapping $\sigma : z \to \overline{z}$ from the field extension $\mathbb{R}(i)$ to the field extension $\mathbb{R}(-i)$, and establish that σ is an isomorphism. The algebra involved in checking this is in fact exactly Lemma 3.6.

To what extent does this proof exhibit the three different kinds of fit?

D_1:*Coherence.* The proof is coherent. The statement of the theorem regards a sort of indistinguishability of z_0 and $\overline{z_0}$, and it is exactly the indistinguishability of i and $-i$ that is the driving idea in the theorem.

D_2:*Specificity.* The proof is specific. A less specific proof could have talked about general mappings preserving certain identities in field extensions, but here only the pertinent observations regarding i and $-i$ are given. Giving the particulars of the mapping $\sigma : z \to \overline{z}$ would have further contributed to the specificity.

P_1:*Level of detail.* The proof has a reasonable level of detail given the level of the theorem, but much more is left to the reader than in the first proof. For instance, the geometric argument is said to follow from the algebraic assumptions without an explicit mapping of one to the other.

P_2:*Transparency.* The proof is not completely transparent since it is only a sketch, but the ideas here are clear and easy to grasp, namely that $-i$ and i function interchangeably as far as the complex numbers are concerned.

F_1:*Generality.* The proof is general, in the sense that the symmetry idea can be employed to prove a range of theorems regarding polynomials in field extensions. One example of such a theorem is the following: 'Let $f(x)$ be a polynomial with integer coefficients. Suppose that $a + b\sqrt{c}$ is a root of the equation $f(x) = 0$, where a and b are rational and $\sqrt{c}$ is irrational. Then $a - b\sqrt{c}$ is also a root of the equation.'

F_2:*Connectedness.* The proof is connected. It is a particular instance of a family of results regarding polynomials in field extensions, as indicated above. It is also an instance of the types of theorems proved in Galois theory on roots of polynomials.

Unlike the other pairs of proofs which seemed very different in nature, these two proofs are fairly similar. Actually, it could be argued that the two proofs are really just the same proof, but presented with different levels of abstraction. However there are several differences between the proofs, which while subtle, could make a difference to their sense of fit. Whereas the first proof emphasizes the calculation aspect, which does not seem to be at the heart of the statement, the second proof emphasizes the symmetry, which *does* seem to be at the heart of the statement. The second proof, by emphasizing the symmetry, helps us

see an underlying mechanism, getting at the heart of why the computational argument works. The difference between the proofs seems to be what aspects get foregrounded and backgrounded. This might, in turn, be connected to what is considered salient about the proofs.[8]

3.4. Pick's Theorem

In this section we will look at two proofs that demonstrate familial fit (and some other aspects of fit) in different ways. The proofs both demonstrate Pick's theorem, which gives a formula for the area of lattice polygons based on the number of lattice points inside and on the boundary of the polygon. Pick's theorem gives a simple formula for calculating the area of a lattice polygon, that is, a polygon constructed on a grid of evenly spaced points. The theorem, first proven by Georg Alexander Pick in 1899, is a classic result of geometry.[9] The first proof gets its sense of family membership through the technical tool used to make the proof tractable (in this case via angle measures). In the second case the sense of family membership comes via its key idea, which places the proof in a family connected by the relationship established in Euler's formula.

We shall restrict ourselves to the lattice $\mathbb{Z}^2$, that is the set of points in the plane with integer coordinates. In this case, a lattice polygon is simply a polygon in the plane, all of whose vertices have integer coordinates. An *interior (lattice) point* is a point of the lattice that is properly contained in the polygon, and a *boundary (lattice) point* is a point of the lattice that lies on the boundary of the polygon.

Theorem 3.7 (Pick's Theorem). *Let A be the area of a lattice polygon, let I be the number of interior lattice points, and let B be the number of boundary lattice points, including vertices. Then $A = I + \frac{1}{2}B - 1$.*

For example, in the lattice polygon given in Figure 2a, there are 10 boundary points and 11 interior points; so the area is $11 + 10/2 - 1 = 15$.

In both proofs, we will draw on the following two lemmas, which we state here without proof, as they will not figure in the analysis and discussion. An *elementary triangle* is a triangle whose vertices are lattice points, and which has no further boundary points and no interior points.

Lemma 3.8. *Any lattice polygon can be triangulated by elementary triangles.*

Lemma 3.9. *The area of any elementary triangle in the lattice $\mathbb{Z}^2$ is $1/2$.*

---

[8] See [Lange, 2014] for a discussion of salience in connection to mathematical explanation.

[9] The original proof is found in [Pick, 1899]. A short historical account is given at http://jsoles.myweb.uga.edu/history.html. An initial analysis of these proofs can be found in [Raman and Öhman, 2011]. The first proof was suggested to us by Bjorn Poonen, and the second appears in [Aigner and Ziegler, 2010].
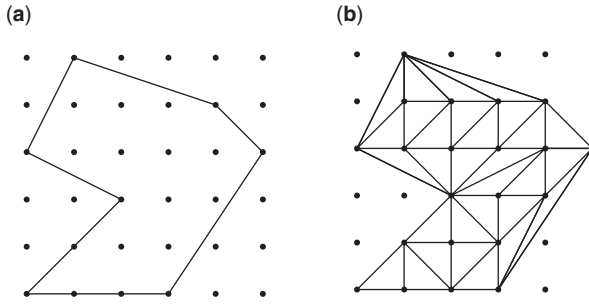
Fig. 2. a. Example of a lattice polygon. b. One possible triangulation.
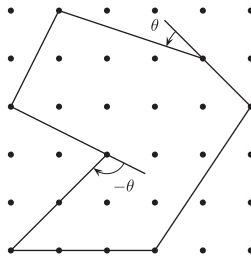

Fig. 3. Lattice polygon, with two exterior angles marked

From Lemma 3.8 and Lemma 3.9, it follows that the number of triangles in any triangulation of a given lattice polygon is the same.

First proof of Theorem 3.7, using angles. We begin by partitioning the polygon P into N elementary triangles, which is possible by Lemma 3.8 (see Figure 2b).

We now sum up the internal angles of all of these triangles in two different ways. On the one hand, the angle sum of any triangle is π; so the sum of all the angles is $S = N \cdot \pi$.

On the other hand, at each interior point i, the angles of the elementary triangles meeting at i add up to 2π. At each boundary point b that is not a vertex, the angles of the elementary triangles meeting at b sum to π. At the vertices, the angles do not add up to π, but if we add the interior angles at all the vertices, we get $k\pi - 2\pi$, where k is the number of vertices, since the sum of the exterior angles is 2π (see Figure 3). One can argue for this result by noting that walking along the perimeter of the polygon, one completes one full turn, that is 2π. Note that some exterior angles contribute a positive term, and others a negative term.

Let I be the number of interior points and B be the number of boundary points. In all, the sum of the angles at boundary points is $B \cdot \pi - 2\pi$, and the sum of the angles at internal points is $I \cdot 2\pi$. Therefore, $S = I \cdot 2\pi + B \cdot \pi - 2\pi$.

We conclude that $N \cdot \pi = I \cdot 2\pi + B \cdot \pi - 2\pi$; so canceling π we get $N = 2I + B - 2$. Since by Lemma 3.9 the area of any elementary triangle is $\frac{1}{2}$, we have $A = \frac{1}{2}N = I + \frac{1}{2}B - 1$. $\qquad\qquad\square$

To what extent does this proof exhibit the three different kinds of fit?

D_1:*Coherence.* The proof is not coherent. The use of angle measures seems rather extraneous.

D_2:*Specificity.* The proof fulfills the criterion of specificity. The introduction of the angle measures turns out to be adequate to prove the result, yet there is a quality of surprise to the proof, arising from the fact that angle measures seem to be extraneous to the question.

P_1:*Level of detail.* The proof has an appropriate level of detail. It gives only information relevant for the level of the theorem, simply and concisely.

P_2:*Transparency.* This proof is not transparent. It is easy enough to follow the steps, but it is not clear why we should wish to investigate the angle sums of a triangulation. Therefore, the proof has the character of 'bear with me for a while', which is not transparent. The step of dissecting the polygon into elementary triangles, however, is rather natural since areas of triangles are more easily calculated than areas of general polygons, and such dissection is commonly used to facilitate area calculations.

F_1:*Generality.* The proof is not general. The conclusion of the theorem holds for other lattices and geometries as well, with appropriate scaling. For instance, if one applies a dilation along an axis, the present proof can be adapted to prove a corresponding theorem for the new lattice (the areas of the elementary triangles will have changed). However, angle measures are not conserved under such a transformation, which detracts from the generality of the proof idea. Also, for area-preserving transformations, the proof idea may not work, since the angle sums of the elementary triangles may no longer be constant.

F_2:*Connectedness.* This proof has some weak features of connectedness. One can see it as an example of a proof that involves double counting, and it is also an example of a proof that involves angle measure (a feature not directly related to the theorem's meaning). It is probably not the case that many mathematicians consider all proofs that involve angle measure to be a natural and/or important way to categorize proofs. The use of double counting is certainly a more natural candidate for a family resemblance, but still, since double counting is such a low-level principle, this would be akin to grouping proofs that make use of, for instance, the distributive law.

For the second proof, we will also need the well-known Euler's formula.

Lemma 3.10. *(Euler's formula) Let f be the number of faces, e be the number of edges and v the number of nodes in a connected plane graph. Then*

$$v - e + f = 2.$$

Second proof of Theorem 3.7, using Euler's formula. We begin by partitioning P into elementary triangles, which is possible by Lemma 3.8 (again, see Figure 2b).

We then interpret the triangulation as a connected plane graph, where nodes in the graph are vertices of the triangles in the triangulation, and edges in the graph are edges of the triangles in the triangulation. This graph subdivides the plane into f faces, one of which is the unbounded face (the area outside the polygon), and the remaining $f - 1$ faces are the triangles inside the polygon. By Lemma 3.9, each triangle has area $\frac{1}{2}$, and thus $A = \frac{1}{2}(f - 1)$. (This of course proves nothing; it is a simple consequence of how we defined f.)

An *interior* edge borders on two triangles, and a *boundary* edge borders on a single triangle and forms part of the boundary of the polygon itself. Let e_{int} be the number of interior edges, and e_{bd} be the number of boundary edges. Counting the number of edges in two different ways, we get

$$3(f - 1) = 2e_{int} + e_{bd}. \tag{1}$$

Note that we are overcounting here, that is, counting each interior edge twice and each bounded face three times. In other words, the left-hand side counts the edges using the fact that each triangle (bounded face) has 3 edges, giving the effect that each interior edge has been counted twice, and each boundary edge has been counted once, which exactly amounts to the quantity on the right-hand side.

We can also observe that the number of boundary edges is the same as the number of boundary vertices, $B = e_{bd}$, and that the number of nodes in the graph is the sum of all the interior and boundary points, $v = I + B$. Euler's formula for the graph at hand states that

$$(I + B) - e + f = 2 \quad \text{or} \quad e - f = (I + B) - 2,$$

where $e = e_{int} + e_{bd}$ is the total number of edges. We aim to use this to express $f - 1$ in terms of I and B. With some algebraic rearrangements and suitable substitutions, starting with (1), we get

$$
\begin{aligned}
f - 1 &= -2f + 2 + 2e_{int} + e_{bd} \\
&= -2f + 2 + 2e - e_{bd} \\
&= 2(e - f) - e_{bd} + 2 \\
&= 2(I + B - 2) - B + 2 \\
&= 2I + B - 2,
\end{aligned}
$$

and consequently $A = \frac{1}{2}(f - 1) = \frac{1}{2}(2I + B - 2) = I + \frac{1}{2}B - 1$. □

To what extent does this proof exhibit the three different kinds of fit?

D_1:*Coherence.* The proof is not coherent. Faces and edges are not mentioned in the theorem.

D_2:*Specificity.* The proof does to some extent fulfill the criterion of specificity. Euler's formula is adequate to prove the theorem, but is decidedly more generally applicable. Arguably, the proof using angle measure makes use of a simpler tool, which still gets the job done, and thus there is a more specific tool available.

P_1:*Level of detail.* The proof has an appropriate level of detail. It gives only information relevant for the level of the theorem, simply and concisely.

P_2:*Transparency.* This proof is not transparent, though it is a little more transparent than the previous proof. Again, it is easy enough to follow the steps, and the dissection into elementary triangles is reasonable. If one has previous knowledge of Euler's theorem, it should at least seem reasonable to try to apply it to the graph and expect something useful to fall out.

F_1:*Generality.* The proof is rather general. The lattice points are inherent in the formulation of the theorem; so the generality of the proof should only measured against similar results on lattice polygons. The exact same proof works for a lattice that results from $\mathbb{Z}^2$ by applying any area-preserving mapping (a shear, for example). The number of lattice points and the number of boundaries between elementary triangles in a triangulation is invariant under such transformations, and it is this fact that allows the proof idea to be applied in general lattices without modification.

F_2:*Connectedness.* This second proof exhibits connectedness. Here, the use of Euler's theorem situates the proof in a natural class of theorems. The feeling is that 'so this works even here'.

While both of these proofs exhibit some aspects of fit, they derive their sense of fit from different sources. The first proof derives its sense of fit by the choice of looking at angle measures. This is a surprising choice, given that angles have nothing to do either with the setup of the problem nor the result. However, this choice leads to the relationship we want. It allows us to match the givens of the situation with the formula we want to establish. The sense of fit feels like finding exactly the right tool (in this case angle measure) which happens to crack open the problem.

The second proof derives its sense of fit by fitting into a family of proofs that are applications of the same theorem (Euler's formula). Unlike angle measure, which seems like a means to an end (we do not tend to group proofs under a heading such as 'those that can use angle measures'), the identity of the second proof as one that uses Euler's formula is enough to place it in a family. Euler's formula is a significant result, and its significance arises in part from how it appears in new and surprising settings.

TABLE 1 *Summary of Proof Evaluations*

		Sqrt1	Sqrt2	Pyth1	Pyth2	CC1	CC2	Pick1	Pick2
D_1	Coherence	X		X		X	X		
D_2	Specificity		X	X	X		X	X	(X)
P_1	Level of detail		X	X	X	X	X	X	X
P_2	Transparency	X		X		X			
F_1	Generality	X		X		(X)	X		X
F_2	Connectedness	X		(X)		(X)	X	(X)	X

4. SUMMARY OF PROOF EVALUATIONS

The Table 1 summarizes the evaluations we have done so far of the different proofs. An X means the proof fulfills the criterion, (X) means it partially fulfills the criterion, and a blank space means the proof does not fulfill the criterion at all.

5. DISCUSSION

For the analysis above to have more general interest beyond a mere exercise of classification, we should point to some ways in which the notion of fit might connect to other properties we might want a proof to have, both cognitive and aesthetic. The comments below are more stubs than full articulations, highlighting a few ways the framework presented in this paper connects to current discussions in the philosophy of mathematics and related fields.

5.1. Relation to Explanation in Mathematics

Two of the most prominent theories in the area of mathematical explanation are due to Steiner and Kitcher. We will very superficially consider the relation between these two theories and our framework on fit. We will also briefly discuss more recent work by Lange.

Steiner [1978] provides an account of mathematical explanation in terms of the *characterizing property*. He described this as 'a property unique to a given entity or structure within a family or domain of such entities or structures'. This description of characterizing property has an obvious parallel with our notion of coherence. Familial membership is central, both in identifying an entity as one that could explain, as well as in finding the grounds for the explanation. As Steiner continues, 'an explanatory proof makes reference to a characterizing property of an entity or structure mentioned in the theorem, such that from the proof it is evident that the result depends on the property'. Similar, but not identical, to our notion of coherence, the relationship between the entity or structure and the result is central for determining whether a proof explains (or has fit). A proof that has the same terms as a theorem, which is how we have characterized coherence, seems similar to a proof that evidently gives rise to a particular result.

Steiner's notion of explanation involves a process he calls *deformation*. A proof that explains can be modified for members of a particular family (*e.g.*, the set of all polygons) while keeping the proof idea the same. An explanatory proof can be deformed to produce a new theorem. While problematic,[10] the idea behind deformation, that an explanatory proof contains an idea that is invariant to certain intra-family sorts of transformations, is not completely counter-intuitive. The set-up of a proof is what gives it its basic character, an intuition that is behind both what makes a proof a member of a family in the case of familial fit, and what brings about the coherence in the case of direct fit.

Kitcher [1981] offers a view of explanation that is considered to be a counter-proposal to Steiner's view, based on the notion of unification. Kitcher says that explanation arises from the use of arguments that have the same form.[11] These explanations can be found in what Kitcher calls *the explanatory store* and the main task of a theory of explanation is to 'specify conditions on the explanatory store' [1981, p. 80]. While the details of what gives rise to an explanation differ greatly in Kitcher's and Steiner's accounts, one similarity seems to be the emphasis on familial membership, or what we would call connectedness. In Steiner's account the membership comes about via characterizing properties, and in Kitcher's account it comes about via the explanatory store. The fact that there is some kind of unification or some sort of family traits that naturally carry over to similar entities or structures seems central both in these two accounts of mathematical explanation and in our account of mathematical fit.

In contrast to Steiner and Kitcher, whose views of explanation seem to have some component similar to that of connectedness, Lange [2014] suggests a view that, at least in part, relates to our notion of coherence. Lange's account of mathematical explanation has three components: unity, salience, and symmetry. While salience and symmetry might have some counterparts in our framework that are a bit harder to see, the relation between unity and coherence seems fairly straightforward. To Lange, 'A proof is unified when it exploits a property that all of the cases covered by the theorem have in common and treats all of those cases in the same way' (personal communication). Unlike Kitcher and Steiner, whose unification and characterizing property ideas involve family membership, Lange's notion of unity, similar to our notion of coherence, is one that is directly related to the proof. Lange's concept of salience might also overlap with our concepts of coherence and/or transparency. Salience is a feature that is 'worthy of attention' [Lange, 2014, p. 27]. Coherence is what could warrant us to focus our attention, while transparency allows us to access the underlying ideas.

5.2. Relation to Beauty

Less clear than the relation between fit and explanation is the relation between fit and beauty (in part because it is difficult to nail down exactly what beauty

---

[10] See [Hafner and Mancosu, 2005].
[11] See [Lange, 2014] for a summary of this view.

is). The criteria in this paper that we guess are most likely to connect to aesthetic properties such as beauty are those of level of detail, transparency, and connectedness. Level of detail is related to brevity, a quality often suggested as a feature of beauty. We know that mathematicians, even after finding a correct proof, will work hard to find one that is shorter and more concise. This drive, while having some cognitive component (a shorter proof might be easier to understand), seems certainly to be aesthetic (hence the terms 'elegant' and 'beautiful'). Simple proofs are nice.

Transparency, which deals with the structure of the proof, may relate to beauty via the idea that it helps make a proof graspable. This intuition is similar to Rota's suggestion [1997] that a beautiful proof is enlightening. Moreover the foregrounding and backgrounding of information, which is easier to do with a transparent proof than one which is not transparent, might be precisely what renders a proof 'salient', to use a term in Lange's [2015] terms. Proofs which are transparent are those in which the key ideas are salient. This salience may in turn give rise to aesthetic features that make a proof attractive.

The feature of connectedness could also be aesthetic. A connection allows you to see a proof or theorem in a new way, as in the second proof of Pick's theorem which allowed us to see how Euler's theorem could be used in a new unexpected setting. The number of connections a result has tends to be some measure of how deep a theorem is (see [Lange, 2015] and [Stillwell, 2015]).

Finally, with a little effort one might also see even specificity and generality as aesthetic. Family memberships are a kind of grouping which simplify complex relations (the feeling of 'All I need to remember is that this is one of those!'). This kind of simplification, in turn, might increase mental processing speed, a factor that has recently been found to play a central role in beauty judgements.[12]

6. FINAL COMMENTS

There are many attributes a proof might have. It might be elegant, clumsy, enlightening, explanatory, deep, simple, and so on. Some of these properties are more cognitive, relating to how we understand a proof. Others are more aesthetic, relating to how we experience the proof, perhaps similar to how one might experience a piece of music or a work of art. In this paper we have chosen to explore the notion of fit, in part because it could relate both to the cognitive and the aesthetic aspects of proof, and in part because it seems more tractable than the more explored notions of mathematical explanation and beauty.

We hope we have shown via the examples above that even a fairly simple framework does real work to distinguish proofs and to specify the extent to which they possess different aspects of fit. The intuitive ideas of direct fit,

---

[12] See [Mason and Hanna, 2014] for a discussion of the relation between transparency and Gowers's [2007] concept of the 'width' of a proof. The width of a proof, which is connected to the number of ideas it contains, may be related to memorability, which in turn could be related to aesthetic judgements (some data suggest that shortness of processing time correlates with positive aesthetic judgement [Reber *et al.*, 2004]).

which describes the relation between a proof and its theorem, presentational fit, which describes the relationship between a proof and a reader, and familial fit, which describes the relation between a proof and the rest of mathematics, capture several central aspects of the feeling that a proof is 'right', or somehow does the job better than another.

While we have attempted to be careful and systematic in our analysis, we have tried wherever possible to choose criteria that are fairly basic, which capture as simply as possible the reasons why a particular proof might appear to fit. It is our hope, perhaps precisely because these criteria are basic, that we might find them lurking near other topics such as mathematical explanation and beauty, which we would also like to understand better.

APPENDIX

In this appendix, we give further technical details on aspects of the two proofs of the Pythagorean theorem given above.

First Proof of the Pythagorean Theorem

The missing algebra, establishing that it is indeed the equation $|a|^2 + |b|^2 = |c|^2$ that follows from the scaling considerations, can be presented in the following manner.

The linear scaling factor from T_1 to T_2 is $|b|/|a|$, from T_2 to T_0 is $|c|/|b|$, and so on. If we let S_i be the area of T_i, for $i = 0, 1, 2$, it follows that $S_0 = S_1 + S_2 = (|a|/|c|)^2 S_0 + (|b|/|c|)^2 S_0$, from which we get $|c|^2 = |a|^2 + |b|^2$ by cancelling $S_0 \neq 0$ and multiplying through by $|c|^2$. This calculation establishes the claim.

One can also derive the conclusion from the fact that the area of each triangle is a constant fraction of the corresponding square, say $T_0 = r|c|^2$, $T_1 = r|a|^2$, and $T_2 = r|b|^2$. The algebra to reach the conclusion is trivial: $r|c|^2 = r|a|^2 + r|b|^2 \iff |c|^2 = |a|^2 + |b|^2$.

Second Proof of the Pythagorean Theorem

Note that the subtraction formulas for sine and cosine are perhaps most commonly proved using the notion of distance, and hence indirectly the Pythagorean theorem itself. It might seem, therefore, that the trigonometric proof of this theorem, presented above, is circular. However, when restricting the angles to $0 < \alpha, \beta < 90°$ these identities may be proven entirely without recourse to the general notion of distance between two points, and hence the second proof is not circular. One appealing way of proving these formulas is indicated in Figure 4, where the triangle ABC has a right angle at B.

We see that the angles BAD and ABF are α. We can express the length of side AD in two ways, as $|AD| = |AE| + |ED| = |FB| + |GC|$. If we consider the triangle BGC, we may note that $|GC| = \sin\alpha\sin\beta$, and similarly, considering the triangle ABF, we note that $|FB| = \cos\alpha\cos\beta$, from which it follows that

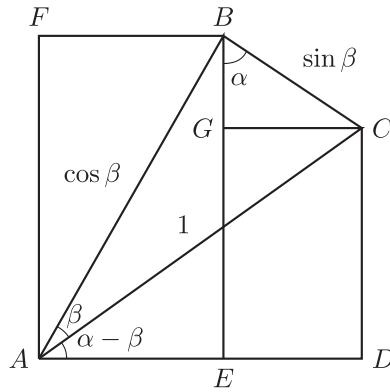$$\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta.$$

Fig. 4. Deriving the subtraction formulas

A similar argument for the lengths of FA, BG, and CD yields the subtraction formula for $\sin(\alpha - \beta)$.

REFERENCES

Agargün, A.G., and E.M. Özkan [2001]: 'A historical survey of the fundamental theorem of arithmetic', *Historia Mathematica* **28**, 207–214.

Aigner, M., and G. Ziegler [2010]: *Proofs from the Book*. 4th ed. Springer-Verlag.

Balacheff, N. [1988]: 'Aspects of proof in pupils' practice of school mathematics', in D. Pimm, ed., *Mathematics, Teachers, and Children*, pp. 216–230. London: Hodder and Stoughton.

Beardsley, M. [1981]: *Aesthetics: Problems in the Philosophy of Criticism*. Indianapolis, Indiana: Hackett Publishing Company.

d'Alembert, J.R. [1746]: 'Recherches sur le calcul intégral', *Histoire de l'Academie Royale des Sciences et Belles Lettres de Berlin* **2**, 182–224. Printed 1748.

Detlefsen, M., and A. Arana [2011]: 'Purity of methods', *Philosopher's Imprint* **11**, 1–20.

Gowers, W.T. [2007]: 'Mathematics, memory and mental arithmetic', in M. Leng, A. Paseau, and M. Potter, eds., *Mathematical Knowledge*, pp. 33–58. Oxford University Press.

Hafner, J., and P. Mancosu [2005]: 'The varieties of mathematical explanation', in P. Mancosu *et al.*, eds, *Visualization, Explanation and Reasoning Styles in Mathematics*, pp. 215–250. Springer.

Hilbert, D. [1899]: *Grundlagen der Geometrie*. Leipzig: Teubner.

Kitcher, P. [1981]: 'Explanatory unification', *Philosophy of Science* **48**, 507–531.

Lange, M. [2014]: 'Aspects of mathematical explanation: Symmetry, unity, and salience', *Philosophical Review* **123**, 485–531.

———— [2015]: 'Depth and explanation in mathematics', *Philosophia Mathematica* (3) **23**, 196–214.

Mason, J., and G. Hanna [2014]: 'Key ideas and memorability in proof', *For the Learning of Mathematics* **34**, No. 2, 12–16.

Pick, Georg [1899]: 'Geometrisches zur Zahlenlehre', *Sitzungsberichte des deutschen naturwissenschaftlich-medicinischen Vereines für Böhmen 'Lotos' in Prag'*, (Neue Folge) **19**, 311–319.

Raman, M., and L.-D. Öhman [2011]: 'Two beautiful proofs of Pick's Theorem', in M. Pytlak, T. Rowland, and E. Swoboda, eds, *Proceedings of Seventh Congress of the European Society for Research in Mathematics Education*, pp. 223–232. Rzeszów, Poland: University of Rzeszów for ESRME.

Reber, R., N. Schwarz, and P. Winkielman [2004]: 'Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?' *Personality and Social Psychology Review* **8**, 364–382.

Rota, G.-C. [1997]: 'Phenomenology of mathematical beauty', *Synthese* **111**, 171–182.

Sinclair, N. [2002]: 'The kissing triangles: The aesthetics of mathematical discovery', *International Journal of Computers for Mathematical Learning* **7**, 45–63.

—— [2004]: 'The roles of the aesthetic in mathematical inquiry', *Mathematical Thinking and Learning* **6**, 261–284.

Steiner, M. [1978]: 'Mathematical explanation', *Philosophical Studies* **34**, 135–151.

Stillwell, J. [2015]: 'What does "depth" mean in mathematics?', *Philosophia Mathematica* (3) **23**, 215–232.

Velleman, J.T. [1999]: 'Love as a moral emotion', *Ethics* **109**, 338–374.

Wechsler, J. [1978]: *On Aesthetics in Science*. Cambridge, Mass.: MIT Press.

Zimba, J. [2009]: 'On the possibility of trigonometric proofs of the Pythagorean Theorem, *Forum Geometricorum* **9**, 1–4.