Beauty as fit: An empirical study of mathematical proofs

Manya Raman
Umeå University

Beauty has been discussed since ancient times, but discussions of beauty within mathematics education are relatively limited. This lack of discussion is surprising given the importance of beauty within the practice of mathematics. This study explores one particular metaphor of beauty, that of beauty as fit, as a way to distinguish between proofs that are considered beautiful and those that are not. Several examples are examined, supported by empirical data of mathematicians and mathematics educators who judged and ranked different proofs in a seminar on mathematical beauty.

Keywords: beauty, fit, proof, mathematician, mathematics

Introduction

The idea of beauty as fit is an ancient one. It was touted by the Stoics, who defined beauty as "that which has fit proportion and alluring color." (Cicero, as quoted in Tatarkiewicz 1972) and the Pythagoreans who claimed, "order and proportion are beautiful and fitting" (Aristotle, as quoted in Tatarkiewicz 1972). The metaphor persists to modern times. Beardsley described one essential characteristic of aesthetic experience to be "a feeling that things are working or have worked themselves out fittingly" (Beardsley 1982).

This metaphor of beauty as fit can be found not only in the arts, but also in mathematics, as Hardy famously asserted, "The mathematician's patterns, like the painter's or the poet's must be beautiful; the ideas like the colours or the words must fit together in a harmonious way" (Hardy 1967). And Sinclair (2002) discusses some of the ways that her sense of fit guided her in the process of discovering a proof of Napoleon's theorem. *That* fit can be used productively as a metaphor seems clear, but we still know little about what fit means in mathematics, whether it has different connotations in different contexts, and why the notion of fit might have anything to do with beauty, or aesthetic preference more generally.

The focus on aesthetics in mathematics education is not new. In the 1970s Papert (1978) suggested that mathematical thinking consists of three processes: cognitive, affective, and aesthetic. At the time of his writing, only the first of these was a serious area of research. Now, there has been substantial process made also on the second. Aesthetics remains under-researched, despite an attempt in the 1980s to jump-start the field (Dreyfus and Eisenberg 1986). In recent years there has been a bit of activity in the field, mostly due to Nathalie Sinclair. One purpose of this paper, part of a larger project conducted in cooperation with Lars-Daniel Öhman, a mathematician at Umeå University, is to try to contribute to the momentum generated from her work.

One of Sinclair's main contributions (see e.g. Sinclair 2004) has been to shift the focus away from judgements of mathematical objects (such as proofs) towards a more holistic account of the role that aesthetics can play mathematical practice, three of which she identifies as generative, motivational, and evaluative. While this is an important shift, the paper here concerns only the evaluative aspect, or in particular what makes a particular proof beautiful. The reason for this focus is that it seems difficult to move forward with any serious research program on aesthetics without nailing down what is meant by mathematical beauty in the first place.

This brings us back to the metaphor of beauty as fit. Fit is by no means the only metaphor used for beauty. Others include unity, perfection, moderation, and metaphor itself (see Tatarkiewicz 1972 for an excellent account of the history of these and other accounts of beauty). Yet, the metaphor seems productive, in ways we will discuss below, and working mathematicians refer to it when making judgements about the beauty of mathematical proofs. In other words the metaphor is persistent enough that it seems important to try to understand exactly what it means.

The goal of this paper is to try to clarify at least some of the roles of fit in the context of mathematical proof. We will discuss below, though two examples of proofs: (i) what does it mean for a proof to fit a theorem, and (ii) what different types of fit a beautiful proof might possess. The analysis of proofs is supported by data collected in a year-long seminar on mathematical beauty, attended by mathematicians and mathematics educators who provided their own subjective judgements about the aesthetic values of the different proofs.

Examples

Pythagorean theorem

> *Let c be the length of the hypotenuse of a right triangle T_0, and let a, b be the lengths of the remaining 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, or symbolically $a^2 + b^2 = c^2$.*

The first example we will consider is the Pythagorean Theorem. This is a familiar theorem, for which most mathematicians will know many different proofs, and most likely have a favourite. Below we present one proof, from Euclid VI. 31, that is fairly familiar and among those that the mathematicians in our seminar preferred, and a second proof which might be new for many people and which proved to be less popular. We begin with the proof from Euclid.

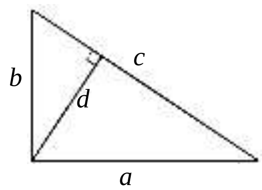

Figure 1

> *First proof.* Consider Figure 1. The line d is perpendicular to c, and intersects the vertex of the triangle. Let T_1 be the right triangle with hypotenuse a and side d, and let T_2 be the right triangle with hypotenuse b and side d. Clearly, by the principle of conservation of area, the sum of the areas of T_1 and T_2 equals that of T_0. We can, of course, consider these three triangles as being constructed on either side of the original triangle. Also, by standard congruences (two shared angles), all the triangles T_0, T_1 and T_2 are congruent. Scaling a figure F in the plane by a linear factor k changes the area of F by a factor k^2. Therefore, if the theorem holds for any set of congruent plane figures constructed on either side of the original triangle, it holds for all such sets of congruent plane figures. As observed above,

> the theorem holds for congruent right-angled triangles, and therefore holds for any set of congruent figures, in particular, squares.

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 and multiplying through by c^2.

> Second proof. Suppose we have the subtraction formulas for sine and cosine:
>
> (1) $\cos(\alpha - \beta) = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)$
>
> (2) $\sin(\alpha - \beta) = \sin(\alpha)\cos(\beta) - \cos(\alpha)\sin(\beta)$.
>
> Suppose that α is the angle opposite to side a, and β is the side opposite to side b, and without loss of generality that $0 < \alpha \leq \beta < 90^\circ$. We now have $\cos(\beta) = \cos(\alpha - (\beta - \alpha)) = \cos(\alpha)\cos(\beta - \alpha) + \sin(\alpha)\sin(\beta - \alpha) = \cos(\alpha)(\cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)) + \sin(\alpha)(\sin(\alpha)\cos(\beta) - \cos(\alpha)\sin(\beta)) = (\cos^2(\alpha) + \sin^2(\alpha))\cos(\beta)$, 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.

In our seminar, these two proofs were presented along with six other proofs of the theorem. Members of the seminar, both mathematicians and mathematics educators were asked to rank the proofs from those they liked best to least, and to write a word to describe what they thought of the proofs (e.g. beautiful, nice, slick). The reason for posing the task as such was to separate the issues of preference and beauty: one might like a proof for other reasons than its aesthetic appeal, and there are words similar to beauty (like elegance) that might have a distinctly different connotation.

The pilot data show that proof 1 above was preferred to proof 2 for all the mathematicians and one of the mathematics educators. The words used to describe the first proof included, “simple”, “beautiful”, and “conceptually correct”, while the words used to describe the second proof included “ugly”, “clever”, and “unnatural”. One of the mathematics educators also preferred proof 1, but the reasons given by the other two for preferring proof 2 was that it was easier for them to follow, having just seen the area argument for the first time and not grasping it entirely.

The point of this first example is to distinguish a proof that our mathematicians agreed fits a theorem (the first proof) from the one that does not (the second). The mathematicians suggested that the reason the first one fits is that it gets directly to what the Pythagorean theorem is about. With a very simple algebraic calculation one can check that the sum of the similar squares behaves the same way as the sum of the similar triangles, which conveniently both lie on the three sides of the triangles and also make up the interior (so one can easily see that the sum of the first two is the same as the second.) The proof is both economical – it doesn’t involve outside information, as the second proof using trigonometry does – and it is transparent– once you see the idea of the proof you immediately see why the conclusion of the theorem follows. In contrast, the second proof, while neat or perhaps new to the reader, involves extraneous information: the Pythagorean relationship falls out of the calculation, but one does not have a sense of why the theorem holds. The proof appears like a trick, a set of algebraic manipulations which give you the result while keeping you in the dark. This proof, while having some aesthetic merits, was not considered by our mathematicians to be beautiful.

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 $A = I + B/2 - 1$.*

The second example we will consider is Pick's theorem, which 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. 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. We will assume two facts as lemmas, first that it is always possible to triangulate a polygon (see Figure 2 as an example), and the second that each of the elementary triangles has area ½. There are many proofs of this theorem, but the one below is considered to be among the most beautiful (see Raman and Öhman (2011) for another beautiful proof).

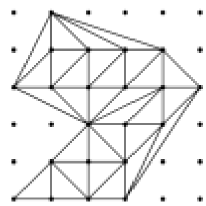

Figure 2: A triangulated lattice polygon

> Proof sketch. For space reasons we sketch the proof below and refer to Aigner and Ziegler (2009) for details. The idea of this proof is to conceive of the triangulated lattice polygon as a polyhedron, with each triangle as a face, and the outer area (outside of the boundary of the polygon) as a face. We can count the number of edges in two different ways: $3N = 2e_{\text{int}} + e_{\text{bd.}}$ where N is the number of triangles, e_{int} is the number of interior edges, and e_{bd} is the number of boundary edges. Note that we are overcounting the edges on both sides, but by the same amount, namely the number of edges that are shared by neighbouring triangles.
>
> Next, we apply Euler's formula, $V + F - E = 2$, where V = number of vertices, F = number of faces, and E = number of edges. For our polyhedron, $V = I + B$, $F = N + 1$, and $E = e_{\text{int}} + e_{\text{bd.}}$ Using substitution and algebra, one can now arrive at the formula $A = I + B/2 - 1$.

The point of this proof is to show that proofs can "fit" in at least two different ways. Proof 1 of the Pythagorean theorem, above, fits the theorem in a way which we will label as "internal fit", meaning that the proof directly illuminates what the theorem is about, providing a sense of why the theorem is true. The proof of Pick's theorem above fits in a way we will label as "external fit", meaning that the proof derives its beauty from the way it is connected to a family of other theorems, in this case the theorems that can be proven using Euler's theorem (and this particular case is a surprising application of that theorem.) This kind of fit does not convey meaning or explanation, *per se*, but links the theorem, through the proof, to a class of theorems not previously thought to be related.

Discussion

We now return to the question of fit: what does it mean for a particular proof to fit a particular theorem? Our analysis above indicates that there are at least two types of fit— internal fit and external fit— that can both potentially give rise to the sensation of beauty. We are far from proving that the relationship always holds, that is to say that fit and beauty are always coupled. In fact, looking at the empirical data we see that some people prefer proofs that we claim do not have fit. The reason for this mismatch might arise from the fact that judgements of mathematical beauty must be linked to understanding. If one does not understand a particular proof, one cannot judge it as beautiful. So it is difficult to say exactly how fit and beauty relate, except to say there seems to be some correlation among people with a particular level of understanding.

Another potential lesson from this short exploration is that we should distinguish between (1) a proof having a particular sort of fit to a theorem; and (2) whether a particular person can see the fit. The first feature could be objective while the second one is subjective. These two features are often confused, giving rise to the knee jerk “beauty is in the eye of the beholder” type attitude. Making a distinction between whether a proof is beautiful and whether a person can grasp that beauty can help explain phenomena such as why mathematicians judge different proofs to be beautiful, or why mathematicians and non-mathematicians do the same, without drawing a necessary conclusion that mathematical beauty is subjective. Moreover, the metaphor of ‘fit’ suggests a more objective view of beauty might be warranted— whether a proof is appreciated as beautiful is a subjective claim, but whether a proof fits a theorem, which relies more on the nature of the proof than our perception of it, is a more objective one.

References

Aigner, M., and G. Ziegler. 2009. *Proofs from THE BOOK* (4th ed.). Berlin, New York: Springer-Verlag.

Beardsley, M. C. 1982. *The aesthetic point of view*. Selected essays. Ithaca: Cornell University Press.

Dreyfus, T., and T. Eisenberg. 1986. On the aesthetics of mathematical thought. *For the Learning of Mathematics*, 6(1): 2-10.

Hardy, G. H. 1967/1999. *A mathematician's apology*. New York: Cambridge University Press.

Papert, S. 1978. The mathematical unconscious. In *On aesthetics and science*, ed. J. Wechsler, 105-120. Boston: Birkhäuser.

Raman, M., and L.-D. Öhman. 2011. Two beautiful proofs of Pick's Theorem. *Proceedings of Seventh Conference of European Research in Mathematics Education*. Rzeszow, Poland. Feb. 9-13.

Raman, M., and L.-D. Öhman. (In prep). Beauty as fit: A mathematical case-study.

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

Sinclair, N. 2004. The roles of the aesthetic in mathematical inquiry. *Mathematical Thinking and Learning*, 6(3): 261-284.

Tatarkiewicz, W. 1972. The great theory of beauty and its decline. The *Journal of Aesthetics and Art Criticism*, 31(2): 165-18.