The Universal Aesthetics of Mathematics

SAMUEL G. B. JOHNSON AND STEFAN STEINERBERGER

Can a proof be objectively beautiful? It is not a surprising claim that the search for beauty, both in theorems and in proofs, is one of the great pleasures of engaging with mathematics.

Quite often the similarity to beauty in the visual arts or music is made explicit:

> The mathematician's patterns, like those of the painter's or the poet's, must be beautiful; the ideas, like the colors or the words, must fit together in a harmonious way (G. H. Hardy [2]).

> Why are numbers beautiful? It's like asking why is Beethoven's Ninth Symphony beautiful. If you don't see why, someone can't tell you. I know numbers are beautiful. If they aren't beautiful, nothing is (Paul Erdős [3]).

> A scientist worthy of the name, above all a mathematician, experiences in his work the same impression as an artist; his pleasure is as great and of the same nature (H. Poincaré [8]).

Theorems can be "deep," "profound," "surprising," or "derivative" and "boring"; conjectures can be "daring," "bold," "natural," and sometimes "false for trivial reasons" [1]. Proofs can be "beautiful," "unexpected," "clean," "technical," "elementary," "lovely," "nifty," "hand-wavy," or even "impudent" (Littlewood's description [6] of Thorin's proof of the Riesz–Thorin theorem). While mathematical tastes are diverse, those working within the same mathematical area tend to have some consensus as to whether a theorem, proof, or conjecture is beautiful (or, say, surprising).

But where do these intuitions come from? Are they the product of mathematical socialization or something deeper about how human beings universally perceive mathematical beauty? Is Thorin's proof *truly* impudent, or does one learn to call it that as part of one's education? This paper describes a psychological experiment designed to give a deeper understanding of the issue; we hope that many more such experiments will follow.

Setting Up the Experiment

There are two main challenges to any such investigation:

1. finding a way to formulate this effect, if it exists, in such a way that it is quantitatively measurable,
2. and ensuring that the effect is authentic and not an artifact of mathematical socialization (something that one is "taught to pretend" during one's education).

The second point alone already rules out a great number of obvious approaches (for example, having students give descriptions of the character of mathematical arguments). We chose to use a comparative approach: participants in our study were shown four mathematical arguments (given below) and then either asked to observe four paintings or listen to four pieces of music; on a scale of 0 to 10, we asked them to rate the similarity between the piece of mathematical reasoning and the work of art. We will now give a more in-depth discussion of the experiment and then describe its results.

© 2018 Springer Science+Business Media, LLC, part of Springer Nature,
https://doi.org/10.1007/s00283-018-09857-5

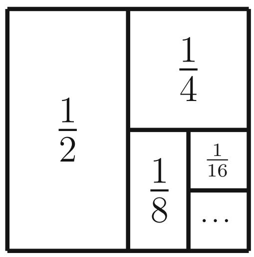
Beautiful Reasoning

We selected four classical pieces of elementary mathematical reasoning based on their beautiful or surprising character and their immediate accessibility to those without mathematical training. Most readers have likely encountered these arguments before:

1. Geometric series.

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}+\cdots=1.$$

We can see this by cutting a square with total area 1 into little pieces.



2. Gauss's summation trick. A quick way to compute

$$1+2+3+4+\cdots+98+99+100=5050$$

is as follows. Write the total sum twice and add the columns:

$$\begin{array}{rrrrrrrr} 1\,+ & 2\,+ & 3\,+ & 4+\cdots+ & 98\,+ & 99\,+ & 100 \\ 100\,+ & 99\,+ & 98\,+ & 97+\cdots+ & 3\,+ & 2\,+ & 1 \\ \hline 101+ & 101+ & 101+ & 101+\cdots+ & 101+ & 101+ & 101 \end{array}$$

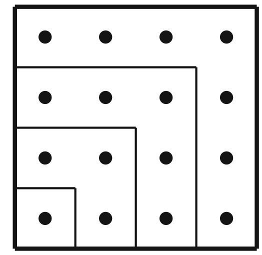
This yields a total of 100 times 101 (giving 10 100), and half of that is exactly 5050.

3. The pigeonhole principle. In any group of five people, there are two who have the same number of friends within the group. We can see this as follows: suppose there exists somebody who is friends with everybody else. Then every person in the group has either one, two, three, or four friends (because everybody has at least one friend). Since there are five people but only four possible numbers of friends, one number has to appear twice. If nobody is friends with all other people, then everybody has either zero, one, two, or three friends; again, since there are five people, one number has to appear twice.

4. Faulhaber's formula. The sum of consecutive odd numbers always adds up to a square number:

$$1=1^2$$
$$1+3=2^2$$
$$1+3+5=3^2$$
$$1+3+5+7=4^2$$

The reason is explained in the picture below: adding the next odd number creates a suitable layer for the next square.



Beautiful Art

We selected superficially similar pieces of art within each category: a "classical" piece for solo piano for music, and nineteenth-century landscape paintings for art. The paintings are displayed throughout the paper; for music, we used the first 20 seconds of the following four works:

1. F. Schubert: *Moment Musical* no. 4, D. 780 (op. 94), played by D. Fray
2. J. S. Bach: Fugue from *Toccata in E Minor* (BWV 914), played by G. Gould
3. A. Diabelli, *Waltz* (the theme of Beethoven's *Diabelli Variations*, op. 120), played by G. Sokolov
4. D. Shostakovich, *Prelude in D-flat Major* (op. 87, no. 15), played by A. Brendel

Though all four pieces are written for solo piano, they have very different characters. Schubert's piece, among the most frequently performed of his works, is often formally compared to Bach, but it is more romantic in style; the Bach fugue, rhythmic and fast-paced, has an "urgent" feel to it and is often grouped with those of Bach's works that show an Italian influence. Diabelli's waltz, famously dismissed by Beethoven as a *Schusterfleck* (literally a cobbler's patch, a disparaging term for a piece of music "cobbled together" by repeating a melody identically one step higher), is a simple classical waltz (William Kinderman [5] speaks of "the banality of the theme," which is "trite" and "insufferably so when repeated"). Finally, Shostakovich's prelude, part of his cycle of 24 preludes and fugues in all keys—inspired, on the 200th anniversary of Bach's death, by that composer's two books of the *Well-Tempered Clavier*—is described by Mark Mazullo [7] as "blatantly insincere," an "artistic non-entity" that is "cracking jokes."

All of the paintings feature realistic romantic landscapes. We invite the reader to repeat the experiment: given these four pieces of music and these four paintings, in which order do they best capture the spirit of the four mathematical arguments? In an informal survey of colleagues, most reported feeling slightly ill at ease and that they considered the question somewhat ill posed (some whom we polled used stronger language). Many of our surveyed colleagues also remarked that matching paintings to math seemed easier than matching music, while others took the

Figure 1. Albert Bierstadt: *Looking Down Yosemite Valley, California* (1865).

opposite point of view (including the second author, who would argue that the Diabelli waltz describes the geometric series perfectly but feels less confident in assigning a painting to it; incidentally, he is also terrible at drawing and has a hard time appreciating the visual arts).

Outcome of the Experiment: Music

Participants were recruited from the online crowdsourcing platform Amazon Mechanical Turk ($N = 299$) and were from the United States. Of these, 90 had taken university-level math courses above calculus. Participants read the four mathematical arguments, were then asked to reflect on the argument and to rate the similarity to the four 20-second clips of music described above on a scale from 0 (not at all similar) to 10 (very similar). The arguments as well as the musical clips were presented in a fully randomized order and on separate pages. After the main task, a memory check question determined whether the participant had paid sufficient attention to the material, and 73 participants were dismissed at that stage. This online sample was complemented by 28 undergraduates and four professional mathematicians. Indeed, most professional mathematicians we asked were slightly puzzled by the experiment; the authors' sanity was questioned more than once.

The results turned out to be far from random: taking all pairwise correlations between the $N = 219$ participants of the MTurk sample who gave nonidentical ratings across the items, correlation with the mean (of all samples but the selected participant) was significantly more often positive than it was negative (145 out of 219, the p-value is $p < 0.001$). This was true for both participants who had taken higher mathematics classes ($p = 0.01$) and those who had not ($p < 0.001$). More drastically, when asked the rather unusual question whether a geometric series sounds more like Bach or Schubert, our subjects responded in a way that was very clearly not random (highly statistically significant, if we want to sound scientific).

Table 1. Amazon MTurk results (N = 226): bold denotes largest or close-to-largest degree of similarity

	Schubert	Bach	Diabelli	Shostakovich
Geometric series	**4.76**	4.39	4.36	4.62
Gauss's summation trick	4.61	**5.11**	4.67	4.31
Pigeonhole principle	4.52	4.42	**4.89**	**4.83**
Faulhaber's formula	4.32	4.59	**5.04**	**5.06**

Outcome of the Experiment: Paintings

We recruited another $N = 300$ participants from Amazon Mechanical Turk (of which 99 had taken higher math classes and 201 had not). Of these, 67 participants were dismissed after failing the memory check question, and one was eliminated due to incomplete ratings. This was supplemented by eight professional mathematicians (a minor byproduct of this study is the insight that mathematicians do not like to fill out psychological surveys). While the connection between mathematics and music is often discussed, the same is not true for paintings, and this clearly shows in the similarity ratings, which were overall lower than for music: apparently, paintings are not as good at capturing the spirit of a proof as music.

However, the association between different mathematical arguments and different works of art is very consistent, indeed, even slightly stronger than for music. The correlation between a participant and the general mean (of all samples except the selected participant) is positive more often than negative (156 out of 211, the p-value is $p < 0.001$). Cronbach's alpha, a statistical test for the reliability of a psychometric test, indicates strongly consistent responses across participants ($\alpha = 0.93$). A much more

Figure 2. Albert Bierstadt: *Storm in the Rocky Mountains* (1866).

Figure 3. John Constable: *The Hay Wain* (1821).

Table 2. **Amazon MTurk results (N = 211): bold denotes largest or close-to-largest degree of similarity**

	Yosemite	Rockies	Hay Wain	Andes
Geometric series	**3.51**	2.99	3.30	3.05
Gauss's summation trick	**2.38**	2.23	**2.43**	1.96
Pigeonhole principle	**2.42**	2.21	2.25	**2.49**
Faulhaber's formula	2.97	2.75	**3.21**	2.44

extensive and complete discussion of the statistical aspects is beyond the scope of this short report and can be found in our paper [4].

What Does This Mean? What's Next?

We have shown that people, both with mathematical education and without, have the ability to recognize aesthetic aspects of mathematical arguments and that these seem to be universal. The second author was once told never to write a paper without a theorem, so here goes:

Theorem. *Proofs have a soul.*

Proof. Empirical. For a nonempirical proof, consult your local Platonist. □

On a more serious note, the results are highly statistically significant and raise quite a large number of interesting questions (which we hope will be answered by many more such experiments in the future).

- How does the effect depend on the type of art? Music seems better at describing proofs than paintings, but the effect is slightly more consistent for paintings; what about sculptures, poems, or possibly even jokes and puns?
- How does the effect depend on the age of the participant? How does it tie in with standard models about the development of capacity for rational thought in children?
- Although these results clearly show that aesthetic intuitions about mathematics are not a product of mathematical socialization, since the effects are robust among laypeople without mathematical training, they may nonetheless interact with culture in intriguing ways. For example, might intuitions about music–math pairings differ in cultures in which the Western canon plays a less dominant role or in which different tuning systems are more widely used?
- Does the effect depend on the type of mathematical argument? Is an analytic inequality close to jazz? Is a combinatorial counting argument more of a waltz?

Figure 4. Frederic Church: *Heart of the Andes* (1859).

Samuel G. B. Johnson
Division of Marketing, Business and Society
School of Management
University of Bath, Bath
UK
e-mail: sgbjohnson@gmail.com

Stefan Steinerberger
Department of Mathematics
Yale University
New Haven, Connecticut
USA
e-mail: stefan.steinerberger@yale.edu

REFERENCES

[1] A. Grothendieck. Hodge's general conjecture is false for trivial reasons. *Topology* 8 (1969), 299–303.
[2] G. H. Hardy. *A Mathematician's Apology*. Cambridge University Press, 1940.
[3] P. Hoffman. *The Man Who Loved Only Numbers: The Story of Paul Erdős and the Search for Mathematical Truth*. Hyperion Books, 1998.
[4] S. G. B. Johnson and S. Steinerberger, The Aesthetic Psychology of Mathematics. arXiv:1711.08376, 2018.
[5] W. Kinderman. The Evolution and Structure of Beethoven's ''Diabelli'' Variations. *Journal of the American Musicological Society* 35:2 (Summer, 1982), 306–328
[6] J. E. Littlewood. *Littlewood's Miscellany,* edited and with a Foreword by Béla Bollobás. Cambridge University Press, 1986.
[7] M. Mazullo. *Shostakovich's Preludes and Fugues: Contexts, Style, Performance.* Yale University Press, 2010.
[8] H. Poincaré. Notice sur Halphen. *Journal de l'École Polytechnique* 6 (1890), 143.