



# Do Mathematicians Agree about Mathematical Beauty?

Rentuya Sa<sup>1</sup> · Lara Alcock<sup>1</sup> · Matthew Inglis<sup>1</sup> · Fenner Stanley Tanswell<sup>2</sup>

Accepted: 5 December 2022 / Published online: 21 February 2023  
© The Author(s) 2023

## Abstract

Mathematicians often conduct aesthetic judgements to evaluate mathematical objects such as equations or proofs. But is there a consensus about which mathematical objects are beautiful? We used a comparative judgement technique to measure aesthetic intuitions among British mathematicians, Chinese mathematicians, and British mathematics undergraduates, with the aim of assessing whether judgements of mathematical beauty are influenced by cultural differences or levels of expertise. We found aesthetic agreement both within and across these demographic groups. We conclude that judgements of mathematical beauty are not strongly influenced by cultural difference, levels of expertise, and types of mathematical objects. Our findings contrast with recent studies that found mathematicians often disagree with each other about mathematical beauty.

Do mathematicians agree about mathematical beauty? Some traditional accounts assume that they do. Under aesthetic realism, mathematical beauty is conceptualized as existing independently of subjective preferences and social contexts. Realist accounts hold that beauty is not reducible to ‘reports of experience in the mind of the observer’ (Simoniti 2017, p.1436). They assume normativity of aesthetic intuition, in the sense that judgements of beauty are either correct or incorrect (Cova and Pain 2012), and that some level of expertise is necessary to make such judgements. Mathematicians who endorse realist accounts, such as Hardy and Erdős, therefore tend to assume aesthetic agreement among mathematicians.

These accounts are, however, open to challenge. Experimental philosophers seeking to empirically investigate assumptions about philosophical intuitions (Knobe 2007) have found widespread aesthetic disagreement within mathematical practice (Wells 1990; Inglis and Aberdein 2015, 2016, 2020). This suggests an alternative position of aesthetic non-realism, according to which mathematical beauty is not an

---

✉ Rentuya Sa  
R.Sa@lboro.ac.uk

<sup>1</sup> Department of Mathematics Education, Loughborough University, Schofield Building, University Road, Loughborough LE11 3TU, UK

<sup>2</sup> Centre for Logic and Philosophy of Science, Vrije Universiteit Brussel, Brussels, Belgium

objective property, but where individual aesthetic preferences might differ idiosyncratically, or perhaps be systematically influenced by social contexts.

As a contribution to the field of experimental philosophy, this paper empirically measures the level of aesthetic consensus among and across British mathematicians, Chinese mathematicians, and British undergraduate mathematics students. We begin by examining philosophical accounts of mathematical beauty and methods previously used to assess aesthetic intuitions. We then introduce comparative judgment as a means of investigating aesthetics, and describe the stimuli, participants, and procedure we employed. Finally, we discuss our empirical findings in the light of current understandings of mathematical beauty, and summarise both our substantive contribution to debates about philosophical intuitions and our methodological contribution to advances in experimental philosophy.

## 1 Agreement and Disagreement about Mathematical Beauty

### 1.1 Assumed Aesthetic Agreement among Mathematicians

One early realist account of mathematical beauty is the Pythagorean ‘cosmocentric’ belief that beauty is an objective property for humans not to invent but to discover (Tatarkiewicz 1963). On this account, mathematics governs the unquestionable essence of physical reality, which can only be grasped through the beauty and harmony of numerical and geometrical patterns (Sinclair and Pimm 2006, p.4). Although this account tends to amalgamate later Pythagorean writers and their modern interpreters (Berghaus 1992, p.44), it initiated a highly influential understanding of mathematical beauty. For instance, according to Plato’s *Philebus* (1993), Socrates states that mathematical objects such as straight lines or circles are not ‘relatively beautiful’ like animals or pictures. Instead, their beauty is eternal and absolute, which means that it does not evolve with changes of social context. This means that the aesthetic dimension of mathematics exists objectively and is not reducible to mathematicians’ subjective perceptions and preferences.

Mathematicians who subscribe to such accounts tend to advocate the existence of fixed aesthetic criteria or a collection of mathematical objects universally agreed to be beautiful. One well-known such account is G.H. Hardy’s (1940, p.29) list of six criteria for beauty in a proof. This list includes economy, which is derived from the classical ideal of beauty as simplicity. This ideal is embodied in different forms, including numerical, explanatory, and logical simplicity (McAllister 1996). For Hardy, economy requires that theorems and proofs have simple lines of argumentation, which facilitate their grasp ‘in a single act of mental apprehension’ (McAllister 2005, p. 19). Identifying economy – and his remaining criteria of significance, depth, generality, unexpectedness, and inevitability – demands, according to Hardy (1940, p.113), ‘a high degree of technical proficiency’ developed through many years of experience as a mathematician. This implies that non-experts, such as undergraduate students, may have difficulty making aesthetic judgements.

Another well-known proponent of aesthetic realism is Paul Erdős, who described ‘a transfinite book of theorems’ where ‘the best proofs are written’ (Alexanderson

1981, p.254). Erdős frequently referred to beautiful proofs as coming ‘straight out of The Book’ (Erdős 1983, p.37), and famously claimed that mathematicians do not necessarily need to believe in the existence of God, but they should believe in the existence of ‘The Book’ (Aigner and Ziegler 2010, p.v). According to Cherniwchan et al. (2010), one of Erdős’s great ambitions in life was to find such beautiful proofs. His students Aigner and Ziegler (2010) published a collection entitled *Proofs from The Book*, based on some of Erdős’s own suggestions. According to Aigner and Ziegler (2010, p.v), a degree of mathematical proficiency is required to comprehend ‘The Book’, although only at the undergraduate level. This suggests that undergraduate mathematical training would suffice for making at least some aesthetic judgments.

The above accounts make two assumptions regarding mathematical beauty that are subject to empirical investigation: (i) there should be agreement about aesthetics among mathematicians, and (ii) judgements about aesthetics require a degree of mathematical proficiency, although it is unclear what level of training is needed. It is worth noting that the theoretical basis for the former assumption is potentially challenged by arguments that mathematical beauty is reducible to non-aesthetic properties that are epistemically centered. Rota (1997, p.175), for instance, agrees that ‘both the truth of a theorem and its beauty are...equally shared and agreed upon by the community of mathematicians’, but suggests that this is due to a sense of enlightenment that is mistakenly referred to as ‘beauty’. Similarly, Todd (2008, pp.71–72) claims that the ‘normative strength of the putative aesthetic claims’ is due to a relationship between truth and epistemic warrants, and Dutilh Novaes (2019) argues that Hardy’s aesthetic criteria are reducible to non-aesthetic properties that facilitate the epistemic function of explanatoriness. These claims highlight ambiguity between aesthetic and epistemic dimensions of mathematics, and we address them further later. However, these authors still assume a degree of agreement between different mathematicians’ aesthetic intuitions, so evidence on (i) pertains to their views too.

## 1.2 Empirical Evidence on Disagreement about Mathematical Aesthetics

Regardless of their basis, claims about agreement concerning mathematical beauty are called into question by recent empirical evidence. In one early study, Wells (1990) asked 68 readers of *The Mathematical Intelligencer* to use a scale to rate the beauty of 24 theorems. He reported that renowned theorems such as Euler’s identity, polyhedron formulas, and the infinity of primes were all rated highly, but other well-regarded theorems were not. This led him to question whether some of the highly rated theorems were genuinely perceived to be beautiful, or whether the ratings were better understood as a product of social influence. Additionally, he detected mixed aesthetic responses to simplicity, as some theorems were rated low because they were not proven in a simple and succinct manner, but other theorems with simpler or easier proofs were also rated low because readers found them too simple and easy.

Agreement, simplicity and social influence have been addressed directly in more recent studies. On agreement, Inglis and Aberdein (2016) asked 112 mathematicians to rate the accuracy of twenty adjectives in describing the proof of Sylvester’s theorem from ‘The Book’ (Aigner and Ziegler 2010). They found a low level of aesthetic agreement,

with 60.4% scoring the proof below the midpoint of the aesthetic scale, and only 31.5% scoring it above. For a ‘Book proof’ to score so low seems to challenge the existence of universal agreement about mathematical beauty; it certainly seems that Aigner and Ziegler’s aesthetic preferences might not be reflected in the wider community.

On simplicity, Inglis and Aberdein (2015) asked 225 mathematicians to rate the extent to which 80 adjectives described a proof that they could think of or had recently read. Using an exploratory factor analysis, they found four main dimensions on which proofs varied: aesthetics, intricacy, precision, and utility. Neither ‘beautiful’ nor ‘elegant’ correlated strongly with ‘simple’. Thus, even if there is agreement, this might not be due to traditionally listed criteria.

On social influence, Inglis and Aberdein (2020) replicated their 2016 study, this time with the manipulation that half of their 203 mathematician participants were told the proof’s source, Aigner and Ziegler’s (2010) attempt to produce a version of Erdős’s ‘The Book’. Pure mathematicians who were given the source rated the proof more highly than those who were not, but applied mathematicians did not show such an effect. This suggests that mathematicians’ aesthetic judgement might indeed be socially influenced: Erdős was most active in pure mathematics, so ‘The Book’ is likely better known among pure than applied mathematicians. This result also relates to the *mere-exposure* account of aesthetics, which suggests that an individual’s aesthetic appreciation is developed through repetitive exposure to the same item (Zajonc 2001). Famous mathematical objects, such as proofs from Aigner and Ziegler’s ‘The Book’, would have more exposure within the field, and so mathematicians’ aesthetic appreciation of such proofs could be socially developed through repetitive exposures. If accounts such as the mere exposure effect, coupled with social conformity effects of the type studied by Inglis and Aberdein, can successfully explain mathematicians’ judgement of mathematical beauty, then its objective existence would be an unnecessary assumption.

Overall, these empirical results show disagreement between different mathematicians’ intuitions of mathematical beauty. The evidence is not decisive, and there are reasons to be cautious about the methodological approaches adopted in these studies, as discussed below. However, it seems reasonable to conclude that investigations into mathematical beauty should not assume that mathematicians all agree. With this in mind, we examine judgments of mathematical beauty in relation to simplicity, cultural context, and proficiency by measuring and comparing the degree of aesthetic agreement among British mathematicians, Chinese mathematicians, and British mathematics undergraduates.

To further situate this work and to raise issues in methodology, we next elaborate on relevant studies involving the study of cross-cultural and cross-expertise philosophical intuitions.

### 1.3 Cross-Cultural Studies of Philosophical Intuitions

Potential cultural influences on perceptions of mathematical beauty have not yet been philosophically discussed or empirically assessed. Indeed, Larvor (2016, p.8) argued that there is ‘a dearth of cultural theory’ in the philosophy of mathematical practice, which is important because mathematical practice – like any other practice – needs to be ‘culturally embedded, manifested, and valued’, to ‘stabilise and reproduce’ its norms and values.

Work in this area would also contribute to wider disputes on the degree of cross-cultural consensus about philosophical intuitions. In early cross-cultural experimental works, Westerners and East Asians were found to have different patterns of epistemic and semantic intuitions (Weinberg et al. 2001; Machery et al. 2004), challenging the normativity of those intuitions and posing a serious problem for the standard philosophical approach of using intuition as evidence (Stich 2001). However, these findings failed to replicate (e.g. Seyedsayamdost 2015; Kim and Yuan 2015), arguably due to methodological weaknesses: Knobe (2019, 2021) pointed out that the number of Asian participants in Weinberg et al.'s study was only 24, and Lam (2010) argued that it was problematic for Machery et al. to present questions to Asian participants in English instead of their native languages. Machery et al. (2017) recently addressed these methodological limitations by having a larger sample ( $N=521$ ) from four different cultural backgrounds, and using materials written in native languages. Contrary to the early experimental results, Machery et al. (2017) identified that people from different cultural backgrounds exhibit similar patterns of epistemic intuitions. Knobe (2019) cited this result in support of the hypothesis that philosophical intuitions are robust across cultures. However, this position was criticized by Stich and Machery (2022), who argued that Knobe had presented an unbalanced summary of the literature: although some studies failed to replicate earlier findings of cross-cultural differences in intuitions, many such findings have successfully replicated (Stich and Machery 2022, Table 1).

Similar points have been raised in work directly related to aesthetics, in the more obvious domain of perceptual beauty (Che et al. 2018). Cross-cultural disagreement was initially detected by McElroy (1952) and Lawlor (1955), but agreement was found in a series of later investigations (Eysenck and Iwawaki 1971; Soueif and Eysenck 1971) and in cross-cultural studies on basic visual features such as symmetry, proportion, curvature, brightness, and contrast (Che et al. 2018). These contrasting empirical results could be influenced not only by the choice of stimuli but also by the types of judgements, since these studies varied in asking participants to judge the stimuli in isolation or in comparison with one another: McElroy and Lawlor asked participants to rank an entire set of artworks presented simultaneously, whereas Eysenck and his collaborators asked participants to make comparative or individual judgements. It is certainly possible that the apparent degree of aesthetic consensus is influenced by methodological approach, and we pick up this point below.

#### 1.4 Beauty, Epistemology, and Expertise

As noted above, there are debates about the extent to which aesthetics overlaps with epistemology. If mathematical beauty requires significant mathematical insight, then recognizing beauty should be impossible for people without adequate training. This would be consistent with an early study by Dreyfus and Eisenberg (1986), who designed a set of problems to evoke elegant solutions and found that undergraduates struggled to come up with such solutions and were unable to distinguish aesthetically pleasing solutions from others when prompted. Dreyfus and Eisenberg interpreted

this as indicating that people need mathematical training beyond undergraduate level in order to appreciate mathematical beauty. Of course, it could be that Dreyfus and Eisenberg's own criteria for elegance or beauty are not widely shared, or that undergraduates can appreciate beauty only in relatively simple mathematical contexts.

It could also be that mathematicians, who do have the potential to make aesthetic judgements based on epistemology, do so only partly on that basis. Starikova (2017) argues in this direction, distinguishing intellectual and perceptual aspects of mathematical beauty. She suggests that intellectual beauty is the aesthetic response to abstract properties of a mathematical object, such as structure or degree of generality; sufficient proficiency is required to detect and appreciate these. Appreciating perceptual beauty, on the other hand, does not necessarily require mathematical understanding. Similarly, Montano (2014) and Pearcy (2020) distinguish the performative appreciation response, which is a reaction of active intellectual engagement, from the basic appreciation response, a passive and automatic reaction. Pearcy (2020, p.59–60) illustrates this with the example of the physicist Richard Feynman and his artist friend who could both see a flower as beautiful, but for different reasons: Feynman has a basic appreciation response in visual aesthetics, but a performative appreciation response in scientific aesthetics, his artist friend vice versa.

These theoretical suggestions are consistent with empirical evidence. Zeki et al. (2014), for instance, asked 15 mathematicians (postgraduate students and postdoctoral researchers) to study 60 equations, rating each for beauty from -5 to 5. After about 2 to 3 weeks, the mathematicians were asked to re-rate the equations as ugly, neutral, or beautiful while their brain activity was fMRI-scanned. A few days after scanning, they were asked to rate their understanding of each equation from 0 to 3. Zeki et al. found a significant positive correlation between understanding and scan-time beauty rating, and a significant difference in brain activity in a region associated with appreciating beauty when participants were viewing equations they rated as beautiful as opposed to ugly or neutral. The latter was driven by beauty ratings after accounting for understanding, so there is room for aesthetic judgements to be based partly on understanding and partly on visual appearance. Consistently with this interpretation, Zeki et al. also found that 12 non-expert participants (educated in mathematics only to the age of 16) indicated that they had no understanding for the vast majority of the equations, but some did give positive beauty ratings for a minority.

For our purposes, this provides evidence of an imperfect overlap between aesthetics and epistemology for experts, but no indication how this develops prior to expertise or of whether or not there is aesthetic agreement. In fact, although Zeki et al. found a highly significant positive correlation between pre-scan and scan-time beauty ratings, the correlation coefficient of  $r = .612$  is some way from perfect, and some large shifts in ratings were seen between the two times. If mathematicians do not always agree with themselves, perhaps it is unreasonable to expect them to agree with one another. The aesthetic judgement of experts and non-experts has been further studied by Johnson and Steinerberger (2019) who asked two groups of experts (mathematicians and mathematics undergraduates) and a group of non-experts to rate the similarity of mathematical arguments to artworks (paintings and classical

music). Perhaps surprisingly, they found that participants could associate each mathematical argument to an artwork, with agreement at above chance levels, suggesting some degree of shared consensus about this kind of aesthetic correspondence.

Using different methods, Hayn-Leichsenring et al. (2021) studied both undergraduate students and aesthetic agreement. They asked twenty mathematics undergraduates and twenty undergraduates without university-level mathematical training to distribute 64 equations into 9 piles ranging from “extremely unaesthetic” to “extremely aesthetic” with predetermined numbers in each pile to form a normally distributed pattern. After participants completed their judgements, they were asked to state which equations they were familiar with, and to indicate the criteria behind their judgements from options including “meaning”. In line with the works of Zeki et al. and Johnson and Steinerberger, Hayn-Leichsenring et al. found a positive relationship between understanding and perceived beauty: in both groups, ratings were significantly higher for familiar equations. This was more pronounced for the mathematics undergraduates, who were familiar with more equations and who more often stated that their aesthetic judgement relied on meaning. This seems to imply that greater understanding would result in greater aesthetic appreciation in mathematics. However, an alternative account would be that understanding is merely an essential pre-condition in making any forms of judgements on equations or proofs. More investigations are needed to examine how intuitions about mathematical aesthetics are related to familiarity and understanding.

Hayn-Leichsenring et al. also looked explicitly at simplicity, finding a significant negative relationship between the number of elements (numbers, letters and mathematical signs) in an equation and its aesthetic rating for the mathematics undergraduates but not the other group. They also found that compared to undergraduates without mathematical training, mathematics undergraduates shared a higher level of aesthetic agreement.

These results suggest that undergraduates with university-level mathematical training have attained sufficient proficiency to share a performative appreciation response to mathematical beauty. And this returns us to questions about methodology. In some of these empirical studies, it seems that mathematicians do not agree about beauty as much as traditional philosophical accounts suppose. In others, it seems that agreement is present even among comparatively inexperienced undergraduates. We suggest that method might be one reason for this. Notably, in Johnson and Steinerberger’s study, participants’ aesthetic judgement was conducted through comparing and contrasting mathematical arguments with artworks. Similarly, in Hayn-Leichsenring et al.’s study, participants also compared and contrasted equations for fine-grained aesthetic classification. In both studies, participants’ aesthetic judgements were relative rather than absolute. We believe it is plausible that mathematicians might have different absolute standards for beauty and thus appear to disagree when asked for absolute judgements, but might nevertheless agree about which objects are more or less beautiful. If this is the case, then such agreement is best sought with methods involving relative judgements. We used one such method, as described below.



## 2 Methodology: Comparative Judgement

We used a comparative judgement (CJ) approach to measure aesthetic judgments about mathematical beauty. Under CJ, participants do not use absolute rating scales, but instead each make multiple pairwise judgements about which of two objects rates more highly in relation to a given quality. The judgements are then used collectively to construct a scaled rank order in which each object is assigned a score (Bisson et al. 2016). This approach is based on the psychological principle that people tend to be better at making relative judgements than at judging one object against a predetermined criterion (Thurstone 1994). This principle is derived from substantial investigation on human judgement of sensory factors such as temperature and audio frequency (Laming 2003; Pollack 1952; Thurstone 1928).

Using CJ has two advantages that are specific in our context, First, it does not require pre-determined criteria for the concept to be measured. Instead, the scores are directly derived from participants' pairwise judgements. This characteristic enables an open-ended approach to measuring concepts that are ambiguous and fuzzy (Bisson et al. 2016, p. 143), as demonstrated by successful use of CJ in measuring students' conceptual understanding (Bisson et al. 2016; Jones et al. 2019), proof comprehension (Davies et al. 2020), and the notion of explanatoriness of proofs (Mejía Ramos et al. 2021). Here, mathematical beauty is conceptualized as an ambiguous concept under dispute among philosophers, and CJ has the advantage of not presupposing any philosophical accounts. Second, CJ directly measures mathematicians' aesthetic conceptions without using absolute scales. It therefore circumvents subjective perceptions of such scales, which could potentially obscure agreement (Heine et al. 2002).

Using CJ, we conducted two studies, in both of which participants were asked to consider pairs of mathematical objects and to judge which is more beautiful. In Study 1, the objects were equations, and participants were from three demographic groups: British mathematicians, Chinese mathematicians, and British mathematics undergraduates. This allowed us to investigate cross-cultural and cross-expertise (dis)agreement about mathematical beauty. The equations were accompanied by brief descriptions, and we also considered factors that might potentially affect aesthetic judgements: number of characters in the equations as a measure of simplicity, number of words in the description, and number of mathematicians' names mentioned in the description as a measure of social influence. In Study 2, the objects were proofs and participants were British mathematicians. This allowed us to investigate whether aesthetic agreement is contingent upon different types of stimuli.

## 3 Study 1: Measuring Mathematicians' and Undergraduates' Aesthetic Judgements

### 3.1 Stimuli

The stimuli for this study were chosen from Zeki et al. (2014)'s list of 60 equations; 20 out of the 60 equations were used (selected by taking every third one), along with the brief descriptions written by Zeki et al. For the Chinese participants, the descriptions were



translated into simplified Chinese. The selected equations and their descriptions were formatted and uploaded to the online CJ platform No More Marking (<https://www.nomormarking.com>). The 20 equations appear in Table 1 in the Results section (along with their CJ scores, to be explained below); their descriptions appear in Table 3 the Appendix.

To assess factors that might predict judgements of mathematical beauty, we counted the number of characters in each equation, the number of words in its description, and the number of mathematicians' names mentioned in the description. For the equations, we counted individual characters ignoring any commas, so that  $\frac{dx}{dt}$  and  $(\alpha - \beta\gamma)$  have 5 and 6 characters respectively. For the descriptions, we counted words ignoring punctuation and brackets; any mathematical element that appeared within the description was counted as one word. For the counts for each equation, see again Table 3.

**Table 1** The ranking and the  $\beta$  score for each equation, separately for each demographic group

| Equation  | British Mathematicians |        | British Undergraduates |        | Chinese Mathematicians |        |
|---|------------------------|--------|------------------------|--------|------------------------|--------|
|   | Ranking                | Score  | Ranking                | Score  | Ranking                | Score  |
| $1 + e^{i\pi} = 0$  | 1                      | 2.391  | 1                      | 1.347  | 1                      | 2.032  |
| $\int_{\partial M} \omega = \int_M d\omega$   | 2                      | 1.147  | 8                      | 0.336  | 4                      | 0.689  |
| $\frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2}$   | 3                      | 1.110  | 6                      | 0.513  | 2                      | 1.212  |
| $a^2 + b^2 = c^2$   | 4                      | 0.764  | 3                      | 0.889  | 3                      | 0.833  |
| $\frac{d}{dx} e^x = e^x$  | 5                      | 0.665  | 4                      | 0.784  | 16                     | -0.598 |
| $a^n + b^n = c^n, n > 2$  | 6                      | 0.602  | 2                      | 0.973  | 5                      | 0.645  |
| $\Delta\varphi = 0$   | 7                      | 0.523  | 10                     | 0.187  | 10                     | -0.045 |
| $e = \lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n$   | 8                      | 0.393  | 7                      | 0.411  | 7                      | 0.362  |
| $\int_M K dA + \int_{\partial M} k_g ds = 2\pi \chi(M)$   | 9                      | 0.083  | 18                     | -1.012 | 12                     | -0.169 |
| $\ x + y\  \leq \ x\  + \ y\ $  | 10                     | -0.046 | 5                      | 0.755  | 6                      | 0.366  |
| $A = \int_{\sigma(A)} \lambda dE_\lambda$   | 11                     | -0.196 | 13                     | -0.256 | 15                     | -0.489 |
| $3^2 + 4^2 = 5^2$   | 12                     | -0.346 | 9                      | 0.303  | 9                      | 0.108  |
| $\prod_{k=0}^{\infty} (1 + x^k) = \sum_{n=0}^{\infty} p(n)x^n$                                    | 13                     | -0.461 | 15                     | -0.445 | 14                     | -0.277 |
| $f(x) = \int_{-\infty}^{\infty} \delta(x - y)f(y)dy$  | 14                     | -0.482 | 11                     | 0.074  | 11                     | -0.143 |
| $1 = \sum_{n=2}^{\infty} (\zeta(n) - 1)$  | 15                     | -0.607 | 14                     | -0.355 | 8                      | 0.287  |
| $V_n(r) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)} r^n$                                  | 16                     | -0.713 | 16                     | -0.741 | 18                     | -0.828 |
| $\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}, s \in \mathbb{C}, \text{Re}(s) > 1$ | 17                     | -0.930 | 20                     | -1.624 | 13                     | -0.212 |
| $\frac{dx}{d} = x(\alpha - \beta y), \frac{dy}{dt} = -y(\gamma - \delta x)$                       | 18                     | -1.048 | 12                     | -0.033 | 19                     | -0.964 |
| $T = de + \omega \wedge e, R = d\omega + \omega \wedge \omega$                                    | 19                     | -1.387 | 17                     | -0.765 | 17                     | -0.729 |
| $R_{\beta[\gamma\delta,\lambda]}^\alpha = 0$  | 20                     | -1.461 | 19                     | -1.341 | 20                     | -2.081 |

## 3.2 Participants and Procedure

British mathematicians and British undergraduates were recruited via an invitation email sent to two UK mathematics departments' mailing lists. Chinese mathematicians were recruited by the same translated invitation email which was sent to the mailing addresses of various mathematics departments in China. The email contained a brief introduction about the research, specific information on what was involved in this study, and a web link to enter the study. Participants who entered the study were assured that it was ethically approved by Loughborough University and that none of their personal data would be collected. Mathematician participants were, however, asked to select their AMS subject classification, their career stage and their number of years working as a mathematician. British undergraduate participants were asked to state their current subject year of study. All then read these instructions:

“We are interested in understanding what mathematicians (undergraduates) mean when they say that certain mathematical objects are “beautiful”. To this end, we are going to ask you about the beauty of various mathematical equations. You will be shown pairs of different equations. Every time you see a pair, we will ask you to choose which equation you think is the most beautiful. Once you have started the judging session...Simply look at the two equations and choose which one you think is more beautiful by clicking either ‘Left’ or ‘Right’. If you are unsure, just go with your instinct.”

Each participant was asked to complete 20 judgements of randomly generated paired equations, with no time limit. In total, 24 British mathematicians completed 480 judgements, 24 Chinese mathematicians completed 480 completed judgements, and 81 British undergraduates completed 1620 judgements.

## 3.3 Results

### 3.3.1 Is There Aesthetic Agreement Among British Mathematicians, Chinese Mathematicians, and British Undergraduates?

For each demographic group, we used the Bradley-Terry Model, which assigns each equation  $i$  a parameter  $\beta_i$  to estimate its beauty. It does this via a process based on using the judgements to iteratively update the probability that equation  $i$  is judged to be more beautiful than equation  $j$  (Bradley and Terry 1952). To check that this yielded meaningful scores, we then calculated inter-rater reliability (IRR) for each demographic group. We randomly split each group into subgroups of 12, splitting the mathematician groups evenly and randomly selecting two groups of 12 from among the undergraduates. We then calculated new estimates of each equation's aesthetic quality from each subgroup's judgements. This process was repeated 1000 times to calculate the average Pearson correlation coefficient between the scores for the two subgroups.

The IRR of the British mathematicians' judgements was  $r=0.721$ , the IRR of the British undergraduates' judgements was  $r=0.701$ , and the IRR of the Chinese mathematicians' judgements was  $r = .722$ . These results indicate relatively consistent

aesthetic agreement within each demographic group: the CJ approach does detect agreement based on relative judgements for all three groups. We thus treat the equation scores for the complete groups as reliable, and these scores are shown in Table 1, ordered from most to least beautiful by the British mathematicians' rankings.

### 3.3.2 Is There Cross-Cultural Agreement?

A first indication of agreement not just within but across the demographic groups is visible in the rankings and scores in Table 1. Both the British and the Chinese mathematicians judged Euler's identity the most beautiful equation and the Second Bianchi Identity the least beautiful. Moreover, although the rankings do not match perfectly, equations judged more beautiful by one group were generally

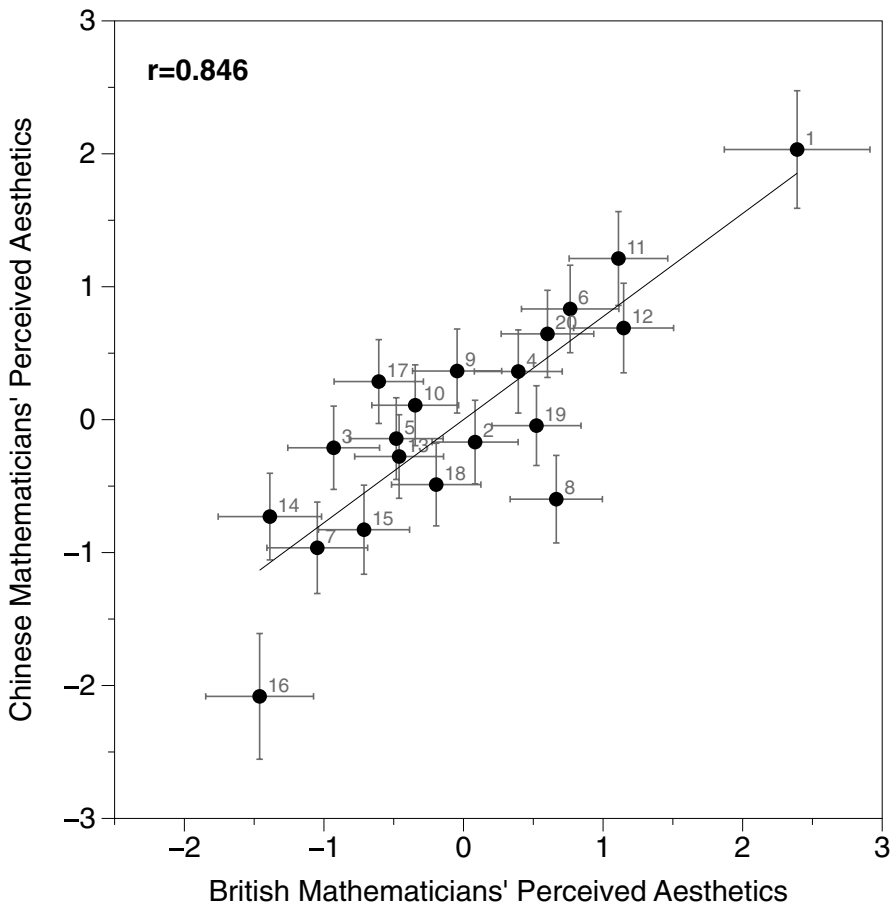
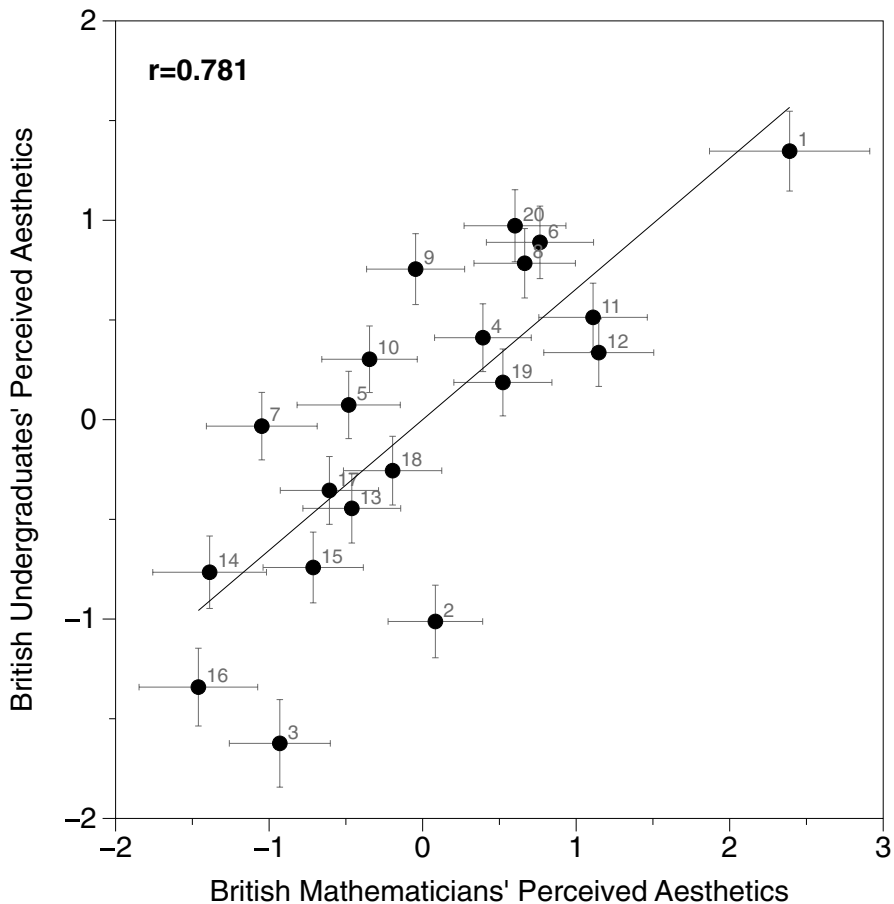


Fig. 1 The correlation between CJ scores derived from British mathematicians' and Chinese mathematicians' aesthetic judgements. Error bars show  $\pm 1$  standard error

judged more beautiful by the other. This is reflected in a statistical analysis: there is a significant and strong positive correlation between the two sets of scores  $r = .846$ , 95% CI [.645, .937]; see Fig. 1. At least for equations, it seems, relative mathematical beauty is judged fairly consistently across these two cultures.

### 3.3.3 Is There Cross-Expertise Agreement?

A first indication of agreement across expertise levels is also visible in Table 1. The British undergraduates, like both groups of mathematicians, judged Euler's identity the most beautiful equation. They also, like the Chinese mathematicians, broadly agreed with the British mathematicians: again we found a significant and strong positive correlation between the two sets of scores  $r = .781$ , 95% CI [.518, .909]; see Fig. 2. Again, at least



**Fig. 2** The correlation between CJ scores derived from British mathematicians' and British undergraduates' aesthetic judgments. Error bars show  $\pm 1$  standard error

**Table 2** Unstandardized regression coefficients (Bs) predicting perceived aesthetics for British mathematicians, British undergraduates, and Chinese mathematicians

| Predictor            | British mathematicians B | British undergraduates B | Chinese mathematicians B |
|----------------------|--------------------------|--------------------------|--------------------------|
| Number of Characters | -0.0890**                | -0.0676*                 | -0.0548                  |
| Number of Names      | 0.1754                   | -0.1367                  | 0.1698                   |
| Number of Words      | 0.0135                   | 0.0164                   | 0.0188                   |
| $R^2$                | 0.3730                   | 0.4450                   | 0.2120                   |

\* $p < .05$ , \*\* $p < .01$ 

for equations, it seems that relative mathematical beauty is judged similarly by British undergraduate mathematics students and more experienced mathematicians.

### 3.3.4 What Predicts Judgements of Beauty?

Finally for this study, Table 1 suggests that characteristics of the equations might predict judgements of beauty: equations judged more beautiful tend to be shorter. To formally investigate this, along with our other possible predictors, we conducted linear regression analyses, separately for each group, predicting CJ scores from the number of characters in the equations and the numbers of words and mathematicians' names in the descriptions.

The results appear in Table 2. For no group did either number of words or number of names predict the beauty scores. Although, in line with earlier studies (Wells 1990), it seems that everyone thinks Euler's identity is beautiful, it appears unlikely that this is due to a generally positive view of equations with names attached (notably, our measure does not capture relative renown). It could well be the case, however, that its perceived beauty is related to its simplicity: for both the British mathematicians and the British undergraduates, the number of characters in an equation did significantly predict beauty score ( $p = .008$  and  $p = .010$  respectively). To contextualize these estimates, for the mathematicians, an extra 10 characters in an equation predicted a drop in beauty score of 0.89, nearly a quarter of the overall range from 2.390 to -1.460; for British undergraduates, it predicted a drop of 0.676 (score range 1.347 to -1.623). For Chinese mathematicians, the number of characters was not a significant predictor of beauty score ( $p = .087$ ), but the analysis nevertheless indicated a similar predictive pattern compared to the previous two demographic groups. Number of characters in an equation is clearly a crude measure of simplicity, but the direction of these results is in line with philosophical claims that simplicity is related to mathematical beauty.

## 4 Study 2: Measuring Mathematicians' Aesthetic Judgements about Proofs

Study 1 found not only consistent aesthetic judgements within our demographic groups of British mathematicians, Chinese mathematicians and British undergraduates, but also agreement across culture and expertise. These findings

contrast with some earlier empirical work (Wells 1990; Inglis and Aberdein 2016, 2020), which found more disagreement than might have been expected based on traditional accounts. One possible explanation was suggested above: ratings against absolute scales might fail to capture underlying consensus on relative mathematical beauty. However, it could also be that level of consensus is affected by the type of stimuli: Inglis and Aberdein (2016, 2020), for instance, found disagreement on proofs rather than equations. Hence, Study 2 aimed to measure mathematicians' aesthetic agreement in relation to proofs, to examine whether a change to different type of stimuli would influence the agreement found in Study 1.

#### 4.1 Stimuli

Eight proofs were employed in Study 2 (we used fewer proofs than equations for the obvious reason that these take longer to read). Five proofs were selected from Aigner and Ziegler's (2010) collection of *Proofs from The Book*, two were from Percy's (2020) *Mathematical Beauty*, and one was from Nelsen's (2000) *Proofs Without Words II: More Exercises in Visual Thinking*. All eight proofs stimuli appear in Table 4 in the Appendix.

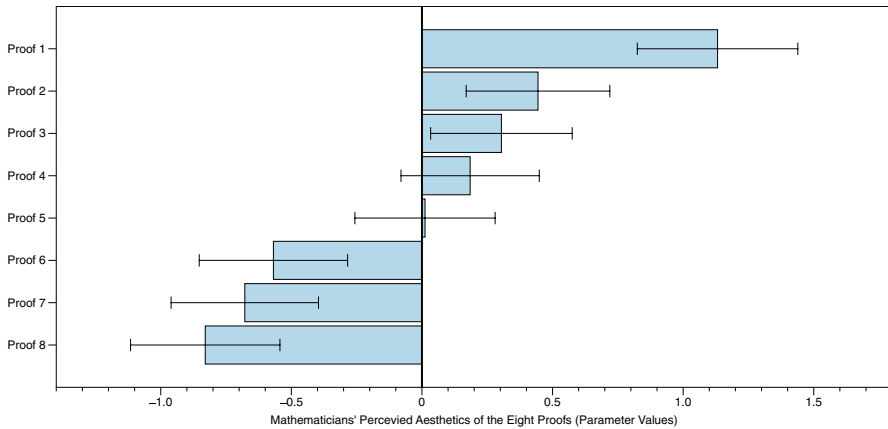
#### 4.2 Participants and Procedure

Thirty-two mathematicians were recruited by an invitation sent to the mailing list of a UK based mathematics department not already contacted in relation to Study 1. Participants were asked to conduct eight pairwise judgments, with the prompt 'Which proof is more beautiful?'. One of the 32 mathematicians was excluded from the analysis because they did not complete their judgements. The remaining 31 participants in total completed 248 pairwise judgements.

#### 4.3 Results: Beauty in Proofs

Following the same procedure as in Study 1, we first considered reliability, finding an IRR of  $r = .643$ . This is slightly lower than the IRRs found for the equations in Study 1, but still means that there was considerable agreement about which proofs were more beautiful. Subsequently, the total 248 aesthetic judgments of proofs were analyzed using the Bradley-Terry Model, which resulted in a scaled rank order in which Euclid's proof of the infinitude of the primes was judged the most beautiful and an algebraic proof that  $p'(x)^2 \geq p(x)p''(x)$  for all  $x \in \mathbb{R}$  was judged the least beautiful. For proofs, rankings and scores, see Fig. 3 and Table 4 in the Appendix.

In sum, Study 2 found similar results to Study 1. Specifically, mathematicians' level of aesthetic consensus was not substantially affected by asking them to consider proofs rather than equations. This makes us more confident that the agreement in Study 1 is not contingent simply upon the fact that equations are simpler objects. Rather, agreement is found for proofs as well.



**Fig. 3** The parameter values of British mathematicians' aesthetic judgements of proofs

## 5 Discussion

### 5.1 Summary

The studies in this paper investigated whether mathematicians agree about mathematical beauty. Using comparative judgement methods, Study 1 found agreement about the aesthetics of equations both within and across three demographic groups: British mathematicians, Chinese mathematicians, and British undergraduate mathematics students. It also found that simplicity – operationalized by counting characters in each equation – predicted collective judgements of beauty. Study 2 broadened the range of stimuli, finding a similar level of between-participant agreement among British mathematicians about the aesthetics of proofs.

Together, these studies constitute evidence that relative judgements about beauty in mathematics are fairly stable and robust within and across cultures, and that undergraduates have learned enough to be able to judge beauty in a way similar to expert mathematicians, at least for the equations we considered. In this last section, we discuss the implications of these findings for views on beauty in relation to agreement and simplicity, to cross-cultural studies, and to epistemology; throughout, we consider issues of methodology in experimental philosophy.

### 5.2 Aesthetic Agreement and Simplicity

Two of the most interesting results of this paper are the level of aesthetic agreement found and the result that short equations tend to be judged more beautiful. Both findings are in line with traditional accounts of mathematical beauty, but they go against prevailing trends in recent empirical work, which has found aesthetic disagreement among mathematicians (Wells 1990; Inglis and Aberdein 2016, 2020) and a lack of relationship between beauty and simplicity (Inglis and Aberdein 2015). We suggest that, in both cases, methodological factors might account for these apparent contradictions.



Regarding agreement, we have discussed two methodological differences across studies: the types of stimuli and the method for collecting aesthetic judgments. The stimuli in many of the studies that have found evidence of aesthetic disagreement were proofs, whether proofs that the participants called to mind (Inglis & Aberdein 2015) or proofs explicitly presented for evaluation (Inglis & Aberdein 2016, 2020). In Study 1, our work followed a different trend in the literature, using equations (cf. Hayn-Leichsenring et al. 2021). Clearly equations and proofs differ enough that they could prompt different degrees of consensus. It is possible that mathematicians tend to agree on shorter objects such as equations, but disagree on more complex objects such as proofs. However, when we examined this potential explanation in Study 2, we found relatively consistent between-participant aesthetic agreement about proofs too. No doubt different types of mathematical objects influence aesthetic judgements to some degree, but our evidence suggests that consensus in aesthetic intuition is not highly contingent upon the type of stimuli.

The other methodological difference is that in most studies that have reported aesthetic disagreement, mathematicians' aesthetic judgements were measured on absolute scales, whereas we used comparative judgement. Since CJ avoids potential subjective interpretations of absolute scales, it is more akin to the sorting method used in Hayn-Leichsenring et al.'s (2021) study. That Hayn-Leichsenring et al. also found agreement suggests that when relative judgements are used to measure mathematicians' sense of aesthetics, underlying agreement can be detected. Hence the disagreement found in Wells' (1990), and Inglis and Aberdein's studies (2016, 2020) could be a result of mathematicians not sharing the same standards in relation to absolute scales, rather than of a fundamental aesthetic dispute.

Regarding simplicity, we found that shorter equations tended to be judged more beautiful by British mathematicians and undergraduates, with a trend in the same direction for Chinese mathematicians. This, too, is consistent with traditional philosophical accounts and with Hayn-Leichsenring et al.'s (2021) study of the aesthetic evaluation of equations, but different from earlier empirical work that did not find such an association (Inglis and Aberdein 2015). A methodological reason for this difference could be that simplicity in our paper was measured through the admittedly crude means of counting the number of characters in each equation, whereas Inglis and Aberdein measured it through mathematicians' use of the adjective 'simple' in proof appraisal. The latter clearly allows for more sophisticated judgements, but it also introduces ambiguity regarding whether it was numerical, explanatory, logical or some other form of simplicity that these mathematicians had in mind. Certainly, further empirical studies on this aesthetic ideal might usefully unpack the notion of simplicity in more detail.

### 5.3 Cultures, Expertise and Epistemology

Our first study deliberately examined judgements of beauty from multiple demographic groups, answering calls to consider culture in mathematical practice (Larvor 2016) and addressing the issue of what expertise is required to exercise aesthetic

judgement. Our finding of a strong degree of aesthetic consensus across the British and Chinese groups suggests that mathematicians' aesthetic judgements are not strongly influenced by cultural differences. This is consistent with the moderate aesthetic agreement on basic visual properties found elsewhere in the field of cross-cultural empirical aesthetics (Che et al. 2018), and is good news for those inclined towards aesthetic realism: although agreement among mathematicians does not imply that mathematical beauty is objective or that aesthetic intuitions can be normatively correct, it provides no reason to reject that position.

Another way of accounting for our finding of aesthetic consensus across the British and Chinese groups would be to suggest that the two groups share a similar mathematical culture. In other words, perhaps mathematics is so interconnected in the modern world, it no longer makes sense (if it ever did) to talk about distinct mathematical cultures. We doubt that this is the case. While it is certainly true that there has been a great deal of interaction between Western and Chinese mathematics, both historically and today, this has led to concerted efforts by some Chinese mathematicians to try to preserve what they consider to be distinctive about Chinese mathematics. For instance, following the arrival in China of Jesuit missionaries with Western mathematical texts in the late seventeenth century, political movements such as "Chinese Origins of Western Science" were founded. These aimed to minimize the significance of Western influence in the development of Chinese mathematics by valuing and maintaining its traditional culture (Bréard 2019, p.82).

The aim to preserve Chinese cultural identity in mathematics was again found during the early development of modern mathematics in China. By the 1930s, the first group of mathematics departments were founded in Chinese universities, which led to exchanges with Western institutions. For instance, mathematicians such as Bertrand Russell and William F. Osgood visited Peking University during the 1930s, and their visits were significantly valued by the Chinese mathematics community (Zong 2020). But a number of well-respected Chinese mathematicians responded by emphasising the need to preserve the Chinese approach to mathematics. For example, Shiing-Shen Chern advocated that "Chinese mathematics must be on the same level as its Western counterpart, though not necessarily bending its effect in the same direction" (Hudecek 2014, p.166). Chern's student Wu Wen Tsun noted that "there is an essentially Chinese mathematical style, and that Chinese mathematicians have a patriotic duty to study it and build upon it" (Hudecek 2014, p.161). After Wu returned to China from France in the 1951, he focused on promoting the ancient Chinese style of mathematics characterised by algorithms and the "mechanisation of mathematics" (Hudecek 2012). In sum, there are reasons to suppose that distinctive cultural aspects of Chinese mathematics were, and still are, valued by Chinese mathematicians, despite the international mobility that characterises modern academia.

The fact that the way mathematics is taught in China contrasts to many Western countries gives us further reasons to suppose that Chinese and Western mathematical practices do not share identical cultural norms. Recent decades have seen the development of international comparison studies where student achievement is compared

between educational jurisdictions. These have tended to find that Chinese students, and indeed students in the Pacific rim more generally, tend to outperform Western students of the same age in mathematics (Fan and Zhu 2004). This, in turn, has led to systematic investigations into how typical pedagogy in Chinese classrooms differs from typical pedagogy in Western classrooms (Fan et al. 2004). A common observation is that mathematics education in China tends to place a relatively stronger emphasis on the acquisition of mathematical content through rote learning and hard work than is normal in the West (Leung 2001). Again, these findings support the view that Chinese and Western mathematical cultures are not identical in general, despite our findings that Chinese and British mathematicians' aesthetic tastes appear to be largely shared.

Moreover, we have narrowed down the way in which other social influences might affect judgements about beauty: if social conformity plays a role in mathematicians' aesthetic judgements, this is not visible in effects of the number of influential names attached to an equation. That said, our findings cannot in this case be said to contradict those of Inglis and Aberdein (2020): Euler's identity, with its longstanding aesthetic status, was consistently judged more beautiful than the rest of our stimuli.

With regard to expertise, we found not only that mathematics undergraduates are capable of making collectively consistent aesthetic judgements, but also that they seem to share aesthetic criteria with mathematicians. This provides evidence against the claims of Hardy that judging mathematical beauty requires advanced mathematical proficiency, at least as applies to equations. However, the nature of the criteria remains unclear. Starikova (2017) or Percy (2020) could argue that this cross-expertise agreement might be derived from either perceptual or basic appreciation responses: perhaps both mathematicians and students have similar responses to the visual appearance of the equations, and only mathematicians go beyond this. Taking our work in conjunction with Johnson and Steinerberger (2019) and Hayn-Leichsenring et al.'s (2021), however, we consider it more likely that epistemology plays a role, that undergraduates have developed sufficient proficiency to engage an intellectual or performative appreciation response at a level that might not match that of professional mathematicians but does reflect shared values beyond those accessible to the general population. In addition, since our stimuli includes some famous equations – such as Euler's identity, the Pythagorean theorem and Fermat's last theorem – the mere-exposure effect could be an alternative explanation of the aesthetic agreement that we have found between mathematicians and undergraduates. However, since we did not empirically measure participants' familiarity with each equation, we cannot test this hypothesis. Further work that directly measures how familiar mathematicians are with stimuli such as ours would be worthwhile.

Finally, we highlight that our use of a comparative judgement method might be useful technique more generally in experimental philosophy. A common goal of experimental philosophers is to empirically assess philosophical intuitions (e.g., Heintz and Taraborelli 2010). The psychophysics literature has established that humans are more reliable at making judgements of physical properties

such as height, weight and brightness when asked to compare stimuli rather than judge them in isolation. Our findings suggest that the same may be true when participants are asked to make philosophical judgements. If this conclusion is correct then the method of comparative judgement might be widely useful to experimental philosophers.

To conclude, we stress that the conflicting empirical findings about aesthetic agreement in mathematics echo the broader evolution of experimental philosophy. Although this paper's finding of cross-cultural aesthetic agreement is consistent with stability of philosophical intuitions found in the more recent works of experimental philosophy, this does not mean that one should disregard findings to the contrary. Rather, it highlights the degree of complexity in constructing measures to operationalize the relevant constructs. This paper's findings that aesthetic judgments in mathematics are relatively stable and robust across expertise and cultures suggest that empirical findings of aesthetic disagreement based on absolute scales do not paint a full picture of the nature of mathematical beauty. As Knobe (2019) has suggested, it is important to seek that full picture, which demands triangulation (Löwe and Kerkhove 2019). In this case, investigations might profitably work towards directly comparing absolute with relative judgement approaches, and towards developing more sophisticated operationalizations of simplicity and understanding. Our work suggests that more nuanced accounts might successfully marry empirical findings with traditional accounts of mathematical beauty.

# Appendix

Table 3 Equation stimuli

| Stimuli ID | Equations  | Description  | Characters | Names | Words |
|------------|--|--|------------|-------|-------|
| 10         | $3^2 + 4^2 = 5^2$  | Pythagoras's theorem for a 3: 4: 5 triangle  | 8          | 1     | 6     |
| 11         | $\frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2}$          | The Basel problem, set as a challenge by Jakob Bernoulli in 1689 and solved by Euler in 1735   | 14         | 3     | 18    |
| 12         | $\int_{\partial M} \omega = \int_M d\omega$                    | Stoke's theorem states that the integral of a differential form $\omega$ over the boundary of some orientable manifold $M$ is equal to the integral of its exterior derivative $d\omega$ over the whole of $M$                   | 9          | 1     | 34    |
| 13         | $\prod_{k=0}^{\infty} (1 + x^k) = \sum_{n=0}^{\infty} p(n)x^n$ | An equation due to Euler: on the left hand side we have an infinite product and on the right hand side appears a power series where $p(n)$ denotes the number of all possible representations of $n$ as a sum of natural numbers | 23         | 1     | 42    |
| 14         | $T = de + \omega \wedge e, R = d\omega + \omega \wedge \omega$ | Cartan structural equations, where $e$ is the frame field, $\omega$ the connection form, $T$ the torsion form and $R$ the curvature form   | 16         | 1     | 22    |
| 15         | $V_n(r) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} r^n$ | Volume of an $n$ -dimensional sphere   | 21         | 0     | 5     |
| 16         | $R_{\beta\gamma\delta\lambda}^{\alpha} = 0$                    | Second Bianchi identity of the Riemann tensor  | 10         | 2     | 7     |
| 17         | $1 = \sum_{n=2}^{\infty} (\zeta(n) - 1)$                       | The integer 1 expressed as an infinite series involving the Riemann zeta function  | 15         | 1     | 13    |
| 18         | $A = \int_{\sigma(\lambda)} \lambda dE_{\lambda}$              | A formulation of the spectral theorem expresses the operator as an integral of the coordinate function over the operator's spectrum with respect to a projection valued measure  | 11         | 0     | 26    |
| 19         | $\Delta\varphi = 0$  | Laplace equation   | 4          | 1     | 2     |
| 1          | $1 + e^{ix} = 0$   | Euler's identity links 5 fundamental mathematical constants with three basic arithmetic operations each occurring once   | 7          | 1     | 15    |
| 20         | $a^n + b^n = c^n, n > 2$                                       | Fermat's Last Theorem states that no three positive integers $a, b,$ and $c$ can satisfy this equation for any integer value of $n$ greater than two   | 11         | 1     | 26    |

Table 3 (continued)

| Stimuli ID | Equations   | Description  | Characters | Names | Words |
|------------|---|--|------------|-------|-------|
| 2          | $\int_M K dA + \int_{\partial M} k_g ds = 2\pi\chi(M)$  | The Gauss-Bonnet theorem connects the geometry of surfaces (curvature) to their topology (Euler characteristic)  | 20         | 3     | 14    |
| 3          | $\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}, s \in \mathbb{C}, \Re e(s) > 1$ | The reciprocal of the zeta function can be expressed as a Dirichlet series over the Möbius function $\mu(n)$   | 29         | 2     | 18    |
| 4          | $e = \lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n$   | An identity for Euler's number $e$   | 14         | 1     | 6     |
| 5          | $f(x) = \int_{-\infty}^{\infty} \delta(x-y)f(y)dy$  | Identity making use of the Dirac delta function  | 21         | 1     | 8     |
| 6          | $a^2 + b^2 = c^2$   | Pythagoras' theorem: In any right-angled triangle, the area of the square whose side is the hypotenuse (the side opposite the right angle) is equal to the sum of the areas of the squares whose sides are the two legs (the two sides that meet at a right angle) | 8          | 1     | 48    |
| 7          | $\frac{dx}{dt} = x(\alpha - \beta y), \frac{dy}{dt} = -\gamma(\gamma - \delta x)$             | Lotka-Volterra equations (predator-prey equations), are a pair of first-order, non-linear, differential equations used to describe the dynamics of biological systems in which two species interact, one a predator and one its prey                               | 27         | 2     | 32    |
| 8          | $\frac{d}{dx} e^x = e^x$  | The derivative of an exponential is an exponential   | 9          | 0     | 8     |
| 9          | $\ x+y\  \leq \ x\  + \ y\ $  | Triangle inequality for a normed vector space, the norm of the sum of two vectors is at most as large as the sum of the norms of the two vectors   | 13         | 0     | 30    |

**Table 4** Proof ranking

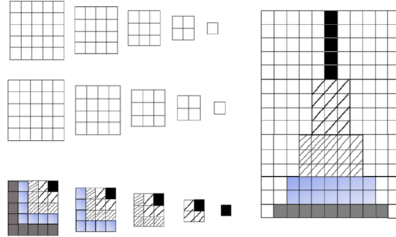
| Stimuli ID | Proof  | Ranking | Score |
|------------|--|---------|-------|
| 1          | <p><b>Theorem.</b> <i>There are infinitely many primes.</i></p> <p><i>Proof.</i> For any finite set <math>\{p_1, \dots, p_r\}</math> of primes, consider the number <math>n = p_1 p_2 \dots p_r + 1</math>. This <math>n</math> has a prime divisor <math>p</math>. But <math>p</math> is not one of the <math>p_i</math>: otherwise <math>p</math> would be a divisor of <math>n</math> and of the product <math>p_1 p_2 \dots p_r</math>, and thus also of the difference <math>n - p_1 p_2 \dots p_r = 1</math>, which is impossible. So a finite set <math>\{p_1, \dots, p_r\}</math> cannot be the collection of all prime numbers. <math>\square</math></p>  | 1       | 1.132 |
| 2          | <p><b>Theorem.</b> <i>The number <math>\sqrt{2}</math> is irrational.</i></p> <p><i>Proof (by contradiction).</i> Suppose toward contradiction that <math>\sqrt{2}</math> is rational. This means that <math>\sqrt{2} = p/q</math> for some integers <math>p</math> and <math>q</math>, with <math>q \neq 0</math>. We may assume that <math>p/q</math> is in lowest terms, which means that <math>p</math> and <math>q</math> have no nontrivial common factors. By squaring both sides, we conclude that <math>2 = p^2/q^2</math> and consequently <math>2q^2 = p^2</math>. This tells us that <math>p^2</math> must be an even number, which implies that <math>p</math> also is even, since the square of an odd number is odd. So <math>p = 2k</math> for some integer <math>k</math>. From this, it follows that <math>2q^2 = p^2 = (2k)^2 = 4k^2</math>, and therefore <math>q^2 = 2k^2</math>. So <math>q^2</math> also is even, and therefore <math>q</math> must be even. Thus, both <math>p</math> and <math>q</math> are even, which contradicts our assumption that <math>p/q</math> was in lowest terms. So <math>\sqrt{2}</math> cannot be rational. <math>\square</math></p> | 2       | 0.444 |
| 3          | <p><b>Theorem.</b> <i>Let <math>\langle a, b \rangle</math> be an inner product on a real vector space <math>V</math> (with the norm <math> a ^2 := \langle a, a \rangle</math>). Then</i></p> $\langle a, b \rangle^2 \leq  a ^2  b ^2$ <p><i>holds for all vectors <math>a, b \in V</math>, with equality if and only if <math>a</math> and <math>b</math> are linearly dependent.</i></p> <p><i>Proof.</i> Consider the quadratic function <math> xa + b ^2 = x^2  a ^2 + 2x \langle a, b \rangle +  b ^2</math> in the variable <math>x</math>. We may assume <math>a \neq 0</math>. If <math>b = \lambda a</math>, then clearly <math>\langle a, b \rangle^2 =  a ^2  b ^2</math>. If, on the other hand, <math>a</math> and <math>b</math> are linearly independent, then <math> xa + b ^2 &gt; 0</math> for all <math>x</math>, and thus the discriminant <math>\langle a, b \rangle^2 -  a ^2  b ^2</math> is less than 0. <math>\square</math></p>  | 3       | 0.304 |



**Table 4** (continued)

**4** **Theorem.** For all  $n \geq 1$ ,  $1^2 + 2^2 + 3^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$  4 0.185

*Proof.* Draw out the sum  $1^2 + 2^2 + 3^2 + \dots + n^2$ , three times as seen on the left of the below diagram. Then arrange it in a rectangle as seen on the right of the diagram with base  $2n + 1$ , and height  $1 + 2 + 3 + \dots + n = \frac{n(n+1)}{2}$



Focus on the rectangle on the right hand side of the diagram.

$$\text{The area of the rectangle} = 3(1^2 + 2^2 + 3^2 + \dots + n^2)$$

Given the rectangle area also equals base multiplied by height, we have:

$$(1 + 2 + 3 + \dots + n)(2n + 1) = 3(1^2 + 2^2 + 3^2 + \dots + n^2)$$

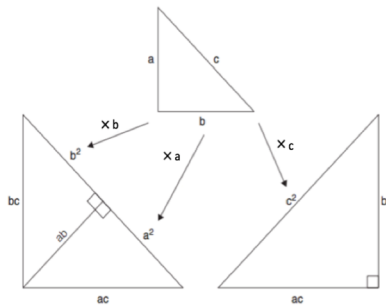
$$\frac{n(n+1)(2n+1)}{2} = 3(1^2 + 2^2 + 3^2 + \dots + n^2)$$

$$1^2 + 2^2 + 3^2 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$$

□

**5** **Theorem.** For a right-angled triangle with side lengths  $a$ ,  $b$  and  $c$  (where  $c$  is the hypotenuse),  $a^2 + b^2 = c^2$  5 0.012

*Proof.*



□

**Table 4** (continued)

|          |   |          |               |
|----------|---|----------|---------------|
| <b>6</b> | <p><b>Theorem.</b> <math>e^{ix} + 1 = 0</math></p> <p><i>Proof.</i> Consider the function, <math>f(x) = e^{-ix}(\cos(x) + i \sin(x))</math>, and differentiate using the product rule.</p> $f'(x) = -ie^{-ix}(\cos(x) + i \sin(x)) + e^{-ix}(-\sin(x) + i \cos(x))$ $= e^{-ix}(-i \cos(x) + \sin(x) - \sin(x) + i \cos(x))$ $= 0$ <p>If <math>f'(x) = 0</math>, then <math>f(x) = k</math>, <math>k</math> being a real number. Inputting <math>x = 0</math> into <math>f(x) = k</math> will find <math>k</math>.</p> $k = f(0) = e^{-ix0}(\cos(0) + i \sin(0))$ <p>So <math>k = 1</math></p> $f(x) = e^{-ix}(\cos(x) + i \sin(x)) = 1$ <p>So <math>\cos(x) + i \sin(x) = e^{ix}</math></p> <p>Substituting <math>x = \pi</math> completes the proof. <span style="float: right;">□</span></p>  | <b>6</b> | <b>-0.569</b> |
| <b>7</b> | <p><b>Theorem.</b> For any natural number <math>n</math>, the number <math>n^2 - n</math> is even.</p> <p><i>Proof (by induction).</i> The claim that <math>n^2 - n</math> is even is true when <math>n = 0</math>, since 0 is even. So it starts out true. To see that the fact is propagated from each number to the next, suppose that <math>n^2 - n</math> is even and consider the next instance, <math>(n + 1)^2 - (n + 1)</math>. By using some elementary algebra, this is equal to <math>n^2 + 2n + 1 - n - 1 = n^2 + n</math>, which can be written as <math>(n^2 - n) + 2n</math>. Thus, in moving from <math>n^2 - n</math> to <math>(n + 1)^2 - n^2</math> we have added <math>2n</math>, an even number. And so if the statement is true at a number <math>n</math>, then it remains true at <math>n + 1</math>, and so the truth of the statement is propagated from each number to the next. So by mathematical induction, it is true for all numbers. <span style="float: right;">□</span></p>   | <b>7</b> | <b>-0.678</b> |
| <b>8</b> | <p><b>Theorem.</b> Let <math>p(x)</math> be a non-constant polynomial with only real roots. We have <math>p'(x)^2 \geq p(x)p''(x)</math> for all <math>x \in \mathbb{R}</math></p> <p><i>Proof.</i> If <math>x = a_i</math> is a root of <math>p(x)</math>, <math>p(a_i) = 0</math> so there is nothing to show.</p> <p>Assume that <math>x</math> is not a root, using The product rule to differentiate <math>p(x) = (x - a_1)(x - a_2)...(x - a_n)</math></p> $p'(x) = (x - a_1) \frac{d}{dx} [(x - a_2)(x - a_3)...(x - a_n)] + 1 \cdot (x - a_2)(x - a_3)...(x - a_n)$ $p'(x) = \frac{(x - a_1)(x - a_2)...(x - a_n)}{(x - a_1)} + \frac{(x - a_1)(x - a_2)...(x - a_n)}{(x - a_2)} + \dots + \frac{(x - a_1)(x - a_2)...(x - a_n)}{(x - a_n)}$ $p'(x) = \sum_{k=1}^n \frac{p(x)}{x - a_k}$ $\frac{p'(x)}{p(x)} = \sum_{k=1}^n \frac{1}{x - a_k}$ <p>Using The quotient rule to differentiate <math>\frac{p'(x)}{p(x)} = \sum_{k=1}^n \frac{1}{x - a_k}</math>.</p> $\frac{p''(x)p(x) - p'(x)^2}{p(x)^2} = - \sum_{k=1}^n \frac{1}{(x - a_k)^2} < 0.$ <p>From this it follows that <math>p'(x)^2 \geq p(x)p''(x)</math> for all <math>x \in \mathbb{R}</math> <span style="float: right;">□</span></p> | <b>8</b> | <b>-0.830</b> |

**Acknowledgements** Many thanks to Dave Sirl, Ou hao Chen, Michael Barany, Andrew Aberdein, Brendan Larvor, Paola Iannone, Paul Hasselkuß, two anonymous reviewers, and attendees at the Novembertagung for helpful discussions about this work.

## Declarations

**Conflict of Interest** None.

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

## References

- Aigner, M., and G.M. Ziegler. 2010. *Proofs from THE BOOK*, 4th ed. Springer.
- Alexanderson, G.L. 1981. An interview with Paul Erdős. *The Two-Year College Mathematics Journal* 12 (4): 249–259.
- Berghaus, G. 1992. Neoplatonic and Pythagorean notions of world harmony and unity and their influence on Renaissance dance theory. *Dance Research: The Journal of the Society for Dance Research* 10 (2): 43–70.
- Bisson, M.-J., C. Gilmore, M. Inglis, and I. Jones. 2016. Measuring conceptual understanding using comparative judgement. *International Journal of Research in Undergraduate Mathematics Education* 2 (2): 141–164. <https://doi.org/10.1007/s40753-016-0024-3>.
- Bradley, R., and M. Terry. 1952. Rank analysis of incomplete block designs the method of paired comparisons. *Biometrika* 39: 324–345.
- Bréard, A. 2019. *Nine Chapters on Mathematical Modernity: Essay on the global historical entanglements of the science of numbers in China*, 1st ed. Cham: Springer.
- Che, J., Sun, X., Gallardo, V., & Nadal, M. 2018. Cross-cultural empirical aesthetics. In J. Christensen, A. Gomila, & V. Walsh (Ed.), *The arts and the brain: Psychology and physiology beyond pleasure: Volume 237(Progress in Brain Research)* (Vol. 237, pp. 77–103). Elsevier B.V.
- Cherniwchan, C., Ghassemi, A., and J. Keating. 2010. *Is a mathematical proof beautiful? Mathematical ethnographies project*. [Video]. YouTube. <https://www.youtube.com/watch?v=tANB6hXmY7U>.
- Cova, F., and N. Pain. 2012. Can folk aesthetics ground aesthetic realism? *The Monist* 95 (2): 241–263.
- Davies, B., L. Alcock, and I. Jones. 2020. Comparative judgement, proof summaries and proof comprehension. *Educational Studies in Mathematics* 105 (2): 181–197.
- Dreyfus, T., and T. Eisenberg. 1986. On the aesthetics of mathematical thought. *For the Learning of Mathematics* 6 (1): 2–10.
- DutilhNovaes, C. 2019. The beauty (?) of mathematical proofs. In *Advances in experimental philosophy of logic and mathematics*, ed. A. Aberdein and M. Inglis, 63–94. Bloomsbury.
- Erdős, P. 1983. Combinatorial problems in geometry. *Mathematical Chronicle* 12 (1): 35–54.
- Eysenck, H.J., and S. Iwawaki. 1971. Cultural relativity in aesthetic judgements: An empirical study. *Perceptual and Motor Skills* 32 (3): 817–818.
- Fan, L., and Y. Zhu. 2004. How have Chinese student performed in mathematics? A perspective from large-scale international mathematics comparisons. In *How Chinese learn mathematics: Perspective from insiders*, eds. L. Fan, W. Ngai-Ying, C. Jinfa, & Li. Shiqi, 3–26. World Scientific.
- Fan, L., Wong, N-Y., Cai, J., and S. Li. 2004. How Chinese learn mathematics: Perspectives . In *How Chinese learn mathematics: Perspective from insiders*, eds. L. Fan, W. Ngai-Ying, C. Jinfa, & Li. Shiqi, pp. vii–xii. World Scientific.
- Hardy, G. 1940. *A Mathematician's Apology*, 1st ed. Cambridge University Press.

- Hayn-Leichsenring, G., O. Vartanian, and A. Chatterjee. 2021. The role of expertise in the aesthetic evaluation of mathematical equations. *Psychological Research Psychologische Forschung*. <https://doi.org/10.1007/s00426-021-01592-5>.
- Heine, S.J., D.R. Lehman, K. Peng, and J. Greenholtz. 2002. What's wrong with cross-cultural comparisons of subjective likert scales?: The reference-group effect. *Journal of Personality and Social Psychology* 82 (6): 903–918.
- Heintz, C., and D. Taraborelli. 2010. Editorial: Folk epistemology. The cognitive bases of epistemic evaluation. *Review of Philosophy and Psychology* 1 (4): 477–482.
- Hudecek, J. 2012. Ancient Chinese mathematics in action: Wu Wen-Tsun's nationalist historicism after the cultural revolution. *East Asian Science, Technology and Society: An International Journal* 6 (1): 41–64.
- Hudecek, J. 2014. *Reviving ancient Chinese mathematics: Mathematics, history and politics in the work of Wu Wen-Tsun*, 1st ed. Springer.
- Inglis, M., and A. Aberdein. 2015. Beauty is not simplicity: An analysis of mathematicians' proof appraisals. *Philosophia Mathematica* 23 (1): 87–109.
- Inglis, M., and A. Aberdein. 2016. Diversity in proof appraisal. In *Mathematical cultures: The London meetings 2012–2014*, ed. B. Larvor, 163–179. Springer.
- Inglis, M., and A. Aberdein. 2020. Are aesthetic judgements purely aesthetic? Testing the Social Conformity Account. *ZDM* 52 (6): 1127–1136.
- Johnson, S.G.B., and S. Steinerberger. 2019. Intuitions about mathematical beauty: A case study in the aesthetic experience of ideas. *Cognition* 189: 242–259.
- Jones, I., M. Bisson, C. Gilmore, and M. Inglis. 2019. Measuring conceptual understanding in randomised controlled trials: Can comparative judgement help? *British Educational Research Journal* 45 (3): 662–680.
- Kim, M., & Yuan, Y. (2015). No cross-cultural differences in Gettier car case intuition: A replication study of Weinberg et al. 2001. *Episteme* 12(3):355–361.
- Knobe, J. 2007. Experimental philosophy. *Philosophy Compass* 2 (1): 81–92.
- Knobe, J. 2019. Philosophical intuitions are surprisingly robust across demographic differences. *Epistemology & Philosophy of Science* 56 (2): 29–36.
- Knobe, J. 2021. Philosophical intuitions are surprisingly stable across both demographic groups and situations. *Filozofia Nauki* 29(2): 11–76
- Lam, B. 2010. Are Cantonese-speakers really descriptivists? Revisiting cross-cultural semantics. *Cognition* 115(2): 320–329.
- Laming, D. 2003. *Human Judgment: The eye of the beholder*, 1st ed. CENGAGE Learning.
- Larvor, B. 2016. What are mathematical cultures? In *Cultures of mathematics and logic*, ed. S. Ju, B. Löwe, T. Müller, and X. Yun, 1–22. Birkhäuser.
- Lawlor, M. 1955. Cultural influences on preference for designs. *The Journal of Abnormal and Social Psychology* 51 (3): 690–692.
- Leung, F.K.S. 2001. In search of an east Asian identity in mathematics education. *Education Studies in Mathematics* 47(1): 35–51.
- Löwe, B., and B. Van Kerkhove. 2019. Methodological triangulation in empirical philosophy (of mathematics). In *Advances in experimental philosophy of logic and mathematics*, ed. A. Aberdein and M. Inglis, 15–37. Bloomsbury.
- Machery, E., R. Mallon, S. Nichols, and S. Stich. 2004. Semantics, cross-cultural style. *Cognition* 92 (3): B1–B12.
- Machery, E., S. Stich, D. Rose, A. Chatterjee, K. Karasawa, N. Struchiner, and T. Hashimoto. 2017. Gettier Across Cultures. *Noûs* 51 (3): 645–664.
- McAllister, J. W. 1996. *Beauty and revolution in science*. New York: Cornell University Press.
- McAllister, J.W. 2005. Mathematical beauty and the evolution of the standards of mathematical proof. In *The Visual Mind II*, ed. M. Emmer, 15–34. The MIT Press.
- McElroy, W. 1952. Aesthetic appreciation in Aborigines of Arnhem Land: A comparative experimental study. *Oceania* 23 (2): 81–95.
- Mejia Ramos, J., T. Evans, C. Rittberg, and M. Inglis. 2021. Mathematicians' assessments of the explanatory value of proofs. *Axiomathes* 31 (5): 575–599.
- Montano, U. 2014. *Explaining beauty in mathematics: An aesthetic theory of mathematics*, 1st ed. Cham: Springer.
- Nelsen, R.B. 2000. *Proofs without words II: More exercises in visual thinking*, 1st ed. The Mathematical Association of America.

- Pearcy, D. 2020. *Mathematical beauty: What is mathematical beauty and can anyone experience it?*, 1st ed. John Catt Educational.
- Plato. 1993. *Philebus*. Hackett Publishing Company.
- Pollack, I. 1952. The information of elementary auditory displays. *Journal of the Acoustical Society of America* 24 (1): 745–749.
- Rota, G.-C. 1997. The phenomenology of mathematical beauty. *Synthese* 111 (2): 171–182.
- Seyedsayamdost, H. 2015. On normativity and epistemic intuitions: Failure of replication. *Episteme* 12 (1): 95–116.
- Simoniti, V. 2017. Aesthetic properties as powers. *European Journal of Philosophy* 25 (4): 1434–1453.
- Sinclair, N., and D. Pimm. 2006. A historical gaze at the mathematical aesthetic. In *Mathematics and the aesthetic: New approaches to an ancient affinity*, ed. N. Sinclair, D. Pimm, and W. Higginson, 1–17. Springer.
- Soueif, M., and H. Eysenck. 1971. Cultural differences in aesthetic preferences. *International Journal of Psychology* 6 (4): 293–298.
- Starikova, I. 2017. Aesthetic preferences in mathematics: A case study. *Philosophia Mathematica* 26 (2): 161–183.
- Stich, S. 2001. Plato's method meets cognitive science. *Free Inquiry* 21 (2): 36–38.
- Stich, S., and E. Machery. 2022. Demographic Differences in Philosophical Intuition: A reply to Joshua Knobe. *Review of Philosophy and Psychology*. <https://doi.org/10.1007/s13164-021-00609-7>.
- Tatarkiewicz, W. 1963. Objectivity and subjectivity in the history of aesthetics. *Philosophy and Phenomenological Research* 24 (2): 157–173.
- Thurstone, L. 1928. Attitudes can be measured. *The American Journal of Sociology* 33 (4): 529–554.
- Thurstone, L. 1994. A law of comparative judgement. *Psychological Review* 101 (2): 266–270.
- Todd, C. 2008. Unmasking the truth beneath the beauty: Why the supposed aesthetic judgements made in science may not be aesthetic at all. *International Studies in the Philosophy of Science* 22 (1): 61–79.
- Weinberg, J., S. Nichols, and S. Stich. 2001. Normativity and epistemic intuitions. *Philosophical Topics* 29 (1): 429–460.
- Wells, D. 1990. Are these the most beautiful? *The Mathematical Intelligencer* 12 (3): 37–41.
- Zajonc, R.B. 2001. Mere Exposure: A gateway to the subliminal. *Current Directions in Psychological Science* 10 (6): 224–228.
- Zeki, S., J.P. Romaya, D.T. Benincasa, and M. Atiyah. 2014. The experience of mathematical beauty and its neural correlates. *Frontiers in Human Neuroscience* 8 (1): 1–12.
- Zong, C. 2020. Blaschke, Osgood, Wiener, Hadamard and the early development of modern mathematics in China. 1–8. <https://doi.org/10.48550/arXiv.2009.13688>

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