Routledge
Taylor & Francis Group

Philosophical Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cphp20

Further exploration of anti-realist intuitions about aesthetic judgment

James Andow

To cite this article: James Andow (2021): Further exploration of anti-realist intuitions about aesthetic judgment, Philosophical Psychology, DOI: 10.1080/09515089.2021.2014440

To link to this article: https://doi.org/10.1080/09515089.2021.2014440

Published online: 27 Dec 2021.

PHILOSOPHICAL PSYCHOLOGY
https://doi.org/10.1080/09515089.2021.2014440

Routledge
Taylor & Francis Group

ORIGINAL PAPER

Further exploration of anti-realist intuitions about aesthetic judgment

James Andow

Department of Philosophy, University of East Anglia, Norwich, UK

ABSTRACT

Experimental philosophy of aesthetics has explored to what extent ordinary people are committed to aesthetic realism. Extant work has focused on attitudes to normativism – a key commitment of realist positions in aesthetics – the claim that aesthetic judgments/statements have correctness conditions, invariant between subjects, such that there is a fact of the matter in cases of aesthetic disagreement. The emerging picture is that ordinary people strongly and almost universally reject normativism and thus there is no strong realist tendency in ordinary people's thinking about the aesthetic. This has been taken to dissolve the traditional puzzle in aesthetics of how to best account for the fact that (a) aesthetic judgments seem intersubjectively valid, while (b) aesthetic experience seems subjective. This paper presents studies which further enrich our understanding of ordinary thinking about the aesthetic: ordinary thinking about the aesthetic may not be so vehement in its rejection of normativism; and where previous results suggested that, in many cultures, the dominant trend is to reject correctness conditions for aesthetic judgments, the current results suggest participants think aesthetic judgments have correctness conditions (albeit perhaps very finely relativized to specific circumstances of judgment).

ARTICLE HISTORY
Received 29 April 2020
Accepted 30 November 2021

KEYWORDS
aesthetics; realism; intuitions; normativism; experimental philosophy; experimental aesthetics; relativism; expressivism

1. The normativism puzzle

It is widely accepted among aestheticians that *normativism* is widely accepted among the population: aesthetic judgments are either correct or incorrect and their correctness conditions are invariant between subjects so that in cases of aesthetic disagreement at most one party can be correct. When I say something is beautiful and you say it isn't, normativism says one of us is wrong. That normativism is widely accepted among the population is taken as a data-point in debate about aesthetic realism. The puzzle is – how best to explain this given the apparent subjectivity of aesthetic experience? A major consideration in favor of realism is thought to be that realism can easily account for why we act as though normativism is true.

CONTACT James Andow jamesandow@gmail.com University of East Anglia, Norwich, UK

What is aesthetic realism? Different authors make the distinction between realism and antirealism in slightly different ways. But Eaton (1998) captures the idea nicely.

> At least some aesthetic judgements of the form 'Object O has property F' are true. There are ways of adjudicating disputes about whether O is F. It is not the case that whatever anybody says about the aesthetic properties possessed or not possessed by O is as good as whatever anybody else says. Radical relativism of this sort is false.

Aesthetic realism is at least committed to aesthetic judgments being in the business of stating facts and there being such facts.[1]

Aesthetic realism thus can neatly explain why we act as if normativism is true: it *is* true; aesthetic judgments state non-relative facts about objects; those facts obtain or don't. Other positions can't neatly explain this fact. If there are no aesthetic facts, or if aesthetic judgments are not in the business of stating non-relative facts, a more complicated story is needed about how aesthetic judgments come to be correct/incorrect, and how the correctness conditions are such that in aesthetic disputes at most one party is correct. It is not impossible to tell such a story, of course, but the availability of a simple explanation has been taken to be a key consideration in favor of some form of realist position.

2. Experimental work claimed to dissolve the puzzle

As we've seen, philosophical discussion about normativism and realism is premised on certain empirical claims. Kant's the usual citation here:

> It would be ridiculous if ... someone who prided himself on his taste thought to justify himself thus: "This object ... is beautiful for me." For he must not call it beautiful if it pleases merely him. ... [I]f he pronounces that something is beautiful, then he expects the very same satisfaction of others: he judges not merely for himself, but for everyone, and speaks of beauty as if it were a property of things. Hence he says that the thing is beautiful, and does not count on the agreement of others with his judgment of satisfaction because he has frequently found them to be agreeable with his own, but rather demands it from them. He rebukes them if they judge otherwise, and denies that they have taste, though he nevertheless requires that they ought to have it; and to this extent one cannot say, "Everyone has his special taste." (Kant, 2000, pp.98, Ak. 5,212–3)

The challenge for aestheticians is then seen to be how best to account for our ordinary ways of thinking about the aesthetic. It is commonsense that these empirical matters should be investigated – simply so that we understand what it is that our philosophical accounts should account for.

This is the motivation for the recent work in experimental philosophy of aesthetics by Cova and colleagues.[2] The first set of studies looking at folk aesthetic realism was Cova and Pain (2012).[3] Their measure had the following form and intended to capture the extent to which participants embrace normativism.[4]

> Agathe and Ulrich are on holidays in the country. While having a walk in the fields, they hear a nightingale singing. Agathe says: "What beautiful singing!" But Ulrich answers: "No. It's definitely not beautiful." According to you:(1) One of them is right and the other is not.(2) Both are right.(3) Both are wrong.(4) Neither is right or wrong. It makes no sense to speak in terms of correctness in this situation. Everyone is entitled to his own opinion.

Only 1 is interpreted as consistent with normativism; 2 & 3 are interpreted as indicating some kind of relativism; 4 is interpreted as indicating some simple form of expressivism.[5] Participants considered 12 scenarios: three in which 'beautiful' was applied to natural objects; three to people (with photos); three involving uncontroversially objective predicates (e.g., 'written by'); and three involving paradigmatic subjective predicates (e.g., 'tastes good'). A second experiment did the same with 'ugly'. The results were striking. For both 'beautiful' and 'ugly', participants selected the normativist option less than once on average per domain. This much more closely resembled participants' pattern of responses for subjective cases than for objective cases (where the average number of normativist responses neared the maximum of three). The results were similar for a third experiment that asked participants to describe something they personally found very beautiful and then to imagine someone disagreeing with them.

More recently, Cova et al. (2018) 'report the results of a cross-cultural study with over 2,000 respondents spanning 19 countries' intending to investigate whether Cova and Pain's findings generalize and so 'provide a much stronger test of the hypothesis that common sense is committed to the intersubjective validity of aesthetic judgments.' Cova et al. (2018) use a similar paradigm to Cova and Pain (2012). Participants were asked about a disagreement case in which they were a protagonist. The disagreement was about something with respect to which participant had a strong aesthetic opinion. The full text presented to participants was as follows:[6]

> (1) Describe something (e.g., a natural object, or some work of art) that you find very beautiful.
>
> (2) Now, imagine that you meet someone and that this person says to you that he does not find this thing beautiful at all. In your personal opinion, which of the following best describes this situation:[7]

(a) One of you is correct while the other is not.

(b) Both of you are correct.

(c) Neither is correct. It makes no sense to talk about correctness in this situation.

(3) How certain are you of your response to Question 2 on a (0–100)% scale, with low numbers indicating that you are not sure and high numbers indicating that you are sure? I am % certain of my response.

Overall, only 7% of participants selected option (a) (the normativist option), and normativist responses were a clear minority across all subgroups. Cova et al. (2018) found cross-cultural variation in whether participants preferred the 'Both are correct' answer or the 'Neither is correct' answer (although overall the latter was most popular).[8]

Andow (2019) found similar results in a study on aesthetic testimony. Attitudes to normativism were measured as an independent variable using a similar disagreement case:

While visiting an art gallery, Alex and Harry look at a painting. Harry says that the painting is beautiful, while Alex says that the very same painting is not beautiful.

Participants then rated the statements as used in Cova and Pain (2012) on a scale (1–7) rather than forced choice. The results paint a similar picture, however. Participants tended to disagree with normativist option ($m = 1.98$) and 'Both are wrong'($m = 2.18$), and to agree with 'Both are right' ($m = 4.84$) *and* 'Neither is right or wrong' ($m = 6.53$).[9]

Cova et al. (2018) want us to take the following lesson from such results.

... one widespread argument for aesthetic realism is the following: ... aesthetic realism ... is true because it provides the best explanation for the fact that people attribute intersubjective normativity to aesthetic judgments. ... However, because our results suggest that what aesthetic realism purports to explain is not the case, so that there is nothing to explain, this argument can no longer provide evidence in favor of aesthetic realism. Thus, by drastically revising our beliefs about the content of common sense, our results undercut some powerful arguments for aesthetic realism. They also change the dialectical equilibrium: It is not true that those who reject aesthetic realism bear the burden of proof because they have to provide a compelling error theory explaining why common sense is committed to aesthetic realism...

They conclude 'the traditional way of approaching the debate over the nature of aesthetic judgment is fundamentally misguided' on the basis of a more empirically-informed understanding of the nature of ordinary thinking about the aesthetic for which the philosophical accounts of the aesthetic need to be able to account.[10]

3. Where next?

A common pattern in various areas of experimental philosophy is that central philosophical problems are characterized by stable tensions between intuitions that pull in directions that are not easily reconcilable (Knobe, 2021). It would thus not be surprising to find that ordinary people's understanding of aesthetic judgments was more equivocal that is reflected in the results of existing studies around the issue of aesthetic realism.

And, in fact, there are some reasons to think that the key question format used by Cova and colleagues may not pick up on the full subtlety of ordinary thinking about the aesthetic. For example, any participant who feels there is an important difference between the aesthetic and everyday descriptive matters can only express this, given a forced choice paradigm, by selecting an option other than the first option, which is automatically interpreted as rejecting normativism.[11] As a result, it is still somewhat tenable to maintain that the above results may reflect only a tendency to recognize that the aesthetic and the descriptive are different in some important respect (rather than any precise claim about the nature of the difference participants are tracking). Consequently, it is important that experimental aesthetics explores a range of question designs to ensure a full understanding of ordinary thinking about the aesthetic and, in particular, to investigate what these previous results really tell us about ordinary people's understanding of aesthetic judgment.[12]

Indeed, there is evidence that different question designs might pick up on different aspects of the way ordinary people think about the aesthetic. Andow (2020) obtains results that are naturally interpreted as suggesting that ordinary thinking about normativism is somewhat more equivocal. He uses the following case and response options:

> John and Fred are in an argument. John says, "Van Gogh's sunflower paintings are beautiful," and Fred says, "No, Van Gogh's sunflower paintings are not beautiful."
>
> (1) Van Gogh's sunflower paintings are beautiful, so John is right and Fred is wrong.
>
> (2) Van Gogh's sunflower paintings are not beautiful, so Fred is right and John is wrong.
>
> (3) I don't know whether Van Gogh's sunflower paintings are beautiful, so I don't know which of them is right and which is wrong, but one of them is wrong.
>
> (4) There is no fact of the matter about unqualified claims like "Van Gogh's sunflower paintings are beautiful." Different people believe different things, and it is not absolutely true or false that Van Gogh's sunflower paintings are beautiful.

Some participants rated each item individually rather than facing a forced choice. Options 1–3 are consistent with normativism. Expressivists, relativists, and error theorists were all expected to agree with 4. Reanalysis of Andow (2020)'s data finds that item 1 received 30% agreement, item 2

received 20% agreement, and item 3 received 46% agreement, with 48% indicating agreement with at least one of these 'normativist' items. Perhaps ordinary thinking about normativism in aesthetics is slightly more divided than Cova and colleagues' results might suggest.

Another piece of evidence, that how disagreement cases and response options are worded can affect sensitivity to normativist strands within ordinary thinking, comes from the experimental philosophy of color. Cohen and Nichols (2010) originally report that participants' responses to disagreement about color are very ambiguous and divided on the basis that 47% of participants selected a similar item to Cova and Pain (2012)'s fourth, 'There is no fact of the matter about unqualified claims like 'the tomato is red'. Different people have different visual experiences when they look at the same object' (call this the 'non-normativist' item'), over normativist items in a forced choice. However, Roberts et al. (2014) then use a different design and obtain results that suggest ordinary thinking about color is less ambiguous and more committed to normativism. Roberts et al. (2014)'s participants received cases such as the following:

> Alex and Harry examine an object. Alex and Harry examine the object in typical lighting from the same position. They are both fluent English speakers and have normal eyesight. Harry says that the object is red, while Alex says that the very same object is green.

Roberts et al. (2014) asked participants to rate seven statements on a ten-point scale. Only the final item was of interest (in primary analyses), the first six statements (randomized) were presented to disambiguate the final statement and to give participants the opportunity to express attitudes which they might otherwise have used the final item to express. The statements were as follows:

> **Epistemic** We could find out who is right about the color of the object.
>
> **Fault** One of them, and possibly both, is at fault for getting the color wrong.
>
> **Appearance** The object may appear in different ways to Alex and Harry, and so, for all we know, both of them could be correctly reporting how the object appears to them.
>
> **Meaning** Alex and Harry may only disagree about what the words "red" and "green" mean, and so given how they may be individually using the words, they could for all we know both be right about the color of the object.
>
> **Verbal** People often disagree about what word best describes how an object appears. For example, people often disagree about whether something should be called "red" or "orange".
>
> **Perceptual** People disagree a lot about what colors things perceptually appear to have.

Target In reality, there is an absolute fact of the matter about the color of the object regardless of how it appears to Alex and Harry and regardless of what they think, say, or do.

In Roberts et al. (2014), 72% of participants indicated agreement with the target statement in the color cases (75% when excluding participants with postgraduate experience in philosophy). So these results also suggests the format used by Cova and colleagues in experimental aesthetics, which is very similar to the question format in Cohen and Nichols (2010), may not capture all aspects of ordinary people's understanding of aesthetic judgment.[13]

There are other interesting questions about how to interpret the results of studies like Cova and Pain (2012) and Cova et al. (2018). As Cova et al. (2018) note, there are interesting questions about what lies behind the apparent variation (including some cultural variation) in preference for the 'both are correct' and the 'neither is correct' responses to disagreement cases. Should we interpret participants who select these options as having specific stances as to whether aesthetic judgments have correctness conditions? One approach that might help us make progress on this point – help us better understand the relation between participants' use of these response options and distinctions relevant in the philosophical discussion – is inspired by an example from the experimental philosophy of pain. The main narrative in this literature is that philosophers have tended to believe, incorrectly, that ordinary thinking about pains was committed to a mental state view of pains. Such views hold there is no appearance-reality gap for pains; it is impossible to feel as if you have a pain, but not have a pain; it is impossible to feel no pain, yet have a pain; the correctness of my pain attributions is determined by the private experiences of the person to whom I am attributing pain, there is no external experience-independent standard; in first personal cases, there is a certain immunity from error when making pain attributions. Various studies have put pressure on this idea in various ways (see Sytsma & Reuter, 2017). One type of question asked of participants is the direct, 'Do you think that it is possible to feel a pain as being hurtful even though it is really not hurtful at all?' Sytsma and Reuter (2017)'s results using such questions suggest a strong tendency to envisage the possibility of an appearance-reality gap and thus of error in first personal pain attributions. The possibility of error is also at stake when it comes to how to interpret previous studies on normativism about aesthetic judgment. Realist positions straightforwardly accept the possibility of error in aesthetic judgment in all circumstances. Some anti-realist positions deny that aesthetic judgments can be correct or incorrect. A different (somewhat extreme) anti-realist position would provide complete immunity from certain kinds of error: that your judgment that something is beautiful is correct

so long as the thing seems beautiful to you. In between, there are various different positions as to what would render one open to the possibility of error. E.g., a cultural relativist will think that error is precluded so long as certain conditions are met, e.g, judgments are formed in line with some culturally relevant standards of taste. So, questions focused on the possibility of error, such as those in Sytsma and Reuter (2017), may provide an alternative way to explore ordinary thinking about the correctness conditions of aesthetic judgment.

4. New empirical studies

4.1. Study 1: Disagreement across cultures

This study's aim was to try out a new question format (based on that developed by Roberts et al., 2014). The hypothesis was that the results would see lower numbers of participants rejecting the idea that there's a fact of the matter in aesthetic disputes than seen using Cova and Pain (2012)-style questions. Before they respond to a target item, in this new format, participants are asked about a series of statements giving them opportunity to express attitudes which they might otherwise have used the target item to express.

4.1.1. Participants

Participants in all studies in this paper were recruited online through Prolific (prolific.ac). Participation was restricted to participants whose first language was English and who were currently resident in the UK or US. No participant participated in more than one of the studies reported in this paper. The number of participants in Study 1 after exclusions was 142.[14] Participants were 44% Male, 55% Female, 1% Other. Participants' mean age was 34.52 years old ($SD = 10.60$). The majority had no philosophical training (85%), but some indicated philosophical training at undergraduate (16%) or postgraduate (4%) level.

4.1.2. Materials

All questions concerned the following scenario (presented at the top of each page).

> Two people from cultures with very different values disagree about whether an object is beautiful. One person, like most people in her culture, judges that it is beautiful. The other person, like most people in her culture, judges that the very same object is not beautiful.

The main questions were presented in two blocks (randomized). Block A contained the new question format modeled on Roberts et al. (2014). In Block A, following two attention checks, participants indicated their agreement with each of the following statements on a seven-point scale with verbal anchors (Strongly Disagree, to Strongly Agree) at each point (3–8 randomized, 9 always last). As in Roberts et al. (2014), only statement 9 is a measure of normativism (3–8 are used only to disambiguate the target item, 9).[15]

(3) One of them, and possibly both, deserves blame for being wrong about whether the object is beautiful.
(4) The object may appear in different ways to the two people and so, for all we know, both of them are correctly reporting how they experience the object.
(5) We could find out who is right about whether the object is beautiful.
(6) The two people may only disagree about what the word 'beauty' means, and so given how they may be individually using the words, they could for all we know both be right about the beauty of the object.
(7) People often disagree about what word best describes how an object appears. For example, people often disagree about whether something should be called 'beautiful' or 'pretty'.
(8) People disagree a lot about the visual appearance of objects. For example, people often disagree about an object's exact shape and color.
(9) In reality, the object is either beautiful or it is not beautiful, regardless of whether these two people think it is beautiful or would describe it as 'beautiful'.

Block B asked about the same scenario and, like Cova and Pain (2012), gave participants a forced choice between:

(1) One of them is right and the other is not.
(2) Both are right.
(3) Both are wrong.
(4) Neither is right or wrong. It makes no sense to speak in terms of correctness in this situation. Everyone is entitled to his own opinion.

4.1.3. Block A results

The mean response to the target item was 2.83 ($SD = 1.67$) which is significantly lower than the midpoint of the scale ($t(141) = 8.36, p < .001, d = .70$) with 17% of participants indicating some level of agreement, 69% some level of disagreement, and 14% at the midpoint. Results for all items are given in Table 1.

Table 1. Descriptive statistics for statements 3–9 (Study 1) including midpoint comparison in a one sample t-test, and predictive value in a simple linear regression model predicting statement 9.

Statement	% Agree	M	SD	$t(141) =$	R^2
3	5%	2.01	1.23	$19.21, p < .001, d = 1.62$	.009
4	8%	1.63	0.94	$30.01, p < .001, d = 2.52$	.009
5	92%	5.96	1.07	$21.78, p < .001, d = 1.83$	.004
6	76%	5.26	1.35	$11.12, p < .001, d = .93$	.019
7	85%	5.46	1.12	$15.49, p < .001, d = 1.30$	.002
8	80%	5.18	1.35	$10.48, p < .001, d = .87$	.000
9	17%	2.83	1.67	$8.36, p < .001, d = .70$	–

Table 2. Proportion of participants selecting each of the options in Block B (Study 1) including the results of a one-sample chi-square test testing the null hypothesis that the probability that participants select the relevant option is 0.25.

Statement	% Agree	$\chi^2(1, 142) =$
1	1%	$44.70, p < .001$
2	10%	$17.36, p < .001$
3	0%	–
4	89%	$314.45, p < .001$

4.1.4. Block B results

Only one participant gave a response which would be interpreted as normativist ('One of them is right and the other is not'). A breakdown of the other responses is given in Table 2.

4.1.5. Effects of question design

Blocks A and B both measure rejection of normativism. As can been seen in Figure 1, the degree of coherence was high: the majority (68%) gave the same response in Block A (disagreement vs agreement with target item) and Block B (selection of 1 vs one of 2–4). Nonetheless, among those participants for whom block made a difference, the direction was clear: only one gave a normativist response in Block B and not in Block A (Mcnemar's: $\chi^2(1) = 39.20, p < .001, OR = 44$).

4.1.6. Demographic effects

There were no gender or age effects, or effects of studying philosophy at either graduate or undergraduate level.

4.1.7. Discussion

The question design has an important influence on results. The forced choice design (as per Cova & Pain, 2012) would lead us to conclude that 99% of participants reject realism (given its commitment to normativism). However, the new design (based on Roberts et al., 2014) would lead to only 69%.[16] So, we should be a little cautious in interpreting previous

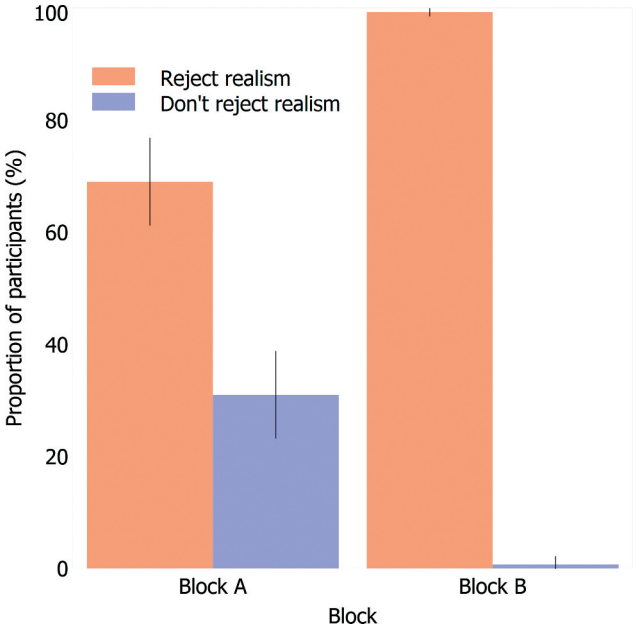

Figure 1. Graph showing proportion of participants giving responses that would be interpreted as rejecting and not rejecting realism in Blocks A and B (Study 1) showing that question design makes an important difference. Error bars indicate 95% Confidence Interval.

results as indicating near unanimous rejection of normativism. Although, the results still suggest a healthy majority of participants rejecting this realist picture of the aesthetic.

Cross-cultural disagreements are a good testing ground for intuitions about normativism because even cultural relativists should reject the idea that one of the parties must be wrong in such cases. However, we can go beyond this to gain a richer picture of what participants might think the correctness conditions of aesthetic judgments are relativized to.

4.2. Study 2: Disagreement within culture

This was as Study 1 except the scenario concerned within-culture disagreement.[17]

> Two people from the same culture disagree about whether an object is beautiful. One person judges that it is beautiful. The other person judges that the very same object is not beautiful.

If responses in Study 1 flow from widespread commitment to cultural relativism, responses should be different now the crosscultural factor is absent.

4.2.1. Participants

The number of participants after exclusions was 112.[18] The gender split was 63% female, 37% male, and 1% other. Participants' mean age was 33.13 years old ($SD = 12.23$). The majority had no philosophical training, 10% had some undergraduate, 1% had some postgraduate.

4.2.2. Block A results

The mean response to the target item was 2.62 ($SD = 1.79$), this is significantly lower than the midpoint ($t(111) = 8.17, p < .001, d = .77$) with 19% of participants indicating some level of agreement, 69% some level of disagreement, and 13% at the midpoint. Full descriptive results for the other statements (3–9) can be found in Table 3.[19]

4.2.3 Block B results

All participants gave a response which would be interpreted as non-normativist/anti-realist. A breakdown is given in Table 4.

4.2.4. Effects of question design

As can be seen in Figure 2, the degree of coherence between Block A and B was high: the majority gave the same response in both (69%) and, among participants for whom block made a difference, all of them gave the normativist response in Block B ($\chi^2(1) = 33.03, p < .001$).[20]

Table 3. Descriptive statistics for statements 3–9 (Study 2) including midpoint comparison and predictive value of a simple regression model predicting statement 9.

Statement	% Agree	M	SD	t(111) =	R^2
3	5%	2.16	1.37	$14.18, p < .001, d = 1.34$	.199
4	7%	1.94	1.43	$15.28, p < .001, d = 1.44$	.162
5	94%	6.22	1.00	$23.48, p < .001, d = 2.21$	.020
6	87%	5.65	1.20	$14.59, p < .001, d = 1.37$	.020
7	85%	5.68	1.14	$15.57, p < .001, d = 1.47$	.004
8	72%	5.09	1.57	$7.35, p < .001, d = .69$	.015
9	19%	2.62	1.79	$8.17, p < .001, d = .77$	–

Table 4. Proportion of participants selecting each of the options in Block B (Study 2) including the results of a one-sample chi-square test testing the null hypothesis that the probability that participants select the relevant option is 0.25.

Statement	% Agree	$\chi^2(1, 136)$
1	0%	–
2	12%	$10.71, p = .001$
3	0%	–
4	88%	$240.05, p < .001$

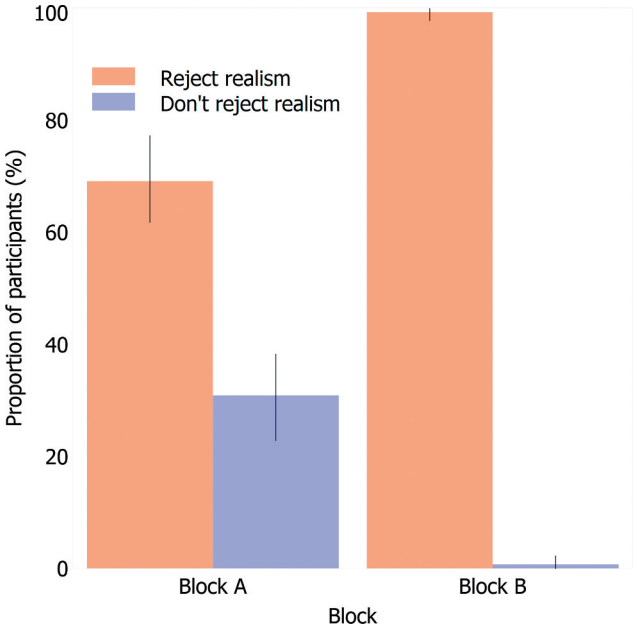

Figure 2. Graph showing proportion of participants giving responses that would be interpreted as rejecting and not rejecting realism in Blocks A and B (Study 2) showing that question design makes an important difference. Error bars indicate 95% Confidence Interval.

4.2.5. Demographic effects

There were no gender or age effects or effects of studying philosophy at either graduate or undergraduate level.

4.2.6. Discussion

The results are very similar to Study 1. The Cova-style design would lead us to report 100% denying that there is a fact of the matter whereas the new design leads to the rather lower figure of 69%. This suggest again that we should be cautious in interpreting results using the Cova-style questions as indicating near unanimous rejection of normativism.

This result also puts pressure on the idea that the results of Study 1 reflect a commitment to cultural relativism. The pattern of results hasn't shifted at all between Studies 1 and 2, suggesting that commitment to cultural relativism is not a big factor in leading participants to reject option 1 in Block B and reject the target option in Block A. Rather what the pattern of results suggests is that what drives the rejection of those options is a tendency to think of the correctness conditions of aesthetic judgments as relativized at a finer grain, e.g., to subcultures, or to individuals. If participants' responses in Study 1 and Study 2 reflect widespread commitment to some kind of subjectivism, for example, we should expect to see responses shift when the interpersonal factor is absent.

4.3. Study 3: Disagreement within subject

Studies 1 and 2 suggest people are typically inclined to deny that there is a fact of the matter about the aesthetic. However, the trend is less strong and the results more ambiguous than might be suggested by previous results. In fact, in light of previous studies, the current results seem to suggest surprisingly high levels of agreement that there is a fact of the matter. Studies 1 and 2 give this result with respect to intercultural and interpersonal disagreements respectively. Study 3 now extends this question to intrapersonal disagreements: a radically changing opinion over a brief period of time. Will we again see surprisingly high levels of agreement with the idea that there is a fact of the matter now that the disagreement is intrapersonal? If participants' responses in Study 1 and Study 2 reflect widespread commitment to some kind of subjectivism, we should expect to see higher levels of agreement with the idea that there is a fact of the matter.

4.3.1. Participants

The number of participants after exclusions was 123.[21] Of the remainder, 39% were male, 60% female, 1% other, mean age was 32.50 ($SD = 11.14$), 15% had some undergraduate training in philosophy, none had postgraduate training.

4.3.2. Materials

Participants read the following scenario.

> Yesterday, John judged that a particular object was beautiful. Today, although the object hasn't changed at all, John judges that the very same object is not beautiful.

Participants then responded to the following two blocks of questions in a random order. After answering comprehension questions, Block A asked participants to rate the following on a seven-point scale as above (3–8 randomized, 9 always last).

(3) We could find out whether John was right yesterday or is right today about whether the object is beautiful.
(4) John deserves blame for being wrong about whether the object is beautiful either yesterday or today, or on both occasions.
(5) The object may appear in different ways to John at different times and so, for all we know, John is correctly reporting how he experiences the object both yesterday and today.
(6) John may be using the word 'beauty' to mean different things at different times, and so given how he may be using the words, for all we know John could both have been right yesterday and be right today about the beauty of the object.

(7) People often change their opinions about what word best describes how an object appears. For example, people often change their opinions about whether something should be called 'beautiful' or 'pretty'.
(8) People often change their opinions about the visual appearance of objects. For example, people often change their opinions about an object's exact shape and color.
(9) In reality, the object is either beautiful or it is not beautiful, regardless of whether John thinks it is beautiful or would describe it as 'beautiful' either yesterday or today.

In Block B, participants had to select one of the following:

(1) John was either right yesterday or is he is right today, but not both.
(2) John was right yesterday and he is also right today.
(3) John was wrong yesterday and he is also wrong today.
(4) Neither of John's judgments is right or wrong. It makes no sense to speak in terms of correctness in this situation. Everyone is entitled to change their mind.

4.3.3. Block A results

The mean response to the target item was 3.08 ($SD = 1.86$), this is significantly lower than the midpoint ($t(122) = 5.47, p < .001, d = .49$) with 20% of participants indicating some level of agreement, 63% some level of disagreement, and 16% at the midpoint. Full descriptive results for all statements can be found in Table 5.[22]

4.3.4. Block B results

All but one participant (99%) gave a response which would be interpreted as non-normativist/anti-realist. A breakdown of non-normativist responses is given in Table 6.[23]

Table 5. Descriptive statistics for statements 3–9 (Study 3) including midpoint comparison and predictive value of a simple regression model predicting statement 9.

Statement	% Agree	M	SD	t(122) =	R^2
3	15%	2.72	1.69	$8.40, p < .001, d = .76$	.02
4	2%	1.65	1.04	$25.06, p < .001, d = 2.26$	.01
5	96%	6.07	.95	$24.18, p < .001, d = 2.18$	.01
6	84%	5.40	1.35	$11.51, p < .001, d = 1.04$	.01
7	88%	5.68	1.18	$15.78, p < .001, d = 1.42$	.00
8	76%	5.19	1.49	$8.84, p < .001, d = .80$	.01
9	20%	3.08	1.86	$5.47, p < .001, d = .49$	–

Table 6. Proportion of participants selecting each of the options in Block B (Study 6) including the results of a one-sample chi-square test testing the null hypothesis that the probability that participants select the relevant option is 0.25.

Statement	% Agree	$\chi^2(1, 123) =$
1	1%	$38.38, p < .001$
2	8%	$18.67, p < .001$
3	0%	–
4	91%	$286.25, p < .001$

4.3.5. Effects of question design

As can be seen in Figure 3, most participants gave consistent responses to Block A and B (64%). Among those whose answer was different, all selected the normativist response in Block B ($\chi^2(1) = 42.02, p < .001$).[24]

4.3.6. Demographic effects

Participants who had studied philosophy before at undergraduate level ($M = 3.89, SD = 2.05$) gave higher ratings than other participants ($M = 2.93, SD = 1.78$) of the target item in Block A ($t(121) = 2.10, p = .038, d = .498$). No other effects of gender, age or studying philosophy were observed.

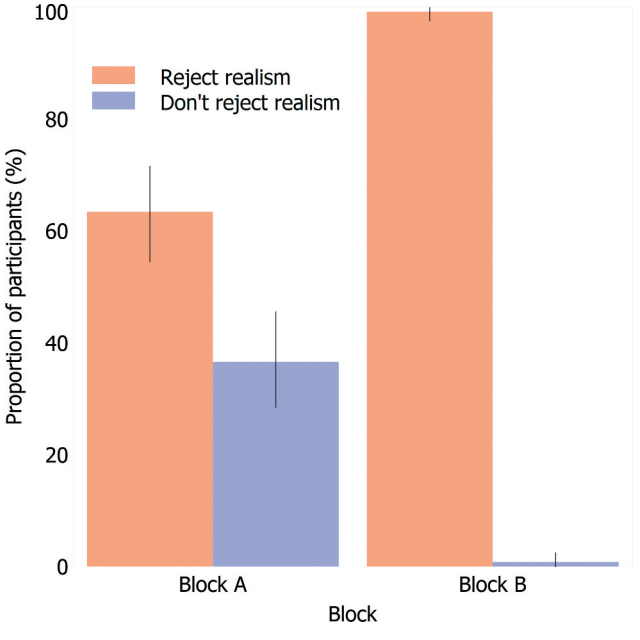

Figure 3. Graph showing proportion of participants giving responses that would be interpreted as rejecting and not rejecting realism in Blocks A and B (Study 3) showing that question design makes an important difference. Error bars indicate 95% Confidence Interval.

4.3.7. Discussion

How anti-realist are ordinary people? To what extent are they committed to there being no fact of the matter? Again, the impression we get is different depending on how the question is asked. The Cova-style question makes it seem the situation is cut and dry. But using the new form of question, although there being a fact of the matter is still a minority position, it is not a dramatic minority and the mean response hangs only around the 'somewhat disagree' anchor on the seven point scale. So again, we should be cautious about interpreting the Block B questions as indicating that a full 99% of participants reject the idea that there is a fact of the matter.

What lies behind the consistent pattern of responses across Studies 1, 2 and 3? The fact the pattern of results in Study 3 is so similar to that seen in Studies 1 and 2 puts pressure on any interpretation of those results in Studies 1 and 2 as being driven by a commitment even to correctness conditions for aesthetic judgments that are relativized to the level of the individual. If this was what Studies 1 and 2 were picking up on, we should have seen participants, in Study 3's case of within-subject disagreements, happier to say there is a fact of the matter. But this is not what we see.

One alternative hypothesis is that responses across these three studies do not reflect an anti-realist sentiment of a *relativist* stripe (according to which aesthetic judgments have correctness conditions which are relativized in some way) but rather an *expressivist* stripe (according to which aesthetic judgments have no correctness conditions). This hypothesis would fit well with the Block B results and the way Cova and Pain (2012) interpret such responses. In Study 1, 89% of participants indicated agreement with the statement 'Neither is right or wrong. It makes no sense to speak in terms of correctness in this situation. Everyone is entitled to his own opinion.' Such a response is interpreted by Cova and Pain (2012) as indicating a commitment to expressivism understood as the position that "aesthetic judgments are neither correct (nor incorrect), nor do they possess any truth-value. They express the state of pleasure or displeasure felt by the one who utters it. In that sense, they are equivalent to expressions such as: 'Yuck' or 'Great!'."

4.4. Study 4: Possibility of error

Reflecting on Studies 1–3, it is plausible to maintain that a dominant tendency in ordinary thinking is to think there is no fact of the matter in aesthetic disputes because aesthetic judgments are not in the business of stating facts and are neither correct nor incorrect. We can explore this hypothesis using a direct form of questions (modeled on Sytsma & Reuter, 2017).

4.4.1. Study 4a

4.4.1.1. Participants. The number of participants who completed the survey was 136.[25] The gender split of participants was 46% Male, 54% Female. Participants' mean age was 34.60 years old ($SD = 10.82$). The majority had no philosophical training (86%), but some indicated philosophical training at either undergraduate (14%) or postgraduate (2%).

4.4.1.2. Materials. Participants were asked 24 questions about the possibility of various kinds of error across various domains: aesthetic (aes), subjective (sub), descriptive (des), and extreme positions (xxx). The questions were presented in a random order. Available responses were 'yes' or 'no':

aes1 Do you think it is possible for a particular painting to seem more beautiful than it really is?
aes2 Do you think it is possible for a particular painting to seem more ugly than it really is?
aes3 Do you think it is possible to experience a particular painting as being ugly when in fact it is beautiful?
aes4 Do you think it is possible to experience a particular painting as being beautiful when in fact it is ugly?
aes5 Do you think it is possible to be wrong about whether a particular painting is beautiful?
aes6 Do you think it is possible to be wrong about whether a particular building is ugly?
sub1 Do you think it is possible for a particular roller-coaster to seem more fun than it really is?
sub2 Do you think it is possible for a particular roller-coaster to seem more boring than it really is?
sub3 Do you think it is possible to experience a particular roller-coaster as being boring when in fact it is fun?
sub4 Do you think it is possible to experience a particular roller-coaster as being fun when in fact it is boring?
sub5 Do you think it is possible to be wrong about whether a particular roller-coaster is fun?
sub6 Do you think it is possible to be wrong about whether a particular roller-coaster is boring?
des1 Do you think it is possible for a particular meal to seem to contain more meat than it really does?
des2 Do you think it is possible for a particular meal to seem to contain less meat than it really does?
des3 Do you think it is possible to experience a particular meal as containing meat when in fact it contains no meat?

des4 Do you think it is possible to experience a particular meal as containing no meat when in fact it contains meat?
des5 Do you think it is possible to be wrong about whether a particular meal contains meat?
des6 Do you think it is possible to be wrong about whether a particular meal contains no meat?
xxx1 Do you think it is possible to take a train to the moon to visit the moon people?
xxx2 Do you think it is possible to complete a task that cannot be completed?
xxx3 Do you think it is possible to ride on a roller-coaster with your friends?
xxx4 Do you think it is possible to eat a meal with your family and friends?
xxx5 Do you think it is possible to be wrong about something which it is possible to be wrong about?
xxx6 Do you think it is possible to be wrong about something which it is impossible to be wrong about?

For each participant, three composite scores were computed for aes, sub, and des items (one point for each 'yes' answer).[26]

4.4.1.3. Results. Descriptive results are available in Table 7and plotted in Figure 4. For all participants, mean scores differed across domains ($F(1.51, 203.93) = 3.37, p = .049, \eta_p^2 = .024$) with descriptive receiving higher scores than aesthetic ($p = .023, d = .26$) but not subjective ($p = .137, d = .17$), and with no difference between aesthetic and subjective ($p = .167, d = .09$).[27]

If we plot the proportion of participants who acknowledge the possibility of error for each of the individual aesthetic items, we can see that there is majority agreement in each case (see Figure 5).[28]

Table 7. Descriptive statistics for domain composite scores (Study 4a) showing result with and without a filter excluding participants with extreme views on possibility. Lowest score possible is zero and the highest six so the notional midpoint is three.

Domain	M (with filter)	SD (with filter)
Aesthetic	4.63 (4.56)	1.61 (1.70)
Subjective	4.78 (4.64)	1.58 (1.78)
Descriptive	5.03 (5.24)	1.42 (1.14)

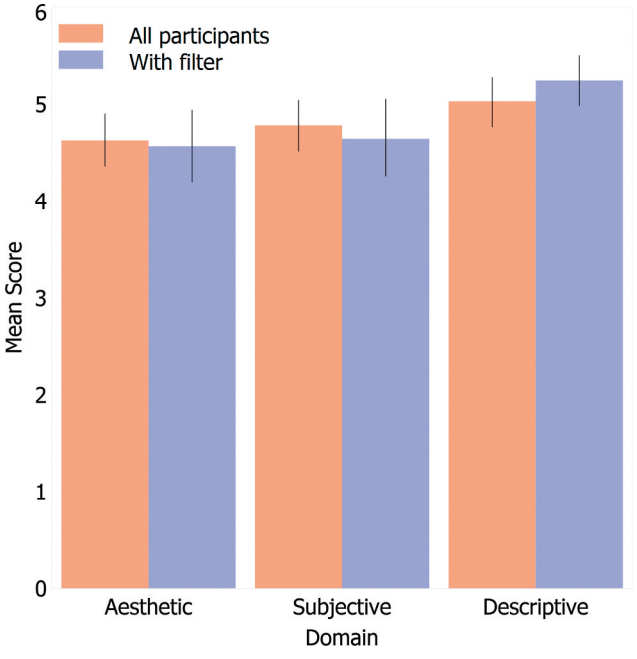

Figure 4. Graph showing mean number of acceptances of the possibility of error across the Aesthetic, Subjective, and Descriptive domains for the whole sample and when participants with extreme opinions about possibility are filtered out (Study 4a). The maximum possible number of acceptances for each domain was six and the minimum zero. Error bars indicate 95% confidence interval.

4.4.1.4. Demographic effects. There were no gender or age effects, or effects of studying philosophy at either graduate or undergraduate level.

4.4.1.5. Discussion. Despite some significant differences between domains, participants seem to entertain the possibility of error across all domains including the aesthetic. This is evidence against the idea that ordinary people are the committed expressivists we might otherwise have assumed on the basis of the Block B results in Studies 1–3. One puzzling aspect of the results in Study 4a is that, for aesthetic and subjective domains, more participants deny the possibility of error for item 5 than item 3, and item 6 than item 4. This pattern is puzzling because, e.g., if you think it is 'possible to experience a particular painting as being ugly when in fact it is beautiful' it would seem inconsistent to not think it is 'possible to be wrong about whether a particular painting is beautiful.' One possibility is that the presence of the word 'wrong' makes participants less likely to accept the possibility of error. Perhaps, for example, 'wrong' is taken to imply blameworthiness. The relevant descriptive items, e.g., about whether a particular meal contains meat, didn't see the same pattern but maybe the content

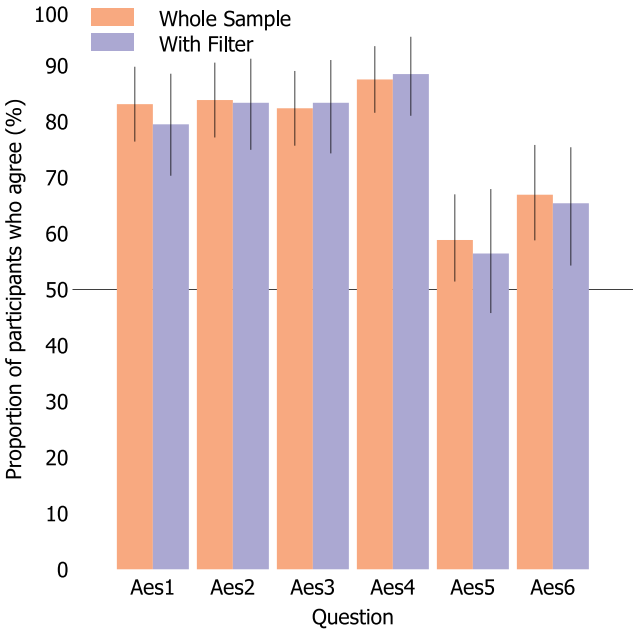

Figure 5. Graph showing proportion of participants accepting the possibility of error across the six aesthetic items for the whole sample and when participants with extreme opinions about possibility are filtered out (Study 4a). Error bars indicate 95% Confidence Interval.

means participants were happy to accept potential blameworthiness – you are to blame if you serve meat to a vegetarian even if you thought you weren't – in a way they were not for the aesthetic and subjective items. In any case, this puzzling pattern might cast doubt on exactly what the results tell us about participants' thinking about the aesthetic.

4.4.2. Study 4b

This was the same as Study 4a except with a few changes to guard against the concerns just highlighted. Items containing 'wrong' were reframed in terms of correctness, e.g., 'Do you think it is possible to make an incorrect judgment about whether a particular painting is beautiful?', and descriptive items were changed to be about 'pasta' rather than 'meat'. Subjective items were dropped.

4.4.2.1. Participants. The number of participants was 137.[29] The gender balance was 31% male, 68% female, and 1% other. Mean age was 33.02 ($SD = 12.40$) years old. Most had no philosophical training (88%), but 12% had some at undergraduate, and 2% some at postgraduate.

Table 8. Descriptive statistics for domain composite scores (Study 4b) showing results with and without a filter excluding participants with extreme views on possibility. Lowest score possible is zero and the highest six so the notional midpoint is three.

Domain	M (with filter)	SD (with filter)
Aesthetic	4.22 (4.05)	1.93 (1.94)
Descriptive	4.31 (4.45)	1.78 (1.60)

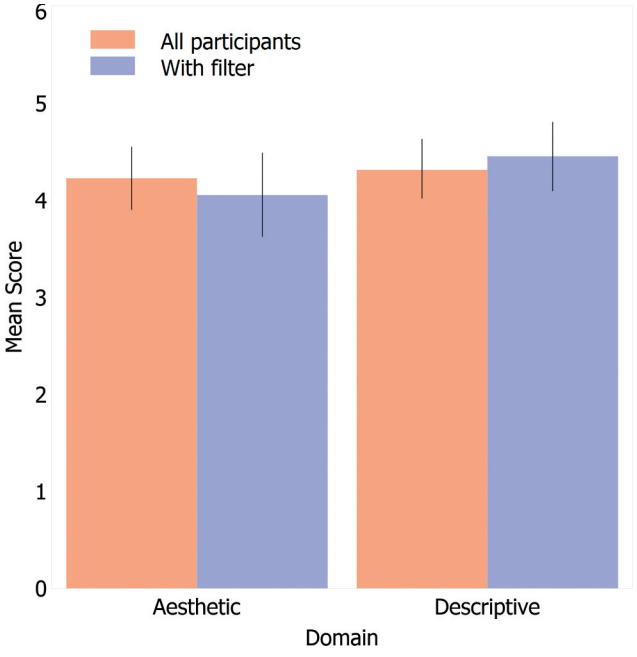

Figure 6. Graph showing mean number of acceptances of the possibility of error across the Aesthetic and Descriptive domains for the whole sample and when participants with extreme opinions about possibility are filtered out (Study 4b). The maximum possible number of acceptances for each domain was six and the minimum zero. Error bars indicate 95% Confidence Interval.

4.4.2.2. Results. Descriptive results are available in Table 8and plotted in Figure 6. A paired samples t-test using all responses found no difference between domains ($p = .687$).[30] If we plot the proportion of participants who acknowledge the possibility of error for each of the individual aesthetic items, we can see there is majority agreement for aes1–4 but not for aes5 and aes6 (see Figure 7).

4.4.2.3. Demographic effects. There was a significant difference between the aes composite scores for men ($M = 4.68, SD = 1.76$) and women ($M = 3.71, SD = 1.98$) when those with extreme views on possibility were excluded ($t(72) = 2.12, p = .037, d = .53$).[31] There was also a significant

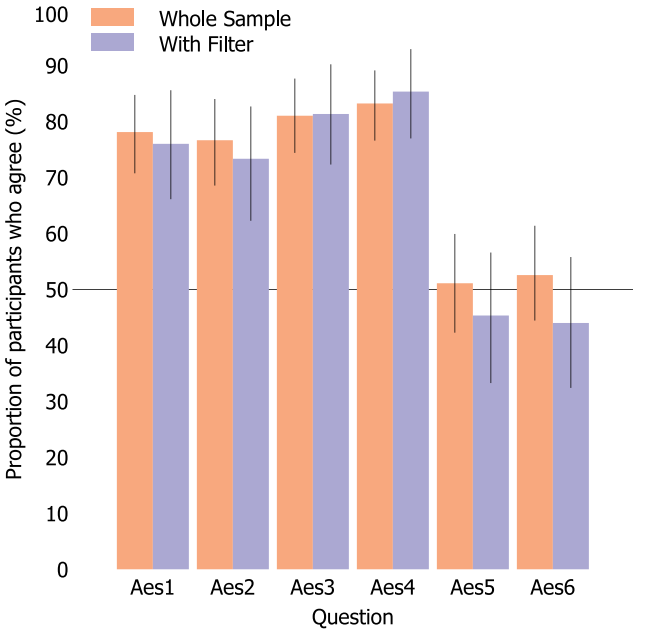

Figure 7. Graph showing proportion of participants accepting the possibility of error across the six aesthetic items for the whole sample and when participants with extreme opinions about possibility are filtered out (Study 4b). Error bars indicate 95% Confidence Interval.

difference between the des composite score for those with philosophical training at undergraduate level ($M = 5.00, SD = 1.22$) and those without ($M = 4.22, SD = 1.83$) when including all participants ($t(27.29) = 2.30, p = .029, d = .50$).[32] No other effects of gender, age, or philosophical training were observed.

4.4.2.4. Discussion. The results of Studies 4a and 4b paint a picture at odds with an interpretation of Studies 1–3 as reflecting a tendency to think of aesthetic judgments lacking truth conditions. The majority of participants seem fairly happy to acknowledge the possibility of error.

There are some minor worries one might have about Studies 4a and 4b. For example, in retrospect, the failure to find a difference between domains in Study 4b may be an artifact of the precise items chosen.[33] One might also question the choice of descriptive properties used in the materials so far. Whether a meal contains pasta or meat, for example, seems to be a straightforwardly binary affair. However, whether something is beautiful or not might be a slightly fuzzier or vague boundary. A more sensible contrast, in retrospect, might be between beautiful and a slightly fuzzier or vague descriptive property. There also remains the puzzling tension between participants' judgments about the possibility of mistaken experience (aes3 and aes4) and their judgments about the possibility of incorrect/wrong

judgments (aes5 and aes6) which the change in wording from 'wrong' to 'incorrect' hasn't reduced. One possibility is that participants are using the item to assess *likelihood* of error rather than possibility. They may then think as follows: although it is quite common to have an erroneous experience, it is quite uncommon to fail to be able to correct for this error at the level of judgment at least when it comes to *whether* an object is beautiful (as measured by aes5–6) if not the *exact degree* of beauty.

There is also a more major concern one might have about the interpretation of the results and the supposed conflict with Studies 1–3. It is possible that participants are willing to countenance only certain kinds of error in aesthetic judgment and that the relevant kinds of error are consistent with expressivism. For example, perhaps the kinds of error participants are envisaging are things like the following: the conditions in which the aesthetic judgment (or the expressivist equivalent of judgment) is made might be really unusual or distorting; the judgment might be based on a misperception of perceptual qualities of the painting (e.g., due to intoxication); the judgment may not be made with serious attention to the nature of the experience; the judgment may be somehow out of character or out of step with the judgments the agent is disposed to make.

4.4.3. Study 4c

If participants are broadly expressivist in their responses to Studies 4a and 4b, in a way that allows for the possibility of error due to factors such as misperception, then their recognition of the possibility of error about the aesthetic should drop away once the situation is fleshed out and a few obvious routes for expressivism-compatible 'error' to creep in ruled out. Study 4c explores this by asking about the possibility of error in a specific case (similar to the disagreement cases from Studies 1–3). Here, John is now described as looking at a *car*, rather than an object, to make the idea that John has some general tastes with which the judgment might be (in)congruent plausible (this is necessary to rule out one potential source of expressivism-compatible 'error'). A change is also made to the design in order to guard against responses reflecting assessments of likelihood rather than possibility of error.

4.4.3.1. Participants. The number of participants was 151.[34] The gender split was 26% male and the rest female. Mean age was 36.02 ($SD = 12.14$). Only 8% had undergraduate training in philosophy, and one had postgraduate.

4.4.3.2. Materials. Participants answered eight pairs of questions about a scenario of the following form.[35]

> John looks at a car. The car is very similar to many cars which John thinks are very [beautiful]. John typically is very predictable in what he will judge to be [beautiful]. John is looking at the car in perfect viewing conditions and there is nothing wrong with John's vision. On the basis of his experience of the car, John judges that this car is [beautiful/ugly].

Within each domain, half the question pairs concerned a scenario in which John forms a judgment which is congruent with John's general dispositions, e.g., the car is beautiful, and half in which the judgment is opposite, e.g., the car is ugly. For the subjective domain, the adjectives used were 'pleasing' and 'displeasing,' and for the descriptive domain, 'large' and 'small.'

Each pair of questions took the following form to help disambiguate likelihood and possibility of error. First, participants were asked how likely it was that X on a seven-point scale (very unlikely to very likely). Second, participants were asked about the same X, 'Is it possible, however likely or unlikely, that X?' The values for X were:

wrong John is wrong about whether the car is [beautiful]
experience valence John experiences the car as being [beautiful] when in fact it is [ugly]
experience strength John experiences the car as being more [beautiful] than it really is
incorrect John has made an incorrect judgment about whether the car is [beautiful]

A composite 'Poss' score was created for each participant from the number of 'yes' responses they gave to the possibility questions.

4.4.3.3. Results. Descriptive results are available in Table 9 and plotted in Figure 8. A one-way ANOVA found no difference between domains ($p = .816$) and the result was the same excluding those with extreme views on possibility ($p = .880$). We can also look at proportion of participants who acknowledge the possibility of error for each of the individual items (Table 10). A clear majority of participants recognize a possibility of all types of error for each domain.

Table 9. Descriptive statistics for composite Poss scores (Study 4c) showing result with and without a filter excluding participants with extreme views on possibility. Lowest score possible is zero and the highest eight so the notional midpoint is 4.

Domain	M (with filter)	SD (with filter)
Aesthetic	6.71 (7.12)	1.85 (1.81)
Subjective	6.46 (7.10)	2.25 (2.17)
Descriptive	6.69 (6.87)	2.31 (2.14)

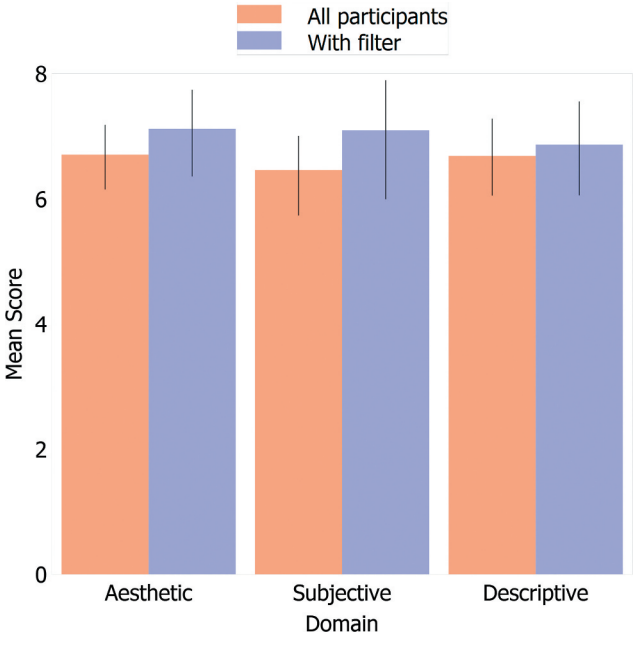

Figure 8. Graph showing mean number of acceptances of the possibility of error across the Aesthetic, Moral, and Descriptive domains for the whole sample and when participants with extreme opinions about possibility are filtered out (Study 4c). The maximum possible number of acceptances for each domain was eight and the minimum zero. Error bars indicate 95% confidence interval.

Table 10. Proportion of all participants agreeing for each possibility question. Item labels are given in the text and the ± indicates the relevant property (+ = Beautiful, pleasing, large. – = Ugly, displeasing, small) (Study 4c).

	% selecting 'Yes' (with filter)		
Item	Aesthetic	Subjective	Descriptive
+wrong	82% (92%)	84% (91%)	82% (87%)
-wrong	86% (92%)	76% (86%)	86% (93%)
+experience valence	80% (88%)	80% (95%)	78% (80%)
-experience valence	92% (92%)	82% (86%)	84% (87%)
+experience strength	82% (92%)	82% (95%)	88% (87%)
-experience strength	84% (84%)	82% (91%)	84% (83%)
+incorrect	86% (92%)	80% (81%)	78% (80%)
-incorrect	77% (80%)	80% (86%)	86% (90%)

4.4.3.4. Demographic effects. There was a negative correlation between age and poss score for descriptive ($r = -31, p = .029$) with older people tending to be less open to the possibility of error. There was a significant difference between the poss scores of those with and without philosophical training at undergraduate level for subjective ($t(48) = 2.66, p = .011$) with

philosophical training being associated with lower scores (although the number of those with training is only 4). There were no other effects of gender, age or of studying philosophy.

4.4.3.5. Discussion. Participants seem very happy to sign up to the possibility of error even in cases where perceptual problems or oddities in viewing conditions are ruled out and whether the judgment seemed to cohere with the protagonist's general tastes seems to make little to no difference. This seems difficult to interpret in a way that would be easy to integrate into a worldview in which aesthetic judgments lack correctness conditions.

4.5. Study 5: Disagreement and possibility of error at an instant

How are we to reconcile the tension in the findings so far? Studies 1–3 suggest that ordinary thinking rejects correctness conditions for aesthetic judgments; the dominant tendency was to select 'Neither is right or wrong. It makes no sense to speak in terms of correctness in this situation. Everyone is entitled to his own opinion' and (albeit less dominantly) to disagree with 'In reality, the object is either beautiful or it is not beautiful, regardless of whether these two people think it is beautiful or would describe it as 'beautiful" and this remains unchanged when we shift from a case of cross-cultural disagreement to within-culture disagreement to within-individual disagreement. It seems participants use both probes to reject correctness conditions on aesthetic judgments – even individually-individuated conditions. On the other hand, Studies 4a–c suggest ordinary thinking is perfectly happy to countenance correctness conditions for aesthetic judgments.

One possibility is that participants see correctness of aesthetic judgment as being relativized to a level finer than the individual judge (accommodating Studies 1–3), but which nonetheless accepts the possibility of error (accommodating Studies 4a–c). If one thought, for example, that correctness conditions were relativized to judge *and* time *and* place, then one might well reject the target item in Study 3 on the basis that the standards for John have shifted between yesterday and today. Study 5 explores this hypothesis by asking participants about a case of within-subject disagreement in which such additional variables as time and place of judgment are held fixed (or as closely as possible). The design brings together elements of Studies 1–3 and Studies 4a–c.

4.5.1. Participants

The number of participants was 61.[36] The gender split was 44% male and the rest female. Mean age was 31.89 ($SD = 11.84$). Most had no philosophical training, 13% had undergraduate, none had postgraduate.

4.5.2. Materials

Participants read the following scenario.

> John is looking at a car. The car is very similar to many cars which John thinks are very beautiful. John typically is very predictable in what he will judge to be beautiful. John is looking at the car in perfect viewing conditions and there is nothing wrong with John's vision.
>
> At first, John experiences the car to be ugly and, on the basis of his experience, John judges that the car is ugly. Then, for onesecond, John closes his eyes. While John had his eyes closed no changes occur to the car, to the viewing conditions, or to his vision.
>
> After one second, John reopens his eyes and looks at the very same car again. John is paying attention to exactly the same features of the car as he was paying attention to before he closed his eyes. Nonetheless, after reopening his eyes, John finds that he now experiences the very same car to be beautiful. On the basis of his new experience, John revises his judgment, and judges the car to be beautiful.

Participants then responded to the following three blocks of questions in a random order. After two comprehension questions, Block A asked participants to rate the following on a seven-point scale as above (3–10 randomized, 11 always last).

(3) We could find out whether John was right at first when he judges the car to be ugly (before closing his eyes)
(4) We could find out whether John was right in the end when he judges the car to be beautiful (after reopening his eyes)
(5) John deserves blame for being wrong when he judges the car to be ugly (before closing his eyes)
(6) John deserves blame for being wrong when he judges the car to be beautiful (after reopening his eyes)
(7) The car may appear in different ways to John at different times and so, for all we know, John is correctly reporting how he experiences the car both before closing his eyes and after reopening his eyes.
(8) John may use the words 'beautiful' and 'ugly' to mean different things at different times, and so given how he may be using the words, for all we know John could have been right about the beauty of the car both before closing his eyes and after reopening his eyes.
(9) People often change their opinions about what word best describes how an object appears. For example, people often change their opinions about whether something should be called 'beautiful' or 'ugly'.

(10) People often change their opinions about the visual appearance of objects. For example, people often change their opinions about an object's exact shape and color.
(11) In reality, the car is either beautiful or it is not beautiful, regardless of whether John thinks it is beautiful or would describe it as 'beautiful' either before closing his eyes or after reopening them.

In Block B, participants were asked 'Which of the following best represents your interpretation of this scenario?' concerning the same scenario, participants had to select one statement only:

(1) John was either right at first before he closed his eyes or is he is right at the end after he reopens his eyes, but not both.
(2) John was right on both occasions, i.e., both at first before he closed his eyes and also at the end after reopening his eyes.
(3) John was wrong on both occasions, i.e., both at first before he closed his eyes and also at the end after reopening his eyes.
(4) Neither of John's judgments is right or wrong. It makes no sense to speak in terms of correctness in this situation. Everyone is entitled to change their mind.

In Block C, participants responded to pairs of questions, as in Study 4c, assessing statements first for likelihood then possibility. The statements were:

wrong1 Before closing his eyes, John was wrong about whether the car is beautiful
experience valence1 Before closing his eyes, John experienced the car as being ugly when in fact it is beautiful
experience strength1 Before closing his eyes, John experienced the car as being less beautiful than it really is
incorrect1 Before closing his eyes, John made an incorrect judgment about whether the car is beautiful
wrong2 After reopening his eyes, John is wrong about whether the car is beautiful
experience valence2 After reopening his eyes, John experiences the car as being beautiful when in fact it is ugly
experience strength2 After reopening his eyes, John experiences the car as being more beautiful than it really is
incorrect2 After reopening his eyes, John made an incorrect judgment about whether the car is beautiful

Table 11. Descriptive statistics for statements 3–11 (Study 5) including midpoint comparison and predictive value of a simple regression model predicting statement 11.

Statement	% Agree	M	SD	t(60) =	R^2
3	20%	3.00	1.60	$4.88, p < .001, d = .624$	.136
4	18%	3.13	1.61	$4.20, p < .001, d = .537$	.164
5	7%	2.16	1.41	$10.21, p < .001, d = 1.307$	.166
6	7%	2.16	1.32	$10.87, p < .001, d = 1.392$	.150
7	75%	4.95	1.37	$5.42, p < .001, d = .693$	.076
8	72%	4.80	1.30	$4.82, p < .001, d = .617$	.002
9	80%	5.15	1.38	$6.51, p < .001, d = .834$	.004
10	84%	5.18	1.30	$7.11, p < .001, d = .910$	.008
11	41%	3.70	1.89	$1.22, p = .228, d = .156$	–

Table 12. Proportion of participants selecting each of the options in Block B (Study 5) including the results of a one-sample chi-square test testing the null hypothesis that the probability that participants select the relevant option is 0.25.

Statement	% Agree	$\chi^2(1, 61) =$
1	7%	$11.07, p = .001$
2	10%	$7.78, p = .006$
3	0%	–
4	84%	$111.743, p < .001$

4.5.3. Block A results

The mean response to the target item was 3.70 ($SD = 1.89$), this is not significantly different from the midpoint ($p = .228$) with 41% of participants indicating some level of agreement, 46% some level of disagreement, and 13% at the midpoint. Full descriptive results for all statements can be found in Table 11.

4.5.4. Block B results

Almost all (93%) gave a response which would be interpreted as rejecting normativism (breakdown in Table 12).

4.5.5. Effects of question design: Block A v Block B

As can be seen in Figure 9, the degree of coherence between Block A and B was not high: a very slim majority of participants gave the same response in both (52%) and among participants for whom block made a difference all of them gave the normativist response in Block B ($\chi^2(1) = 27.034, p < .001$).[37]

4.5.6. Block C results

The minimum possible number of 'yes' answers across the possibility questions was 0 and the maximum 8 with a notional midpoint of 4. The mean number of 'yes' answers was 6.57 ($SD = 2.36$) which is significantly higher than the midpoint ($t(60) = 8.508), p < .001, d = 1.089$). We can look at the

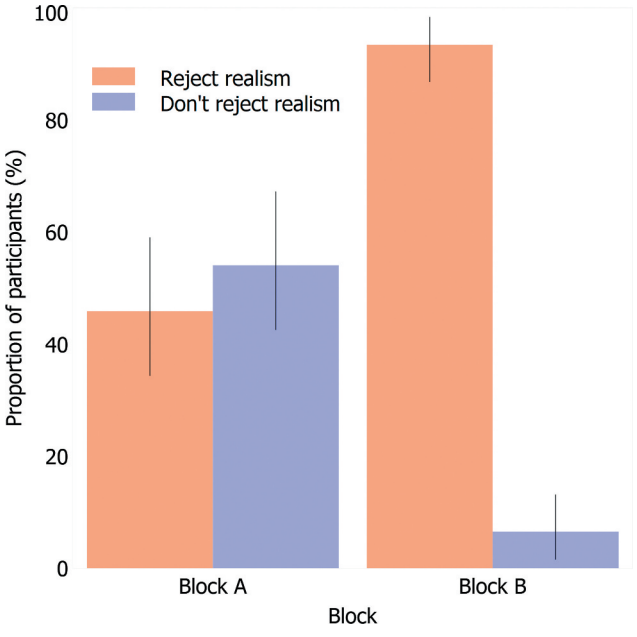

Figure 9. Graph showing proportion of participants giving responses that would be interpreted as rejecting and not rejecting realism in Blocks A and B (Study 5) showing that question design makes an important difference. Error bars indicate 95% Confidence Interval.

Table 13. Proportion of all participants agreeing for each possibility question (Study 5). Item labels are given in the text.

Item	% selecting 'Yes'
wrong1	85%
valence1	85%
strength1	87%
incorrect1	85%
wrong2	72%
valence2	77%
strength2	87%
incorrect2	79%

proportion of participants who acknowledge the possibility of error for each of the individual items (Table 13). A clear majority recognize a possibility for each type of error. The majority (57%) selected 'yes' for all possibility items.[38]

4.5.7. Demographic effects

No effects of gender, age or studying philosophy were observed.

4.5.8. Discussion

The hypothesis was that participants think of aesthetic judgments having correctness conditions (allowing for an appearance-reality gap even when stipulated, as in Study 4c, that conditions are perfect, etc.) that are more fine-grained than the level of the individual.

In this study, participants were asked about a scenario in which a protagonist changes their aesthetic judgment in a blink of an eye – despite the fact that the viewing conditions and vision are stated to be perfect throughout and no changes occur to the object – from a judgment which is out of line with their typical tastes to a judgment that is in line with their typical tastes. The results were that participants, in the Block B design, largely seemed to reject the idea that 'John was either right at first before he closed his eyes or is he is right at the end after he reopens his eyes, but not both' in favor of 'Neither of John's judgments is right or wrong. It makes no sense to speak in terms of correctness in this situation. Everyone is entitled to change their mind.' Nonetheless, this apparent strong anti-realism tendency is vastly eroded in the Block A question design in which only a minority of participants reject 'In reality, the car is either beautiful or it is not beautiful, regardless of whether John thinks it is beautiful or would describe it as 'beautiful' either before closing his eyes or after reopening them.' Moreover, participants clearly acknowledge the possibility that John makes an error, in Block C.

Do participants continue to give a response to Block A and Block B questions that might be taken to indicate a rejection of the idea that aesthetic judgments have correctness conditions? The answer is that the results are much more ambiguous in this study that in previous studies. The Block B results would again lead us to think that the vast majority of participants reject the idea there is a fact of the matter. This result would be truly difficult to square with the judgments about the possibility of error. It is thus particularly notable that the Block A question design produces results here that don't simply errode the extent to which 'there is no fact of the matter' is the dominant position (as in Studies 1–3) but in fact make it disappear.[39] This suggests that participants are truly divided and more so than in the previous studies (see Figure 10).

Do participants acknowledge that someone in John's situation might make an error in their aesthetic judgment or experience? Yes. Replicating Study 4c, the dominant trend is for participants to recognize the possibility of error even when an individual judges in line with their experience in perfect conditions and in ways that are congruent with their typical dispositions (see Figure 11 for comparison with previous studies). Again, this makes it difficult to interpret the results as reflecting some expressivist-compatible sense of the possibility of error.

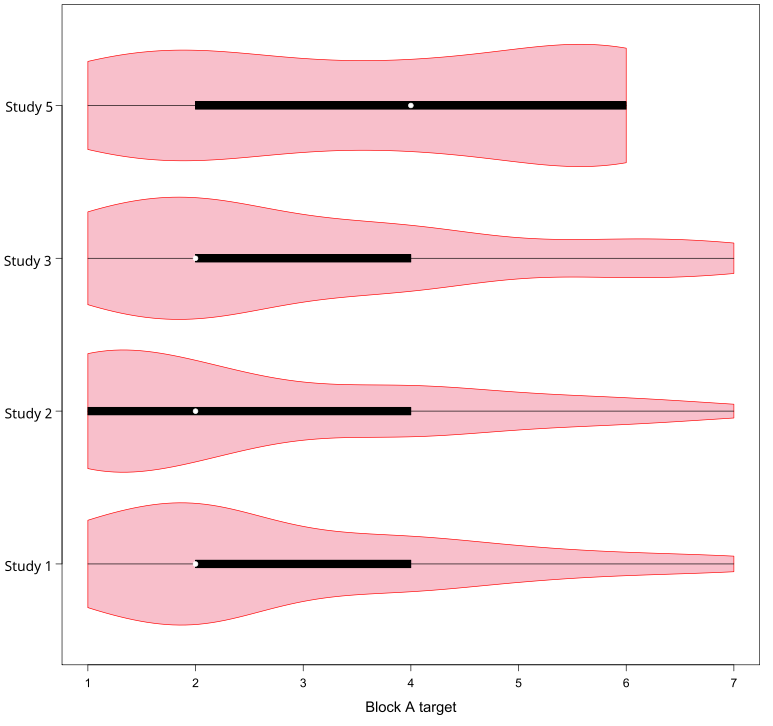

Figure 10. Violin plot of responses to the Block A target item for all participants for all studies in this paper involving disagreement cases. The black box indicates the interquartile range, and the shaded area is a density plot showing the distribution of the data. Study 1 involved cross-cultural disagreement, Study 2 involved within-culture disagreement, Study 3 involved within-subject disagreement, and Study 5 involved within-subject disagreement in a small time interval.

5. General discussion and conclusions

The idea that the ordinary understanding of the aesthetic assumes normativism has been significant in aesthetics. Importantly, it has been taken to be a major consideration in favor of realism in light of the fact that realism can easily explain why we act as it normativism is true. This argument has recently been claimed to be exploded by the work of Cova and colleagues who present evidence that ordinary people reject normativism about the aesthetic. However, the question of whether ordinary thinking assumes normativism is just one interesting question about the shape of ordinary thinking about the aesthetic. It is important to know whether ordinary thinking about the aesthetic tends to think aesthetic judgments have no correctness conditions at all or if it tends to think that aesthetic judgments do have correctness conditions but that they are relativized in some way. The studies in this paper explored these issues. The picture that emerges is a little complex and casts doubt on exactly how to interpret Cova and colleague's instrument and results.

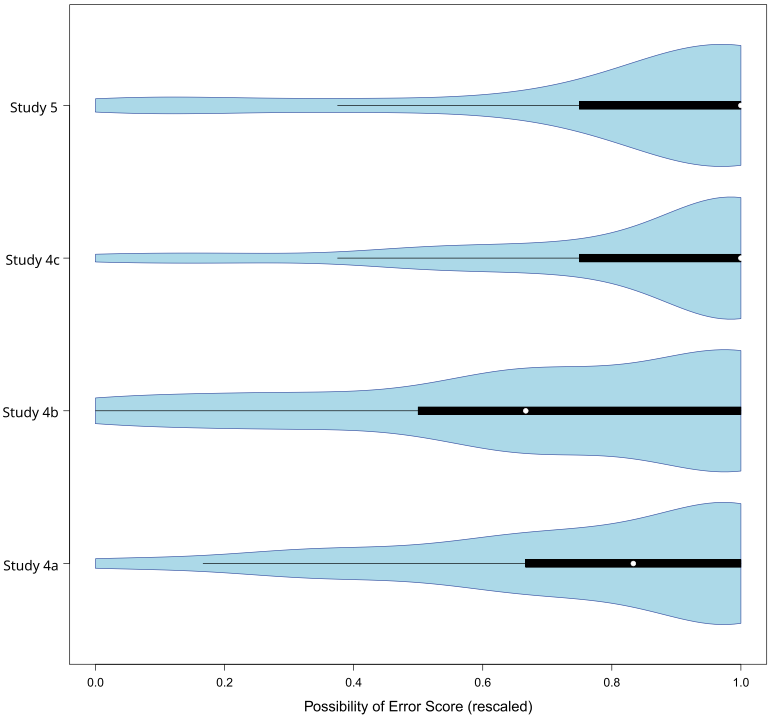

Figure 11. Violin plot of number of positive responses to possibility of error items for all participants across all studies in this paper involving possibility of error questions. These have been scaled such that '1' represents the maximum possible number of positive responses. The black box indicates the interquartile range, and the shaded area is a density plot showing the distribution of the data. Study 4a involved the basic questions, Study 4b removed the wording around 'wrong', Study 4c had participants first assess likelihood before possibility, Study 5 involved possibility items relating to two distinct time-points.

It seems there is a dominant tendency to think of aesthetic judgments as having correctness conditions. While the majority of participants in my studies select the final option in Cova-style probes – 'Neither of judgment is right or wrong, it makes no sense to speak in terms of correctness in this situation' – we shouldn't assume that this indicates a rejection of the idea that aesthetic judgments have correctness conditions, because when asked directly participants are very willing to say that it is possible for aesthetic judgments and experiences to be in error. Supposing that these responses reflect a tendency to regard aesthetic judgments as having correctness conditions, we can ask about the shape of those correctness conditions. They don't seem to be thought to be absolute as the dominant tendency is for participants to think that in cross-cultural disagreements there is no fact of the matter. They don't seem to be thought to be relativized to culture either since the pattern of results doesn't shift at all when participants are asked about cases of within-culture disagreement. Likewise the correctness conditions don't even seem to be thought to be relativized to the individual

as the pattern of responses doesn't shift when participants are asked about within-subject disagreements. The only stage at which the pattern of results changes and at which participants seem to be notably more willing to accept the idea that there is a fact of the matter – although in fact the pattern of results really only suggests that participants are very divided or unsure – is in the final study. In this final study, participants are asked about a case in which the protagonist's judgment about the beauty of a car shifts over the course of one second despite the fact that there is nothing wrong with their vision, the viewing conditions are perfect, the car doesn't change, and the features to which they attend don't change. Only in this case does the dominant trend to deny there is a fact of the matter disappear. One way to make sense of this result is to interpret it as telling us something key about the correctness conditions that ordinary thinking about the aesthetic tends to ascribe to aesthetic judgment: that there is perhaps is some tendency to think of correctness conditions of aesthetic judgments as very finely relativized to very specific circumstances of judgment (although, given the pattern of results, the this tendency isn't a dominant or unequivocal tendency).

One caveat here: another possibility is that the difference in Block Aresults between Studies 1–3 and Study 5 reflects something different.[40] For independent reasons explained in the above, whereas Studies 1–3 asked participants about cases of disagreement concerning an unspecified 'object,' Study 5 asked participants about acase involving acar. This change might have been significant. Perhaps Block Aresults in Studies 1–3 would have been equally as ambiguous as those in Study 5 had the cases used in Studies 1–3 also concerned adisagreement about the properties of acar or another more concrete example rather than an unspecified 'object'.[41] So it is possible that the current results under-represent the reasons for caution around interpreting previous results, and that future work using more concrete disagreement cases might find higher levels of responses reflecting endorsement even of *inter*subjective correctness conditions for aesthetic judgments.

Either way, these current studies provide reason for caution around interpreting the results of previous studies.[42] The central point is this: insofar as there is a tendency to deny the idea that there is a fact of the matter about the aesthetic, (a) it doesn't seem to be as dominant or unequivocal as previous studies have been interpreted as showing, and (b) that tendency seems to be compatible with the recognition of the possibility of error in aesthetic judgment. This is contrary to how results using the Cova-style question from previous studies are naturally interpreted. One would naturally interpret these results as showing that the dominant trend was to reject the idea that aesthetic judgments were correct and incorrect. But the

results of the current studies above suggest participants may be using the final option in Cova-style questions to express more subtle and less unequivocal understandings of aesthetic judgment.

These current studies don't pretend to be the last word in pinning down the dynamics of ordinary thinking about the aesthetic. But they do advance our understanding of ordinary thinking about aesthetic matters a little. The current results offer very little comfort for the realist. Although things are not quite as bad for the realist as Cova and colleagues suggest, it is certainly not the case that participants tend to embrace normativism. So, there is no fact about ordinary thinking about the aesthetic which realists are particularly well placed to explain. The current results do, however, potentially provide support for a certain kind of theory which holds that aesthetic judgments have very finely relativized correctness conditions over (i) theories which hold that aesthetic judgments have more coarsely relativized correctness conditions (although bear in mind the caveat mentioned above), and (ii) kinds of expressivist view which deny aesthetic judgments have correctness conditions at all.

Notes

1. Other ideas about distinguishing features of realism include: the relevant facts must be robustly mind-independent (Hanson, 2018); the realist believes in 'a property of beauty independent of judgments which ascribe it' (Goldman, 1993); the realist thinks aesthetic judgments state objects in the external world possess aesthetic properties and that '[a]ttributions of aesthetic properties are not reducible to reports of experiences in the mind of the observer' (Simoniti, 2017); realism just is normativism (Schafer, 2011).
2. Experimental aesthetics in Xphi is a relatively new field. It has a broad scope, dealing with diverse topics, e.g., imaginative resistance (Liao et al., 2014), art ontology (Kamber, 2011), aesthetic predicates (Liao et al., 2016). For an early survey, see Cova et al. (2015). For a recent collection, see Cova and Réhault (2018). Here, I focus solely on research involving attitudes toward aesthetic realism.
3. The question format employed by Cova and colleagues originates in metaethics (Nichols, 2004) and variations on it have been widely used since (see, e.g., aBeebe et al., 2015; aBeebe & Sackris, 2016; Goodwin & Darley, 2008; Khoo & Knobe, 2018; Sarkissian et al., 2011; Wright, 2013). See Pölzler (2017) for a recent critical survey.
4. Originally in French.
5. Although 3 might also betray error theoretic sentiments.
6. In translation for non-English-speaking samples.
7. Cova and Pain's third response option was dropped.
8. In Cova et al. (2018), the 'neither is correct' option is interpreted as 'nihilist' rather than 'expressivist' (as in Cova & Pain, 2012), but it amounts to the same key point in relation to correctness conditions: 'aesthetic judgments cannot be ascribed correctness or incorrectness.'
9. Reanalysis of Andow (2019)'s data finds only 3% of participants indicate any level of agreement with the normativist option.

10. The idea isn't that esthetic matters are completely subjective (or intuitively so) in the sense that all aesthetic judgments are on a par. Indeed, researchers have explored various factors that play a role in determining a 'hierarchy' of aesthetic judgments (see Cova, 2018; Goffin & Cova, 2018).
11. Thanks to a referee for this journal who drew my attention to Rabb et al. (2020) (which came out almost exactly when this paper was originally submitted). Rabb et al. pursue a related project, exploring the extent to which participants can be pushed toward objectivism. Their conclusion is rather different from that of this paper; as their title has it, 'Expressivist to the core: Metaaesthetic subjectivism is stable and robust.' Ultimately, I think the spirit of the concerns raised in this section – around intepreting studies using Cova-style questions – extend to the form of question used Rabb etal. But I don't have space to discuss the issue in depth here (for some extended critical discussion of Rabb et al., 2020's design and the bearing of their results on metaaesthetics, see; Moss & Bush, 2021).
12. It would be a mistake to try to invent a new form of survey question to be a perfect, unquestionable litmus test for meta-a
esthetic sensibilities. The search for *the right way* to formulate a question is typically futile. We build up the richest and most informative picture, instead, by approaching the issue from many different angles.
13. Roberts and Schmidtke (2016) used a similar target item to investigate folk understandings of color, shape, sound, taste and likability, and report what we might think of as normativist tendencies for shape, color, and sound.
14. A power analysis determined the sample size necessary to detect a medium effect ($OR > 3.47$) (Chen et al., 2010) when comparing responses across the two conditions using a McNemar's one tailed test (power = .95, proportion of discordant pairs = .3, two-tailed). The required sample size was 134. 174 entered the survey. 5 failed to complete it. Of the remainder, 27 failed to respond correctly to at least one of two comprehension questions and were excluded from the analysis. Exclusions made no qualitative differences to results.
15. As an anonymous reviewer pointed out, it is perhaps unclear whether statements 3–8 in fact serve to disambiguate 9 as 9 may not need disambiguating. It would be interesting to investigate whether the presence of 3–8 affects participants' responses to 9. However, the issue doesn't affect the interpretation of the results in this paper.
16. It is, of course, possible some other difference between the two designs produces this difference, e.g., rating vs forced choice, and future work could explore this further. However, both tasks should be straightforward for any participant who rejects normativism (in Block Arejecting the target statement, in Block Brejecting the first option). Thanks to areviewer for asking about this.
17. With appropriate changes to comprehension questions.
18. Using the effect size and proportion of discordant pairs observed in Study 1, a power analysis determined a sample size of only 30 was required to achieve power of .95 for the between block comparison using McNemar's test (two tailed). A larger sample was required to detect a Cohen's d of .561 in single sample t-tests (N = 44). 139 entered the survey, 3 failed to complete. Of the remainder, 24 failed to respond correctly to at least one of two comprehension questions and were excluded from the analysis.
19. Unlike in Study 1, two of statements 3–8 were significant predictors in a simple regression model predicting response to the target statement. When entered as predictors in a multiple regression model ($F(2, 109) = 20.45, p < .001, R^2 = .273$), both statement 3 ($\beta = .353, p < .001$) and statement 4 ($\beta = .287, p = .001$) were significant predictors

20. Odds ratio isn't calculable due to the number of participants in one cell being zero.
21. Using the effect size observed in Study 1 (not calculable in Study 2), and the smallest proportion of discordant pairs observed in Studies 1 and 2 (31% in Study 2), a power analysis determined that a sample size of only 30 was required to achieve power of .95 for the between block comparison using McNemar's test (two tailed). A larger sample was required to detect a Cohen's d of .410 (as observed in Study 2) in the single sample t-tests (N = 80). 138 entered the survey. 2 failed to complete. 13 failed comprehension checks and were excluded.
22. None of 3–8 were significant predictors of 9 in simple linear regression models ($ps < .08$).
23. We can explore the relationship between participants' responses to the Block A question (full scale responses) and the individual anti-realist options presented in Block B among participants who select an anti-realist option in Block B (only statements 2 and 4 were chosen by participants). Participants who selected option 2 gave slightly lower ratings of the target item ($m = 2.70, sd = 1.89$) than participants who selected option 4 ($m = 3.09, sd = 1.85$), although note that this difference was not significant ($t(10.61) = .626, p = .526, d = .208$) and the number of participants selecting option 2 was low ($n = 10$).
24. The odds ratio isn't calculable due to the number of participants in one cell being zero.
25. A power analysis determined a sample size of 54 was required to detect a medium effect (dz = .5) in pairwise comparisons of domains in paired t-tests (power = .95, two-tailed). 139 entered the survey. 3 failed to complete. There were no exclusions on the basis of comprehension questions (although note the use of a filter in the results).
26. Answers to xxx items were used to filter out participants who seem to have extreme positions on the possibility of error. Giving unexpected answers to xxx items might also result from a lapse in attention but in any case it is useful to compare results with and without these participants.
27. Filtering out those with extreme views on possibility, mean scores still differed across domains ($F(1.55, 119.61) = 6.29, p = .005, \eta_p^2 = .076$) with descriptive receiving higher scores than both aesthetic ($p = .004, d = .45$) and subjective ($p = .015, d = .40$), but with no difference between aesthetic and subjective ($p = .593, d = .05$).
28. Although for aes5 the 95% confidence interval overlaps with 50% once participants with extreme views about possibility are filtered out.
29. A power analysis determined that a sample size of 54 was required to detect a medium effect (dz = .5) in pairwise comparisons of domains in paired t-tests (power = .95, two-tailed). In total, 142 entered the survey, 137 completed it. There were no exclusions on the basis of comprehension.
30. The result was the same excluding those with extreme views on possibility ($p = .156$).
31. This was trending but not significant at the .05 level when including all participants.
32. The number of those with training in this sample is low ($n = 17$) and when excluding those with extreme views on possibility the effect isn't observed ($p = .84$).
33. Participants may be inclined to disagree with des3 and des4 on the grounds that they think people are typically pretty good pasta detectors (relative to their abilities in meat detection).
34. A power analysis determined a sample size of 54 was required to detect a medium effect (dz = .5) in pairwise comparisons of domains in paired t-tests (power = .95, two-tailed) (although note that pairwise comparisons were not performed as no

difference between conditions was found in a one-way ANOVA). 153 entered the survey, two failed to complete. No rejections were made on the basis of comprehension.

35. One pair was presented on each page. The relevant scenario for each pair was presented at the top of the page. The participants also saw the six xxx questions used above each on a new page. The order of all pages was randomized.
36. Using the effect size observed in Study 1 (not calculable in Studies 2–3), and the smallest proportion of discordant pairs observed in Studies 1–3 (31% in Study 2), a power analysis determined that a sample size of only 30 was required to achieve power of .95 for the comparison between Block A and B using McNemar's test (two tailed). Using the smallest effect size observed comparing possibility of error scores for the aesthetic domain to the scale midpoint in Studies 4a-c ($d = .63$), a power analysis determined that a similar sample size ($N = 34$) was required for the single sample t-tests to be used in analyzing the results from Block 3 (power = .95, two-tailed). 80 entered the survey, 9 failed to complete. 10 failed comprehension checks and were excluded from the analysis.
37. The odds ratio isn't calculable due to the number of participants in one cell being zero.
38. There is a clear pattern that participants are less likely to acknowledge the possibility of error for John's judgment at the second time point. Indeed, if we compare mean number of yes-answers across the first four items and the second four items in a paired samples t-test we see this difference is significant ($t(60) = 2.289, p = .026, d = .293$).
39. As a reviewer noted, it is possible that the change from 'object' to 'car' had some effect here. This possibility and its implications are discussed in the general discussion below.
40. Thanks to a reviewer for highlighting this issue.
41. This isn't implausible. As described in Section 3, Andow (2020) presented participants with a intersubjective (but presumably within culture) aesthetic disagreement case with a specified object (Van Gogh's sunflowers), and gave participants the chance to rate a series of positions relating to normativism (although identical neither with the Block A or Block B questions here), and in the results around half of participants agreed with at least one normativist position.
42. This includes studies such as Rabb et al. (2020) (published as this paper was submitted) whose main form of question asks participants about whether something is amatter of fact or of opinion/preference/taste and whose results the authors interpret as potentially reflecting aform of expressivism. The results here, e.g., concerning the possibility of error, give cause for caution about that interpretation. Thanks to a reviewer for drawing my attention to this paper.

Acknowledgments

Thanks to the reviewers for this journal and to participants at the Future of Aesthetics workshop at the University of Leeds in 2019 for very helpful discussion.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

Support provided by the UEA HUM Faculty Small Awards Scheme;University of East Anglia, Humanities Faculty Research Small Awards;

Notes on contributor

James Andow is Lecturer in Philosophy at the University of East Anglia

ORCID

James Andow http://orcid.org/0000-0002-5760-0475

References

Andow, J. (2019). Aesthetic testimony and experimental philosophy. In F., & Cova, and S. Réhault, (Eds.), *Advances in experimental philosophy of aesthetics* (pp. 33–58). Bloomsbury Academic. http://dx.doi.org/10.5040/9781350038950.ch-002

Andow, J. (2020, September). Why don't we trust moral testimony? *Mind & Language*, *35* (4), 456–474. https://doi.org/10.1111/mila.12255

Beebe, J. R., & Sackris, D. (2016, August). Moral objectivism across the lifespan. *Philosophical Psychology*, *29*(6), 912–929. https://doi.org/10.1080/09515089.2016.1174843

Beebe, J., Qiaoan, R., Wysocki, T., & Endara, M. A. (2015). Moral objectivism in cross-cultural perspective. *Journal of Cognition and Culture*, *15*(3–4), 386–401. https://doi.org/10.1163/15685373-12342157

Chen, H., Cohen, P., & Chen, S. (2010, March). How big is a big odds ratio? interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation*, *39*(4), 860–864. https://doi.org/10.1080/03610911003650383

Cohen, J., & Nichols, S. (2010). Colours, colour relationalism and the deliverances of introspection. *Analysis*, *70*(2), 218–228. https://doi.org/10.1093/analys/anp161

Cova, F., Garcia, A., & Liao, S. (2015, December). Experimental philosophy of aesthetics. *Philosophy Compass*, *10*(12), 927–939. https://doi.org/10.1111/phc3.12271

Cova, F., Olivola, C. Y., Machery, E., Stich, S., Rose, D., Alai, M., . . . Zhu, J. (2018, August). De pulchritudine non est disputandum? a cross-cultural investigation of the alleged intersubjective validity of aesthetic judgment. *Mind & Language*, *34*(3), 317–338. https://doi.org/10.1111/mila.12210

Cova, F., & Pain, N. (2012). Can folk aesthetics ground aesthetic realism? *Monist*, *95*(2), 241–263. https://doi.org/10.5840/monist201295214

Cova, F., & Réhault, S. (2018). *Advances in experimental philosophy of aesthetics* (1st ed.). Bloomsbury Academic.

Cova, F. (2018). Beyond intersubjective validity: Recent empirical investigations into the nature of aesthetic judgment. In F. Cova, and S. Réhault (Eds.), *Advances in experimental philosophy of aesthetics*(pp. 13–32). Bloomsbury. https://doi.org/10.5040/9781350038950.ch-001

Eaton, Marcia Muelder. (1998). Intention, supervenience, and aesthetic realism. *British Journal of Aesthetics*, *38*(3), 279–293. https://doi.org/10.1093/bjaesthetics/38.3.279

Goffin, K., & Cova, F. (2019). An empirical investigation of guilty pleasures. *Philosophical Psychology*, *32*(7), 11129–1155. https://doi.org/10.1080/09515089.2019.1646897

Goldman, A. I. (1993). Epistemic folkways and scientific epistemology. *Philosophical Issues*, *3*, 271–285. https://doi.org/10.2307/1522948

Goodwin, G., & Darley, J. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, *106*(3), 1339–1366. https://doi.org/10.1016/j.cognition.2007.06.007

Hanson, L. (2018). Moral realism, aesthetic realism, and the asymmetry claim. *Ethics*, *129* (1), 39–69. https://doi.org/10.1086/698732

Kamber, R. (2011, May). Experimental philosophy of art. *The Journal of Aesthetics and Art Criticism*, *69*(2), 197–208. https://doi.org/10.1111/j.1540-6245.2011.01461.x

Kant, I. (2000). First section, first book: Analytic of the beautiful. In P. Guyer, and E. Matthews, Trans., P. Guyer (Ed.), *Critique of the power of judgment (the Cambridge edition of the works of Immanuel Kant)*. Cambridge University Press.

Khoo, J., & Knobe, J. (2018, March). Moral disagreement and moral semantics. *Noûs*, *52*(1), 109–143. https://doi.org/10.1111/nous.12151

Knobe, J. (2021). Philosophical intuitions are surprisingly stable across both demographic groups and situations. *Filozofia Nauki*, *29*(2), 1–66. https://doi.org/10.14394/filnau.2021.0007

Liao, S., McNally, L., & Meskin, A. (2016, August). Aesthetic adjectives lack uniform behavior. *Inquiry*, *59*(6), 618–631. https://doi.org/10.1080/0020174X.2016.1208927

Liao, S., Strohminger, N., & Sripada, C. S. (2014). Empirically investigating imaginative resistance. *British Journal of Aesthetics*, *54*(3), 339–355. https://doi.org/10.1093/aesthj/ayu027

Moss, D., & Bush, L. S. (2021). Measuring metaaesthetics: Challenges and ways forward. *New Ideas in Psychology*, *62*, 100866. https://www.sciencedirect.com/science/article/pii/S0732118X21000155

Nichols, S. (2004). After objectivity: An empirical study of moral judgement. *Philosophical Psychology*, *17*(1), 3–26. https://doi.org/10.1080/0951508042000202354

Pölzler, T. (2017, June). Revisiting folk moral realism. *Review of Philosophy and Psychology*, *8*(2), 455–476. https://doi.org/10.1007/s13164-016-0300-9

Rabb, N., Han, A., Nebeker, L., & Winner, E. (2020). Expressivist to the core: Metaaesthetic subjectivism is stable and robust. *New Ideas in Psychology*, *57*, 100760. https://www.sciencedirect.com/science/article/pii/S0732118X19301990

Roberts, P., Andow, J., & Schmidtke, K. (2014). Colour relationalism and the real deliverances of introspection. *Erkenntnis*, *79*(5), 1173–1189. https://doi.org/10.1007/s10670-014-9600-6

Roberts, P., & Schmidtke, K. A. (2016, September). Relationalism about perceptible properties and the principle of charity. *Synthese*, *193*(9), 2779–2803. https://doi.org/10.1007/s11229-015-0886-y

Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). Folk moral relativism. *Mind & Language*, *26*(4), 482–505. https://doi.org/10.1111/j.1468-0017.2011.01428.x

Schafer, K. (2011). Faultless disagreement and aesthetic realism. *Philosophy and Phenomenological Research*, *82*(2), 265–286. https://doi.org/10.1111/j.1933-1592.2010.00391.x

Simoniti, V. (2017). Aesthetic properties as powers. *European Journal of Philosophy*, *25*(4), 1434–1453. https://doi.org/10.1111/ejop.12224

Sytsma, J., & Reuter, K. (2017, September). Experimental philosophy of pain. *Journal of Indian Council of Philosophical Research*, *34*(3), 611–628. https://doi.org/10.1007/s40961-017-0121-y

Wright, J. (2013). Tracking instability in our philosophical judgments: Is it intuitive? *Philosophical Psychology*, *26*(4), 485–501. https://doi.org/10.1080/09515089.2012.672172