

Frankfurt am Main, 07.08.2024

Dear Dr. Chris Chambers, dear Dr. Patrick Savage,

We are very grateful to our reviewer, Dr. Patrick Savage, for his valuable feedback. We have addressed all mentioned points in our revised manuscript. In what follows, we briefly respond to individual points.

Best wishes,

Camila Bruder,

On behalf of all authors

Review by Patrick Savage, 12 Jun 2024 01:12

I think the manuscript mostly meets the key criteria for Stage 2 acceptance: the authors appears to have conducted the study as described and interpreted the results sensibly according to their pre-specified Stage 1 criteria, with appropriate caveats in the Discussion and appropriate weighting in the abstract.

The only minor question I have about following protocol is why the final participant total was 62 instead of the 60 participants proposed. Was this because of collecting more than 60 in case of exclusions? I'd just suggest adding a sentence somewhere to make this explicit.

Once again, thank you very much for the constructive feedback!

Because of our test-retest design, we anticipated a higher number of incomplete datasets. Therefore, we collected data from 62 participants to ensure we would have complete datasets for 60 participants. Surprisingly, the expected dropouts did not occur, and we decided not to randomly exclude the additional participants. We have added an explicit sentence about this to the manuscript (lines 267-269).

I do think, however, that there are a few structural changes required to confirm with PCI-RR policies:

I note that the authors appear to have altered the introduction from the version that received IPA to add recently published references (e.g., Ostrega et al., 2024). This contradicts PCI-RR policy (https://rr.peercommunityin.org/PCIRegisteredReports/about/full_policies):

“Aside from changes in tense (e.g. future tense to past tense), correction of typographic and grammatical errors, and correction of clear factual errors, the introduction, rationale and hypotheses of the Stage 2 submission must remain identical to those in the approved Stage 1 manuscript. To make any changes clear, authors are required to submit a tracked changes version of the manuscript at Stage 2.”

It is commendable to incorporate recent studies, but I'd suggest that the revised version cites new references not cited in the Stage 1 protocol in the Discussion section instead. (I think it is OK to update references for preprints already cited at Stage 1 to their recently published final versions - e.g., Albouy et al., 2024; Ozaki et al., 2024).

Thank you for the clarification. We included the new reference in the Introduction to be thorough, as it did not alter the meaning of the paragraph. We did not realize this would be problematic. Since it is not particularly relevant to our discussion, we have now removed it from the Introduction.

I note the authors have also included exploratory analyses in the results section before the official "Exploratory analysis" section ("Liking ratings differed for each of the styles in pairwise comparisons with all other styles (all p s < .001; based on average ratings of sessions 1 and 2 and adjusting p-values for multiple comparisons with the Holm method. Note that these comparisons were not preregistered; they were included for completeness, since it seemed reasonable to first present the distribution of our dependent variable).") These (and any other analyses not part of the Stage 1 confirmatory analyses) should be moved to the "Exploratory analyses" section.

Thank you for pointing that out. We have moved that paragraph into the “Exploratory analyses” section.

Finally, the manuscript refers to some supplementary figures but these are not visible in the manuscript. I suggest the supplementary figures be merged with the main manuscript (after the reference section).

Thank you for this suggestion. We aimed to keep the manuscript as concise as possible by moving those figures to a separate Supplementary Information document. Following your advice, we have now merged that file with the manuscript, appending it to the end.

Other points:

In general I recommend changing the title at Stage 2 submission to something that is more informative about the actual results

Thank you for this suggestion, which we have followed.

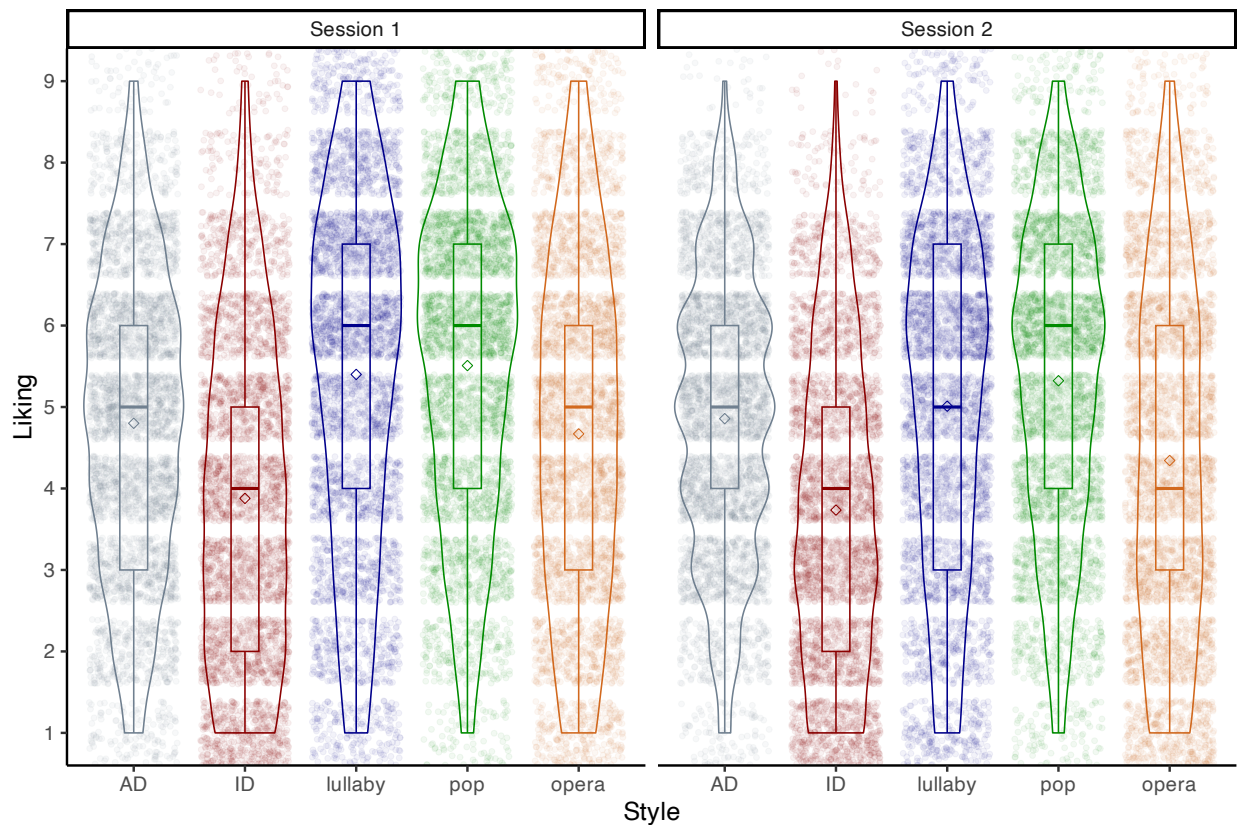
I recommend including more details about the sample in the abstract (e.g., 62 experiment participants, lyrics in Brazilian Portuguese) [sorry I didn't catch this in Stage 1!]. I recommend this editorial for thinking about how to make abstracts and titles more informative: <https://www.nature.com/articles/s41562-023-01596-8>

Thank you for this suggestion, which we have also followed.

Fig. 1: I recommend visualizing individual datapoints in addition to averages/distributions (<https://www.nature.com/articles/s41551-017-0079>)

Thank you for this suggestion, which we assume refers to Figure 1 (right), regarding MM1 values. We added the individual datapoints of MM1 values to that figure.

For Figure 1 (left), which displays overall liking ratings, there are too many individual datapoints to plot clearly ($N = 4092$ trials per style in each session), so we kept the boxplots and mean values but added information on the distribution by adding violin plots. Only for comparison, here is an updated version of Figure 1(left), including individual datapoints of liking ratings using the 'geom_jitter(alpha = .05)' argument in ggplot2:



Line 549: “ $r(20) = .37(0.087)$; and $r(20) = .37 (p = .089)$ ”: is it missing a “ $p =$ ” before “0.087”? Personally I’m not sure it is helpful to even report p-values at all for exploratory analyses, but if you do want to report them you should fix that typo.

Thank you for this suggestion. You were correct, that was a missing “ $p =$ ”. We corrected that.

Line 571: “mirrored” typo

Thank you for pointing that out. We corrected that.

I’d remove “highly” from the abstract - feels a bit strong

We assume the reviewer is referring to the sentence: “We found highly idiosyncratic preferences across all styles.” We also mention that the singers were highly trained, which we believe is not controversial.

In this context, we'd argue that describing preferences across all styles as highly idiosyncratic is grounded. If we restrict ourselves to our registered analyses, inspection of MM1 values, which ranged from .36 to .54, supports that interrater agreement was limited (or that shared taste was low, or preferences highly private or idiosyncratic) for all styles. If we inspect the beholder index, which has been used in studies of aesthetic preferences in the visual domain (Hönekopp, 2006; Kramer et al., 2018; Leder et al., 2016; Specker et al., 2020) and quantifies the time-stable variance that is accounted for (or not residual) to quantify private taste, the values ranging from .59 to .75 for bi_1 (or from .8 to .9 for bi_2 , which also considers the variance in the rater cluster as contributing to private taste) indicate a preponderance of private or idiosyncratic taste for all vocalizations styles. Therefore, we believe that describing preferences as "highly idiosyncratic" is accurate and well-justified. This is reflected in our revised manuscript title.

Line 588: I agree that Cronbach's alpha is "an inadequate measure of interrater agreement", but you might want to support this with a reference

Thank you for this suggestion. We have now included references to support this claim, which were already cited elsewhere in the manuscript – i.e., the Hönekopp (2006) and Kramer et al. (2018) papers.

Line 672: "(and actually equivalent to lullaby singing)" - remove parentheses and "actually"

We followed your suggestion.

Line 675: "definetively" typo

Thank you, this is now corrected.

Line 721: "the prediction of some consistency of average preferences for some voices across styles was supported by the found interstyle agreement of .52, which, according to the specified threshold of .8, is not considered highly consistent" - this wording is confusing - perhaps instead of "some consistency", "limited consistency" (as in the abstract) would be better? Grammar of "found interstyle..." also feels a bit awkward.

Thank you for this suggestion. We rephrased this section to "Concerning our second hypothesis, the prediction of limited consistency of average preferences for some voices across styles was supported by an interstyle agreement of .52. According to the specified threshold of .8, this is not considered highly consistent.

References

- Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgments of facial attractiveness. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 199–209. <https://doi.org/10.1037/0096-1523.32.2.199>
- Kramer, R. S. S., Mileva, M., & Ritchie, K. L. (2018). Inter-rater agreement in trait judgements from faces. *PLOS ONE*, *13*(8), e0202655. <https://doi.org/10.1371/journal.pone.0202655>
- Leder, H., Goller, J., Rigotti, T., & Forster, M. (2016). Private and shared taste in art and face appreciation. *Frontiers in Human Neuroscience*, *10*. <https://doi.org/10.3389/fnhum.2016.00155>
- Specker, E., Forster, M., Brinkmann, H., Boddy, J., Immelmann, B., Goller, J., Pelowski, M., Rosenberg, R., & Leder, H. (2020). Warm, lively, rough? Assessing agreement on aesthetic effects of artworks. *PLOS ONE*, *15*(5), e0232083. <https://doi.org/10.1371/journal.pone.0232083>