**Reviewer 1:**

Using textures from several wallpaper groups, which are defined by different combinations of the 4 fundamental symmetries, the authors asked how similar exemplars from the same group are to each other. As highlighted by the authors, symmetries from those wallpaper groups are interesting stimuli to study the processing of symmetry at the brain level (fMRI, EEG studies) and the perceptual level (psychophysics experiments). However, the perceptual categorisation of those different wallpaper groups doesn’t necessarily match their mathematical definition; and some exemplars within one same group can also vary perceptually. Starting from this latter observation, the authors used a behavioural sorting task to determine the amount of variability in self-similarity among 5 chosen wallpaper groups. They found that, except for the simplest group that is only composed of translational symmetry, the self-similarity of exemplars belonging to the same wallpaper group doesn’t directly reflect the established wallpaper groups.  
This is an interesting result that brings novel questions related to the processing and perception of similarity of symmetrical patterns. The finding also further suggests some discrepancy between mathematically defined groups of symmetry and the perceived similarity of such stimuli.

**We thank the reviewer for these helpful comments and have addressed them below.**

Comments:

- 5 wallpaper groups were selected (P1, P6, P3M1, P31M, and P6), based on a previous study (Clarke et al. 2011) that reported those groups (except P1) to be high in self-similarity. I feel the justification for choosing such a subset of wallpaper groups is a bit weak and the study would benefit from adding extra groups that contain fewer symmetries. For instance, groups with an intermediate number of rotations, such as P2 or PMM, could provide valuable information, notably given the link between similarity relationships and the number of rotations (Clarke et al. 2011).  
> line 166: “We speculate that this lack of further differentiation is a result of an upper limit on how additional complexity can influence perceptual self-similarity.” this could shed light on this speculation as we would in that case expect a linear relationship between the perceived self-similarity and the stimulus complexity.

**P1 was also reported as high in self-similarity, in Clarke et al. (2011; see Figure 7). The selection of P6, P3M1, P31M and P6M was also partially motivated by the fact that these four groups share the same lattice shape, and by the fact that P6, P3M1, P31M are all subgroups of P6M (see Kohler & Clarke, 2021), but differ in their symmetry content. We have elaborated on this in the text.**

**Our motivations for selecting the groups notwithstanding, we agree with the reviewer that the inclusion of groups like P2, PMM and P4M, that were also reported as being high in self-similarity by Clarke et al., could potentially have offered a better sense of the relationship between complexity and self-similarity. We agree that this is a limitation of the current study and have added a sentence in the Discussion to indicate this.**

- the introduction could benefit from more explicitly stating what the current study adds compared to Clarke et al. (2011)’s study.

**We agree with the reviewer and have added language in the Introduction that spells this out more clearly.**

- line 243: "Upon completion of each sorting task, participants were asked to verbalize which features they used to sort the exemplars. After completion of all five sorting tasks, participants were asked if they had a distinct method for sorting the images, and if any wallpaper group was particularly easy or difficult to sort"  
> Unless I missed it, there is nothing reported about this in the results section or elsewhere. I suggest adding a few words on the outcome of those points, as potential differences in the sorting difficulty between wallpaper groups could be informative.

**We have added a sentence about our participants’ comments regarding the difficulty in sorting P1 exemplars.**

Results section:  
- line 101: "Our pairwise t-tests also showed that P31M had lower Jaccard indices than P6 (p=0.037). This effect is relatively weak."  
> there is no reported statistic that tests the strength of such an effect (e.g. the effect size) to support the claim that the effect is relatively weak.

**We now report the complete statistics for this comparison, including Cohen’s *D* as a measure of effect size. Our results show that the effect described above does not survive Bonferroni correction for multiple comparisons and has a fairly weak effect size (*D* = 0.215). We thank the reviewer for pushing us to be more precise about this.**

- figure 3 and figure 4 could appear on the same panel to ease comparison between measures.

**We have merged Figures 3 and 4 to address this comment.**

Discussion section:  
- the authors mention previous neuroimaging and psychophysics work that investigated the distinction between wallpaper groups at different levels. Could they add a word on how the variability of the perceptual self-similarity they found within wallpaper groups might impact cortical representations or psychophysical thresholds?

**We have added language to that effect in the third paragraph of the Discussion.**

**Reviewer 2:**

In their article “Perceptual Similarities Among Wallpaper Group Exemplars”, Kohler and colleagues investigate the perceived similarity of different exemplars of the same wallpaper groups in a group of untrained participants. Results showed that P1 was the most self-similar group among the others, which did not differ much in terms of self-similarity.

The article is clearly written, and the provided figures help the reader to follow the text. Nevertheless, more background is needed in the Introduction section to introduce the topic. The adopted methods are appropriate to investigate the research question. However, a subsection in which authors report the statistical approach of the study in the Materials and Methods is needed. Regarding the statistical approach of the study, a statistical correction of the significance level should be adopted when conducing post-hoc pairwise comparisons among groups. Finally, results should be reported in standard format throughout the whole manuscript.

**We thank the reviewer for these helpful comments and have addressed them below.**

**Major Issues**

* The article is missing keywords, which are an important hint for the reader to create a better picture of the context in which the article is rooted. Please add them.

**We agree with the reviewer that keywords are important and have added them.**

* The Introduction section does not provide enough background to the reader. What is the definition of symmetry? What is a wallpaper group? What does each mentioned symmetry type mean?

**The reviewer raises an important point. We have extensively revised the Introduction to provide more background, and now include a figure showing the four fundamental symmetries. We think this will make the paper more accessible to non-experts, and hope the reviewer agrees.**

* In the Results section, the authors referred to group P1 without providing its features. How is P1 different from other groups? Please clarify.

**In fact, a general introduction to the features/symmetry content of the five wallpaper groups used in the study is missing. In our view, this belongs in the Introduction, and we have added it there. We appreciate the reviewer bringing this to our attention.**

* Line 87: Please report results in the standard format. When the authors posit “[…] other groups (median = 4-5)”, which groups have median 4 and which groups have median 5? Please clarify.

**We are not sure what is meant by “the standard format” here. We are reporting that the median number of sets generated was 3 for group P1, and 4-5 for the four other groups. The intention is to provide a general overview of our results, and we refer the reader to a figure where exact data from each group can be read out. We have minimally updated the sentence, and hope this will adequately address the reviewer’s concern.**

* Line 90: “Post-hoc pairwise t-tests showed that the mean number of sets was lower for P1 than all other groups, but no other means differed.” Please report the results. Also, was any statistical correction adopted to adjust the significance level based to control for repeated measures? If not, please adopt a correction (e.g., Bonferroni correction) and add it to the manuscript every time that post-hoc pairwise comparisons are conducted.
* Lines 96-97: “Post-hoc pairwise t -tests showed that P1 had higher Jaccard indices than all other groups”. Please add the results and not only p-value.
* Lines 101-102: “Our pairwise t-tests also showed that P31M had lower Jaccard indices than P6”. Please add the results and not only p-value.

**We are now reporting corrected significance. All statistics are for the two sets of post-hoc pairwise t-tests are now reported in Table 1.**

* How were participants recruited? Please clarify.

**Participants were recruited from the undergraduate participant pool in the Department of Psychology at the Pennsylvania State University. We now note this in the Methods.**

* Please add a Statistical Analysis subsection in the Materials and Methods section in which the authors report the statistical approach to data of their study.

**We have added a Statistical Analysis subsection and moved most of our discussion of ANOVAs and post-hoc t-test there.**

**Minor Issues**

* Lines 35-39: the authors posited “Brain imaging studies using functional MRI (Kohler et al., 2016) and EEG (Kohler et al., 2018; Kohler and Clarke, 2021) have shown that the human visual system carries detailed and precise representations of the symmetries within the individual wallpaper groups.” The sentence seems a bit too generic and an example of what the authors mean for “the human brain carries detailed and precise representation of the symmetries” is needed.

**We have added another sentence that hopefully provides the needed example.**

* Lines 39-42: the authors posited that: Functional MRI evidence from macaque monkeys reveal similar representations in analogous areas of the macaque visual system”. To which brain areas are the authors referring?

**We have modified the sentence to be clearer about the brain areas we are referring to.**

* Lines 43-45: “These representations, complex as they are, […]”, which representations? Please clarify.

**We have modified the sentence to be clearer about the representations we are referring to.**

* Figures and tables should be inserted in the main text and not on the side of the text. Please revise Figure 1 and Table 21 accordingly.

**We have revised accordingly.**

* Lines 62-83: this part does not seem to fit the Introduction section and it regards more the Methods and the Results of the study. Please move it to the Materials and Methods and Results sections.

**While we think it is appropriate to briefly discuss the Methods in the Introduction, we agree with the reviewer that this section had excessive detail for this section of the paper and have revised accordingly.**

* Figure 4: “unlike Figure 3 were each data point is a participant” should be “unlike Figure 3 where each data point is a participant”.

**We have corrected this error.**

* Figure 4: “The box boundary and whiskers follow the same logic as in Figure 3”. Figures’ captions should be self-explanatory. Please clarify the logic of box boundary and whiskers also here, without referring to Figure 3 again.

**We have now merged Figures 3 and 4, so this problem has been addressed.**

* “We will explore this idea more in depth shortly, but for now we can conclude that […]”. There are no temporary conclusions in a paper. Please delete this part of the sentence, stick to data, and state the conclusions.

**We have deleted the sentence as requested.**

**Final comments**

I would recommend the article to undergo Major Revision.