**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. The pdf file “wallpaper\_sort\_symmetry\_revised.pdf” holds the new version of the manuscript and the file “wallpaper\_sort\_symmetry\_revised\_changes.pdf” holds the same manuscript but with changes from the initial submission highlighted for the reviewer’s convenience.**

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 Introduction and in the Stimulus Generation section of the Methods.**

**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 to the Discussion about the 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.**