Perceptual Similarities Among Wallpaper Group Exemplars

Peter J. Kohler[1,2], Shivam Vedak[3], and Rick O. Gilmore[3]

[1] *York University, Department of Psychology, Toronto, ON M3J 1P3, Canada*
[2] *Centre for Vision Research, York University, Toronto, ON, M3J 1P3, Canada*
[3] *Department of Psychology, The Pennsylvania State University, Pennsylvania, USA*

Abstract

Symmetries are abundant within the visual environment, and many animals species are sensitive to visual symmetries. Wallpaper groups constitute a class of 17 regular textures that each contain a distinct combination of the four fundamental symmetries, translation, reflection, rotation and glide reflection, and together represent the complete set of possible symmetries in two-dimensional images. Wallpapers are visually compelling and elicit responses in visual brain areas that precisely capture the symmetry content of each group in humans and other primates. Here we ask to what extent exemplars from the same wallpaper group are perceptually similar. We algorithmically produce a set of well-matched exemplars from 5 of the 17 wallpaper groups and instructed participants to freely sort the exemplars from each group into as many subsets as they wished based on any criteria they saw appropriate. *P1*, the simplest of the 17 groups, was consistently rated more self-similar than any other group, while the other four groups, although varying in symmetry content, were comparable in self-similarity. Our results suggest that except for the most extreme case (*P1*), perceived self-similarity of wallpaper groups is not directly tied to categories of symmetry based on group theory.

Introduction

Symmetry has been recognized as important for human visual perception since the late 19th century (Mach, 1959). In the two spatial dimensions relevant for images, symmetries can be combined in 17 distinct ways, *the wallpaper groups* (Fedorov, 1891; Polya, 1924; Liu et al., 2010). Wallpaper groups are different from stimuli typically used to probe the role of symmetry in visual perception in two ways: First, they contain combinations of the four fundamental symmetry types translation, reflection, rotation and glide reflection, rather than just reflection or mirror symmetry, which have been the focus of most vision research. Second, the symmetries in wallpaper groups are repeated to tile the plane, rather than positioned at a single image location as is usually the case. These differences, and the fact that wallpaper groups together form the complete set of symmetries possible in the two-dimensional image plane, make wallpapers an interesting stimulus set for studying perception of visual symmetries.

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. Functional MRI evidence from macaque monkeys reveal similar representations in analogous areas of the macaque visual system (Audurier et al., 2021).

These representations, complex as they are, do not appear to be readily available for driving conscious behaviour: Humans have limited intuitive sense of group membership for wallpaper group exemplars, as evidenced by behavioral experiments showing that although naïve observers can distinguish many of the wallpaper groups (Landwehr, 2009), they tend to sort exemplars into fewer (4-12) sets than the number of wallpaper groups, often placing exemplars from different wallpaper groups in the same set (Clarke et al., 2011). Wallpaper groups are nonetheless visually compelling, and anecdotally we have observed that exemplars from a given group can be quite perceptually diverse. This observation inspired the current study. Here, we use behavioral sorting, a common technique to study perceptual categorization (Milton et al., 2008; Pothos et al., 2011), to probe the perceptual self-similarity of different exemplars from the same wallpaper group, and assess the extent to which self-similarity varies across five groups.

We algorithmically generated 20 well-matched exemplars from each group (see Figures 1 and 2 for a selection of the exemplars, and the Materials and Methods section for details on how they were generated) and printed them out on white cardstock. We then gave participants the 20 cards with exemplars from each wallpaper group, and asked them to freely sort them into as many subsets as they wished based on any criteria they saw appropriate. This approach allowed us to compare the five wallpaper groups, both in terms of how many subsets participants generated, and also in terms of the *Jaccard index*, a summary statistic capturing the similarity across exemplar pairs for each group. Within each group, we were also able to identify exemplar pairs that were rated

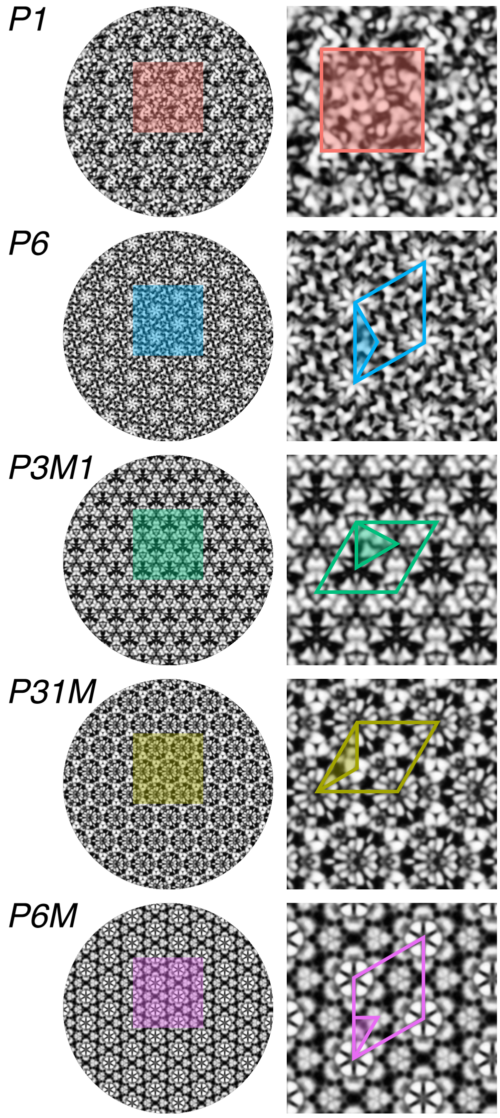

Figure 1: The fundamental region and lattice structure of the five wallpaper groups used in the study. The complete wallpaper is shown in the left-hand column with a shaded region that is repeated and enlarged in the right-hand column. The colored outline in the enlarged region indicates the repeating lattice for each group, while the shaded area indicates the fundamental region (see text). For *P1* the fundamental region covers the entire lattice. Note that even though *P6* and *P31M* have the same fundamental region and lattice shapes, they differ in terms of the symmetries present within the lattice - most notably, *P31M* contains reflection symmetry while *P6* does not. The symmetry content of each group is detailed on the wallpaper group wikipedia page.

as highly similar and highly dissimilar. Our main conclusion is that *P1* was systematically more self-similar than the any other groups, while the other four other groups could not be distinguished on these measures. We also show that for all five groups, participants consistently group certain pairs of exemplars together, although the number of consistent pairs varies among groups. Our results open the door to further investigations into the factors that drive perceptual similarity among wallpaper group exemplars, and indeed among exemplars from different classes of structured patterns.

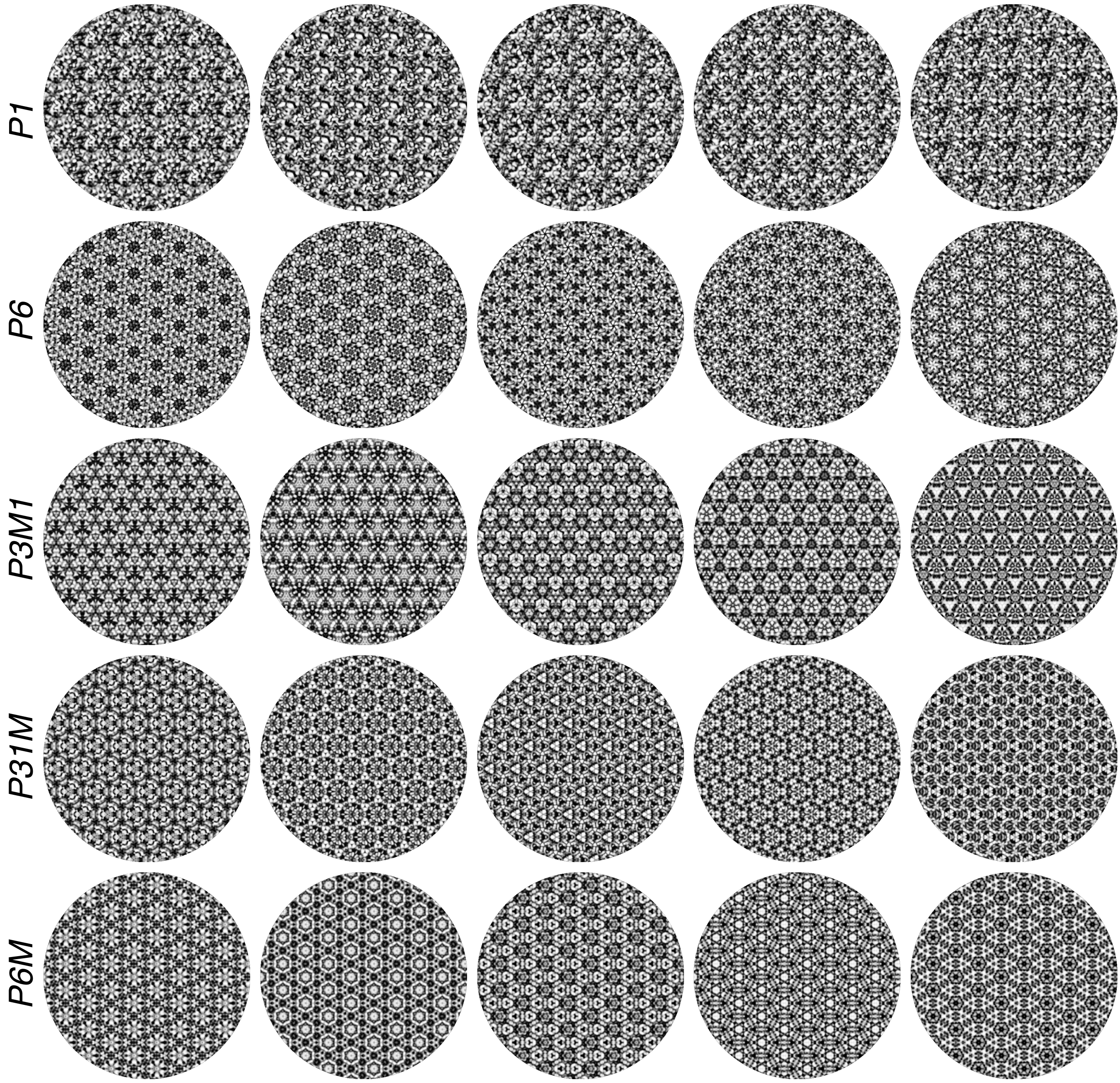

Figure 2: 5 of the 20 exemplars used for each group are shown to highlight the diversity among exemplars.

Results

Wallpaper group *P1* was more self-similar than the other four groups. This was evident in the number of sets generated for this group across participants, which was lower for *P1* (median = 3) than for the other groups (median = 4-5, see Figure 3). We confirmed this observation statistically by running a repeated measures analysis of variance (ANOVA) with group as a fixed

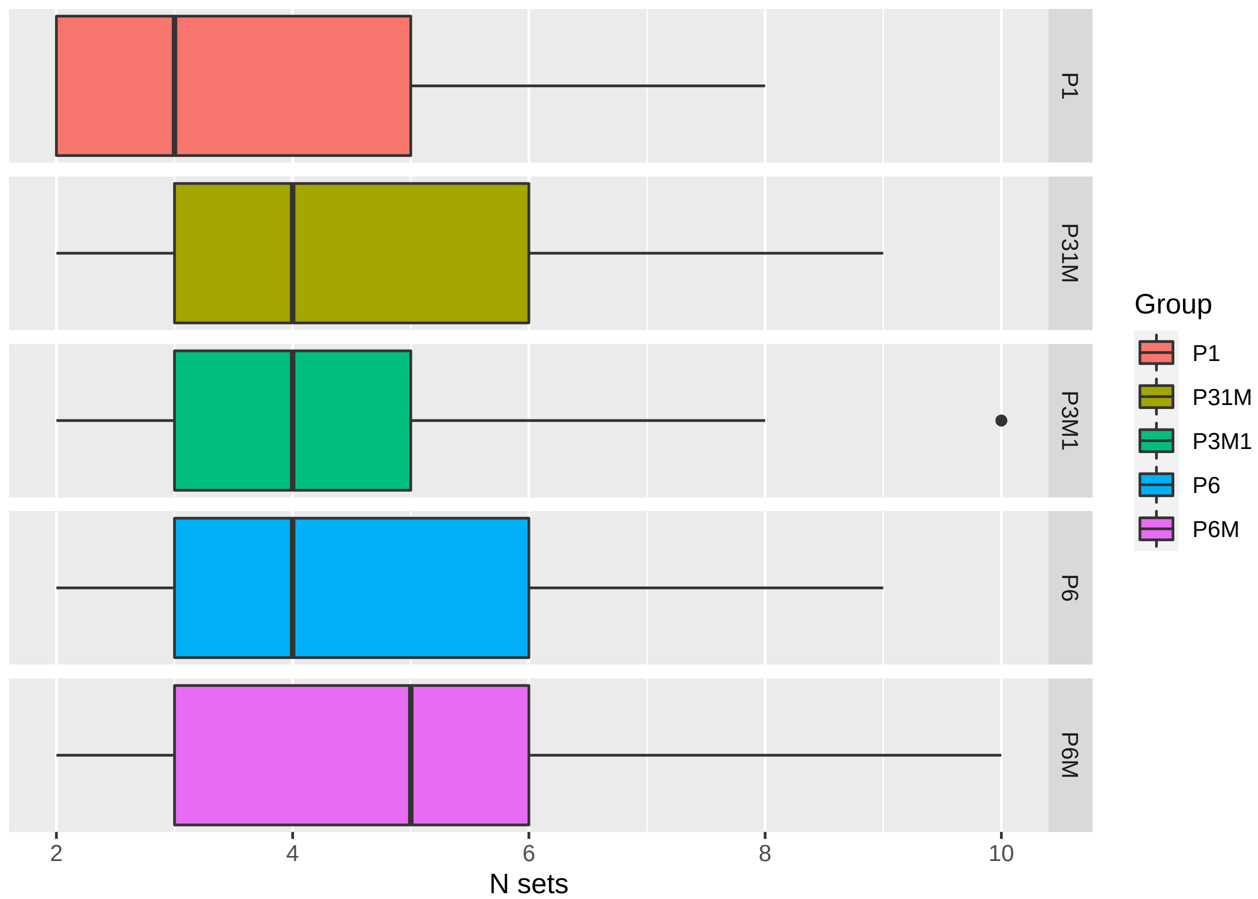

Figure 3: Boxplots showing the number of subsets generated by participants for each of the wallpaper groups. The lower box boundary is the 25th percentile. The dark line in the box is the median. The upper box boundary is the 75th percentile. The "whiskers" show -/+ the interquartile range * 1.5.

factor and participant as a random factor, which revealed a significant effet of group ($F(4,124) = 7.330, p < 0.0001$)). 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. Next, we computed the Jaccard index (see Materials and Methods) across participants for every pairwise combination of exemplars in each group. This provides a measure of the similarity between exemplars within each group. *P1* had sysmetically higher Jaccard indices than the four other groups (see Figure 4), as confirmed by an ANOVA with wallpaper group as a factor. The analysis revealed a statistically signficant effect of group ($F(4,495) = 20.178, p < 0.0001$). Post-hoc pairwise t-tests showed that *P1* had higher Jaccard indices than all other groups ($ps < 0.0001$). The fact that the group (*P1*) for which fewer subsets were generated also had higher Jaccard indices than the other groups illustrates the inherent link between the two measures. For wallpaper groups where the 20 exemplars are sorted into fewer subsets, each individual exemplar pair is more likely to be a member of the same subset, and less likely to be a member of distinct subsets. This in turn leads to higher Jaccard indices. Our pairwise t-tests also showed that *P31M* had lower Jaccard indices than *P6* ($p = 0.037$). This effect is relatively weak, but may reflect real differences in how consistently exemplars were grouped together across participants. We will explore this idea more in depth shortly, but for now we can conclude that out of the five groups tested, *P1* is the only one that can be reliably differentiated based on our measures, being higher on self-similarity among the exemplars, and thus lower on

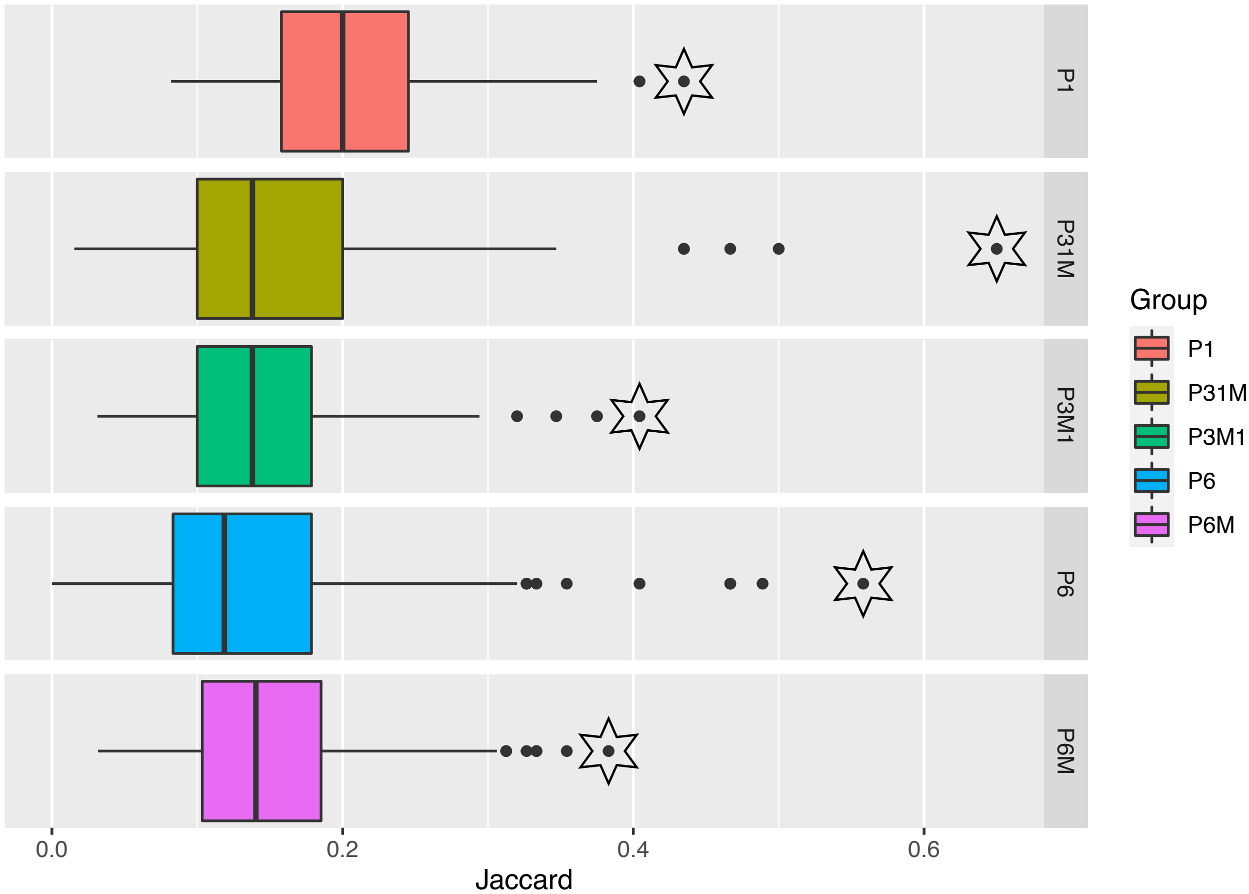

Figure 4: Boxplots showing Jaccard indices for every pairwise combination of exemplars in each of the wallpaper groups. Note that each data point here is the Jaccard index for a particular exemplar pair calculated across participants, unlike Figure 3 were each data point is a participant. The box boundary and whiskers follow the same logic as in Figure 3. The exemplar pairs with the highest Jaccard indices have been highlighted with stars. Those outlier pairs are explored further in Figure 6.

diversity among exemplars.

In order to quantify the extent to which exemplars were consistently grouped together, we ran a permutation analysis in which exemplar labels were shuffled among the sets generated for each participant (see Materials and Methods). This provides, for each group, the expected distribution of Jaccard indices for every pairwise combination of exemplars, if exemplars were assigned randomly to subsets. And the analysis allows us to compute an empirical z-score that expressed the extent to which a given pair of exemplars deviates from random assignment. Because the random distribution is generated by shuffling exemplars across the specific sets generated by each participant for each group, this z-score is independent of the number of sets. If for a given group, none of the pairs deviate significantly from the random distribution, it would indicate that no exemplar pairs were consistently grouped together across participants. To estimate the extent to which this is the case, we look at the distribution of z-

	consistent pairings	
group	$p < 0.01$	$p < 0.0001$
P1	6	1
P31M	17	10
P3M1	12	3
P6	17	11
P6M	15	4

Table 1: Number of consistent pairings at two different α-levels for the five groups.

scores across the pairs for each group, as plotted in Figure 5, and count the number of pairs for each group for which the p-value associated with the threshold exceeds a given α value. At a threshold of $\alpha = 0.01$, several pairs survive for all groups, and even at a much more conservative criterion of $\alpha = 0.0001$ most groups have more than one pairing that survives (see Table 1). It is worth noting that the latter threshold ($\alpha = 0.0001$) is lower than the α associated with a Bonferroni correction within group, given that there are 190 pairs per group:

$$\alpha = \frac{0.05}{190} = 0.0003$$

. So we conclude that for several exemplar pairs, participants are consistent in how they tend to pair the exemplars. It is interesting to consider that this measure of consistency might provide another way of differentiating wallpaper groups in terms of perceptual self-similarity. While groups *P31M*, *P3M1*, *P6* and *P6M* have comparable Jaccard scores (see Figure 4), they differ in the number of consistent pairings, with *P31M* and *P6* producing more consistent pairs than the other two (see Figure 5).

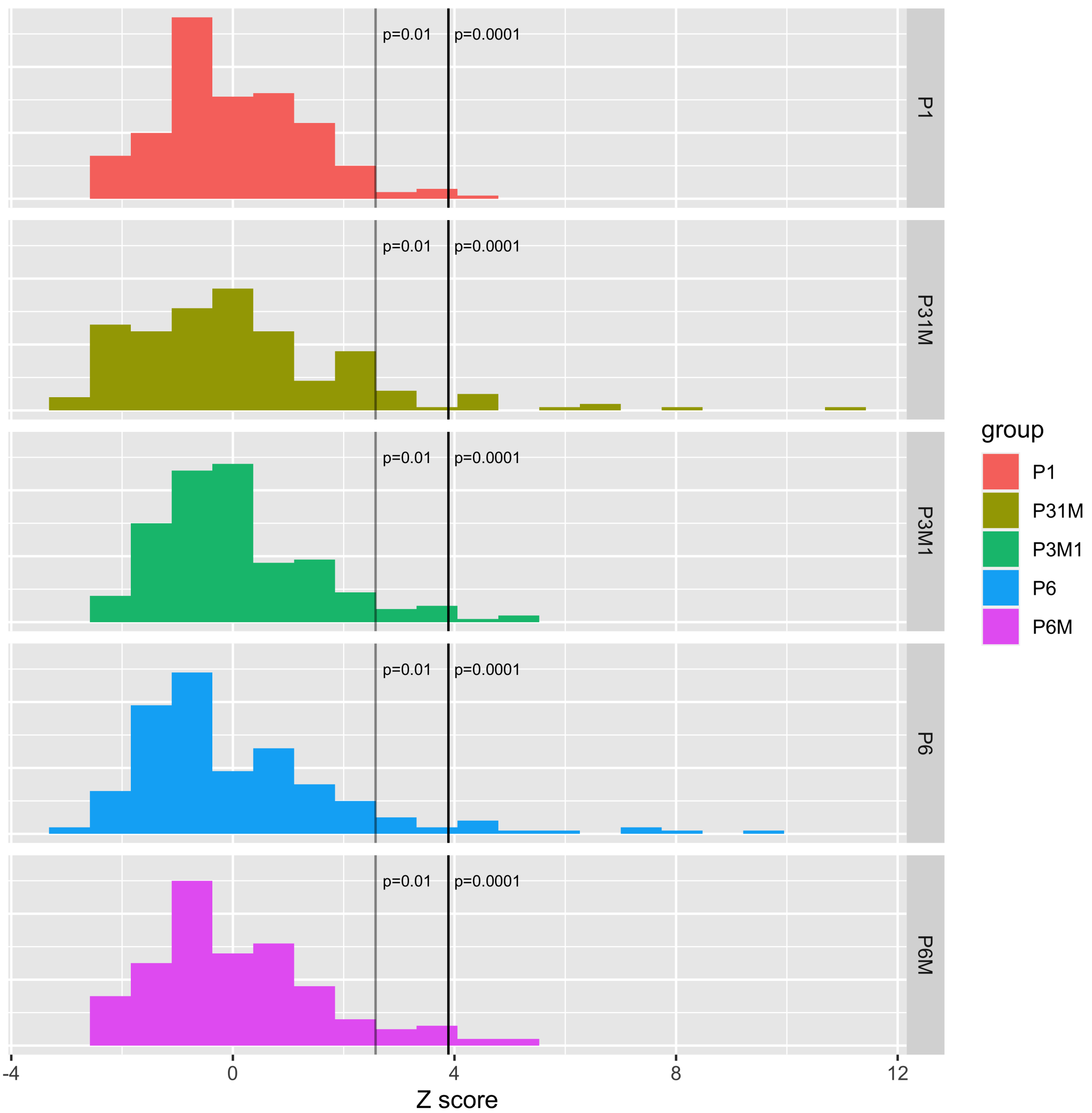

Figure 5: Distribution of z-scores across the 190 pairs in each of the five wallpaper groups. The two lines indicate the z-scores associated with α of 0.01 and 0.0001, respectively.

The Jaccard indices also allow us to focus on exemplar pairs that have a high level of similarity relative to the rest of the pairs in the set. We do this by identifying outliers pairs from each group in term of Jaccard indices, as identified with stars in Figure 4. Because the Jaccard indices are computed across participants, these outliers are also among the pairs most consistently sorted together, as identified in Figure 5. For each exemplar in each outlier pair, we can visualize the pairwise similarity (as measured by the Jaccard index) to every other exemplar in the set (see Figure 6). That is, we can visualize portions of the network of perceived similarity within a set of exemplars. Future work could probe the extent to which networks of perceived similarity have similar structure across wallpaper groups and examine what perceptual features best account for participants' perceptions of exemplar similarity.

Discussion

Previous work has demonstrated that visual cortex of both humans and macaque monkeys carries highly detailed representations of the symmetries within wallpaper groups, as evidenced by systematic differences in the magnitude of the response elicited by different groups (Kohler et al., 2016; Kohler and Clarke, 2021; Audurier et al., 2021). This distinction between groups can also be observed in psychophysical threshold measurements (Kohler and Clarke, 2021), although observers may not have a strong awareness of the wallpaper group membership of individual exemplars (Clarke et al., 2011). In the current study, we explored a new piece of the story about how wallpaper groups are processed by the visual system, namely the issue of how self-similar different exemplars from the *same* wallpaper group appear to untrained observers. We tested this by asking participants to spontaneously sort 20 exemplars from each of five wallpaper groups into different subsets.

Our first finding concerns the number of subsets generated for each group. We find that *P1* is divided into fewer subsets than the other four groups. This indicates that the limited complexity of this group, which contains only translation symmetry, has a direct effect of the number of distinct subsets. The relationship between complexity / symmetry content and number of subsets produced is not straightforward, however, as indicated by the fact that *P6M* is not consistently grouped into more subsets than *P6*, *P3M1* and *P31M*, despite the fact that these other groups all contain fewer symmetries than *P6M* - and generate weaker brain activity (Kohler and Clarke, 2021). We speculate that this lack of further differentiation is a result of an upper limit on how additional complexity can influence perceptual self-similarity.

We also computed Jaccard indices that, for every possible exemplar pair, expresses the frequency of those two exemplars being grouped together. As described above, the average Jaccard index for a group is inherently linked to the number of subsets produced for that group, because fewer subsets mean that exemplars are more likely to be made members of the same pair, and less likely to be made members of the same pair. It is therefore not surprising that we find the same general pattern for Jaccard indices and number of subsets, namely that *P1* has higher indices than the other groups. The advantage of the Jaccard indices, however, is that they allow us to conduct a permutation analysis that quantifies the extent to which pairs of exemplars are

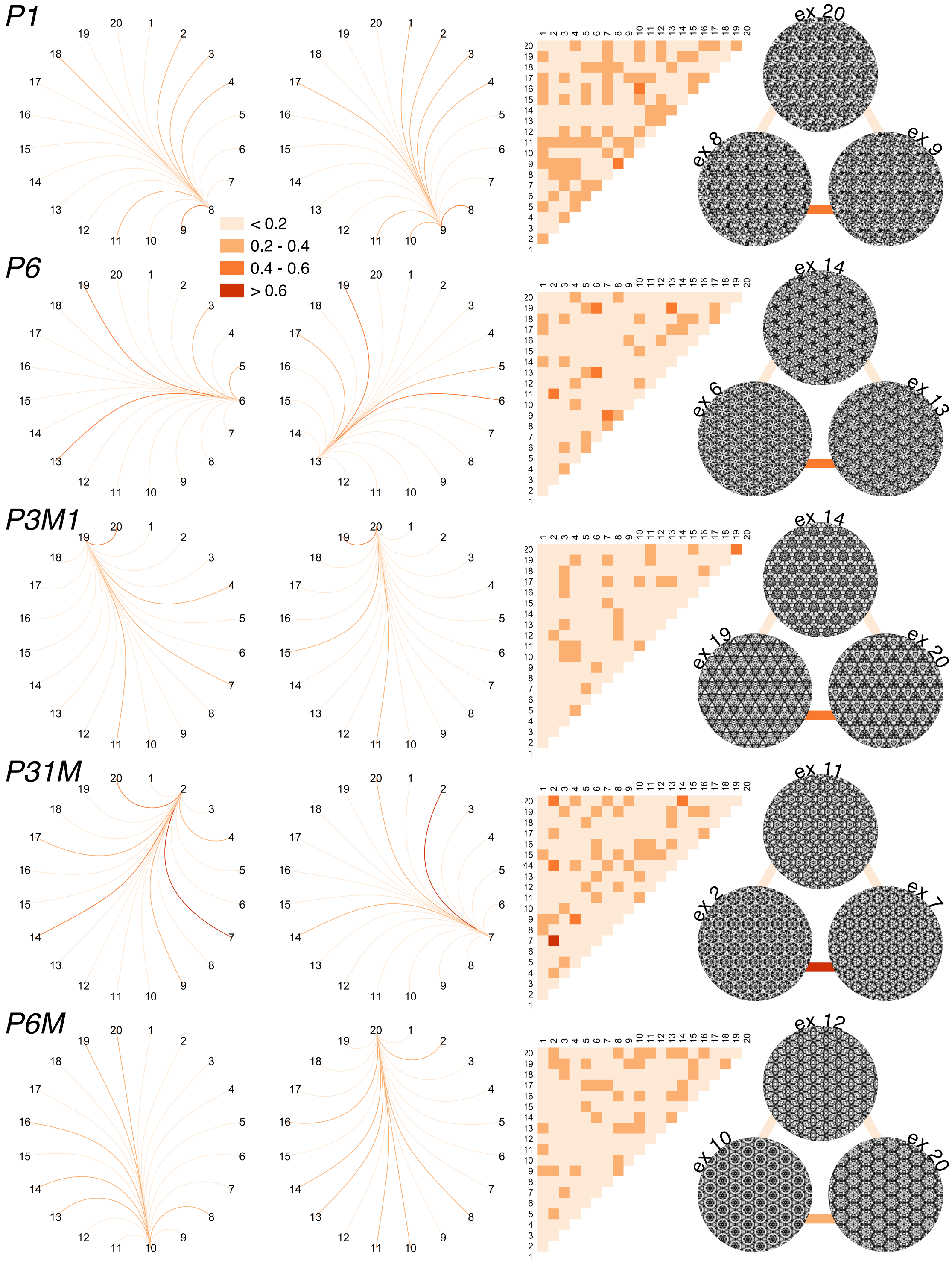

Figure 6: For each wallpaper group, we identified the two most self-similar exemplars, the same pair that is indicated by the right-most datapoint for each group in Figure 5. The two circular network plots are showing the pairwise similarities between those two exemplars and every other exemplar in the set. The pairwise similarities across all exemplars are plotted as a similarity matrix and on the rightmost side of the plot, the two most self-similar exemplars (bottom) are plotted with the exemplar that was least similar to both (top). The connecting lines between the exemplars indicate the similarity.

consistently grouped together across participants, independent of the number of sets produced for a given group. It is important to note that consistency in the choice of which exemplars to group into subsets is not an unavoidable consequence of our experimental design, and it does not follow naturally from the results described so far. It would be perfectly possible for participants to group the sets together, producing fewer subsets for *P1* as observed, but exhibit no consistency across participants at all. That is not what we see, however. Even when setting a conservative threshold, all five groups produce one or more pairs that are consistently grouped together, demonstrating that the sorting of exemplars into subsets is not done randomly or in an arbitrary across the participants. Rather, different individuals agree to some extent on which exemplars belong together. Because our measure of consistency is independent of the number of subsets produced for a given group, it allows us to show that although *P1* has the highest overall Jaccard indices (as a result of the fewer sets produced for this group), it in fact produces fewer consistent pairs than other groups (see Table 1).

In sum, we find consistencies in the way that untrained human observers sort wallpaper images. Observers sort exemplars with translational symmmetry alone (*P1*) into smaller numbers of sets than exemplars with rotational or reflection symmetry. On average, pairs of *P1* exemplars are sorted together more often than exemplars from other wallpaper groups. At the same time, some specific exemplar pairs from wallpaper groups with 3- or 6- fold rotational or reflection symmetry are sorted together substantially more often than predicted by chance.

Taken together, our results suggest that human observers are sensitive to the dimensions of 2D symmetry (translation, rotation, and reflection) embedded in wallpaper exemplars. However, their behavior shows only weak evidence that group-theoretic measures of symmetry influence the perception of self-similarity. These results contribute to a small, but growing literature on the perception of visual aesthetics (Carneiro et al., 2012; Graham et al., 2010; Friedenberg, 2012; Laine-Hernandez and Westman, 2008; Richards, 1972) where symmetry is one of many contributing factors.

Materials and Methods

Participants

33 participants (9 Male, 24 Female), ranging in age between 18 and 35 completed this study. All participants had self- reported 20/20 or corrected to 20/20 vision. We obtained written consent to participate from all participants under procedures approved by the Institutional Review Board of The Pennsylvania State University (#38536). The research was conducted according to the principles expressed in the Declaration of Helsinki. Participants include n=11 collected and described in (Vedak, 2014), plus an additional group collected at a later date using the same protocol.

Stimulus Generation

Five wallpaper groups (*P1*, *P3M1*, *P31M*, *P6* and *P6M*) that has previously been shown to be high in self-similarity (Clarke et al., 2011), were selected. 20 exemplars from each of these five wallpaper groups were generated using a modified version of the methodology developed by Clarke and colleagues (Clarke et al., 2011) that we have described in detail elsewhere (Kohler et al., 2016). Briefly, exemplar patterns for each group were generated from random-noise textures, which were then repeated and transformed to cover the plane, according to the symmetry axes and geometric lattice specific to each group. The use of noise textures as the starting point for stimulus generation allowed the creation of an almost infinite number of distinct exemplars of each wallpaper group. To make individual exemplars as similar as possible we replaced the power spectrum of each exemplar with the median across exemplars within a group. These images were printed onto white cardstock and cut into squares, allowing participants to manipulate the orientation of the images during the sorting tasks. Five exemplars from each group are shown (in reduced size) in Figure 2.

Procedure

Participants were presented with the 20 exemplars of a single wallpaper group (i.e. P1, P3M1, P31M, P6, P6M) and instructed to sort them into subsets by placing them into piles. Participants were advised to sort the exemplars into as many piles as they deemed necessary based on whatever criteria they desired. There were no time constraints placed on this sorting task, and the participants were allowed to move exemplars between piles until they were satisfied with their classification. This method was then repeated for the remaining four wallpaper groups for each participant, with group presentation order randomized between participants. These tasks were carried out on a large table with sufficient space to randomly lay out all twenty exemplars of each set, illuminated by normal overhead room lighting. 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 which if they had a distinct method for sorting the images, and if any wallpaper group was particular easy or difficult to sort.

Generating the Jaccard Index

The data was prepared for analysis by creating one binary variable for each subset created by each participant within a sorting task. Then, each exemplar was assigned a value of one (1) if it was included in a subset, or a value zero (0) if it was not. Next, the similarity of each pair of exemplars within a sorting task was calculated using the Jaccard index, a measure of similarity and diversity for binary data. This index is calculated by the equation

$$J = \frac{x}{x + y + z}$$

with x representing the number of subsets that contained both exemplars, and y and z the number of subsets that contain only one exemplar of the pair (Capra, 2005), across participants. Thus, the Jaccard index is the ratio of the number of subsets containing both exemplars of a pair to the number of subsets containing at least one of the exemplars of a pair, thereby excluding subsets with joint absences.

Permutation Analysis

The permutation analysis involved generating a randomized dataset, as follows. For each participant and wallpaper group, we randomized which specific exemplars were sorted together. This retained the basic structure of each participants' sorting data–the number of subsets created–but randomized the relationship between specific wallpaper exemplars that were sorted together across the participants. We then created 1,000 such permuted datasets, and calculated the Jaccard index for each exemplar pair within each group for each of the permuted datasets. This permitted the calculation of an *empirical* Jaccard index based on the permuted data from which distributional statistics like z could be calculated. The *observed* Jaccard indices for each exemplar pair were then compared to the empirically-derived reference distribution to determine which exemplar pairs were sorted together more frequently than chance would predict.

References

Audurier, P., Héjja-Brichard, Y., Castro, V. D., Kohler, P. J., Norcia, A. M., Durand, J.-B., and Cottereau, B. R. (2021). Symmetry processing in the macaque visual cortex. *bioRxiv*, page 2021.03.13.435181. Publisher: Cold Spring Harbor Laboratory Section: New Results.

Capra, M. G. (2005). Factor Analysis of Card Sort Data: An Alternative to Hierarchical Cluster Analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(5):691–695. Publisher: SAGE Publications Inc.

Carneiro, G., da Silva, N. P., Del Bue, A., and Costeira, J. P. (2012). Artistic image classification: An analysis on the PRINTART database. In *Computer Vision – ECCV 2012*, pages 143–157. Springer Berlin Heidelberg.

Clarke, A. D. F., Green, P. R., Halley, F., and Chantler, M. J. (2011). Similar symmetries: The role of wallpaper groups in perceptual texture similarity. *Symmetry*, 3(4):246–264.

Fedorov, E. (1891). Symmetry in the plane. In *Zapiski Imperatorskogo S. Peterburgskogo Mineralogichesgo Obshchestva [Proc. S. Peterb. Mineral. Soc.]*, volume 2, pages 345–390.

Friedenberg, J. (2012). Aesthetic judgment of triangular shape: compactness and not the golden ratio determines perceived attractiveness. *i-Perception*, 3(3):163–175.

Graham, D. J., Friedenberg, J. D., Rockmore, D. N., and Field, D. J. (2010). Mapping the similarity space of paintings: Image statistics and visual perception. *Visual cognition*, 18(4):559–573.

Kohler, P. J., Clarke, A., Yakovleva, A., Liu, Y., and Norcia, A. M. (2016). Representation of maximally regular textures in human visual cortex. *The Journal of Neuroscience*, 36(3):714–729.

Kohler, P. J. and Clarke, A. D. F. (2021). The human visual system preserves the hierarchy of two-dimensional pattern regularity. *Proceedings of the Royal Society B: Biological Sciences*, 288(1955):20211142. Publisher: Royal Society.

Kohler, P. J., Cottereau, B. R., and Norcia, A. M. (2018). Dynamics of perceptual decisions about symmetry in visual cortex. *NeuroImage*, 167(Supplement C):316–330.

Laine-Hernandez, M. and Westman, S. (2008). Multifaceted image similarity criteria as revealed by sorting tasks. *Proceedings of the American Society for Information Science and Technology*, 45(1):1–14.

Landwehr, K. (2009). Camouflaged symmetry. *Perception*, 38:1712–1720.

Liu, Y., Hel-Or, H., Kaplan, C. S., and Van Gool, L. (2010). Computational symmetry in computer vision and computer graphics. *Foundations and Trends® in Computer Graphics and Vision*, 5(1–2):1–195.

Mach, E. (1959). The Analysis of Sensations (1897). *English transl., Dover, New York*.

Milton, F., Longmore, C. A., and Wills, A. J. (2008). Processes of overall similarity sorting in free classification. *Journal of experimental psychology. Human perception and performance*, 34(3):676–692.

Polya, G. (1924). Xii. Über die analogie der kristallsymmetrie in der ebene. *Zeitschrift für Kristallographie-Crystalline Materials*, 60(1):278–282.

Pothos, E. M., Perlman, A., Bailey, T. M., Kurtz, K., Edwards, D. J., Hines, P., and McDonnell, J. V.

(2011). Measuring category intuitiveness in unconstrained categorization tasks. *Cognition*, 121(1):83–100.

Richards, L. G. (1972). A multidimensional scaling analysis of judged similarity of complex forms from two task situations. *Perception & psychophysics*, 12(2):154–160.

Vedak, S. (2014). The salience of lower-order features in highly self-similar wallpaper groups. *Honors thesis, The Pennsylvania State University*.