Studying Aesthetics With the Method of Production: Effects of Context and Local Symmetry

Gesche Westphal-Fitch, Jinook Oh, and W. Tecumseh Fitch
University of Vienna

We investigated the role of local and global context on visual patterns produced by normal participants, examining the effects of both top-down context (framing) and bottom-up content (element-internal symmetry) in a computer-based experimental framework. In the first study, we allowed participants to generate rectangles of arbitrary proportions and found an effect of framing on width-to-height ratios of rectangles produced, demonstrating the importance of taking visual framing into account when discussing human shape preferences. In a second study, using FlexTiles, an interactive pattern-generation framework, we showed that the patterns humans produce are influenced by local symmetrical properties of pattern elements. Participants also had to indicate preferences between pairs of pattern variants. We found that in some cases, pattern preferences and pattern production lead to different results. We conclude that visual context, either in the form of visual framing or local symmetries, changes aesthetic patterns that humans produce and prefer in predictable ways. These differences between the productive and perceptual preferences highlight the importance of using multiple methods when studying the human aesthetic sense.

Keywords: pattern, production, symmetry, preference, aesthetics
research has been conducted on how people produce such two-dimensional (2D) patterns.

We recently introduced such a method of production in Westphal-Fitch, Huber, Gómez, & Fitch (2012) using a computer interface to provide a relatively complex stimulus and an open-ended production task, with tightly constrained operation(s) available to the participant. Our first study in this vein used a software interface called FlexTiles, in which a matrix of square tile images can be manipulated by clicking them, which rotates the targeted tile by 90°. Participants in this study, presented with an initial random array, were very consistent in their behavior. Although they were instructed simply to click as often and wherever they wanted, and then to indicate when they were finished, virtually all subjects clicked in such a way as to produce highly ordered and structured stimuli. They thus reliably manifested Gombrich’s sense of order in the laboratory. Furthermore, one particular pattern, typical of real-world tilings, was often produced first, and was produced at least once by all subjects. Despite this consistency, participants were not uncreative: When given the opportunity to produce a second or third pattern with the same tile array, they were creative and often generated quite different and more unusual arrangements. Thus, this novel approach to empirical aesthetics combines a desirable stimulus complexity and freedom of expression with constraints, leading to replicable results amenable to statistical analysis.

The studies reported here build on these initial findings in two ways, with the general goal of understanding the various factors that influence the patterns produced by normal participants. First we use a novel computer-graphics framework to investigate the role of global context on the production of a simple single-element pattern: a rectangle whose height and width could be freely varied. This experiment revisits the very old and controversial issue of the aesthetic significance of the golden section by allowing participants to generate rectangles by dragging a computer mouse. We analyze how visual context (specifically, the frame within which the rectangle appears) affects participants’ productions.

In the second study, we take a bottom-up approach, examining the role of element-internal symmetry on global pattern production in the FlexTiles interface discussed above. Here, we explore the specific effect of tile-internal symmetry (diagonal, orthogonal, or none) on the patterns produced and preferred. Both of these studies clearly illustrate the value of production-based approaches to a richer understanding of basic aesthetic proclivities.

Experiment 1: Rectangle Production

In our first experiment, we used a simple computer interface to allow participants to draw rectangles of a size and proportion of their liking. The overall issue we hoped to explore was the preference (or lack thereof) for simple ratios, and particularly for the golden section ($\phi = 0.618$ or $1.618$), which has attracted great attention in aesthetics since the ancient Greeks proposed it as an “ideal” and pleasing proportion. This “golden section hypothesis” has accumulated what is probably the largest literature in empirical aesthetics, dating from the seminal works by Fechner (1871, 1876), which report a strong preference for “golden rectangles” with width/height ratios around $\phi$. Since then, interest in this topic has waxed and waned, but the overall literature is full of contradictory findings, with some studies reporting strong preferences for golden rectangles (or other shapes), some finding weak preferences, and some finding none at all (for an authoritative review see Green, 1995). Although most of these studies were based on choice and/or preference ratings, several studies also used the method of production by having participants draw a rectangle on paper (e.g., Davis, 1933). Based on this prior literature, we developed hypotheses that might help explain this lack of consistency.

First and foremost, we hypothesized that the context in which the rectangles are presented and/or drawn may play a major role in determining the experimental outcome. For example, in drawing studies we may expect the proportions of the paper to influence the rectangle drawn, whereas for computer-based presentations, the screen dimensions might be important. When actual cardboard rectangles are presented, we may expect the proportions of the table upon which they are laid to play a role. Such contextual details are often omitted from the published reports, making this hypothesis difficult to evaluate in most of the literature. To test it, we asked our participants to generate rectangles in four different contexts: no frame, square frame, and either horizontal or vertical rectangular frames. Interestingly, a similar experiment was already proposed by Fechner in 1865, but to our knowledge, the effect of framing rectangles within larger rectangles has never been tested empirically.

The other issue we hoped to examine was the influence of culture, in a broad sense, on aesthetic preferences. Berlyne (1971) found some rather subtle differences in rectangular preferences between Japanese and Canadian schoolgirls, and in particular, found that the Japanese group tended to prefer squarish rectangles (and that both groups frequently chose squares as their best-liked shape). Canadians were about twice as likely as Japanese to initially choose the golden rectangle. Furthermore, there is some empirical evidence that East Asians tend to have a more holistic visual perception, for example paying more attention to background information in a scene than do Westerners (Ji, Peng, & Nisbett, 2000; Nisbett, 2003; Nisbett & Miyamoto, 2005). To gain first insights into the importance of this issue, our participants were roughly half Korean and half Western. There are of course many other possible contributing factors to context (see Discussion), some of which are easier to exclude or vary than others, but these two provide a start to evaluating the strength of these various contextual effects.

Method

Participants. We recruited 12 Korean (6 female, mean age 27.7 years, range: 25–31 years) and 11 Western participants (6 female, mean age 31.8 years, range: 24–44 years). The Western participants originated from Austria, Germany, Canada, Russia, Croatia and Italy and were recruited at the University of Vienna. On average, the Korean participants had spent 2.9 years in Austria (range: 0.1–7 years). The data of one Western participant had to be excluded due to a software problem. The Korean participants were recruited through informal connections to the Korean community in Vienna. A Korean native speaker was present before and after the experiment to provide clarifications for the Korean participants.

All participants gave their written informed consent prior to taking part. They received chocolate for participating. The Korean
participants also received €5 as compensation for traveling to the University of Vienna especially for the experiment.

**Materials and procedure.** This experiment had four conditions: (a) no frame, (b) square frame, (c) horizontal frame (1.6 width/height ratio), and (d) vertical frame (0.6 width/height ratio). The images were projected on a blank white wall (4.84 × 2.7 m) in an empty room. In the no-frame condition, the participant saw only a white starting point (5 × 5 pixels) on a black background, whereas in the other conditions, a gray frame (2 pixels in width, red, green, blue (RGB) values: 0.5, 0.5, 0.5) was shown on the center of the projector screen, within which the white starting point was centered. On an LCD projector, colors are projected with additive light (e.g., white is a mix of red, green and blue light). Black is shown by an absence of light, that is, black, when projected, blends into a white background with no visible boundary, in particular when the ambient light in the room is bright. Hence, the ceiling lights remained switched on during the experiment. The area of the shape that the participants could change was constrained by a visible frame in the case of the framed conditions, and by an invisible square boundary in the unframed condition. The available area for the rectangles was the same for the framed conditions, as closely as could be achieved using whole pixel units (see Table 1 below for an overview of the frame conditions). Each condition comprised 10 trials presented in a block, and the conditions were presented in random order. The unframed condition was shown twice in the course of the experiment, and all others were shown once, leading to a total of 50 trials per participant.

Participants were seated directly behind the projector, approximately 374 cm away from the wall. They used a wireless mouse to change the dimensions of the shape, which always remained centered on the screen. By pressing and holding the left mouse button, the dimensions of the rectangle (width and height) could be changed. For every pixel that the mouse moved, two pixels were added to the shape in the direction that the mouse moved. For example, if the mouse moved one pixel vertically, 2 pixels were added to the height of the shape, and the center of the shape was recalculated on the fly by the software, and aligned with the center of the screen. Consequently, regardless of how its dimensions were changed, the shape remained centered on the screen for the participants. To finish, the participants clicked on the right mouse button, which ended the trial, and they had to click an “OK” button on a pop-up window to continue to the next trial. The experimenter was present in the room and sat behind an opaque barrier so that neither the participant nor the projected image was visible to her. Western participants pooled.

This experiment had four conditions: (a) no frame, (b) square frame, (c) horizontal frame (1.6 width/height ratio), and (d) vertical frame (0.6 width/height ratio). The images were projected on a blank white wall (4.84 × 2.7 m) in an empty room. In the no-frame condition, the participant saw only a white starting point (5 × 5 pixels) on a black background, whereas in the other conditions, a gray frame (2 pixels in width, red, green, blue (RGB) values: 0.5, 0.5, 0.5) was shown on the center of the projector screen, within which the white starting point was centered. On an LCD projector, colors are projected with additive light (e.g., white is a mix of red, green and blue light). Black is shown by an absence of light, that is, black, when projected, blends into a white background with no visible boundary, in particular when the ambient light in the room is bright. Hence, the ceiling lights remained switched on during the experiment. The area of the shape that the participants could change was constrained by a visible frame in the case of the framed conditions, and by an invisible square boundary in the unframed condition. The available area for the rectangles was the same for the framed conditions, as closely as could be achieved using whole pixel units (see Table 1 below for an overview of the frame conditions). Each condition comprised 10 trials presented in a block, and the conditions were presented in random order. The unframed condition was shown twice in the course of the experiment, and all others were shown once, leading to a total of 50 trials per participant.

Participants were seated directly behind the projector, approximately 374 cm away from the wall. They used a wireless mouse to change the dimensions of the shape, which always remained centered on the screen. By pressing and holding the left mouse button, the dimensions of the rectangle (width and height) could be changed. For every pixel that the mouse moved, two pixels were added to the shape in the direction that the mouse moved. For example, if the mouse moved one pixel vertically, 2 pixels were added to the height of the shape, and the center of the shape was recalculated on the fly by the software, and aligned with the center of the screen. Consequently, regardless of how its dimensions were changed, the shape remained centered on the screen for the participants. To finish, the participants clicked on the right mouse button, which ended the trial, and they had to click an “OK” button on a pop-up window to continue to the next trial. The experimenter was present in the room and sat behind an opaque barrier so that neither the participant nor the projected image was visible to her.

The experiment was run on an Apple MacBookPro (Cupertino, CA) using custom Python software (Beaverton, OR; Version 2.6.4; www.python.org). The images were projected with a NEC NP-M350X projector (Tokyo, Japan) placed on a table, 293 cm from the wall, resulting in an overall screen size of 222.5 cm (width) × 166 cm (height), which, however was not visible to the participants. We used SPSS (Versions 17 and 19, Armonk, NY) and R (Version 2.15.1, Vienna, Austria, r-project.org) for statistical tests and graphs.

### Results

To ensure that participants had made appreciable changes in both dimensions, we excluded those trials in which the height and/or width of the 5 × 5 pixel starting point had been changed by less than 5 pixels. Furthermore, we excluded those trials in which both parameters were at their maximum value (i.e., the participants had filled the available space completely) because this simply resulted in rectangle proportions identical to the frame, inevitably confirming our framing hypothesis. These two exclusions reduced the number of trials from 1100 to 919. The rectangles can be fully characterized by two values: size (area) and proportions (width/height ratio).

We found no significant differences between the mean area or width/height ratio for the two unframed conditions, area: paired-samples t test, *t*(21) = −.48, *p* = .638; width/height ratio: paired samples *t* test, *t*(21) = .78, *p* = .48, so we pooled the no-frame data.

We used an alpha level of .05 for all statistical tests, and applied a Bonferroni correction when multiple tests were conducted. As the histograms of the width/height ratios (Figure 1A) show, the distributions are spread widely. There were a nontrivial number of outliers with large ratios (horizontal stripes). Rather than excluding outliers based on an arbitrary cutoff point, we log-transformed the data, which reduced the influence of these outliers considerably (Figure 1B). Results are reported for both raw and log-transformed width/height ratios. In the histograms of the raw ratios, there are peaks around 1.5 and 1 width/height ratio for the horizontal frame; in contrast, there is a single clear peak around 0.6 for the vertical frame, demonstrating an effect of framing. To evaluate this rigorously, we conducted repeated-measures ANOVAs on the mean values of three measures: area, area fraction (proportion of the area of the produced shape relative to available area) and width/height ratio (see Table 2 for an overview). One Korean participant was excluded from the repeated-measures analysis because in one condition, he only produced stripes of less than 10 pixels width or height. By our exclusion criteria, these all had to be excluded, leaving no valid trials. For all ANOVAs performed, we found no significant main effect or interaction of ethnicity. Hence, except where explicitly noted, we report the results with Korean and Western participants pooled.

### Table 1

**Overview of the Four Frame Conditions in Experiment 1**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Frame dimensions (Width × Height, in pixels)</th>
<th>Frame dimensions (Width × Height, in cm)</th>
<th>Frame area (in cm²)</th>
<th>Visual angle (width of stimulus)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No frame</td>
<td>800 × 800*</td>
<td>173 × 173*</td>
<td>29,929*</td>
<td>19.92*</td>
</tr>
<tr>
<td>Square</td>
<td>474 × 474</td>
<td>102 × 102</td>
<td>10,404</td>
<td>12.26</td>
</tr>
<tr>
<td>Horizontal</td>
<td>600 × 375</td>
<td>130 × 80</td>
<td>10,400</td>
<td>15.63</td>
</tr>
<tr>
<td>Vertical</td>
<td>375 × 600</td>
<td>80 × 130</td>
<td>10,400</td>
<td>26.67</td>
</tr>
</tbody>
</table>

* In the “no frame” condition, the specifications refer to the largest possible shape that could be produced.
We conducted a one-way repeated-measures ANOVA analysis with the frame conditions as a within-subjects factor. We found that there were significant differences in the areas of the shapes produced, \( F(1.23, 24.62) = 25.7, p < .001 \), Greenhouse–Geisser correction for spherical data; the area of shapes for the unframed condition was significantly larger than the other three conditions, paired-samples \( t \) test, \( t(20) = 5.08–5.67, p < .001 \), Bonferroni-corrected significance threshold = .0083), and the

**A: Width/Height Ratio (untransformed, outliers not illustrated)**

![Histogram of width/height ratios (untransformed, outliers not illustrated)](image)

**B: Width/Height Ratio (log transformed, outliers shown)**

![Histogram of width/height ratios (log transformed, outliers shown)](image)

**Figure 1.** Histograms of width/height ratios (all participants included). A: Distribution of untransformed ratios. Because there were many large outliers, only ratios between 0 and 3.0 are shown; however, the means shown are for all data. B: Distribution of natural log-transformed ratios (all data are shown).
Table 2
Overview of the Mean Area, Mean Area Fraction, Width/Height Ratio of Shapes Produced in Experiment 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean area (in total pixels)</th>
<th>Mean area fraction (in %)</th>
<th>Mean width/height ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>No frame</td>
<td>175,383</td>
<td>27.4</td>
<td>1.68</td>
</tr>
<tr>
<td>Square</td>
<td>54,044</td>
<td>24.1</td>
<td>1.38</td>
</tr>
<tr>
<td>Horizontal</td>
<td>64,013</td>
<td>28.5</td>
<td>2.41</td>
</tr>
<tr>
<td>Vertical</td>
<td>57,164</td>
<td>25.4</td>
<td>1.52</td>
</tr>
</tbody>
</table>

three framed conditions did not differ significantly from each other in area, paired sample $t$ tests, $t(20) = -1.24–1.42$, $p > .17$. Bonferroni-corrected significance threshold = .0083.

Area fraction. However, this size difference between the unframed condition and the framed conditions might be due to the fact that the overall size available was constrained to a smaller area in the framed conditions. To evaluate this, we calculated the fraction of the total available area that the produced shape occupied and conducted a one-way repeated-measures ANOVA analysis on this, with frame type as a within-subjects factor. We found that there were no significant differences, $F(3, 60) = 0.83$, $p = .481$ in the area fractions between conditions, confirming our guess that larger shapes in the unframed condition are due to the fact that participants had a larger area to work with.

Width/height ratio. A repeated-measures ANOVA on untransformed ratios, with frame type as a within-subjects factor, showed that there were significant differences in the width/height ratios of shapes produced in the different conditions, $F(2.32, 46.31) = 5.13$, $p = .007$, Greenhouse–Geisser correction. Post hoc tests revealed a significant difference between the horizontal and square conditions: paired $t$ test, $t(20) = -2.99$, $p = .007$. Though the difference between horizontal and vertical conditions approached significance, it was not below the Bonferroni-corrected threshold: paired-samples $t$ test, $t(20) = 2.8$, $p = .011$. All other pairings did not differ significantly (all $p > .0083$). Thus, there was a clear effect of frame condition, even with the untransformed data.

Based on the histograms, it seemed likely that a high number of outliers were obscuring differences between the conditions. We thus ran another repeated-measures ANOVA using logtransformed width/height ratios as the dependent variable. We again found a significant effect of frame condition: $F(2.23, 42.35) = 12.54$, $p < .001$, Greenhouse–Geisser correction. However, now post hoc tests showed that four of the six possible comparisons of the conditions differed significantly from each other, $t(20) = -3.98–4.6$, $p < .004$, with the exception of square versus vertical: paired-samples $t$ test, $t(20) = 1.73$, $p = .1$ and square versus unframed: paired-samples $t$ test, $t(20) = 2.1$, $p = .049$, not significant after Bonferroni correction. All other comparisons of framing contexts showed significant changes in the proportions of the rectangles produced.

Shape categories. We classified the shapes as either horizontal (width > height), vertical (height > width) or square (width = height). Given the difficulty of matching height and width exactly by eye, we allowed up to 5% discrepancy between width and height in the square category. Horizontal shapes were the most frequent shape type for all conditions except the vertical frame condition, in which vertical shapes were most common (see Table 3).

The distribution of shape types for the various frame conditions deviated significantly from random: Pearson $\chi^2(6, N = 399) = 68.63$, $p < .001$. Our results also show that Koreans produced slightly higher numbers of squares than Westerners (69 vs. 39), providing some support for Berlyne’s ethnicity findings. However, though the number of produced squares did depend on frame type—repeated-measures ANOVA with Greenhouse–Geisser correction, frame condition as within-subjects factor, ethnicity as between-subjects factor, $F(1.7, 32.24) = 10.07$, $p < .001$—there was no significant effect of ethnicity across conditions, $F(1.7, 32.24) = 2.34$, $p = .12$. That is, the same basic pattern is true for both cultural groups: Horizontal shapes are produced most frequently, with the exception of vertical frames, in which vertical shapes are most frequent, as Figure 2 shows.

Discussion

This study revealed a strong and consistent effect of framing on the rectangles that our participants produced. First and most obvious, significantly larger shapes were produced in the no–frame condition, but this seems to result simply from the interface allowing the possibility of larger rectangles in this condition. When looking at the area of the shape compared with the overall available area for the different conditions, we found that on average, between 24.1 and 28.5% of the area was filled (see Table 2). We interpret this as a somewhat trivial type of context dependence: The area of produced shapes is dependent on the area available.

More intriguing are the pronounced effects of frame context on proportions. First, we found an effect of framing between square and horizontal contexts using the raw ratios. Second, when logtransformed to diminish the effect of numerical outliers, the framing effect was even stronger, with four out of six condition pairs differing significantly from each other. One exception makes sense: The square versus unframed conditions did not differ significantly. Since the unframed condition was invisibly constrained by an 800 × 800 square, it is unsurprising that it did not differ from the square condition. The reason for the second exception, square versus vertical frame, is less obvious, but it may result from the multipeaked nature of vertical distribution, which creates difficulties for standard statistical techniques such as ANOVA or generalized linear models.

This robust effect of framing may help to explain the many inconsistent findings in the previous literature, since some implied frame always exists (even if only the viewer’s own visual field), but the specific proportions of the framing context are rarely specified in previous work. In contrast, we found no significant effect of the cultural background of the participants (Korean vs.

Table 3
Influence of Frame on Proportions of Shapes: Percentages (Rounded to the Next Full %) of Vertical, Horizontal, and Square Shapes Produced in the Four Different Shape Conditions

<table>
<thead>
<tr>
<th>Condition</th>
<th>% Horizontal</th>
<th>% Vertical</th>
<th>% Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>No frame</td>
<td>54</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>Square</td>
<td>48</td>
<td>34</td>
<td>18</td>
</tr>
<tr>
<td>Horizontal</td>
<td>69</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>Vertical</td>
<td>37</td>
<td>57</td>
<td>5</td>
</tr>
</tbody>
</table>
25% of his participants exhibited a negative rectangular preference with peaks roughly at two consistent underlying factors: a preference for squares, and a preference rather than a mathematical artifact. Based on a factor analysis of individual preference distributions, McManus found dual preference peaks at the 1:1 (square) ratio and a broader peak near the golden rectangle. Green (2012) has recently shown that McManus & Wu (2012) tested rectangle preferences in Western and Chinese participants and found no consistent preference pattern between the two cultural groups. Green (1995) suggested that much previous research had unfairly pitted 1:1 against the golden ratio, as if a preference for squares would nullify Fechner’s hypothesis. Instead, Green suggested that both gestalt approaches (e.g., Arnheim, 1954) or the broader Pythagorean approach to empirical aesthetics would incorporate the possibility of multiple preferences. Furthermore, he cited the possibility originally proposed by Davis (1933) that the frequent finding of a broad peak in the rough vicinity of 1.6 actually represents the summation of several independent peaks at other nearby complex ratios, such as \(\sqrt{2} (= 1.414\ldots)\) and \(\sqrt{3} (= 1.73\ldots)\) as well, possibly, \(\phi\) itself.

After his painstaking analysis of all possible pairings of stimuli from a pregenerated stimulus set, McManus (1980) concluded that, despite “moderately good evidence” for Fechner’s golden rectangle hypothesis, “techniques at present [are not] accurate enough” (p. 523) to discriminate the golden section from other nearby ratios (such as 1.5 or 1.75). Our results show that the method of production presented here provides a potential solution to this problem, allowing any normal participant to easily generate a large number of rectangles (or any other simple figure) quickly and with pixel accuracy. With a larger number of trials than employed here, we can, in principle, discriminate between these hypothesized preferred ratios statistically. The key point from our results is that future work must control for, or experimentally evaluate, the effect of framing context for these accurate values to be meaningful.

Given the strong effect of rectangular framing found here, the most obvious next step in this line of research will be to evaluate the extent to which the specific ratio of the framing rectangle influences the rectangle produced. We used rectangles of 0.6/1.6, which is the proportion of many modern computer and TV screens and very close to \(\phi\), but providing a series of frame proportions around this ratio (e.g., \(\sqrt{2}, 1.5, \phi,\) and \(\sqrt{3}\)) would allow us to evaluate the specificity (or lack thereof) of this effect. Given the powerful effect of framing, it would also be valuable to follow up the suggestion of Hintz and Nelson (1970) that individual differences in rectangle preferences may result from differences in the viewer’s own visual field dimensions. An obvious example of this effect (again apparently ignored in previous work) would be to analyze the effect of wearing eyeglasses, and of eyeglass shape, on rectangle productions and preferences.

In summary, our results are consistent with much previous research in finding preference peaks around 1:1 and \(\phi\), but also provide some evidence for other peaks, for example at \(\sqrt{2}\). However, we discovered a powerful effect of framing, the possibility of which has gone essentially unconsidered in the previous literature. Given the much higher likelihood that a subject will produce a golden rectangle in a rectangular context, this framing effect may go a long way toward explaining the inconsistency in the prior literature. Thus, these results show the importance of controlling this aspect of context (and certainly the necessity of reporting it) in future work.
Experiment 2: Pattern Generation With Tile Matrices

In this experiment, we approach the question of context from a different angle. Whereas Experiment 1 showed that the properties of a frame influence the proportions of rectangles produced, here we investigate the bottom-up effect of local structure, specifically symmetry within pattern elements, on global patterns that participants produce. Experiment 2 involves a 6 × 6 matrix composed of square-patterned tile elements. When observing tile patterns in real life (see Figure 3), it is striking that (a) most tiles have some degree of internal symmetry, along one or more axes and that (b) tiles with different symmetries are used for different patterns. An informal survey of the usage of diagonal and orthogonal tiles (e.g., Pepin Press tile-pattern compendium, Hernández Navarro, 2006) shows that orthogonal axis tiles are used mainly on borders in a repetitive translational fashion (i.e., the motive is repeated along a horizontal or vertical axis without rotation or reflection). Diagonal axis tiles are almost never used in a translational fashion, but are instead usually arranged in groups of 2 × 2 tiles, with each tile rotated, leading to a larger diamond or circle figure.

Based on these informal observations, we hypothesized that humans are sensitive to the symmetry of local pattern elements, and adjust both their pattern-making strategies and their preferences for certain tile/pattern combinations accordingly, without explicit instructions to do so. Thus, local tile properties, that is, the presence or absence of symmetry, as well as the orientation of any symmetry, will influence the preferences for patterns, but will also bias the production of patterns in favor of certain constellations.

More specifically, we predicted that tiles with orthogonal symmetry would lead to more translational patterns, mirroring real-life usage. Tiles with diagonal symmetry will lead to patterns in which the tiles are arranged in groups of 2 × 2 rotated tiles. Concerning choice, we expect congruent combinations (i.e., grouped rotation in combination with a diagonal tile, translational with a horizontal tile) to be preferred over incongruent combinations (grouped rotation with a horizontal tile, translational pattern with a diagonal tile).

Method

Participants. Nine participants (five women, mean age: 27.7 years, range: 23–37 years) took part in the experiment. All had normal color vision and normal or adjusted-to-normal visual acuity. Participants gave their written informed consent prior to participating and were paid for taking part.

Stimuli. For both the production and preference tasks, we used digital images of tiles from Barcelona Tile Designs and Havana Tile Designs (Hernández Navarro, 2006; Hernández Navarro, 2007). We chose images that were available in two variants within a pattern: diagonal symmetry (mirror symmetry along one diagonal axis), and orthogonal symmetry (i.e., mirror symmetry along either the vertical or horizontal axis). The orthogonal symmetry is typically used in tile borders, and tiles with diagonal symmetry are typically arranged in the main tile pattern. We specifically chose tiles that were available in both variants to ensure that the basic color scheme, complexity, and style, were very similar. We created a third, nonsymmetrical tile type from the two tile variants by dividing both tiles along the diagonals and using two of the four resulting triangles from both images to create a tile that had 50% of its pixels from the diagonal and orthogonal tile respectively, but crucially did not have perfect symmetry along either the horizontal or diagonal symmetry axis. The tile manipulation was done in Adobe Illustrator (14.0.0). We considered these three variants together to be a “tile family.”

Procedure. The experiment consisted of two tasks: a pattern production task and a two-alternative forced choice (2AFC) task. In the production part, participants were presented with a 6 × 6 grid of identical tiles (100 × 100 pixels). The tiles could have one of four possible orientations (0, 90, 180 or 270°), which were initially random. To change the orientation, the participant clicked on an individual tile, upon which the tile rotated incrementally by 90° clockwise. The initial orientation of the tiles in the array was randomized in each trial and differently for each participant. There was no time limit for the production trials, and the participants received no specific task instructions, other than to change as little

![Symmetry along orthogonal axis:](image)

![Symmetry along diagonal axis:](image)

*Figure 3.* Photograph of a floor tiling in a 12th-century building in Florence, Italy, 2011 by Gesche Westphal-Fitch. The tiles are shown as line drawings on the right to illustrate the internal symmetry types, orthogonal (top) and diagonal (bottom). Tiles with orthogonal symmetry are used on the border, whereas tiles with a diagonal symmetry are mainly arranged to create 2 × 2 diamond figures.
or as much as they liked in the initial random pattern, and click on a “Finish” button under the tile array when they were ready to move on to the next trial (Westphal-Fitch et al., 2012). In total, there were 18 production trials, which were divided into blocks of six.

After each block of production trials, a 2AFC preference session was interspersed. Here, two patterns were presented side by side and participants had to indicate their preference for one of the two images by clicking on the preferred one. We compared three patterns: translational symmetry, grouped symmetry (for examples see Figure 5) and random orientations. We did a complete pairwise comparison between the three patterns and three tile types for six tile families, yielding 216 comparisons per participant (see description of tile families below and Figure 4). We also included a distractor task of comparing the same pattern made of the same tile type from different tile families; however these preferences are not part of the present analysis. In total, there were 351 comparison trials, broken down into three blocks with 117 trials. Participants had 3 sec to make their choices, and were encouraged in the instructions to make their choices as quickly as possible. If a trial timed out, no preference was recorded, and the trial was not repeated. Participants had an opportunity to take a short break between each block. For the production task, we recorded the initial and final orientation of each tile, as well as the clicks with time information for each tile.

Data was analyzed with SPSS (Versions 17 and 19), as well as custom R scripts (Version 2.15.1) using the package prefmod (Hatzinger & Dittrich, 2012). In the preference task, the image pair and the chosen image were recorded. We used Circos (http://circus.ca, Krzywinski et al., 2009) for the circular graph.

Images were presented on a 17-in. touch screen (ELO Intellitouch, Menlo Park, CA) using custom Python software running on an Apple MiniMac. The participants indicated choices and changed the patterns by touching the screen with their fingertip. The experimenter was present in the room but was visually separated from the participant by an opaque barrier. Instructions were given in writing and did not contain words alluding to patterns, beauty, or symmetry.

Results

Analysis of production results. We analyzed how frequently grouped or translational patterns were produced, and for which tile type these patterns were most common. In addition to the basic grouped pattern, patterns were categorized as grouped if the 2×2 figure was inverted or offset. A translational pattern could be horizontal or vertical, and the stripes could be inverted by 180°. We allowed maximally one error to occur in a pattern (but this happened in only one instance).

Of the 162 production trials, 10 patterns were excluded because no element in the original pattern array had been changed. Of the remaining patterns, 58 were grouped or translational patterns (26 grouped, 32 translational), constituting 38% of all patterns. The other patterns produced were also often highly ordered and symmetrical in more complex ways, but their analysis is beyond the scope of this paper (see Westphal-Fitch et al., 2012). We found a clear effect of tile type on the patterns produced (see Figure 6 below): Of the 26 grouped patterns, 21 were made with diagonal tiles (80.8%) and only five (19.2%) were made from nonsymmetrical tiles. None were made with orthogonal tiles. In the case of the translational patterns, only three (9.38%) were made of diagonal tiles; 16 (50%) were made with orthogonal tiles and 13 with nonsymmetrical tiles (40.6%).

The distribution of these two patterns by tile type differed significantly from values expected if there had been no association, Pearson χ²(2, N = 58) = 32.77, p < .001. The 58 arrays of interest were produced over all three production blocks (first block: 36.2%, second block: 29.3%, third block: 34.48%). Grouped patterns were produced with on average 81 clicks in 69.4 seconds, while the translational patterns were produced with on average 89.6 clicks in 105.3 seconds. Based on number of clicks, the two patterns seem to be roughly equivalent in effort, although the translational patterns did take longer to produce. All but one participant produced both translational and grouped patterns.

Analysis of paired comparisons. Of the 3,159 (9×351) choice trials run, 25 timed out (0.79%), leaving 3,134 valid choices, 1,207 in the distractor task and 1,927 in the real task. For all but two participants, the diagonal grouped pattern was clearly the most preferred pattern. One participant had a strong preference for grouped patterns with nonsymmetrical tiles and another had a weak preference for translational patterns with nonsymmetrical tiles (see Figure 7).

Choices were not distributed evenly across pattern and tile types: Random patterns were only chosen 22.8% of the time (regardless of tile type); patterns with a grouped pattern were chosen 42.7% of the time. The grouped rotations were most popular with diagonal (360 choices, 18.7% of all trials) and non-
symmetrical tiles (276 choices, 14.3%), and less so with horizontal tiles (187 choices, 9.7%). A full overview of choice distributions is given in Table 4.

A chi-squared test confirmed that the differences in number of choices between the tile types and pattern types were highly significant, Pearson $\chi^2(4, N = 1927) = 22.29$, $p < .001$. A comparison with the production data is particularly interesting: The preference for grouped patterns (rather than translational patterns) with diagonal tiles is similar in both production and preference tasks (see Figure 6). However, for orthogonal tiles, the preference for the translational pattern over the grouped pattern is markedly weaker in the preference task than in the production task. Graphing all possible choices and the number of choices favoring the diagonal grouped pattern confirms an overwhelming preference of this tile/pattern combination. Figure 7 represents the choices visually: Each band connects an image pair. The width of the band ends represent how often that image was chosen over the other—that is, the wider the band is, the more frequently that image was chosen over its partner. Moving around the circle from the top in a clockwise fashion, the image categories are ordered by descending popularity: Diagonal grouped patterns were chosen more often than any other image type, and nonsymmetrical grouped patterns were chosen more often than all other remaining categories. The choices are also summarized in Table 4. Of the regular patterns, the patterns exhibiting local incongruity between tiles (i.e., horizontal grouped and diagonal translational) are least popular.

We ranked the nine pattern comparisons based on the participants’ decisions and fitted a log-linear Bradley Terry model (a form of generalized linear model) to the data, assuming an underlying Poisson distribution. We found that the model that fitted best on both the group and individual levels, as measured by Akaike’s information criterion (AIC), included all 9 stimulus categories ($df = 28$, residual deviance: 26.82, AIC: 475.45). We determined

![Figure 5](image1.png)

**Figure 5.** Examples of the grouped and translational patterns with simplified tiles for clarity (tiles used in the experiment were colored and structurally more complex). 1–3: grouped pattern with diagonal, orthogonal and nonsymmetrical tiles. 4–6: translational pattern with diagonal, orthogonal and nonsymmetrical tiles. We also created random patterns with the tiles, which are not shown here.

![Figure 6](image2.png)

**Figure 6.** Perception/production differences. Left: The number of translational (black bars) and grouped-rotation (white bars) patterns made with each of the three tile types: diagonal, horizontal, and nonsymmetrical. Right: The number of patterns chosen in the preference task for each of the tile types (choices of random patterns are not shown here).
the fit to the data by conducting a chi squared test of the residual deviance and degrees of freedom and subtracting that value from 1. A nonsignificant value is taken to be an indicator of a good fit of the model to the data. This was the case here, suggesting that the fit is good ($1-p = .53$, residual deviance: 26.82, $df = 28$). A further reduction in stimulus categories, resulting in a simpler decision model, led to far higher residual deviances and rises in AIC values of over 100 (all $1-p < .001$), suggesting that such a reduction in stimulus categories is not advisable (Burnham & Anderson, 1998).

From the coefficients of the models we calculated the worth values for both the group as a whole and individual participants for each stimulus category. The worth value corresponds to probability of being the chosen pattern in a particular pair. That is, when a pair of images is given, the one with the higher worth will be chosen, according to the following equation:

$$d_g \cdot n_g \cdot h_t \cdot n_t \cdot h_g \cdot d_t \cdot d_r \cdot n_r$$

$\text{d}_g$: diagonal grouped  $\text{d}_t$: diagonal translational  $\text{d}_r$: diagonal random

$\text{h}_g$: horizontal grouped  $\text{h}_t$: horizontal translational  $\text{h}_r$: horizontal random

$\text{n}_g$: non symmetrical grouped  $\text{n}_t$: non symmetrical translational  $\text{n}_r$: nonsymmetrical random

**Figure 7.** Overview of preferences in the paired-comparisons task: All possible pairs are presented here, connected by a band that varies in width, representing the number of choices. The first letter of the category corresponds to the tile type, the second to the pattern. The pattern examples are exemplified using simplified tiles.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Diagonal tile (%)</th>
<th>Horizontal tile (%)</th>
<th>Nonsymmetrical tile (%)</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>230 (11.9)</td>
<td>219 (11.4)</td>
<td>215 (11.2)</td>
<td>664 (34.5)</td>
</tr>
<tr>
<td>Grouped</td>
<td>360 (18.7)</td>
<td>187 (9.7)</td>
<td>276 (14.3)</td>
<td>823 (42.7)</td>
</tr>
<tr>
<td>Random</td>
<td>173 (9)</td>
<td>117 (6)</td>
<td>150 (7.8)</td>
<td>440 (22.8)</td>
</tr>
<tr>
<td>Total</td>
<td>763 (39.6)</td>
<td>523 (27.1)</td>
<td>641 (33.3)</td>
<td>1927 (100)</td>
</tr>
</tbody>
</table>
As worth is expressed as a probability, the sum of all categories adds up to 1. As Figure 8 shows, seven out of nine participants had the highest values for diagonal–grouped. Participant 2 had a strong preference for nonsymmetrical–grouped, and Participant 1 did not show very consistent preferences, with a slight bias toward nonsymmetrical–translational.

Discussion

We found compelling evidence for a strong relationship between local tile symmetry and global pattern in both the production and preference tasks. Despite the overwhelmingly large number of possible tile constellation (there are \(4^{36} = 5 \times 10^{21}\) possible tile constellations in a \(6 \times 6\) grid), over a third of all patterns produced fell into two simple pattern categories: translational and grouped. However, the occurrence of the patterns was not distributed evenly between the tile types. Rather, the grouped pattern occurred most often with diagonal tiles, and the translational pattern mainly occurred with horizontal tiles. That is, to a significant extent, creative output can be predicted based on formally describable symmetries within local elements in this constrained production environment.

Concerning the preference task, we found a strong preference for grouped patterns with diagonal tiles, matching the production result. However, the preference for translational patterns with horizontal tiles was far weaker in the perceptual choice task than in the production task (see Figure 6). This may be because the participants were fatigued by the large number of choices to be made. We cannot rule out that this may have been a contributing factor. However, if that were the only underlying factor, then we would expect preferences to be weaker overall; yet the preference for grouped over translational patterns for diagonal tiles remains strong throughout. We suggest that the production task biases participants to rely on local cues. It is well known that global properties of compound visual stimuli typically take precedence over local features in normal humans (Navon, 1977). Given the short available time that our participants had to make their perceptual decisions, it is highly likely that global features played a larger role than local features in the perceptual choice task. In the case of the production trials however, participants had no time limit, but were constrained to modify local elements by rotating single tiles in a piecemeal fashion. Thus, it is not surprising that the local compatibility of neighboring tiles has a stronger effect during production, resulting in more translational patterns with horizontal tiles and grouped patterns with diagonal tiles. The fact that nonsymmetrical grouped patterns were the second most popular category supports this idea: The local features of the tile were ambiguous (hybrids between diagonal and orthogonal tiles), and might have been perceived as a quasi-diagonal tile, which might have enabled a fast parsing of the pattern as a good grouped pattern. By using both the method of production in combination with more traditional preference methods, we were able to detect this contrast, which would have gone unnoticed if only preference methods were used.

General Discussion

With these results, we hope to have illustrated the effectiveness of the method of production as a powerful technique for exploring aesthetic preferences for a top-down effect of framing, but also for a bottom-up effect of tile-internal symmetries on the patterns produced with them. The disparity between the production and preference tasks just mentioned suggests that the context of active production versus passive perception can have strong effects on preferences. These findings support Fechner’s original argument for using a combination of generative and perceptual tasks. If
it is indeed the case that pattern production lays a stronger emphasis on local relations and rapid preference tasks rely more heavily on global, easily scannable relations, then it would be particularly interesting to test human groups in which the usual “global first” perceptual principle is not as pronounced. For example, individuals with autism have been shown to have a local bias in perception, which in some cases leads to enhanced performance in visual tasks, due to fewer distractions from global information (Brosnan, Scott, Fox, & Pye, 2004; Bölte, Holtmann, Poustka, Scheurich, & Schmidt, 2007; O’Riordan & Plaisted, 2001; Mottron, Belleville, & Menard, 1999). In our pattern-preference task, we would predict that individuals with autism would maintain the preference for orthogonal tiles arranged in the translational pattern rather than the grouped pattern, as shown by our participants in the production task. Further, we would predict that the dispreference for nonsymmetrical tiles in a grouped pattern, relative to diagonal tiles, would be stronger in autists than normals.

We end by briefly considering the cognitive implications, and possible origins, of the most pervasive patterns occurring in our tile-manipulation experiment: translational and grouped. For clarity and concision we will explicitly notate each pattern using a widely accepted crystallographic symbolic notation (cf. Washburn & Crowe, 1988). This system uses a series of four alphanumeric symbols to indicate all possible symmetries that can be generated by rigid motions in the plane (translation, rotation, reflection, and glide). Translational symmetry, in which all tiles are oriented identically, is simply denoted “p1”, where the 1 indicates no rotational symmetry. The 4-tile grouped rotation is denoted “p4” and our participants generated this with roughly equal frequency to p1. These notations only indicate the overall symmetry, for those cases in which the tiles have no internal symmetry. However, if the tiles themselves have symmetry, the overall pattern that results from these operations can have additional mirror symmetries along the vertical and horizontal axes (p4mm = “p4m”). The rotational plus mirror symmetry pattern p4m is of particular interest because in addition to being the typical pattern for laying tiles in the European context, it is the most common underlying pattern in Islamic art (Wichmann, 2008): 48% of 644 Islamic patterns Wichmann examined exhibit this pattern of combining mirror and rotational symmetry.

It is not difficult to understand why our participants often chose to orient all tiles identically, yielding translational symmetry, since this is in some sense the simplest operation to provide any order whatsoever. Thus, if as predicted by Gombrich’s hypothesis of a human sense of order (Gombrich, 1984), participants felt an urge to create some order, this is the least that they could do. The frequent generation of p4m is more interesting, especially given the pervasiveness of this symmetry pattern in several cultures. This pattern occurred by far most frequently (80% of cases) when the tile possessed a diagonal mirror symmetry (45° axis) and when this tile type is arranged into the p4 rotational groups, it yields an overall pattern with both rotational and mirror symmetries (p4m). These multiple symmetries lead to an ambiguity of interpretation: Such patterns could be generated by either a series of mirror reflections (pmm) or a series of rotations (p4). Either way, if the tiles are diagonally symmetrical, the same pattern will result. Our current implementation of the FlexTiles interface allows only rotation, so participants in this study were forced to implement the p4m pattern via rotations. But perhaps the appeal of this arrangement is that it also satisfies a preference for reflectional symmetry (particularly along a vertical axis). Indeed, it is possible that the deeper appeal of these p4m patterns results mainly, or solely, from the resulting reflectional symmetry, and that the rotations are simply a means to this end, of little or no aesthetic significance themselves. Alternatively, it may be the summation of both rotational and reflectional symmetries that gives the p4m pattern its special appeal.

Recognizing this underlying generative ambiguity allows us to restate the general question in more specific, testable terms—why do people like this particular pattern so much? That is, of the various classes of symmetry, which are most perceptually accessible, and which are aesthetically preferable? There are several ways to evaluate these possibilities empirically. The first is by extending our current FlexTiles interface to allow both rotations and reflections (e.g., using right vs. left mouse clicks). If the underlying perceptual bias driving participants to produce p4m is for reflectional symmetry, as seems plausible, we would expect

---

**Four possible symmetry classes**

![Image of four symmetry classes](image)

**Figure 9.** Four symmetry classes in crystallographic notation. A corresponds to translational patterns. Concerning our grouped patterns, horizontal grouped and nonsymmetrical grouped corresponds to B, and diagonal grouped could formally be B, C or D. P4mm is often abbreviated to p4m.
reflectional mouse clicks to dominate. However, if a rotational bias also exists and is important, we would not expect users generating this pattern to be strongly biased toward reflectional operations, and they might instead use a mix of rotations and reflections. Finally, if production tasks lead to a greater reliance on local cues, it may be that rotation is in some sense more predictable than reflection, making success easier to evaluate. This would then predict rotation as the preferred operation.

Further tests would involve subtly manipulating the symmetries of the tiles themselves, so that the two sets of operations yield different outcomes (see examples in Figure 9). Using the method of production, and allowing both reflection and rotation, a strong rotational preference should clearly yield figures like those in 9B, whereas reflectional preference should yield 9C. Purely perceptual preferences can also be tested through the method of choice by pitting patterns like those in examples 9A–9D against one another. Combining all these approaches could yield new insights into the more abstract principles that underlie human pattern preferences, as well as the specific conditions that might bias preferences one way or another.

An interesting issue left open by our current results concerns the degree to which the proclivities documented here are biologically based, and if so, how deep these biological roots might be. This question is best approached empirically through comparisons across species, as well as across different human cultures. Regarding nonhuman animals, we suspect that getting meaningful results from the method of production would be difficult (but it would be interesting to see what patterns would be produced by chimpanzees or other apes working on touch screens). Regarding perception, our previous work (Westphal-Fitch et al., 2012) suggests that even very simple relational patterns on grids are quite difficult for pigeons to perceive, suggesting that the biological roots of our abilities to perceive and create patterns on a matrix are not shared among all vertebrates. With respect to different cultures, we think the method of production offers a powerful and straightforward way to explore the roles of experience and culture in shaping preferences. More cross-cultural studies, like the Korean/European comparison reported here, are clearly necessary to determine whether aesthetic proclivities have pan-human roots. Although the current data on rectangle production suggest that aesthetic preferences may be widespread or even universal in our species, many more studies are required to test this hypothesis. For example, it would be very interesting to deploy our FlexTiles software in cultures in which tiles in general, and the p4/pnm pattern in particular, are rare (e.g., many East Asian countries, and all traditional hunter–gatherer cultures). Is this type of symmetry strongly preferred only by those exposed to it from an early age, or is it based on such powerful biological predispositions that little or no previous exposure is necessary? Today, such questions are relatively easy to pose via cross-cultural computer-based comparisons, but their answers can be expected to have deep implications for our understanding of aesthetics and the human sense of order.

References


Received November 13, 2012
Revision received December 6, 2012
Accepted December 20, 2012