Supplementary information

Aesthetic Preferences among Spatial Patterns: Large-Scale Experiment, Comprehensive Exploration, and a Three-Component Regularity-Based Model

Liqiang Huang (lqhuang@cuhk.edu.hk)

Department of Psychology, The Chinese University of Hong Kong, Hong Kong, China

Table of Contents

1. Summary of files

All the data and scripts used for data analysis in this study are available on the Open Science Framework (https://doi.org/10.17605/OSF.IO/S9Y32), bundled in a zip file named "data and scripts.zip." These files are organized into four folders:

The "data" folder contains all relevant data files (.mat, .csv, .xlsx) for this study. Two files, "SWM_data.csv" and "Huang_2023_transformed_patterns.csv," are identical to Huang (2023)'s original "data.csv" and "transformed_patterns.csv," storing data and pattern information, respectively, from that study. The remaining files — "raw_data.xlsx," "formatted_data.mat," and "aesthetics_data.csv," — provide data specific to the current study, representing different stages of processing.

The "QCE_ASP" folder contains MATLAB scripts for the QCE-ASP model and related analyses, run on R2022b. Scripts are found in this folder unless stated otherwise.

The "preliminary" and "CNN" folders will be discussed below.

Additionally, another zip file named "stimuli.zip" is available via the same OSF link. This file contains a video titled "sample trials.mp4" and a folder named "all patterns." The folder includes 10,000 images, each showing the eight patterns derived from one of the 10,000 prototypes, along with their aesthetic ratings and predicted ratings. Please note that the predicted aesthetic ratings for the eight patterns of each prototype are identical. This is because the QCE-ASP model is invariant to rotation and reflection, as discussed in the main text.

2. Supplementary information on the method

The patterns (i.e., the locations of black disks) utilized were the same 80,000 patterns[1] used in Huang (2023). These patterns, along with their aesthetic ratings and predicted ratings, can be found in the file "stimuli.zip" via OSF (https://doi.org/10.17605/OSF.IO/S9Y32).

Rewards for participants were primarily based on their level of participation, supplemented by a lottery system. Regarding participation, participants earned a single credit after completing a 100-trial session. They were subsequently ranked based on their cumulative credits. At the conclusion of each week-long period, a cutoff point was determined based on the cumulative credits of the participant ranked 200th among all participants. Those with credits surpassing 75% of this cutoff point, typically encompassing 300-500 participants, received a prize of 25 Chinese Yuan.

[1] The appearance of the patterns, including both the target and other items, differs between this study and Huang (2023). This difference arises because the experiment in Huang (2023) was integrated into the mobile app "Light of the Future", requiring a style adapted for compatibility with the app. In contrast, the present study was conducted using our lab's own data-collection platform, where I aimed to maintain a simple design to minimize extraneous factors.

As for the lottery system, during each week-long period, two participants were randomly chosen from the top 200 to receive special prizes, with each prize of 500 Chinese Yuan.

The data collection was conducted in 9 week-long sessions, from September 12, 2022 to November 14, 2022.

3. Supplementary information on the preliminary analysis

The raw data of this study is saved in the file “raw_data.xlsx”, consisting of two columns × 424,851 rows, with each row representing the data of a 100-trial session. The first column “ID” shows the participant for each session. For privacy, the ID shown is a unique internal identifier, distinct from the participant’s screen ID but mapped one-to-one to it. The second column, “save_data,” displays data for each session, comprising 600 numbers divided by “/” into 100 subsets of 6 numbers each. Each subset represents one trial, showing the prototype and transformation for the first (numbers 1-2) and second patterns (numbers 3-4), the response (number 5), and the response time (number 6). The prototype, ranging [1, 10,000], matches Huang (2023)’s prototype patterns. The transformation, between [0, 7], denotes the eight possible transformations (four orientations and their mirror reflections) in the same way implemented by Huang (2023). The response, within the range of [0, 71], indicates the tapped item among the 72 items of the two patterns. Only the distinction between the two patterns (0-35 vs. 36-71) was analyzed. Response time, measured in milliseconds, was primarily used as a criterion for data exclusion.

The script “preliminary/raw_data_to_formatted_data.m” converts raw data from “raw_data.xlsx” to “formatted_data.mat” and takes up to an hour to run.

The script “preliminary/formatted_data_to_summary.m” processes formatted data to calculate 80,000 aesthetic ratings as the percentage of responses favoring a pattern, as discussed in the main text. These aesthetic ratings are saved in “aesthetics_data.csv,” the format of which will be explained below. This script also computes the average working memory accuracy for the 80,000 patterns from Huang (2023) and calculates the R^2 value for the correlation between this study’s aesthetic preference and Huang (2023)’s working memory performance.

The summarized data (“aesthetics_data.csv”) includes 10,000 rows, with each row representing data for eight patterns generated from one of the 10,000 prototype patterns (see Huang 2023 for details). The file has 297 columns. Columns 1-36 show whether an item is present in each of the 36 cells in the 6 × 6 array (1 = present vs. 0 = absent) for the transformation “0”. Columns 37-72 do the same for the transformation “1”, and so on for all eight transformations. Together, these make up columns 1-288. Columns 289-296 of “aesthetics_data.csv” record aesthetic ratings for the transformation “0” to “7”. The last column labels these prototype patterns (1-10,000). This summarized data (“aesthetics_data.csv”) was used in both the CNN and the QCE-ASP models.

4. Supplementary information on the CNN

The CNN model implemented in "CNN/CNN.py" was run on PyTorch version 1.12.0.

The network was trained to maximize the correlation (R^2) between the predicted and actual aesthetic ratings. Specifically, "1 - R^2" was used as the loss function, although R^2 is referenced here for clarity. The Adam optimizer was used.

During the training of potential CNN networks, a validation-based early stopping strategy was implemented to prevent overfitting. Specifically, each configuration (i.e., each possible combination of the number of layers and channels, as shown in Figures 2b and 2c) was tested 10 times. For each test, the dataset was randomly divided into training and validation sets, each comprising 5,000 prototypes (i.e., 40,000 patterns). The network was trained on the training set, and the average R^2 of the validation set was evaluated every 100 epochs. Training terminated once the R^2 declined from the previous evaluation. The roles of the training and validation sets were then swapped, and the process was repeated.

5. Step-by-step explanation of the QCE-ASP model

The QCE-ASP model is implemented in the "QCE_ASP.m" function. To see the output of the model, set the "i_model" as "1" and run the script "show_model_results.m".

This function consists of six parts. In the following descriptions, parameters are denoted as $P_{parameter\_name}$, while all other variables in the script are denoted as $V_{variable\_name}$.

1. **Parameters and data size**: The parameters and $V_{pattern\_database}$ are inputted.
2. **Locations of items**: The $V_{pattern\_database}$ is used to calculate the locations of items. The location information is represented in two formats for convenient use: V_{x_item} and V_{y_item} represent the x and y coordinates of the 8 items for each pattern, while V_{check_array} represents the distribution of items (1 for presence, 0 for absence) in an 8×8 array (the 6×6 pattern with margins on each side).
3. **Proximity**: First, a look-up table ($V_{proximity\_look\_up\_table}$) is generated to represent the portion of a normal probability distribution function (normpdf) that falls within each cell of an array. The standard deviation (SD) of this distribution is determined by $P_{proximity\_sd}$, the first parameter of the QCE-ASP model. Then, for each item in a pattern, a $V_{proximity\_distribution\_of\_an\_individual\_item}$ is created by extracting the relevant subset from the $V_{proximity\_look\_up\_table}$, which is used to compute two indices: $V_{proximity\_for\_all\_56\_item\_pairs}$, which sums the proximity values for all 8×7 = 56 item pairs, and $V_{total\_proximity\_distribution\_across\_all\_cells}$, which sums the proximity values for all items across all cells, including blank ones. Finally, the $V_{proximity\_scores}$ for a pattern are calculated by taking the sum of the logarithms of these two indices.
4. **Continuity**: For each item in each pattern, $V_{the\_3\_by\_3\_patch\_surrounding\_an\_item}$ is generated by reading the appropriate subset from the V_{check_array}. A $V_{list\_of\_all\_neighboring\_items}$ is then created by identifying the non-blank cells within $V_{the\_3\_by\_3\_patch\_surrounding\_an\_item}$. Next, all possible neighboring pairs are

evaluated to calculate the $V_{continuity\_value\_for\_each\_pair}$, determined by the $V_{angle\_of\_alignment}$ between the two neighbors in the pair, using a sine function (see Figure 3d). Finally, the $V_{continuity\_scores}$ is obtained by summing the $V_{continuity\_value\_for\_each\_pair}$ of all possible neighboring pairs across all items.

5. **Linearity**: Linearity scores ($V_{linearity\_scores}$) are calculated based on the ability of each item in a pattern to extend in four orthogonal directions (leftward, rightward, upward, and downward) and four diagonal directions (top-left, top-right, bottom-left, and bottom-right) as a straight line until an empty cell is encountered. For each successful extension in an orthogonal direction, one point is added to $V_{linearity\_scores}$. Diagonal extensions are awarded half a point each, accounting for the fact that diagonal lines may appear visually weaker due to items being spaced less tightly than in orthogonal lines. Finally, a quadratic function centered around $P_{optimal\_linearity}$—the second parameter of the QCE-ASP model—is applied to $V_{linearity\_scores}$. This adjustment captures the observation that while some degree of linearity is generally aesthetically appealing, excessive linearity can diminish visual appeal.
6. **Components for predictions**: The $V_{proximity\_scores}$, $V_{continuity\_scores}$, and $V_{linearity\_scores}$ are concatenated into $V_{components\_for\_predictions}$, which is the output of the “QCE_ASP.m” function. This output is then utilized by other scripts (e.g., “show_model_results.m” and “optimization.m”) to generate predicted aesthetic ratings through linear combination, which are subsequently compared to the actual data to calculate the model’s fit (R^2).

6. Supplementary information on iterative development of the model

6.1. Details of iterative refinement

In this study, the iterative refinement process unfolded as follows: Building on prior research suggesting that regularity enhances aesthetic appeal (e.g., Enquist & Arak, 1994; Reber, Schwarz, & Winkielman, 2004), I hypothesized that regularity would play a significant role. I began by examining classic regularity-related mechanisms from earlier studies, such as symmetry, but these proved ineffective in accounting for the current dataset. As a result, I focused on identifying regularity-related mechanisms from observational insights.

Among the mechanisms tested, the “proximity” mechanism yielded the highest R^2 value. Following this, I evaluated several potential distributions to capture the concept of “proximity,” including the normpdf of distance, 1/distance, and 1/(distance^2), with the normpdf approach achieving the highest R^2 value. Then, multiple ways of summarizing the proximity values were explored, and both the “proximity values for all item pairs” and the “proximity values distributed across all cells of the 6×6 array” were found to enhance model performance; thus, both were included.

Using the “scientific regret minimization” method, it was observed that the “angle of alignment” of neighboring items and the presence of “lines” also contributed significantly, leading to the incorporation of “continuity” and “linearity” mechanisms. For example, Figure S1 displays 100 patterns where an under-development QCE-ASP model diverges most from the CNN, at a stage where proximity and continuity—but not linearity—had been included (i.e., the model 3, “proximity + continuity”, in Table S1). These patterns exhibit noticeable “lines,” prompting the addition of the “linearity” mechanism.

The above descriptions focus on successful elements and omit other candidates that were tested and rejected. In total, the iterative refinement of the QCE-ASP model took two weeks and tested dozens of candidate models. Compared to other CE models, this process was relatively brief. For instance, the QCE-SWM in Huang (2023) and the QCE-VWM in Huang (2025) took ten and eight months to develop, respectively, with each model testing nearly a thousand candidate models. This brevity may stem from the nature of aesthetic judgment, which seems to have avoided the deliberate strategies, as discussed in the main text.

6.2. Criteria for adding/removing a mechanism

In developing the QCE-ASP model, decisions to retain or remove a mechanism are guided by its conceptual plausibility and a statistical index called CAD (complexity-adjusted Cohen’s d), introduced by Huang (2025). To calculate CAD for a mechanism, two models—one incorporating the mechanism and one excluding it—are optimized. The benefit of including the mechanism (i.e., the improvement in model fit[2]) is then evaluated across all 80,000 patterns, with a t-test applied to assess the consistency of these benefits across the patterns. From this t-test, Cohen’s d is derived.

In the simplest case, where the two models differ by only one parameter, the CAD value is equivalent to Cohen’s d. However, as its name suggests, CAD incorporates complexity-based adjustments when models differ by more than one parameter. Specifically, in this adjustment, Cohen’s d is divided by the square root of Δ_{param} (the difference in the number of parameters between the two models).

While Huang (2025) used a minimum CAD threshold of 0.2, this threshold is unnecessarily strict in the present study (see Table S1). Instead, a minimum CAD threshold of 0.1 is adopted for retaining a mechanism. Table S1 lists the CAD values for comparisons between the QCE-ASP model and several alternative models.

6.3. Optimization algorithm

The QCE-ASP model parameters were optimized using a random search algorithm, chosen for its adaptability across various mechanisms without requiring gradient-based methods. This optimization procedure was identical to that of Huang (2025), except that the goal was changed to maximizing R^2. The optimization process involved two stages: mini-batch and whole-set optimizations. In mini-batch optimization, the data was divided into smaller groups for iterative adjustment, while whole-set optimization used the entire dataset to refine parameters until improvements stabilized. Step sizes were

[2] This study employs R^2 as the fitting index, which serves as an overall measure but cannot be applied directly to individual patterns. For CAD calculations, prediction errors for individual patterns are used instead.

adjusted based on parameter performance to ensure optimal effect, and the model cycled through these processes until achieving stable results. For further details, please refer to Huang (2025).

6.4. Mechanisms not explained by QCE-ASP

As shown in Figure 5, there is a gap between the QCE-ASP and CNN models, suggesting that there are still mechanisms yet to be captured by QCE-ASP. To identify these mechanisms, Figure S2 presents the 100 patterns where the finalized QCE-ASP model diverges most from the CNN. Although clear regularities are absent, there are some subtle clues. After examining several hundred additional patterns along similar lines (not shown here), several hypotheses were formed and tested. None of these met the above-mentioned criteria (CAD > 0.1), so they were ultimately not included in the QCE-ASP. However, two hypotheses resulted in improvements with CAD values around 0.05, making them noteworthy.

One of these is “spread,” where the variance in the spatial distribution of items slightly enhances the model’s performance. The other is the “number of neighbors.” Here, incorporating the number of neighbors in the “continuity” mechanism as a supplementary factor also slightly improves the model’s performance. Statistical details of the models associated with these mechanisms (models 10-12) are provided in Supplementary Information 7.

7. Statistical evidence for the QCE-ASP model

7.1. The alternative models

During the iterative refinement of the QCE-ASP model, a few dozen models have been tried, and 11 of them are presented here because of their conceptual relevance to the QCE-ASP. To view the results of an alternative model, run the “show_model_results.m” script with the corresponding “i_model” from Table S1.

As listed in Table S1, Model 1 is the QCE-ASP model itself. Models 2–7 represent ablated versions of the QCE-ASP model, each with one or two of the model’s three components removed. Models 8–9 involve modifications to key aspects of the QCE-ASP model: Model 8 removes logarithmic transformations in the proximity component, and Model 9 omits quadratic transformations in the linearity component. Models 10–12 introduce additional factors to the QCE-ASP model: Model 10 accounts for item spread, Model 11 considers the number of neighbors, and Model 12 incorporates both factors (see Supplementary Information 6.4 for discussions on both item spread and number of neighbors).

7.2. Model comparison

Each of the 11 aforementioned alternative models was optimized to achieve its own maximum R^2, with the optimized parameters saved in the “saved_parameters.m” script. In the “show_model_results.m” script, the results of these 11 alternative models were compared with those of the QCE-ASP model using 11 t-tests, which assessed the consistency of differences in prediction errors across 80,000 patterns.

As shown in Table S1, all t-values are substantial, with correspondingly small p-values, primarily due to the large sample size (80,000 patterns). Consequently, Cohen's d, an effect size measure unaffected by sample size, and its complexity-adjusted version (CAD) are used as the primary statistical indices in this study (see Supplementary information 6.2).

As shown in Table S1, Models 2–9 performed significantly worse than the QCE-ASP model, with a CAD[3] greater than 0.1. In contrast, Models 10–12 demonstrated only a slight improvement over the QCE-ASP model, with a CAD less than 0.1, which does not justify the added complexity.

Overall, the QCE-ASP model captures the essential mechanisms while omitting non-essential ones, striking an optimal balance between effectiveness and simplicity (see also Figure 5b).

8. Supplementary information on various analysis

8.1. Psychometric function

The psychometric function of aesthetic judgment was calculated by the script "psychometric_function.m". Specifically, both the responses and reaction times (RTs) from all 41,716,400 trials were categorized according to the "aesthetic difference" (i.e., the difference between aesthetic ratings of the top and bottom patterns). These aesthetic differences were then rounded to discrete levels in increments of 0.01, which are presented in Figures 1d and 1e. To mitigate the effects of extreme values and non-normal distributions, RTs were capped within the range of [500msec, 5000msec] and subsequently log-transformed before analysis.

Since aesthetic ratings are derived directly from the responses, it is self-evident that this analysis would reveal a generally increasing function. However, the exact shape of this function is not immediately apparent. Interestingly, the data fit very well to a "ranged logistic curve." Let ***x*** represent the aesthetic difference. The aesthetic preference function can be expressed as:

$$Aesthetic\ preference\ (x) = \left(\frac{1}{1 + e^{-k(x-x0)}} - 0.5\right) \times r + 0.5$$

In this equation:

- ***k*** is the logistic growth rate, which determines the steepness of the curve, with a value of 7.78.
- ***x0*** is the "aesthetic difference" value at the function's midpoint, with a value of -0.0473. This indicates that when the aesthetic rating of the top pattern is 0.0473 lower than that of the bottom pattern, the two patterns are considered a tie.

[3] For Model 8, the CAD is inapplicable because it has the same number of parameters as the QCE-ASP.

- ***r*** defines the range between the ceiling and floor of the function, with a value of 0.579. This means the ceiling and floor of the psychometric function are approximately 0.790 and 0.210, respectively.

8.2. Individual differences analysis

An individual differences analysis was carried out by the script "individual_difference.m." The early parts of the script calculates the aesthetic ratings, and also split-half reliability, for each individual participant. These calculations largely follow the same procedures as in the preliminary analysis except performing the calculations separately for individual participants. Several points require clarification:

First, this analysis is limited to a group of "active participants". Unlike usual lab experiments that recruits a group of participants and ask each of them to complete an equal amount of experimental trials, the present experiment is presented as an online game that anyone can participate for as much or as little as they wish. So the amount of participation vary greatly across participants. As shown in Figure S3a, some participants completed for over 1000 sessions, whereas some others completed only 1 session. To ensure sufficient data for each individual participant, only those who completed at least 500 sessions were included in the individual-difference analysis. This resulted in 341 of the 1060 participants being included, but retained 380,850 (91.3%) out of the total 417,164 sessions, which makes the data reasonably representative of the entire dataset.

Second, to ensure sufficient data for each individual participant at the individual level, the analysis was conducted at the level of the 10,000 prototypes rather than the 80,000 individual patterns. As discussed in the main text, prototype-level variations account for almost all (99.91%) of the explainable pattern-level variations, making this a reasonable approximation.

Third, even after applying the first and second criteria, there were still instances in the calculation of split-half reliability where a participant had zero trials at a prototype level. In such cases, a value of 0.5 (i.e., the midpoint of all aesthetic ratings) was assigned. These cases were rare (0.05% of all cases), so their influence is negligible.

After obtaining the aesthetic ratings and split-half reliability for each individual participant, three indices are calculated:

1. **Portion of Variance Explainable**: This is defined as the mean split-half reliability across all participants, with a value of 0.448[4].

[4] This number appears unusually low for two main reasons. First, this study was not specifically designed for individual-difference analysis, making it suboptimal for this purpose. In particular, there are often very few observations for each level, with 45.6% of levels having 10 or fewer observations. Second, the current aesthetic task is subjective and lacks a standard answer, making it challenging to identify unengaged participants. Although measures were implemented to exclude participants who did not take the task seriously (see the Method section in the main text), it is still possible that some participants either genuinely lacked a reliable aesthetic preference among these patterns or employed strategies that bypassed the screening criteria. Future studies will require more sophisticated screening methods to address these limitations.

2. **Portion of Variance Explained by the Primary Component**: This is defined as the portion of variance explained by the first component of a principal component analysis performed on the aesthetic ratings of all participants. Its value is 0.242.
3. **Portion of Variance Explained by QCE-ASP**: This is defined as the mean R^2 of all correlations between each participant's aesthetic preferences and the QCE-ASP model's predictions. Its value is 0.222.

Based on these results, QCE-ASP's predictions explain 91.8% (0.222/0.242) of the variance in the "primary component" of the individual-level data but only 49.5% (0.222/0.448) of the total explainable variance in the individual-level data itself.

8.3. Practice effect

Following the preceding section, the practice effect on aesthetic preferences was assessed in the final part of the script "individual_difference.m." Specifically, within the subset of data used for the individual differences analysis (comprising participants who participated in at least 500 sessions), their first 500 sessions were divided into five batches of 100 sessions each. The prototype-level variance explained by the QCE-ASP model was calculated for each batch, along with the split-half reliability.

The results, shown in Figure S3b, indicate that the portion of variance explained by the QCE-ASP model slightly increases with practice, while the split-half reliability slightly decreases. Overall, these findings suggest a subtle tendency for observers to gravitate toward the regularity-based strategy described by the QCE-ASP model over time.

8.4. Cross-validation

The QCE-ASP model was developed using the entire dataset, which raises the potential concern of overfitting. However, this model is simple and based on a very large dataset, making overfitting unlikely to play a significant role. To evaluate the extent of potential overfitting, cross-validation was performed.

The cross-validation was conducted using the script *"cross_validation.m."* This script is essentially identical to *"optimization.m"* except that the data is divided into "training data" and "validation data." The parameters of the QCE-ASP model were optimized on the training data and subsequently tested on both datasets. The discrepancy between the R^2 values of the training and validation data served as an index of the "magnitude of overfitting."

Cross-validation was performed across seven levels of data partitioning: 1/2, 1/5, 1/10, 1/20, 1/50, 1/100, and 1/200 of the data were used as training data. For each partition level, the procedure was repeated across all possible subsets (e.g., iterating through 200 subsets for the 1/200 level), and the mean of all "magnitudes of overfitting" was calculated.

The results of this analysis are presented in Figure S3c. As expected, the magnitude of overfitting decreases as more data are used for training. Extrapolating from the tested levels, the estimated magnitude of overfitting for the whole-set model (i.e., the expected decrease in R^2 when applied to future datasets using the same type of patterns) is approximately 0.00007.

8.5. Symmetric patterns

As described in the method section, 10,000 prototype patterns were generated by randomly selecting 8 out of 36 cells. Some of these patterns happened to exhibit symmetry in various forms.

The role of symmetry was examined in the script "check_symmetry.m." Specifically, prototype patterns that are identical to any of their transformed versions were classified as symmetric. In total, 11 such prototype patterns were identified, including: 2 mirror-symmetric along a cardinal axis (778, 5272), 7 mirror-symmetric along a diagonal axis (2795, 3898, 4472, 8576, 8767, 9316, 9802), and 2 rotationally symmetric (2047, 2528). Six of these 11 prototype patterns are shown in Figure S4.

Critically, the aesthetic ratings of the 88 patterns associated with the 11 prototypes were compared to the predictions of the QCE-ASP model. The results indicate that the aesthetic ratings of symmetric patterns do not significantly exceed the QCE-ASP model's predictions (difference = 0.00072, $t = 0.345$, $p = 0.731$, Cohen's $d = 0.037$)[5].

8.6. Most and least appealing virtual patterns

This study is data-driven in that the aesthetic appeal of the patterns is not manipulated along a predetermined dimension but instead emerges from variations among randomly generated patterns. This raises the question of whether these random variations encompass a sufficiently large range of aesthetic appeal. To address this, the script "extreme_patterns.m" examines all C(36,8)=30,260,340 possible patterns and calculates the predicted aesthetic appeal of these "virtual" patterns. They are referred to as virtual because the majority of them have not been tested in this study.

Two key points are worth noting. First, the range of aesthetic ratings predicted by the QCE-ASP model for all these virtual patterns [0.247,0.697] is only moderately broader than the range for the 80,000 tested patterns [0.307,0.654]. This indicates that the 80,000 tested patterns have covered the majority of the full possible range of aesthetic appeal.

Second, the most and least appealing patterns, as predicted by the QCE-ASP model, are shown in Figure S5. These patterns align well with the regularity-based conceptual interpretation of the QCE-ASP model described in the main text. The predicted most appealing patterns are the most regular, while the predicted least appealing patterns are the least regular (i.e., scattered).

8.7. Spatial resolution of aesthetic judgment and working memory

As discussed in the main text, the spatial resolution in aesthetic judgment is notably lower than in working memory. This refers to the resolution of the most fundamental elements within both the QCE-ASP and QCE-SWM models. Specifically, both models are built around a set of normal distributions, each stemming from an item. In the QCE-ASP model, these distributions represent proximity scores (see Figure 3b), while in the QCE-SWM model, they characterize spatial uncertainty in working memory.

[5] Of course, the null effect does not prove that the symmetry effect is absent; it only indicates that the present study did not find sufficient evidence to support the presence of a symmetry effect.

In both models, the precision of these normal distributions is influenced by other mechanisms. To enable a fair and accurate comparison of spatial resolutions in aesthetic judgment and working memory, we examine the standard deviations (SDs) of these normal distributions within simplified models that include only these distributions, excluding other mechanisms. Specifically, this comparison uses the proximity-only model from the present study (see supplementary information 7) and the baseline model from the spatial working memory study (Huang, 2023). In this comparison, processing in aesthetic judgment is 2.02 times as diffused (SD = 1.29 vs. 0.64) as in the working memory task (see Figure 8a).

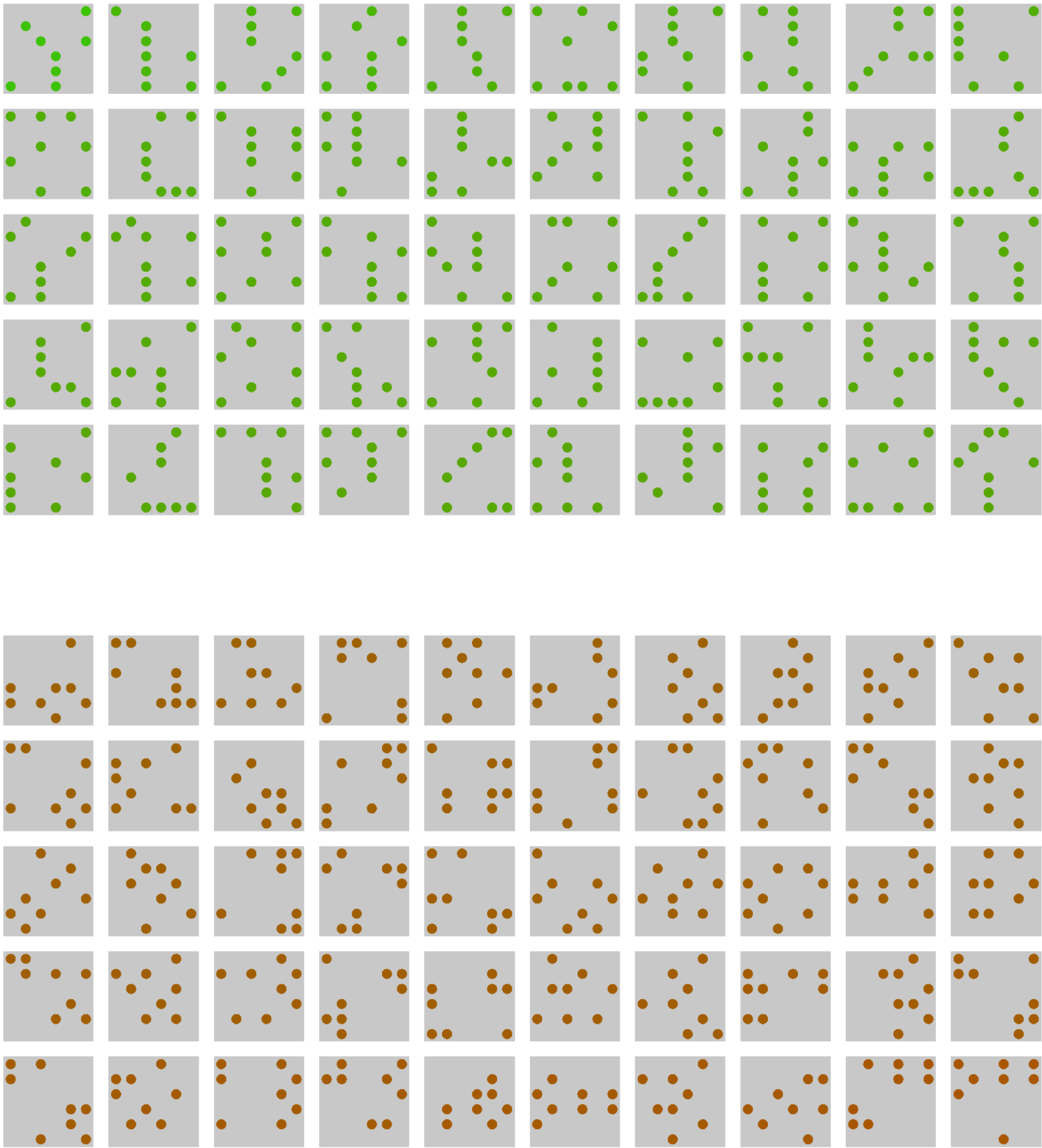

Figure S1. Comparison between the "Proximity+Continuity" Model and CNN

This Figure illustrates the patterns where the CNN and "Proximity+Continuity" models produce the most divergent predictions. The 50 green and red patterns represent instances where CNN predicts higher and lower aesthetic appeal, respectively. The green patterns appear to contain more "lines," suggesting that incorporating a "linearity" mechanism could enhance the model.

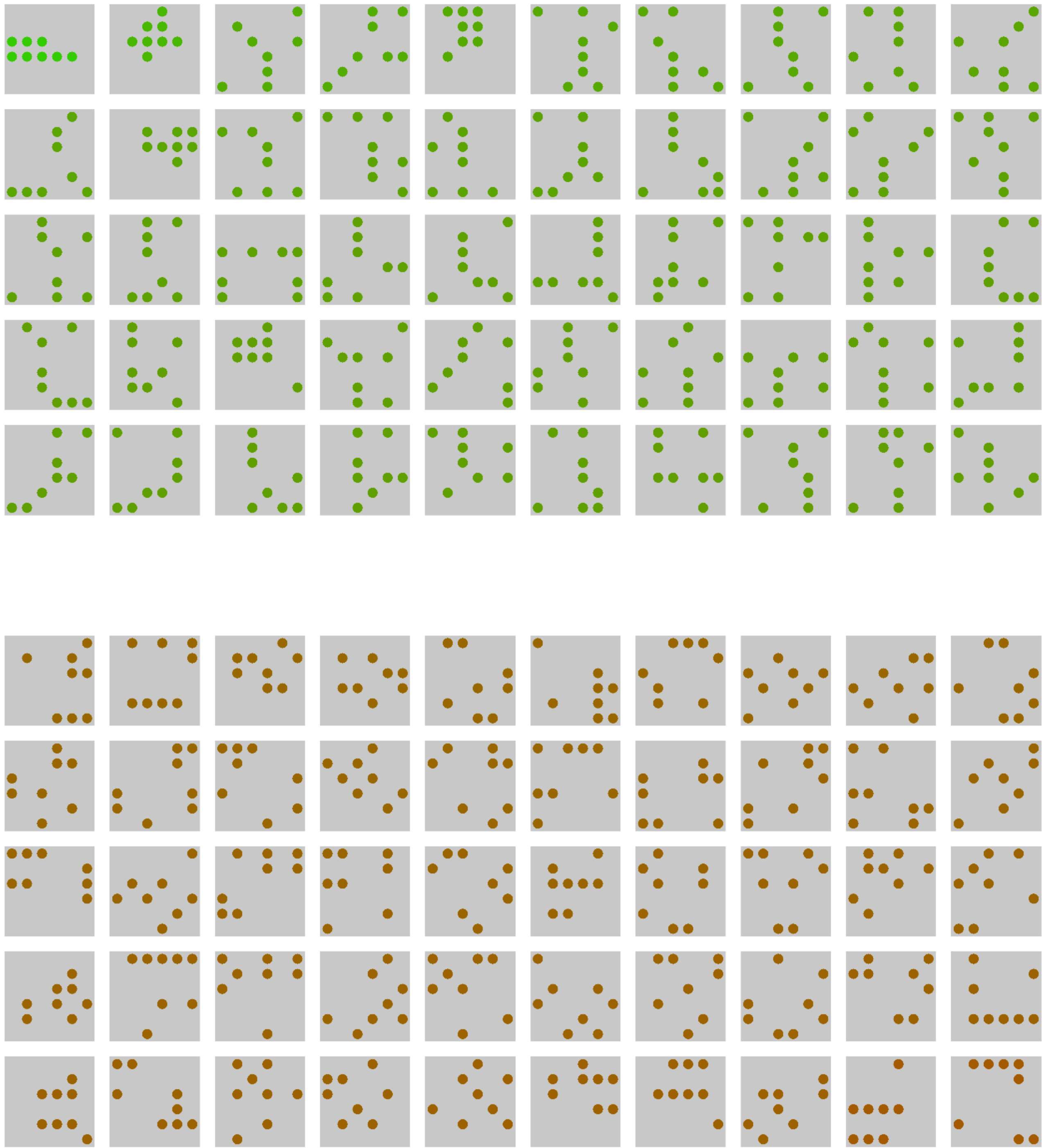

Figure S2. Comparison between the QCE-ASP and CNN

This Figure illustrates the patterns where the CNN and QCE-ASP models produce the most divergent predictions. The 50 green and red patterns represent instances where CNN predicts higher and lower aesthetic appeal, respectively. Although clear regularities are absent, subtle clues suggest that more spread-out patterns and patterns with more neighboring elements tend to be more appealing.

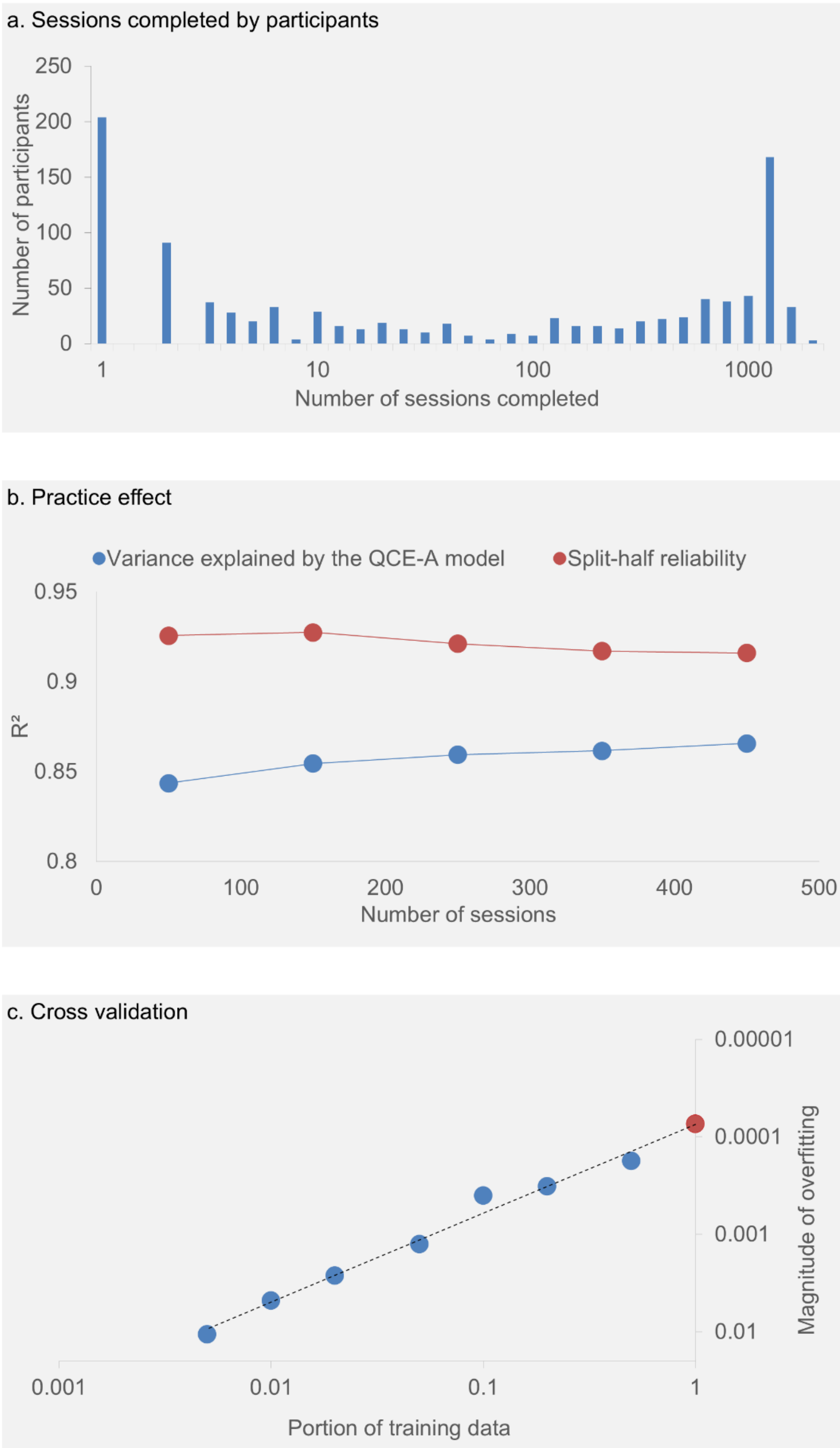

Figure S3. Various Analyses.

Panel a illustrates the distribution of completed sessions among the observers in this experiment. While some observers completed over 1,000 sessions, others completed as few as a single session.

Panel b demonstrates the practice effect. The variance explained by the QCE-ASP model increases slightly with practice, whereas the split-half reliability decreases slightly. This indicates a subtle tendency for observers to gradually adopt the regularity-based strategy described by the QCE-ASP model over time.

Panel c presents the results of cross-validation. As anticipated, the magnitude of overfitting decreases as the training dataset size increases. The blue dots represent the actual tested levels, while the red dot indicates the extrapolated overfitting magnitude for the QCE-ASP model.

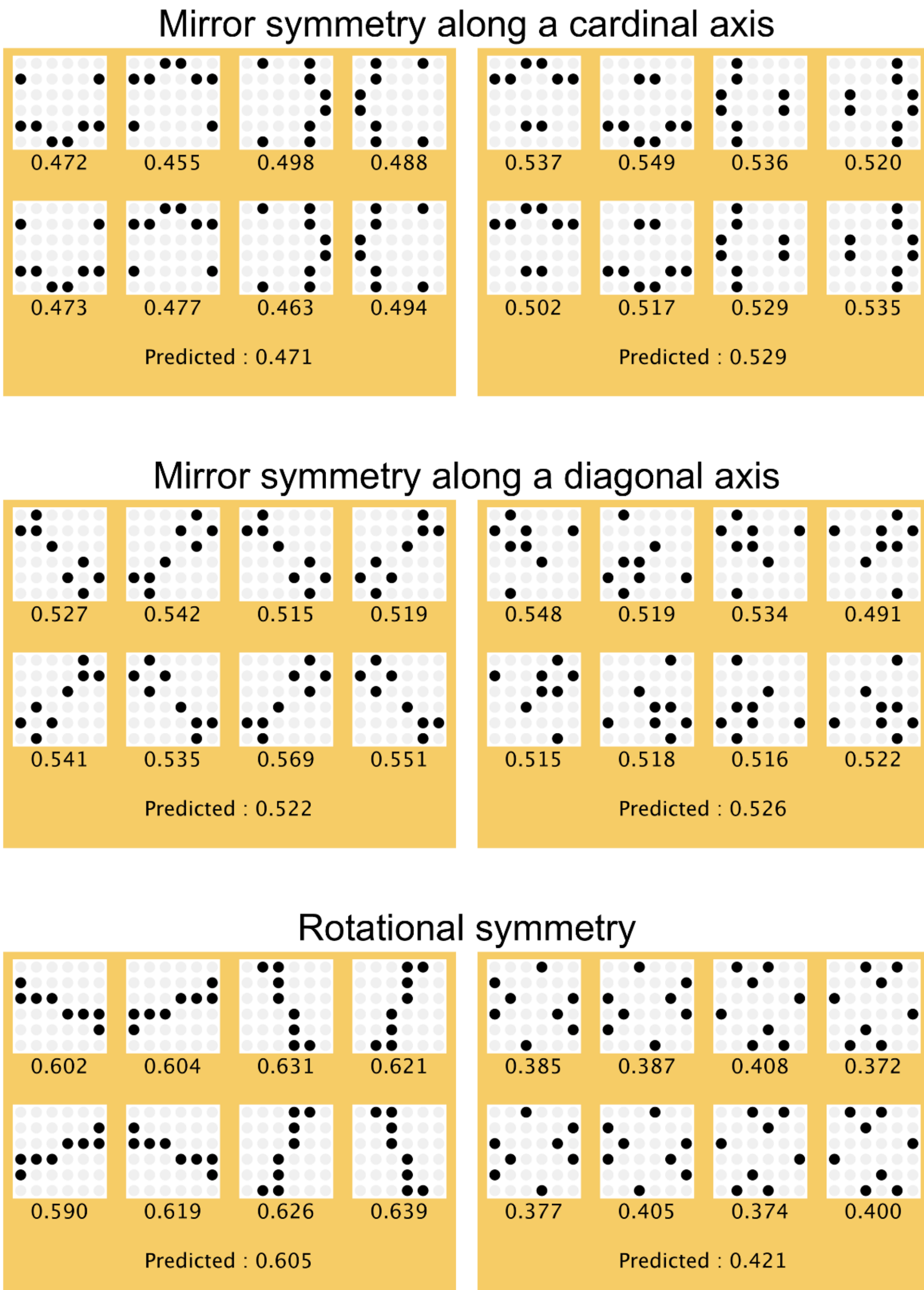

Figure S4. Symmetric patterns.

The aesthetic ratings of these symmetric patterns do not exceed the predictions of the QCE-ASP model, suggesting no intrinsic preference for symmetry in the aesthetic judgments involved in this experiment.

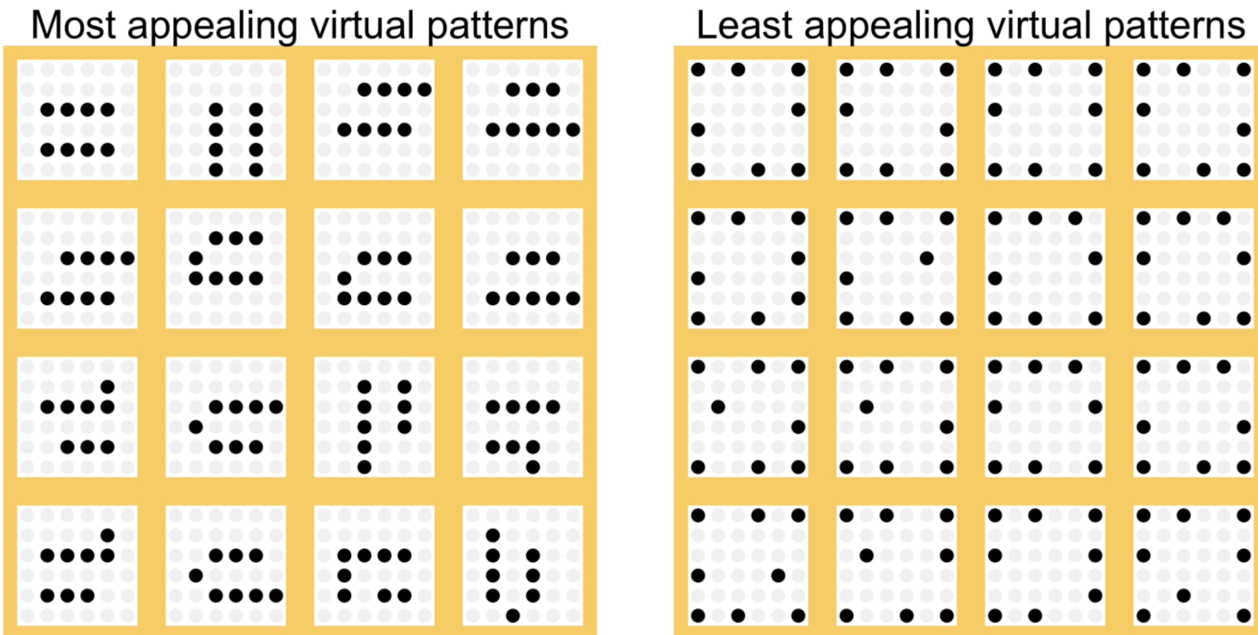

Figure S5. Most and least appealing virtual Patterns.

The QCE-ASP model was used to predict the aesthetic ratings of all possible patterns. The most and least appealing patterns are presented here and align well with the regularity-based interpretation of the QCE-ASP model discussed in the main text. The most appealing patterns are highly regular, whereas the least appealing patterns are characterized by minimal regularity (i.e., scattered).

i_model / name	N_{param}	R^2	Comparisons between the QCE-ASP and the alternative models					
			Change of R^2	t-value	p-value	Cohen's d	Δ_{param}	CAD
1 / QCE-ASP	6	0.8765						
2 / proximity + linearity	5	0.8593	0.0172	52.94	<1.E-300	0.1872	1	0.1872
3 / proximity + continuity	4	0.8372	0.0393	77.04	<1.E-300	0.2724	2	0.1926
4 / continuity + linearity	4	0.6254	0.2511	138.06	<1.E-300	0.4881	2	0.3451
5 / proximity only	3	0.8095	0.0670	96.49	<1.E-300	0.3412	3	0.1970
6 / continuity only	2	0.3115	0.5650	179.36	<1.E-300	0.6341	4	0.3171
7 / linearity only	3	0.6252	0.2513	137.96	<1.E-300	0.4878	3	0.2816
8 / no log in proximity	6	0.8625	0.0140	42.12	<1.E-300	0.1489	0	N.A.
9 / no quadratic in linearity	5	0.8626	0.0139	41.98	<1.E-300	0.1484	1	0.1484
10 / plus spread	7	0.8778	-0.0013	-14.59	3.50E-48	-0.0516	1	-0.0516
11 / plus num neighbor	7	0.8774	-0.0009	-11.56	7.33E-31	-0.0409	1	-0.0409
12 / plus both	8	0.8784	-0.0019	-17.14	8.99E-66	-0.0606	2	-0.0429

Table S1. Model comparison.

Model 1 is the full QCE-ASP model. Models 2–7 are ablated versions of the QCE-ASP model, with one or two of its three components removed. Models 8 and 9 introduce specific modifications to the QCE-ASP model: Model 8 excludes logarithmic transformations in the proximity component, while Model 9 removes quadratic transformations in the linearity component. All these modified models (2–9) perform substantially worse than the full QCE-ASP model.

Models 10–12 extend the QCE-ASP model by adding extra factors: Model 10 considers the spread of items, Model 11 considers the number of neighbors in the "continuity" mechanism, and Model 12 incorporates both factors. However, these extensions (Models 10–12) offer only slight improvements over the QCE-ASP model, insufficient to justify the added complexity. Overall, the QCE-ASP model achieves the best balance of effectiveness and simplicity (see also Figure 4b).

Please note that the p-values for models 2–9 are smaller than the smallest positive number representable in double-precision floating-point format and are therefore displayed as "<1.E-300."