Supplementary Materials

Methods

Statistics of Repeated Measures Spearman's Rank correlations

We have twenty participants with forty different conditions. Therefore, the usual correlation analysis is not applicable as the data violate the assumption of independence. We followed the suggestions of Mohr and Marcon (2005) and Bakdash and Marusich (2017) and used a randomization test (Edgington, & Onghena, 2007) to acquire the significance of the correlations. Spearman's rank correlation was utilized to account for the ordinal data. All the analysis codes are available upon request.

Significance test (p-values)

The significance of the correlation was tested with the randomization test. The procedure was as follows.

1. Compute the Spearman correlation per participant (with 40 conditions)
2. Average the Spearman correlation coefficient across participants
3. Randomly permute the rank within each participant (1,000,000 times)
4. Compute the mean Spearman correlation coefficient of each randomized set
5. Count the number of coefficients larger than the originally acquired coefficient
6. Draw a histogram and fit a Gaussian function to get a probability distribution function (p-values)

Confidence interval (CI)

We followed the bootstrapping method utilized in Bakdash and Marusich, 2017 (rmcorr) to obtain 95% CIs for mean Spearman's rank correlation coefficients. There are numerous analytic methods to estimate CIs (e.g., Fisher, 1921; Woods, 2007), but the methods are not appropriate for our dataset since our data are not parametric. Instead, we used bootstrapping, which does not require distributional assumptions and uses resampling to estimate parameter accuracy (Efron & Tibshirani, 1994).

Here, we resampled the data of each participant 10,000 times (bootstrap samples) and computed the mean Spearman correlation coefficient of each resampled data. Then, we constructed an empirical sampling distribution of the mean Spearman's Rank correlation coefficients. We obtained the CIs from the empirical sampling distribution (CI_{emp}) and with the percentile bootstrap method ($CI_{percent}$; Efron & Tibshirani, 1994). As CI_{emp} and $CI_{percent}$ were comparable, we only presented $CI_{percent}$.

References

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated Measures Correlation. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00456

Edgington, E. & Onghena, P. (2007) Randomization tests, New York, Chapman & Hall.

Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. CRC Press.

Fisher, R. A. (1921). On the’probable error’of a coefficient of correlation deduced from a small sample. *Metron*, *1*, 1–32.

Mohr, D. L., & Marcon, R. A. (2005). Testing for a ‘within-subjects’ association in repeated measures data. *Journal of Nonparametric Statistics*, *17*(3), 347–363. https://doi.org/10.1080/10485250500038694

Ruscio, J. (2008). Constructing confidence intervals for Spearman’s rank correlation with ordinal data: a simulation study comparing analytic and bootstrap methods. *Journal of Modern Applied Statistical Methods*, *7*(2), 7.

Woods, C. M. (2007). Confidence intervals for gamma-family measures of ordinal association. *Psychological Methods*, *12*(2), 185.

Models

Model simulations

Two model were tested. The first model tests whether the number of direct neighboring squares can explain uncrowding. The second model tests whether the number of white pixels within the crowding window determine crowding.

1) Directly connected squares' and flankers' euclidean distance

Model performance was predicted by the inverse of the averaged Euclidean distance from the center square to each of the directly connected square. Therefore, the closer the connected square is, the better performance is. The performance s_i of configuration (condition) i was computed as follow.

$$s_i = M_i \quad \text{(1: Num\_sq)}$$

$$s_i = M_i \times id_i,\ where\ \ id_i = \sum_j^{M_i} \frac{1}{\sqrt{r_{ij}^2 + c_{ij}^2}} \quad \text{(2: Num\_sq\_group)}$$

Or $$s_i = -T_i \times f_i,\ where\ \ f_i = \sum_j^{T_i} \frac{1}{\sqrt{r_{ij}^2 + c_{ij}^2}} \quad \text{(3: Num\_sq\_flankers)}$$

Where $M_i \in [1, 10]$ denotes the number of directly connected squares to the center square. For configuration (condition), i, j encodes each element within configuration i, and r_{ij} and c_{ij} encodes row and column numbers from the center; $T_i \in [15, 24]$ encodes the number of flankers except the directly connected squares $(35 - M_i)$ for each configuration (i); the center square is annotated as $(0,0)$, $r_{ij} \in [-2, 2]$ and $c_{ij} \in [-3, 3]$. We tested three variants of the number of directly connected squares model. 1) Num_sq: only the number of squares as the predictor; 2) Num_sq_group: the number of squares discounted by the inverse of the sum of distances; 3) Num_sq_flankers: minus the number of non-connected square flankers discounted by the inverse of the sum of distances.

2) Pixel-wise euclidean distance

As a control, we also tested whether flankers' pixel values all over the configuration or within a fixed crowding window (1/2 of eccentricity) predict performance, according to a traditional view of crowding (pooling). Similar to Eq. 1, the performance s_i was computed as follow.

$$s_i = M_i \times id_i,\ where\ \ id_i = \sum_j^{M_i} \frac{1}{\sqrt{x_{ij}^2 + y_{ij}^2}} \qquad \text{(4: Pix\_n\_all)}$$

$$of\ which \quad \frac{x_i^2}{a^2} + \frac{y_i^2}{b^2} = 1, \quad a = \frac{1}{2}ecc, b = \frac{1}{4}ecc \qquad \text{(5: Pix\_n\_bouma)}$$

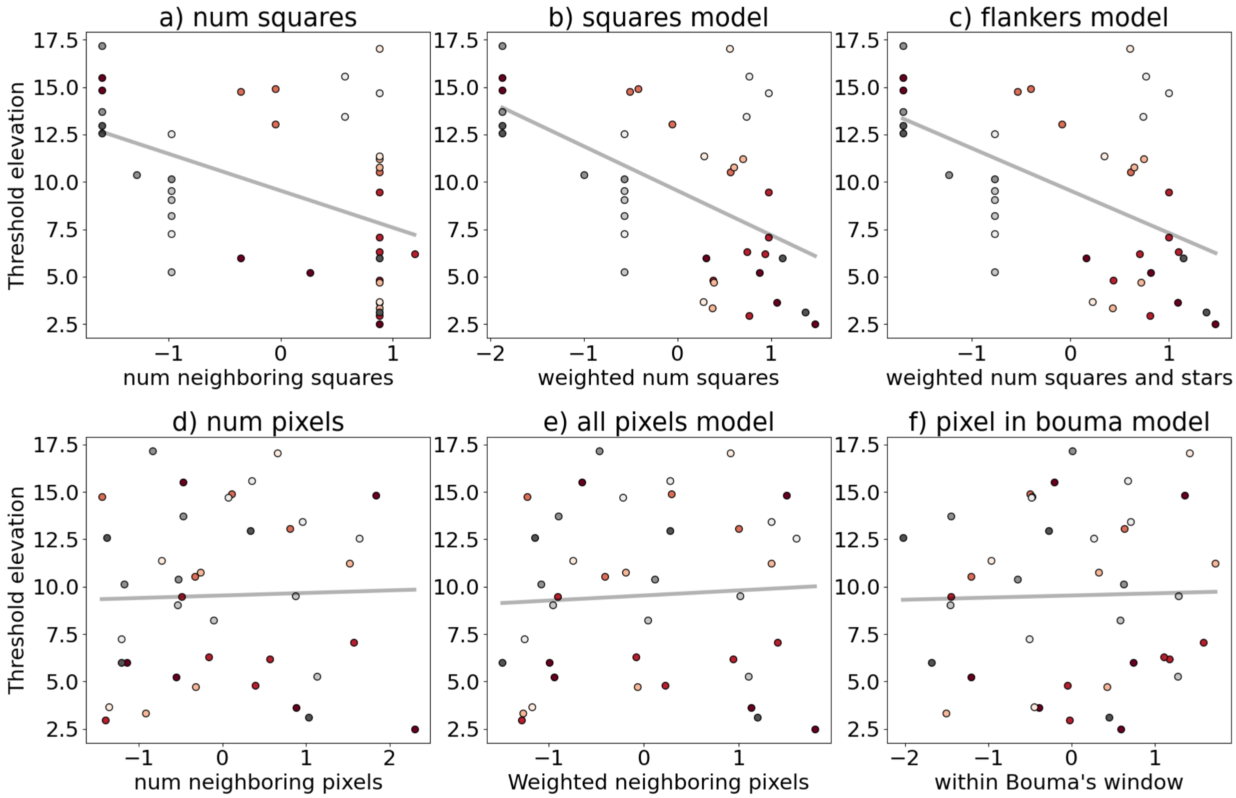
M_i encodes the number of flanker pixels in configuration (condition) i, and x_i and y_i are the pixel positions from the screen center$(0,0)$. The pixel-wise distance measure was computed within Bouma's window, which is an ellipse with ½ of eccentricity ($a = 4.5deg$) and the other focal point of ¼ eccentricity ($b = 2.25deg$). Three variants of the model were tested; 1) Pix_num: only the number of pixels; 2) Pix_n_all: the number of pixels discounted by the inverse of the sum of pixel distances; 3) Pix_n_Bouma: same as Pix_n_all only pixcels within the Bouma's window.

3) Model significance test

We analyzed the predictability of the models using two methods. First, we used LMMs which had each of the model estimates as fixed effects. For each LMM, the fixed effect was model estimates for each configuration, and each subject was considered as random intercepts.

Next, we used a leave one out cross validation (LOOCV) method to determine the explained variance of participants' performance. We linearly fitted the model estimates to the crowding performance of 19 participants behavioral data. Then, we obtained the r squared value (explained variance) by using the last participants' data (data points are not included in the linear regression). We repeated the compuation 20 times (for each participant), then averaged the r squared values from 20 iterations to get the final explained variance of each model.

Model comparisons



Supp. Figure 1. Correlations between model estimates and mean crowding levels for each configuration. The y-axis shows the mean threshold elevation, and the x-axis is the model estimates for each model (both axes have arbitrary units). Each dot represents each configuration, and the color means corresponding Gestalt principles, the same as in Figure 4.

Supp. Figure 1 shows the correlations between the mean performances across the participants and model predictions. The correlations between crowding level and the number of connected squares and that discounted by the distance showed strong correlation ($r_{\text{num\_sq}}$ (38)= - 0.50, $CI_{95\%}$ = [-0.70, -0.23], $p_{Bonf} < 0.01$; $r_{\text{num\_sq\_group}}$ (38)= - 0.60, $CI_{95\%}$= [-0.75, -0.33], $p_{Bonf} < 0.001$; $r_{\text{num\_sq\_flankers}}$ (38)= - 0.58, $CI_{95\%}$ = [-0.71, -0.23], $p_{Bonf} < 0.001$). However, flanker pixel values, regardless of the local crowding window restriction, showed weak correlation ($r_{\text{pix\_num}}$ (38)= - 0.05, $CI_{95\%}$ = [-0.36, 0.26], p = 0.75; $r_{\text{pix\_num\_all}}$ (38)= -0.02, $CI_{95\%}$ = [-0.33, 0.29], p = 0.90; $r_{\text{pix\_num\_bouma}}$ (38)= - 0.03, $CI_{95\%}$ = [-0.35, 0.29], p = 0.87).

To examine predictability further, we analyzed the predictability of the models using two methods. First, we used LMMs which had each of the model estimates as the fixed effects. We found that the number of connected squares and the number of squares with distance discount have a significant effect on the crowding level, but not for the number of pixels. For each LMM, the fixed effect was model estimates for each configuration, and each participant was considered as random intercepts. There were significant fixed effects for the number of directly connected square models, but not for the pixel value models (details in Table 1). Although the effects could only explain 6.0 % of the variances (r_m^2, Num_sq_group; for the other models, see Supp. Table 1), it was still better than the pixel estimators (0.0 %, Pix_num_bouma). Note that explained variances including the random intercept across all the models were comparable, 40% - 45% (r_c^2). This result clearly indicates that none of the models can truly explain crowding and uncrowding, there were rather large performance variances across participants and across configurations.

Supp. Table 1. LMM model likelihood test results. Detailed estimates for each model are in Supp. Table xxx.

Model	Likelihood ratio test	Significance (p)	Explained variance (r^2)
Num_sq	$\chi^2(1) = 57.077$	$p < 0.001$	$r_m^2 = 0.042,\ \ r_c^2 = 0.433$
Num_sq_group	$\chi^2(1) = 83.155$	$p < 0.001$	$r_m^2 = 0.060,\ \ r_c^2 = 0.452$
Num_sq_flankers	$\chi^2(1) = 76.264$	$p < 0.001$	$r_m^2 = 0.055,\ \ r_c^2 = 0.447$
Pix_num	$\chi^2(1) = 0.602$	$p = 0.438$	$r_m^2 = 0.000,\ \ r_c^2 = 0.390$
Pix_n_all	$\chi^2(1) = 0.097$	$p = 0.756$	$r_m^2 = 0.000,\ \ r_c^2 = 0.390$
Pix_n_bouma	$\chi^2(1) = 0.157$	$p = 0.692$	$r_m^2 = 0.000,\ \ r_c^2 = 0.390$

Next, we tested with the leave one out cross validation (LOOCV) method. Hence, here we tested the explained variance of a participants' performance from the other remaining participants' performances. We fitted the model estimates to the crowding performance of 19 participants behavioral data. We obtained an r^2 value (explained variance) by using the last participants' data (data points are not included in the linear regression). We repeated the computation 20 times (for each participant), then averaged the r squared values from 20 iterations to get the final explained variance of each model. As a result, similarly, despite the low correlation, the number of directed squares discounted by their distances could predict the crowding level partially ($r_{\text{LOOCV-num\_sq}}^2$=0.121, $r_{\text{LOOCV-num\_sq\_group}}^2$=0.164, $r_{\text{LOOCV-num\_sq\_flankers}}^2$=0.154), whereas pixel values could not ($r_{\text{LOOCV-pix\_num}}^2$=0.013, $r_{\text{LOOCV-pix\_n\_all}}^2$=0.013, $r_{\text{LOOCV-pix\_n\_bouma}}^2$=0.015).

Tables

Parameter estimates of Linear Mixed Effects Models (LMMs)

Table 2. Estimates from the linear mixed-effects model of the VCrowd task, with the Gestalt principles as predictors and individual participants and flanker configurations as random intercepts.

Fixed Effects	β estimate	β standard error	t-value
(Intercept)	1.1685	0.221	5.288
Symmetry - Closure	0.8129	0.2007	4.049
Symmetry - Continuous	0.2913	0.2007	1.451
Symmetry - Random	0.3219	0.2539	1.268
Symmetry - Repetition	0.4807	0.2007	2.395

Table 3. Tukey's HSD posthoc comparison results

Fixed Effects	Upper bound	Lower bound	p-value	signicance
closure - symmetry	0.267	1.359	<0.001	***
continuous - symmetry	-0.255	0.837	0.59	
random - symmetry	-0.369	1.012	0.707	
repetition - symmetry	-0.065	1.027	0.114	
continuous - closure	-1.119	0.076	0.121	
random - closure	-1.223	0.241	0.356	
repetition - closure	-0.930	0.266	0.551	
random - continuous	-0.702	0.763	1	
repetition - continuous	-0.409	0.787	0.909	
repetition - random	-0.573	0.891	0.976	

Table 4. Estimates from the linear mixed-effects model of subjective groupind and segmentation measures. Gestalt principles are considered as predictors and individual participants and flanker configurations as random intercepts.

GlobRank

Fixed Effects	β estimate	β standard error	t-value
(Intercept)	15.325	1.254	12.225
Symmetry - Closure	8.775	1.982	4.427
Symmetry - Continuous	6.187	1.982	3.122
Symmetry - Random	6.675	2.507	2.662
Symmetry - Repetition	7.575	1.982	3.822

VStandRate

Fixed Effects	β estimate	β standard error	t-value
(Intercept)	3.4	0.1776	19.141
Symmetry - Closure	-0.2813	0.1097	-2.565

Symmetry - Continuous	-0.1844	0.1097	-1.681
Symmetry - Random	-0.175	0.1387	-1.262
Symmetry - Repetition	-0.2313	0.1097	-2.109

GStandRate

Fixed Effects	β estimate	β standard error	t-value
(Intercept)	4.1625	0.1635	25.454
Symmetry - Closure	-0.4375	0.1231	-3.553
Symmetry - Continuous	-0.3	0.1231	-2.437
Symmetry - Random	-0.6625	0.1557	-4.254
Symmetry - Repetition	-0.2156	0.1231	-1.751

GGroupRate

Fixed Effects	β estimate	β standard error	t-value
(Intercept)	3.9625	0.1681	23.576
Symmetry - Closure	-0.5031	0.1795	-2.803
Symmetry - Continuous	-0.325	0.1795	-1.811
Symmetry - Random	-0.8313	0.227	-3.662
Symmetry - Repetition	-0.4875	0.1795	-2.716

Table 5. Tukey's HSD posthoc comparison results

GlobRank

Fixed Effects	Upper bound	Lower bound	p-value	significance
closure - symmetry	3.383	14.167	< 0.001	***
continuous - symmetry	0.795	11.580	0.015	*
random - symmetry	-0.146	13.496	0.058	
repetition - symmetry	2.183	12.967	0.001	**
continuous - closure	-8.494	3.319	0.753	
random - closure	-9.334	5.134	0.933	
repetition - closure	-7.107	4.707	0.981	
random - continuous	-6.747	7.722	1.000	
repetition - continuous	-4.519	7.294	0.968	
repetition - random	-6.334	8.134	0.997	

VStandRate

Fixed Effects	Upper bound	Lower bound	p-value	significance
closure - symmetry	-0.580	0.017	0.076	
continuous - symmetry	-0.483	0.114	0.441	
random - symmetry	-0.552	0.202	0.711	

repetition - symmetry	-0.530	0.067	0.213	
continuous - closure	-0.230	0.424	0.928	
random - closure	-0.294	0.506	0.951	
repetition - closure	-0.277	0.377	0.994	
random - continuous	-0.391	0.410	1.000	
repetition - continuous	-0.374	0.280	0.995	
repetition - random	-0.456	0.344	0.995	

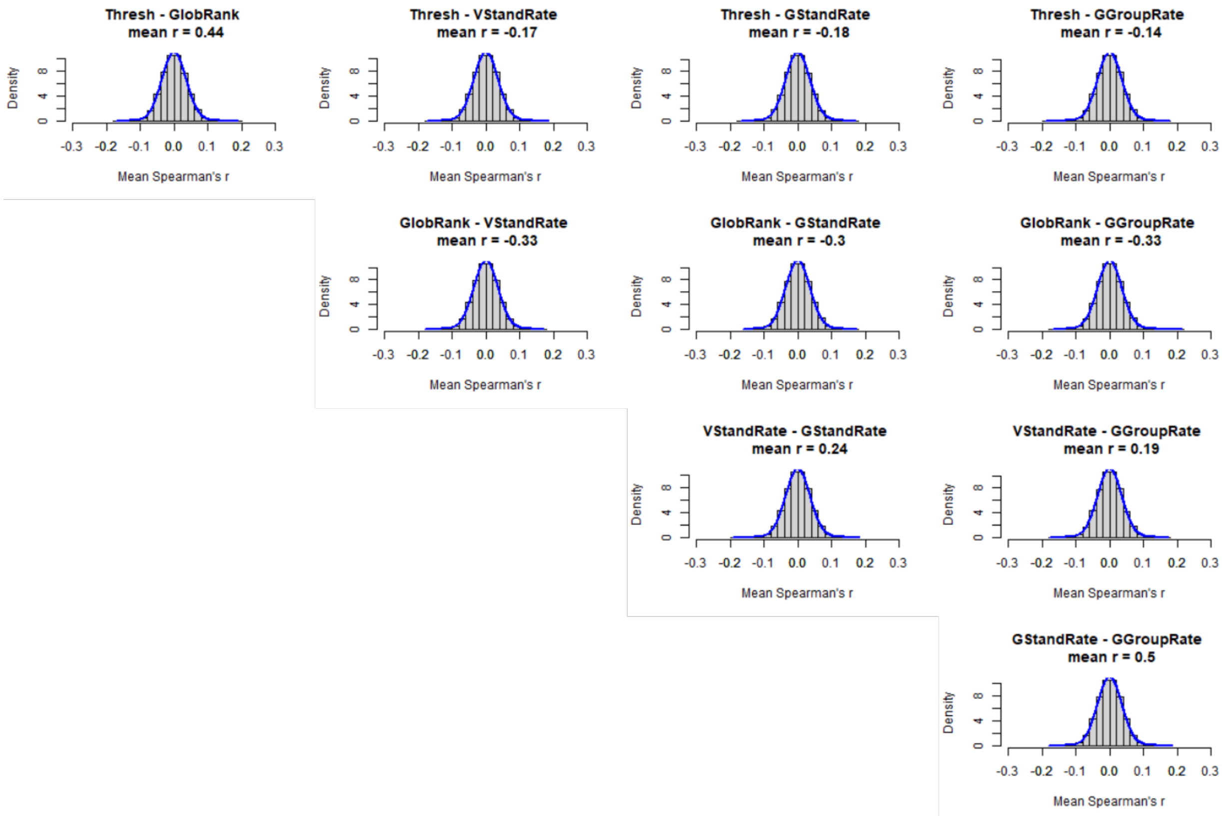
GStandRate

Fixed Effects	Upper bound	Lower bound	p-value	significance
closure - symmetry	-0.772	-0.103	0.003	**
continuous - symmetry	-0.635	0.035	0.104	
random - symmetry	-1.086	-0.239	< 0.001	***
repetition - symmetry	-0.551	0.119	0.398	
continuous - closure	-0.229	0.504	0.844	
random - closure	-0.674	0.224	0.648	
repetition - closure	-0.145	0.589	0.464	
random - continuous	-0.812	0.087	0.179	
repetition - continuous	-0.282	0.451	0.970	
repetition - random	-0.002	0.896	0.052	.

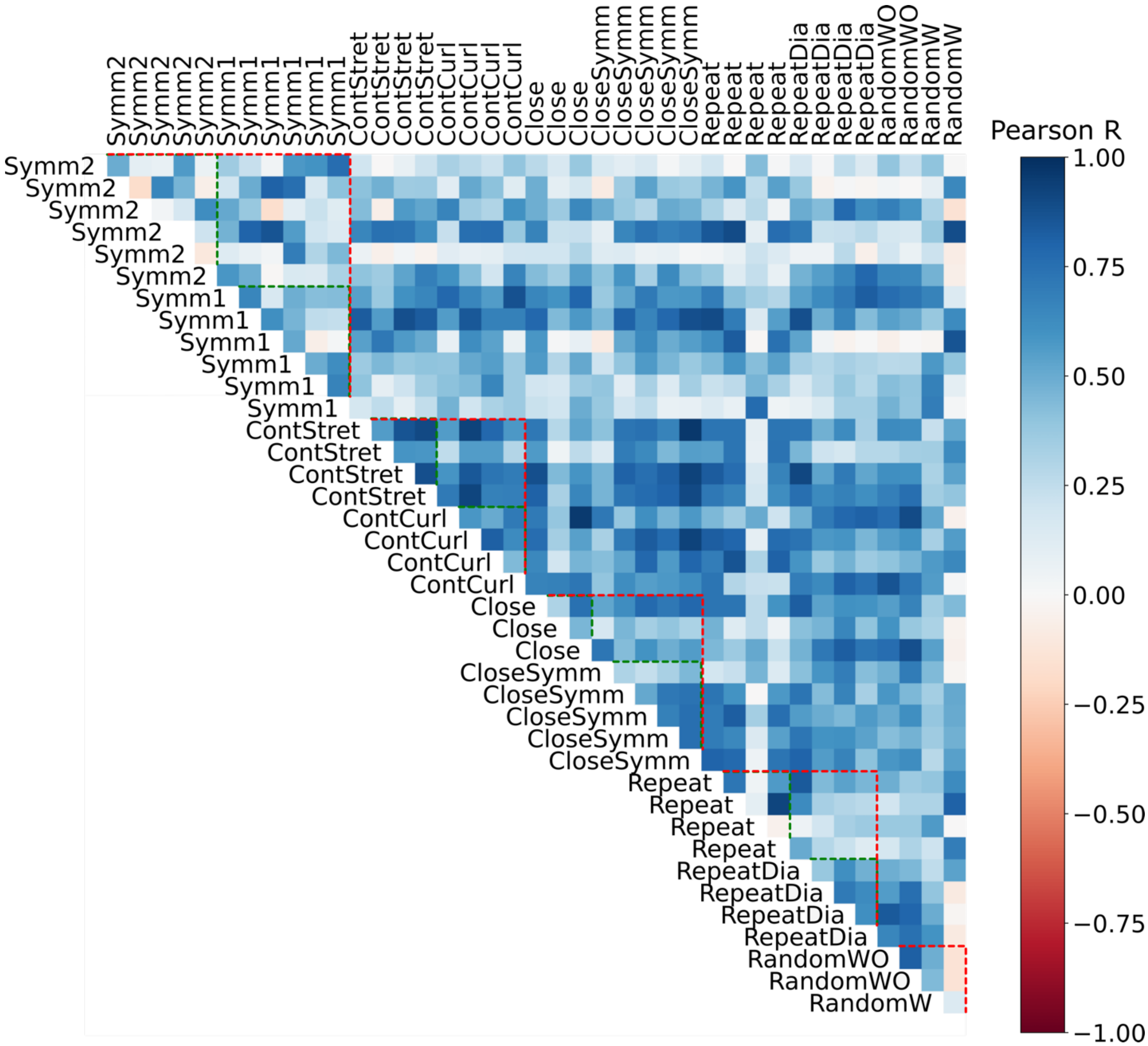
GGroupRate

Fixed Effects	Upper bound	Lower bound	p-value	significance
closure - symmetry	-0.991	-0.015	0.040	*
continuous - symmetry	-0.813	0.163	0.363	
random - symmetry	-1.449	-0.214	0.002	**
repetition - symmetry	-0.976	0.001	0.051	
continuous - closure	-0.356	0.713	0.893	
random - closure	-0.983	0.327	0.647	
repetition - closure	-0.519	0.550	1.000	
random - continuous	-1.161	0.149	0.215	
repetition - continuous	-0.697	0.372	0.921	
repetition - random	-0.311	0.999	0.605	

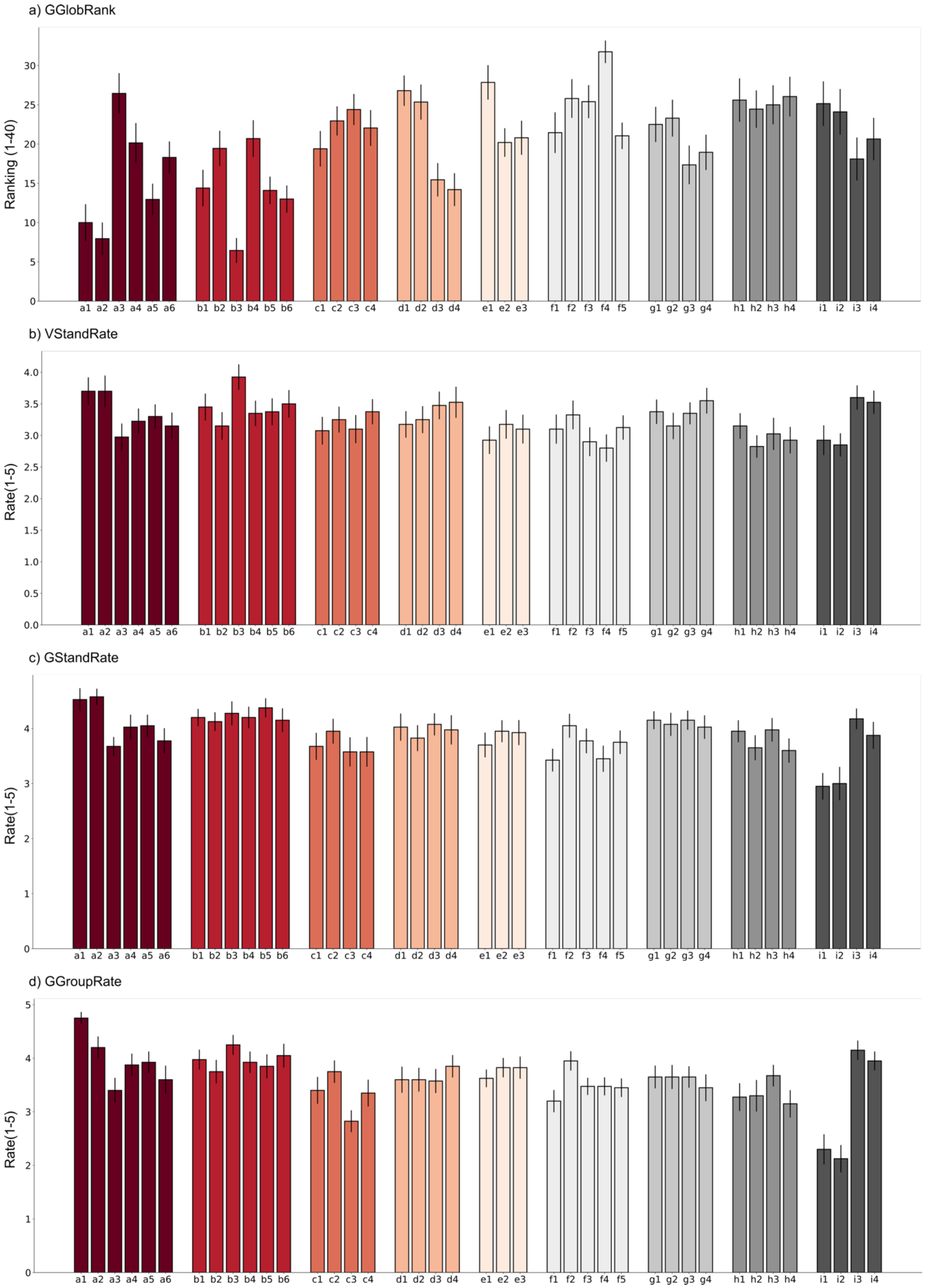
Figures



Supp. Figure 2. Spearman's Rank coefficient randomization test histogram.



Supp. Figure 3. Pearson correlation r coefficients. Dotted red lines show the boundary for the same Gestalt principle, and dotted green lines show the boundary for the same nuanced Gestalt principle. The correlations between the same principles were not higher than the correlations across the principles.



Supp. Figure 4. a) GGlobRank, b) VStandRate, c) GStandRate, and d) GGroupRate for each configuration. Each color represents each Gestalt principle, as in Fig. 4.