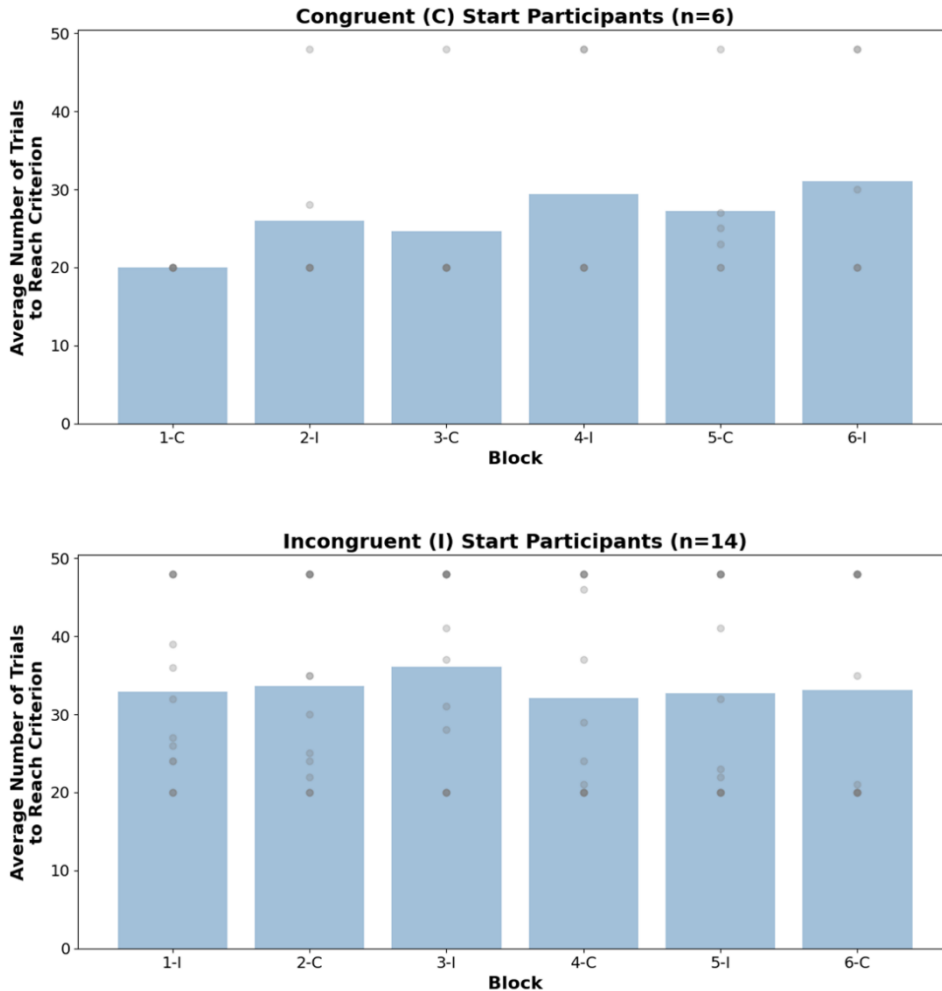




1

2 **Supplemental Fig 1. On average, 31 trials are required to reach 75% accuracy within each**

3 **block.** Pilot data from online participants (n=20) in which participants had to achieve 75%

4 accuracy within the last 10 trials of each block (after a minimum of 20 trials) before a block

5 reversal occurred. The average number of trials required by participants to reach this learning

6 criterion for a block reversal is shown for each block. The Congruent-start group (starting in the

7 Congruent block; n=6) is shown in the top panel, whereas the Incongruent-start group (starting in

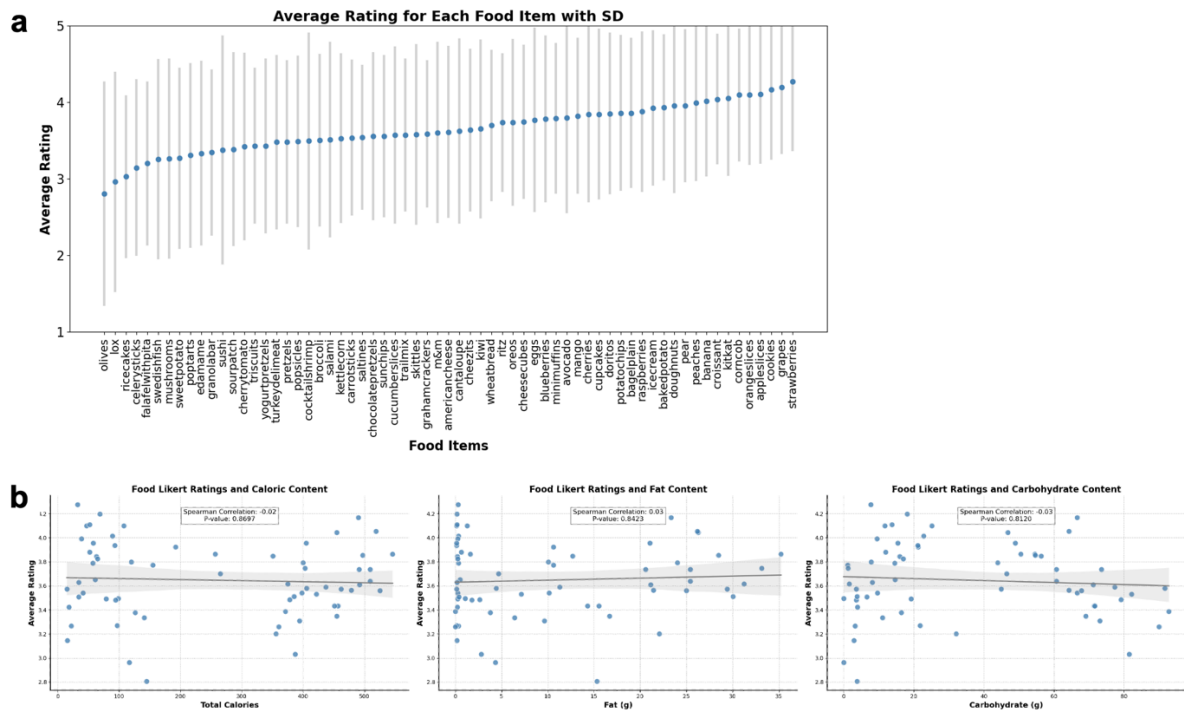
8 the Incongruent block; n=14) is shown in the bottom panel. Data is further categorized by block

9 type (Congruent and Incongruent). For Congruent-start participants, 26.36 ± 3.86 trials were

10 required on average, across blocks of both types, to achieve learning criterion for a reversal; for

11 Incongruent-start participants, 33.42 ± 1.40 trials were required, across blocks of both types. For
12 all participants, regardless of starting group, the average number of trials required to meet
13 learning criterion for a reversal was 31.30 ± 1.31 .

14



15

16 **Supplemental Fig 2. Food item ratings are heterogeneous and not correlated with caloric or**

17 **nutritional content. a** Distribution of Likert ratings for the 65 food stimuli, obtained from the

18 full recruited sample (n=415). Likert values are coded as follows: 1=strongly disliked,

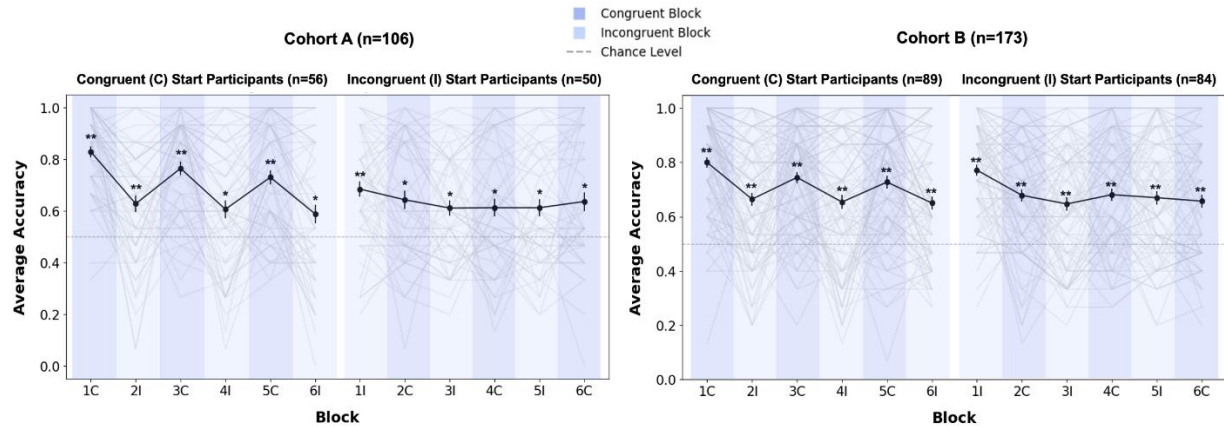
19 2=disliked, 3=neutral, 4=like, 5=strongly like. Values are means $\pm$ s.d., as indicated by the

20 surrounding light grey bars. **b** Spearman correlations among average food item rating and

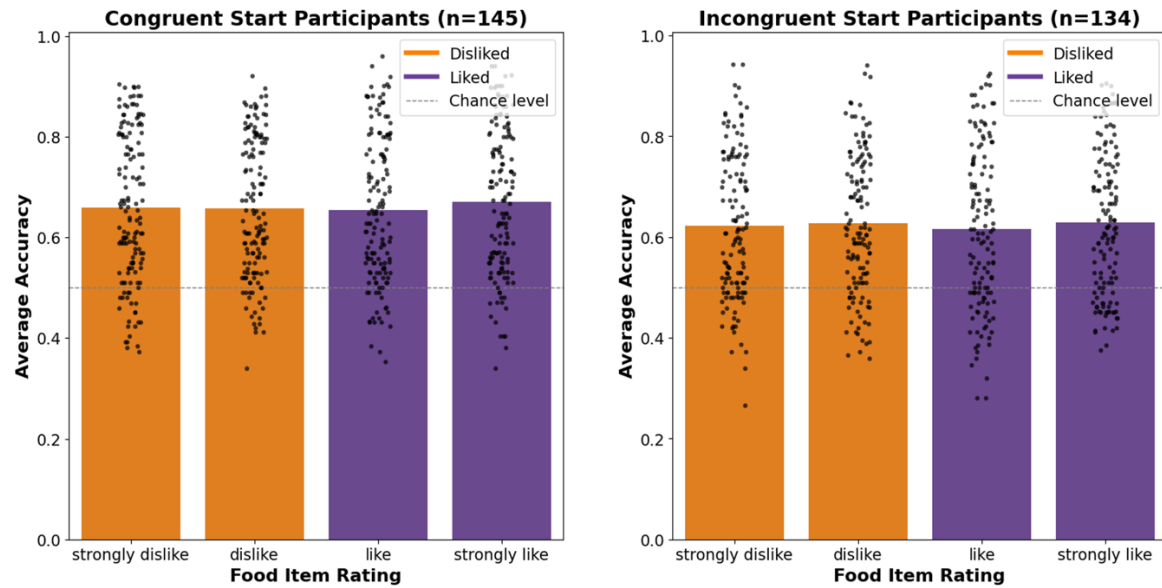
21 caloric content (left panel), fat content (in grams; middle panel), and carbohydrate content (in

22 grams; right panel). Each data point represents a single food item from the 65-item stimuli set.

23

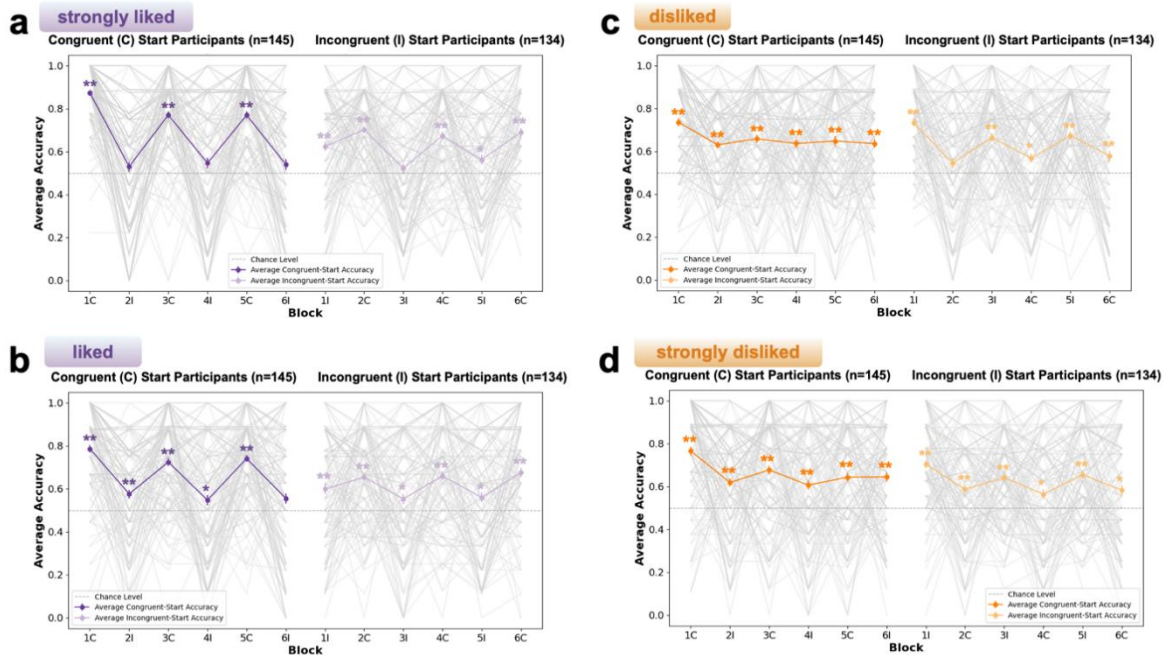


Supplemental Fig 3. Average accuracy of last 15 trials per block, by start group and block type, over time. Data were collected at two separate time points, indicated here as Cohort A (n=106) and Cohort B (n=173). Accuracy data are shown block-wise, as a function of block type (Congruent and Incongruent) and the condition participants started in (Congruent-start and Incongruent-start). Values are means $\pm$ s.e.m. A Wilcoxon signed-rank test was used to determine if the average accuracy of the last 15 trials in each block was significantly above chance level. End-of-block performance significantly greater than chance level is denoted by ** ($p < 0.001$) and * ($p < 0.05$). The grey dashed line at $y = 0.5$ denotes the chance level prediction accuracy.



Supplemental Fig 4. Average accuracy does not differ as a function of food image rating

alone. Accuracy data are categorized by food item rating. The Congruent-start group (starting in the Congruent block; n=145) is shown in the left panel, whereas the Incongruent-start group (starting in the Incongruent block; n=134) is shown in the right panel. P-values displayed above the bars indicate the significance of the difference between food items of different ratings within each block, as determined by the Mann-Whitney U test. The grey dashed line at $y=0.5$ denotes the chance level prediction accuracy.



43

44 **Supplemental Fig 5. Block-wise average accuracy by food item rating, start group, and**

45 **block type.** Average prediction accuracy is shown block-wise by collapsed food item rating, for

46 (a) strongly liked, (b) liked, (c) disliked, and (d) strongly disliked food items. Accuracy data are

47 additionally categorized by block type (Congruent and Incongruent) and the condition

48 participants started in (Congruent-start and Incongruent-start). Within each block, each food item

49 (strongly liked, liked, disliked, and strongly disliked) was shown on approximately 8-9 trials.

50 Values are means $\pm$ s.e.m. A Wilcoxon signed-rank test was used to determine if the accuracy in

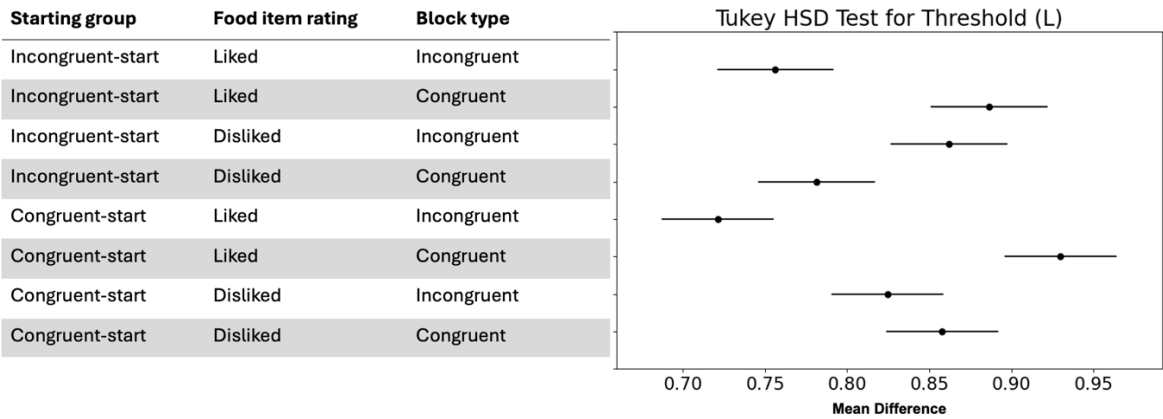
51 each block was significantly above chance level. Block performance significantly greater than

52 chance level is denoted by ** ($p < 0.001$) and * ($p < 0.05$). The grey dashed line at $y = 0.5$ denotes

53 the chance level prediction accuracy. Grey solid lines represent average accuracy for individual

54 participants.

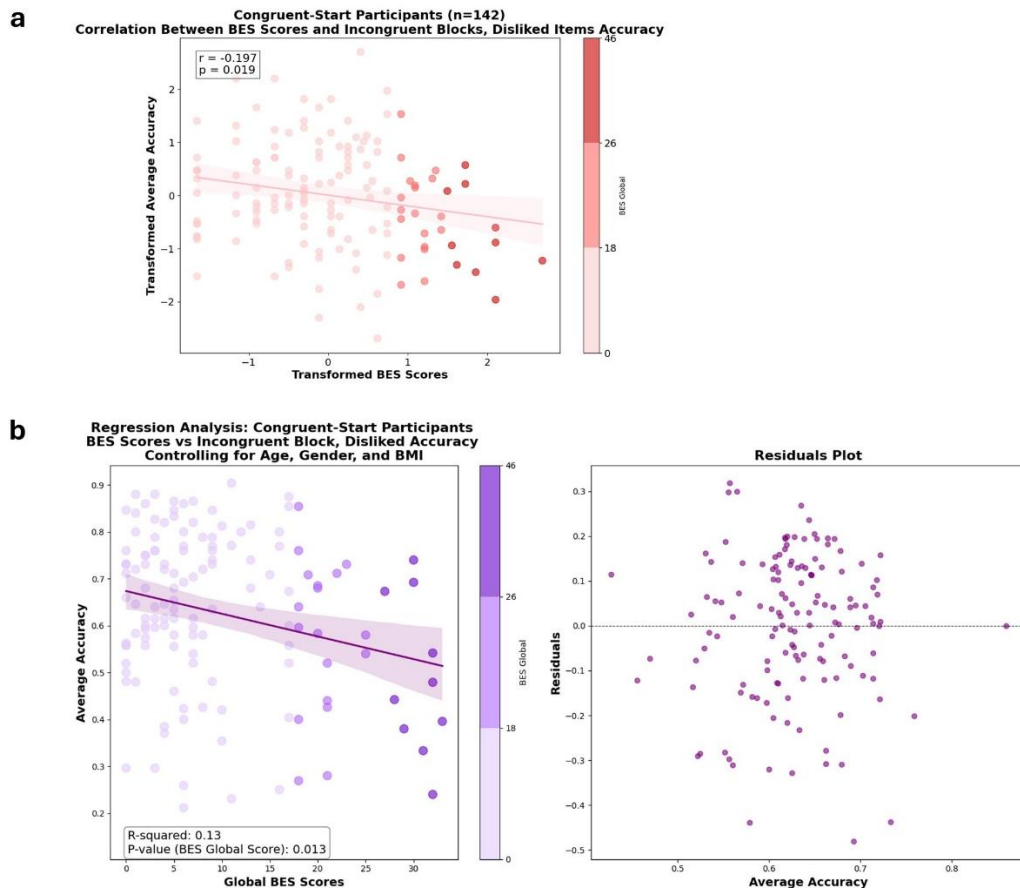
55



56

57 **Supplemental Fig 6. Sigmoid function fit to within-block average accuracy by food item**
 58 **rating, start group, and block type.** An ANOVA was performed to investigate how model-
 59 derived thresholds (Ls) vary across task conditions. Tukey's HSD post hoc test results are
 60 displayed to showcase comparisons of threshold values across conditions. Values are mean
 61 differences $\pm$ 95% CI.

62



Supplemental Fig 7. For Congruent-start participants, binge eating severity is negatively associated with accuracy for Disliked items on Incongruent blocks. a Accuracy data of Congruent-start participants with survey data (n=142) on Incongruent blocks for Disliked items and Binge Eating Scale (BES) Global Scores underwent rank-based inverse normal transformation, followed by Pearson correlation. BES Global Scores are color-categorized according to severity, with increasingly dark shades of the color scale indicating Low, Mild to Moderate, and Severe binge eating, respectively. **b** Linear regression analysis examining the relationship between BES Global Scores and accuracy on Incongruent blocks with Disliked items among Congruent-start participants. The left panel shows a scatter plot of BES Global Scores against average accuracy, with a regression line, R-squared value, and p-value for the

74 BES Global Score coefficient indicating the relationship while controlling for age, gender, and
75 BMI (kg/m^2). BES Global Scores are color-categorized according to severity, with increasingly
76 dark shades of the color scale indicating Low, Mild to Moderate, and Severe binge eating,
77 respectively. The right panel displays a residuals plot.

78

79

Cohort	Start Group	Block	Wilcoxon signed-rank statistic (W)	p -value	Rank- biserial correlation (r)	95% CI for r
A	Congruent- start	1-C	1589	<0.001	0.991	[0.966, 1.000]
A	Congruent- start	2-I	1159.5	0.002	0.453	[0.174, 0.706]
A	Congruent- start	3-C	1401	<0.001	0.958	[0.880, 0.997]
A	Congruent- start	4-I	940.5	0.023	0.314	[0.016, 0.592]
A	Congruent- start	5-C	1371	<0.001	0.916	[0.801, 0.979]
A	Congruent- start	6-I	901.5	0.005	0.414	[0.114, 0.679]
A	Incongruent- start	1-I	904	<0.001	0.673	[0.423, 0.873]
A	Incongruent- start	2-C	850.5	<0.001	0.574	[0.280, 0.790]
A	Incongruent- start	3-I	747.5	0.002	0.510	[0.215, 0.758]
A	Incongruent- start	4-C	872.5	0.005	0.424	[0.120, 0.672]
A	Incongruent- start	5-I	799.5	<0.001	0.545	[0.249, 0.786]
A	Incongruent- start	6-C	962	<0.001	0.571	[0.301, 0.781]
B	Congruent- start	1-C	3888.5	<0.001	0.986	[0.961, 0.998]
B	Congruent- start	2-I	2859.5	<0.001	0.565	[0.360, 0.743]
B	Congruent- start	3-C	3715.5	<0.001	0.941	[0.874, 0.984]
B	Congruent- start	4-I	2702	<0.001	0.514	[0.298, 0.700]

B	Congruent-start	5-C	3314	<0.001	0.901	[0.807, 0.966]
B	Congruent-start	6-I	2762.5	<0.001	0.548	[0.333, 0.723]
B	Incongruent-start	1-I	3199	<0.001	0.927	[0.849, 0.982]
B	Incongruent-start	2-C	2508.5	<0.001	0.760	[0.601, 0.891]
B	Incongruent-start	3-I	2687	<0.001	0.618	[0.426, 0.777]
B	Incongruent-start	4-C	2636.5	<0.001	0.756	[0.590, 0.877]
B	Incongruent-start	5-I	2401	<0.001	0.641	[0.442, 0.798]
B	Incongruent-start	6-C	2588.5	<0.001	0.724	[0.555, 0.853]

Supplementary Table 1. Block-wise prediction accuracy relative to chance by cohort, block

type, and start group. This table reports the results of one-sided Wilcoxon signed-rank tests shown in Fig. 2a assessing whether mean prediction accuracy in each block was significantly above chance level (0.5; dashed line in Fig. 2a). Tests are presented separately for Cohorts A (n=106) and B (n=173), and are further stratified by block type (Congruent, Incongruent) and by participants' starting condition (Congruent-start, $n_A=56$, $n_B=89$; Incongruent-start, $n_A=50$, $n_B=84$). For each block $\times$ start group $\times$ cohort combination, the Wilcoxon signed-rank statistic (W), p -value, rank-biserial correlation (r), and the 95% confidence interval for r are reported.

Food Item Rating	Start Group	Block	Wilcoxon signed-rank statistic (W)	p -value	Rank-biserial correlation (r)	95% CI for r
Liked	Congruent-start	1-C	9865.5	<0.001	0.999	[0.996, 1.000]
Liked	Congruent-start	2-I	5475.5	0.002	0.286	[0.104, 0.478]
Liked	Congruent-start	3-C	9367	<0.001	0.953	[0.876, 0.995]
Liked	Congruent-start	4-I	5143.5	0.013	0.227	[0.027, 0.413]
Liked	Congruent-start	5-C	9542.5	<0.001	0.961	[0.911, 0.992]
Liked	Congruent-start	6-I	5423.5	0.023	0.199	[0.016, 0.379]
Liked	Incongruent-start	1-I	6198.5	<0.001	0.549	[0.394, 0.704]
Liked	Incongruent-start	2-C	7031	<0.001	0.874	[0.784, 0.936]
Liked	Incongruent-start	3-I	4969	0.023	0.204	[0.022, 0.389]
Liked	Incongruent-start	4-C	6817.5	<0.001	0.817	[0.701, 0.912]
Liked	Incongruent-start	5-I	4929	0.002	0.293	[0.105, 0.480]
Liked	Incongruent-start	6-C	7422	<0.001	0.826	[0.714, 0.915]
Disliked	Congruent-start	1-C	9422	<0.001	0.909	[0.844, 0.961]
Disliked	Congruent-start	2-I	7306.5	<0.001	0.665	[0.503, 0.792]
Disliked	Congruent-start	3-C	7758.5	<0.001	0.741	[0.625, 0.856]
Disliked	Congruent-start	4-I	7003	<0.001	0.596	[0.440, 0.737]

Disliked	Congruent-start	5-C	7828.5	<0.001	0.632	[0.485, 0.759]
Disliked	Congruent-start	6-I	7230	<0.001	0.672	[0.537, 0.791]
Disliked	Incongruent-start	1-I	7579.5	<0.001	0.956	[0.902, 0.992]
Disliked	Incongruent-start	2-C	4889.5	<0.001	0.347	[0.154, 0.539]
Disliked	Incongruent-start	3-I	6841.5	<0.001	0.824	[0.726, 0.908]
Disliked	Incongruent-start	4-C	5076	<0.001	0.331	[0.122, 0.513]
Disliked	Incongruent-start	5-I	6570	<0.001	0.81	[0.691, 0.894]
Disliked	Incongruent-start	6-C	5271	<0.001	0.382	[0.173, 0.572]

Supplementary Table 2. Block-wise prediction accuracy relative to chance by rating valence, block type, and start group. This table reports the results of one-sided Wilcoxon signed-rank tests shown in Fig. 5 assessing whether mean prediction accuracy in each block was significantly above chance level (0.5; dashed line in Fig. 5). Tests are presented separately for food item rating valence, (a) Liked items (liked and strongly liked) and (b) Disliked items (disliked and strongly disliked), and are further stratified by block type (Congruent, Incongruent) and by participants' starting condition (Congruent-start, n=145; Incongruent-start, n=134). For each block $\times$ start group $\times$ valence combination, the Wilcoxon signed-rank statistic (W), p -value, rank-biserial correlation (r), and the 95% confidence interval for r are reported.