

**Contour junctions underlie neural representations of scene categories  
in high-level human visual cortex**

Contour junctions underlie neural code of scenes

Heeyoung Choo<sup>1</sup>, & Dirk. B. Walther<sup>1</sup>

<sup>1</sup>Department of Psychology, University of Toronto

100 St. George Street, Toronto, ON, M5S 3G3 Canada

Correspondence should be addressed to Heeyoung Choo,

100 St. George Street, Toronto, ON, M5S 3G3 Canada.

Phone: +1-416-946-3200

Email: [heeyoung.choo@utoronto.ca](mailto:heeyoung.choo@utoronto.ca)

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

1 **Abstract**

2

3 Humans efficiently grasp complex visual environments, making highly consistent judgments of entry-  
4 level category despite their high variability in visual appearance. How does the human brain arrive at the  
5 invariant neural representations underlying categorization of real-world environments? We here show that  
6 the neural representation of visual environments in scenes-selective human visual cortex relies on statistics  
7 of contour junctions, which provide cues for the three-dimensional arrangement of surfaces in a scene. We  
8 manipulated line drawings of real-world environments such that statistics of contour orientations or  
9 junctions were disrupted. Manipulated and intact line drawings were presented to participants in an fMRI  
10 experiment. Scene categories were decoded from neural activity patterns in the parahippocampal place  
11 area (PPA), the occipital place area (OPA) and other visual brain regions. Disruption of junctions but not  
12 orientations led to a drastic decrease in decoding accuracy in the PPA and OPA, indicating the reliance of  
13 these areas on intact junction statistics. Accuracy of decoding from early visual cortex, on the other hand,  
14 was unaffected by either image manipulation. We further show that the correlation of error patterns  
15 between decoding from the scene-selective brain areas and behavioral experiments is contingent on intact  
16 contour junctions. Finally, a searchlight analysis exposes the reliance of visually active brain regions on  
17 different sets of contour properties. Statistics of contour length and curvature dominate neural  
18 representations of scene categories in early visual areas and contour junctions in high-level scene-selective  
19 brain regions.

20

21 **Keywords:** Neural representations of scenes; encoding of scene structure; contour junctions; scene  
22 categorization; parahippocampal place area; occipital place area; visual cortex; fMRI; multi-voxel pattern  
23 analysis

## 24 **1. Introduction**

25 When humans view their complex natural environment, they have rapid access to several aspects of its  
26 content, such as identity and category of the scene, presence of particular object categories, or global  
27 layout (Potter & Levy, 1969; Thorpe, Fize, & Marlot, 1996; Fei-Fei, Iyer, Koch, & Perona, 2007; Greene  
28 & Oliva, 2009a, 2009b, 2010). We here investigate the neural mechanisms for the detection of cues to the  
29 three-dimensional structure of complex real-world scenes in human visual cortex: We show that the neural  
30 representation of scene categories in several high-level visual brain regions, but not in early visual cortex,  
31 critically depends on contour junctions.

32 Entry-level category of a scene is a central aspect of the human visual perception of real-world  
33 environment (Tversky & Hemenway 1983). Recent evidence suggests that humans compulsively  
34 categorize scenes even if it is detrimental to their task (Greene & Fei-Fei, 2014). Scene categories can be  
35 decoded from the parahippocampal place area (PPA) of humans passively viewing scene images (Walther,  
36 Caddigan, Fei-Fei, & Beck, 2009; Park, Bradly, Greene, & Oliva, 2011; Walther, Chai, Caddigan, Fei-Fei,  
37 & Beck, 2011). Moreover, error patterns for category decoding from the PPA match the pattern of human  
38 errors during a rapid scene categorization task (Walther et al., 2009).

39 There has been considerable debate over the visual properties that underlie the neural  
40 representation of scene categories. According to one popular hypothesis, statistics of orientations at  
41 different scales as captured by the Fourier amplitude spectrum make accurate computational predictions  
42 about entry-level categories of real-world scene images. Subsequent principal component analysis  
43 revealed that several diagnostic structures in the Fourier amplitude spectrum are directly related to global  
44 scene properties, such as openness, naturalness, or mean distance (Oliva & Torralba, 2001; Torralba &  
45 Oliva, 2003). These global properties, in turn, are thought to give rise to a representation of scene  
46 categories (Greene & Oliva, 2009a, 2009b, 2010). For instance, the “beach” category is represented as an  
47 open natural environment whereas the highway category is represented as an open and man-made  
48 environment. In support of this hypothesis, several global scene properties have been found to be  
49 represented in activity patterns in the PPA (Harel, Kravitz, & Baker, 2013; Kravitz, Peng, & Baker, 2011;  
50 Park et al., 2011, Park, Konkle, & Oliva, 2015).

51           Here, we posit that real-world scene categories are built by recovering the three-dimensional  
52 shape of the visual world rather than relying on orientation statistics. The visual world can be described  
53 by the relations of surfaces and shapes in space (e.g., by the 2½-dimensional sketch of a scene; Marr,  
54 1982). Contour junctions available in the two-dimensional scene images diagnostically describe these  
55 spatial relations. For instance, L-junctions indicates points of termination of surfaces, T-junctions signify  
56 occlusion in depth, and Y- and arrow-junctions indicate corners facing toward or away from the viewer.  
57 Angles of contour junctions indicate the extent to which depth changes over surfaces (Biederman, 1987).  
58 The diagnostic value of contour junction properties holds for simple artificial scenes consisting of  
59 geometric objects (Guzman, 1968) as well as for real-world object recognition (Biederman, 1987).  
60 Furthermore, a computational model based on category-specific statistics of contour junction properties  
61 explained human errors in rapid categorization of real-world scene images (Walther & Shen, 2014).  
62 According to this structural representation hypothesis, contour junctions should be tied to the neural  
63 representation of complex real-world scenes.

64           Line drawings are a powerful tool to investigate scene recognition, even though they are  
65 impoverished depictions of scenes, compared to full-textured color photographs. In fact, line drawings can  
66 be categorized or recognized as quickly and accurately as full-textured color photographs (Biederman &  
67 Ju, 1988). Line drawings contain sufficient visual information to allow humans to rapidly judge perceptual  
68 and semantic aspects of scenes (Biederman, Mezzanotte, & Rabinowitz, 1982; Biederman, Teitelbaum, &  
69 Mezzanotte, 1983; Kim & Biederman, 2010; 2011). In addition to resulting in similar behavioral error  
70 patterns (Walther and Shen, 2014), color photographs and line drawings of natural scenes also elicit  
71 similar neural representations of scene categories in the PPA (Walther et al., 2011). More importantly,  
72 line drawings provide explicit descriptions of several informative contour properties not readily accessible  
73 in full-textured color photographs, such as contour orientation, length, curvature, and types and angles of  
74 junctions created by multiple contours (Walther & Shen, 2014). The current study benefits from this  
75 direct access to important contour properties by manipulating predictability of either orientation or  
76 junction statistics for scene categories.

77           We tested the causal role of these two sets of candidate features, orientation statistics and  
78 junction properties, for the neural representation of scene categories in the human brain. Scene categories

79 were decoded from the fMRI activity of participants, who passively viewed blocks of line drawings of six  
80 scene categories: beaches, forests, mountains, city streets, highways, and offices. A block design was  
81 employed for its robust signal and for its proven capability to detect category-specific signals in common  
82 to multiple stimuli in a block (Cox & Savoy, 2003; Epstein & Kanwisher 1998; Haxby, Gobbini, Furey,  
83 Ishai, Schouten, & Pietrini, 2001; Kim & Biederman, 2010; Park et al., 2011; Walther et al., 2009). We  
84 devised two image manipulations that allowed for selective disruption of orientation or junction statistics:  
85 One is to rotate line drawings by random angles, which selectively disrupts orientation statistics. The other  
86 is to shift randomly contours of line drawings, which disrupts contour junctions. We then attempted to  
87 decode scene categories from the brain activity of participants while they viewed these manipulated  
88 images. Comparing the results to decoding scene categories from intact images allowed us to assess the  
89 causal involvement of the respective scene properties in the representation of scene categories. Note that  
90 these manipulations *alter* category-specific statistics of the targeted property to be spurious and  
91 uninformative. Although deletion of junctions could be a direct manipulation, pixel removal around  
92 junction locations inevitably affects non-targeted contour properties, such as statistics of orientation,  
93 curvature and length of contours. Our image manipulations, therefore, were to ensure that only one of the  
94 two candidate properties was disrupted by each of the manipulations.

95 We found that the category representation in two high-level visual areas involved in scene  
96 processing, the PPA, and to some extent, the occipital place area (OPA) and the lateral occipital complex  
97 (LOC) relies heavily on junction properties. It was not possible to decode scene categories from these  
98 areas when junctions were disrupted. Disrupting orientation statistics, on the other hand, did not affect the  
99 representation of scene categories. By contrast, scene categories could be decoded from neural activity  
100 patterns in early visual cortex well above chance for images with disrupted junction or orientation  
101 statistics just as from intact images. We further found that correlation of decoding error patterns from the  
102 PPA with behavioral error patterns was contingent on the preservation of junction properties, whereas  
103 disrupting orientations had no effect on error correlations. Finally, we mapped the reliance of the neural  
104 representations of scene categories on several visual properties throughout visual cortex by matching  
105 patterns of neural decoding errors to error patterns from five computational models of scene categorization  
106 from Walther and Shen (2014).

107

## 108 **2. Materials and Methods**

109

110 **2.1 Participants** Sixteen healthy participants (5 females; mean age = 21.6, Standard Deviation (*SD*) =  
111 2.8; one left-handed) were recruited from The Ohio State University community for the functional  
112 magnetic resonance imaging (fMRI) experiment, for which they received monetary compensation of \$15  
113 per hour. A separate group of 49 undergraduates at the Ohio State University participated in the behavioral  
114 experiment for course credit. Participants gave written informed consent. Both experiments were approved  
115 by the institutional review board of The Ohio State University. All participants had normal or corrected-  
116 to-normal vision and normal color vision and reported no history of neurological abnormalities. One  
117 participant (female, right-handed) was excluded from further analysis of the fMRI experiment due to  
118 excessive head movement during scans. Three participants were excluded from the analysis of the  
119 behavioral experiment because they did not complete the entire experiment or did not comply with the  
120 instructions.

121 **2.2 Stimuli & Apparatus** Stimuli consisted of 475 vectorized line drawings of six categories of real-  
122 world scenes (beaches, forests, mountains, city streets, highways, and offices) from Walther et al. (2011).  
123 We generated three versions of each line drawing. Intact line drawings were generated by applying a  
124 circular aperture (Figure 1A). Rotated line drawings were rotated the whole image by an angle randomly  
125 selected from 10 – 340° with 30° increments (Figure 1B). Distributions of contour orientation peaked at 0°  
126 and 90° in most of the six scene categories (Walther & Shen, 2014). Thus, rotation by 90°, 180° and 270°  
127 were deliberately avoided. Contour-shifted line drawings were generated by randomly translating  
128 individual contours within the circular aperture (Figure 1C). This manipulation ensured the disruption of  
129 the relations between contours, represented by contour junctions, while keeping all other contour  
130 properties constant. **Note that both image rotation and random contour-shifting change local contour**  
131 **property statistics. Random image rotation not only systematically alters the original orientation statistics**  
132 **within a local image patch, but it also changes the contour junction statistics within that local patch (i.e.,**  
133 **the contour junction statistics of a different portion of the image will substitute the original local**  
134 **statistics). Similarly, random contour-shifting alters orientation statistics within a local image patch, and**

135 also generates new spurious contour junctions in the patch. The difference between these manipulations is  
136 their effect on *global* image statistics. Rotation preserves the 3D relationships between surfaces and  
137 objects, and contour shifting does not.

138 As a result of these two manipulations, we obtained a total of 475 triplets, each of which  
139 consisted of an *intact*, a *rotated*, and a *contour-shifted* line drawing derived from the same color  
140 photograph of a real-world scene<sup>1</sup>. The same triplets were used for all participants in both fMRI and  
141 behavioral experiment.

142 Both experiments were controlled using Python 2.5 with VisionEgg 1.2 on a PC with Microsoft  
143 Windows XP. Stimuli for the fMRI experiment were back-projected onto a screen mounted in the back of  
144 the scanner bore with a DLP projector (Christie DS+6K-M 3-chip SXGA+) at a resolution of 1280 x 1024  
145 pixels. Participants viewed stimuli through a mirror mounted on the head coil. Line drawings were  
146 rendered as black lines on a white background (2 pixels width) at a resolution of 1023 x 1023 pixels,  
147 which subtended approximately 17° x 17° of visual angle. Line drawings were seen through a circular  
148 aperture of 1023 pixels diameter. The part of the screen outside the circular aperture was 50% gray. A  
149 white fixation cross with a black outline was drawn at the center of the screen and subtended 0.5° x 0.5° of  
150 visual angle. Stimuli for the behavioral experiment were displayed on a CRT monitor with 1024 x 768  
151 pixels resolution and 150 Hz refresh rate. Line drawings were rendered as black lines of 1-pixel width on a  
152 white background at a resolution of 600 x 600 pixels (approximately 18° x 18°), and seen through a  
153 circular aperture of 600 pixels diameter. The fixation cross had a size of 1° x 1°.

154 **2.3. Experiment Design** In the fMRI experiment, participants were asked to attentively view the line  
155 drawings while fixating on the central cross. To ensure that participants followed the instruction, we  
156 monitored participants' eye-movements in real-time using an MR compatible Eyelink 1000 system, but  
157 eye-movements were not recorded. Each participant viewed a total of 384 triplets (64 per scene category)  
158 randomly chosen from the 475 triplets. All participants had eight runs, 6 min and 12 sec in length. Each

---

<sup>1</sup> 2D-fast Fourier transform analysis of the triplets confirmed that our manipulation targeting the orientation statistics held in the Fourier space as well. Random contour-shifting had little impact on the Fourier amplitude spectrum, and the correlation in the Fourier amplitude spectrum were high between intact and rotated line drawings, Fisher's  $z = 3.278$  ( $r = .997$ ). In contrast, the average correlation between the intact and rotated line drawings was relatively low, Fisher's  $z = 1.475$  ( $r = .901$ ). The difference becomes even more pronounced when adjusting for the average correlation between images from different triplets; between intact and contour-shifted:  $z_{adj} = 1.073$  ( $r_{adj} = .791$ ), between intact and rotated:  $z_{adj} = .030$  ( $r_{adj} = .030$ ).

159 run included 18 blocks of all possible combinations of the three line drawing types (intact, rotated, and  
160 contour-shifted), and the six basic scene categories (beaches, forests, mountains, city streets, highways,  
161 and offices). During each block, eight line drawings from the same image type and scene category were  
162 shown for 800 ms, followed by a 200 ms blank per scene image. 12 sec of a blank fixation periods were  
163 inserted between blocks as well as at the beginning and the end of each run. The order of blocks within  
164 runs was counterbalanced across runs and participants according to image type and scene category.

165 In the behavioral experiment images from all six scene categories were randomly interleaved.  
166 Participants were asked to indicate the categories of scene images by pressing one of six keys (s, d, f, j, k,  
167 and l) on a computer keyboard. The mapping between categories and keys was assigned randomly to each  
168 participant. Each trial started with a fixation period of 500 ms, followed by a line drawing for a variable  
169 amount of time (250 ms initially), which was followed by a texture mask for 500 ms and a blank period  
170 for another 2000 ms. The texture mask was derived from a mixture of textures synthesized from all six  
171 scene categories (Loschky, Hansen, Sethi, & Pydimarri, 2010; Portilla & Simoncelli, 2000). Participants'  
172 key responses were recorded from the onset of the image until the end of the blank period. If no response  
173 was made by the end of the blank period, the trial was recorded as incorrect. In the first phase of the  
174 experiment, participants practiced the response mapping until they achieved 90% accuracy. In the  
175 following stair-casing phase the stimulus onset asynchrony (SOA) was adjusted to 65% accuracy using the  
176 QUEST algorithm (Watson & Pelli, 1983) for each participant. **By using the stair-cased SOA and the**  
177 **perceptual masking procedure, we aimed to provoke erroneous responses, so that we could compare the**  
178 **error patterns between behavior and neural decoding reliably.** A randomly selected subset of twelve  
179 triplets of each category was shown for practice and stair-casing, leaving 60-68 triplets per category for  
180 testing. In practice and stair-casing, images were presented in their intact version, and participants were  
181 alerted to their mistakes by a beep.

182 During the testing phase, 60 triplets per category were randomly selected from the unused sets.  
183 Each image was shown only once during testing, either as an intact line drawing, a rotated line drawing, or  
184 a line drawing with randomly contour-shifted contours. 360 trials were grouped into 18 blocks of 20  
185 images. All three line drawing types and six scene categories were presented intermixed within a block.  
186 Participants no longer received feedback during the testing phase of the experiment. The SOA was fixed



187 to the final SOA of the stair-casing procedure. We excluded data from the seven participants with SOAs  
188 exceeding 100 ms from further analysis<sup>2</sup>. As a result, the final SOAs of the remaining 39 participants  
189 ranged from 17 – 87 ms ( $M = 34$  ms,  $SD = 19$  ms). The behavioral data were recorded in three confusion  
190 matrices, one for each image type. The rows of a confusion matrix indicate the scene categories presented  
191 to a participant, and the columns indicate the participant's response. Cells contain the relative frequency of  
192 participants responding with the category indicated by the column, given that the presented image was of  
193 the category indicated by the row. Thus, diagonal entries contain correct responses of a scene  
194 categorization task, and off-diagonal entries contain errors in scene categorization.

195 **2.4. fMRI Data Acquisition and Preprocessing** MRI images were recorded on a 3 Tesla Siemens  
196 MAGNETOM Trio MRI scanner with a 12-channel head coil at the Center for Cognitive and Behavioral  
197 Brain Imaging (CCBBI) at The Ohio State University. High-resolution anatomical images were obtained  
198 with a 3D-MPRAGE (magnetization-prepared rapid acquisition with gradient echo) sequence with sagittal  
199 slices covering the whole brain; inversion time = 930 ms, repetition time (TR) = 1900 ms, echo time (TE)  
200 = 4.44 ms, flip angle = 9°, voxel size = 1 x 1 x 1 mm, matrix size = 224 x 256 x 160 mm. Functional  
201 images were recorded with T2\*-weighted echo-planar sequences with coronal slices, covering  
202 approximately the posterior 70% of the brain: for the main experiment, TR = 2000 ms, TE = 28 ms, flip  
203 angle = 72°, voxel size = 2.5 x 2.5 x 2.5 mm, matrix size = 90 x 100 x 35 mm. fMRI data were registered  
204 to a reference volume (the first volume of the fourth run) using AFNI to correct for head motion during  
205 the experiment. Then, fMRI data were smoothed using a 2 mm full-width-at-half-maximum (FWHM)  
206 Gaussian filter and converted to percentage signal change with respect to the mean of each run.

## 207 **2.5. ROI-Based Neural Decoding**

208 **2.5.1. Decoding Accuracy** As a preprocessing step for neural decoding we regressed out nuisance  
209 parameters using a general linear model (GLM) with regressors only for head motion and scanner drift.  
210 The residuals of the GLM analysis were averaged over the durations of individual blocks, subject to a

---

<sup>2</sup> Including these participants to the analysis did not change the pattern of results. However, as they acquire experience on the task, their relatively long presentation time reduced erroneous observations. Since our objective of having a brief presentation was to compare error patterns between neural decoding and behavior, we decided a priori not to include participants with SOAs < 100 ms to maximize the variance in error patterns.

211 hemodynamic delay of 4 sec. The resulting 144 brain volumes (one for each block) were used as input for  
212 multi-voxel pattern analysis (MVPA).

213 MVPA was performed within pre-specified ROIs using a linear support vector machine (SVM)  
214 classifier (linear kernel, using LIBSVM, Chang & Lin, 2001). The classifier was trained to associate the  
215 correct category label to the blocks in seven of the eight runs, leaving out one run for testing. The  
216 categories for the blocks in the left-out run were predicted by the trained classifier. This leave-one-run-out  
217 (LORO) cross-validation was repeated until each of the eight runs was left out once. The fraction of  
218 blocks with correct test predictions was recorded as accuracy, and misclassifications were recorded in a  
219 confusion matrix (for example see Figure 2B). To investigate the effect of image rotation and contour-  
220 shifting on the representation of scene categories, LORO cross validation was performed both within  
221 image type (using the same image type for training and testing) and across image types (training on one  
222 and testing on another image type). In each case, accuracy was compared to chance (1/6) at the group  
223 level using one-tailed one-sample t-tests.

224 **2.5.2. Error Pattern Correlation between Brain and Behavior** We measured the similarity of  
225 underlying categorical representations by correlating decoding error patterns from each of the ROIs to  
226 behavioral error patterns from the rapid scene categorization experiment (Walther, Beck, & Fei-Fei,  
227 2012). The vector consisting of the 30 off-diagonal entries of the confusion matrix from an ROI was  
228 correlated to the vector of 30 off-diagonal entries of the confusion matrix from behavioral scene  
229 categorization. Statistical significance of the correlation was established non-parametrically against the  
230 null distribution of all error correlations obtained from jointly permuting rows and columns of the  
231 behavioral confusion matrix.

232 How does disruption of contour orientation or junction properties modulate error correlations  
233 between neural decoding and behavior? Since we had the same three image types (i.e., intact, rotated, and  
234 contour-shifted) for both neural decoding and behavioral rapid scene categorization, we could examine  
235 effects of property disruption in a 3-by-3 pattern of correlations. We modeled these correlation patterns as  
236 a linear combination of three models of the interaction between behavior and patterns of brain activity  
237 (Figure 4C). The first model is straightforward in that error correlations between neural decoding and  
238 behavior for the same image types are correlated (*same type* model). The second model states that error

239 patterns from neural decoding for intact line drawings correlate to error patterns from the behavioral  
240 categorization of rotated line drawings and vice versa. This *intact-rotated* model prediction is based on  
241 the idea that junction properties are necessary for maintaining the error pattern similarities between neural  
242 decoding and behavior, which had been reported using intact line drawings (Walther et al. 2011). The last  
243 model assumes that orientation statistics leads to similarities in error patterns between neural decoding and  
244 behavior. This *intact-contour-shifted* model predicts high error correlation between neural decoding and  
245 behavior between intact and contour-shifted line drawings. The weights belonging to the three models  
246 were obtained by linearly regressing the error correlation patterns for all MRI participants onto the model  
247 regressors shown in Figure 2C.

248 We tested which of the three models explain the relationships in the 3-by-3 error correlation  
249 patterns using a linear mixed-effects model (lme4 package in R, Bates, Maechler, Bolker, & Walker,  
250 2014), which included the three idealized models as fixed effects, and participants as random effects. Prior  
251 to the regression analysis, error pattern correlation values were normalized using Fisher's  $z$   
252 transformation. Thus, the coefficients of the predictors provided estimates of how well each of the three  
253 models predicts the error correlations between neural decoding and behavior.

## 254 **2.6. Searchlight Analysis**

255 **2.6.1. Decoding Accuracy** We explored how the human visual cortex outside of the pre-defined  
256 ROIs represents categorical information about scenes using the Searchlight toolbox (Pereira & Botvinick,  
257 2011). Searchlight analysis was performed with partial coverage in the coronal direction, which was  
258 sufficient to encompass approximately the posterior 70% of the brain on average across the participants.  
259 The same block-averaged data used in the previous ROI-based analysis entered the searchlight analysis.  
260 We defined a cubic "searchlight" of 125 voxels, whose size was matched to the average size of unilateral  
261 PPA across participants (142.9 voxels,  $SD = 66.8$  voxels) as closely as possible. The searchlight was  
262 centered on each voxel at a time (Kriegeskorte, Göbel, & Bandettini, 2006), and LORO cross-validation  
263 analysis was performed within each searchlight location using a Gaussian Naïve Bayes classifier until all  
264 voxels served as the center of the searchlight. Decoding accuracy, as well as the full confusion matrix at a  
265 given searchlight location, were assigned to its central voxel. We performed the searchlight analysis

266 separately for decoding of each of the three line drawing types, resulting in three individual accuracy maps  
267 for each participant.

268 To examine the agreement between the searchlight and the ROI-based analysis, we counted  
269 overlap between searchlight results and the areas VI-4, PPA, OPA, RSC, LOC, and FFA. Accuracy maps  
270 were thresholded at  $p < .005$  (one-tailed p-values were obtained by the analytic methods provided by the  
271 Searchlight toolbox), and cluster-corrected using  $\alpha$  probability simulation separately per each participant.  
272 Overlap with ROIs was computed as a percentage of ROI voxels that were included in the thresholded  
273 accuracy maps.

274 For group-analysis, we first co-registered each participant's anatomical brain to the Montreal  
275 Neurological Institute (MNI) 152 template (Fonov, Evans, McKinstry, Almlil, & Collins, 2011) using a  
276 diffeomorphic transformation as calculated by AFNI's 3dQWarp. We then used the same transformation  
277 parameters to register individual decoding accuracy maps to MNI space using 3dNwarpApply, followed  
278 by spatial smoothing with a 2 mm FWHM Gaussian filter. To identify voxels with decodable categorical  
279 information, we performed one-tailed t-tests to test whether decoding accuracy at each searchlight location  
280 was above chance (1/6). After thresholding at  $p < .005$  (one-tailed) we conducted a cluster-level correction  
281 for multiple comparisons, applying a minimum cluster size of 12 voxels, the average cluster size obtained  
282 from the  $\alpha$  probability simulations conducted individually ( $SD = 0.7$  voxels). After thresholding by  
283 decoding accuracy, error pattern correlations were computed between decoding from a searchlight  
284 location and each of the five computational models to determine max property of each searchlight  
285 location.

286 The group-level ROIs were drawn by registering ROIs of individual participants to MNI space  
287 using the same transformation parameters and overlaying them. Voxels counted in at least four  
288 participants were defined as group-level ROIs. This decision was made to ensure reasonably sized group-  
289 level ROIs while minimizing overlap between them. Finally, we excluded any voxels counted in more  
290 than a single group-level ROI.

291 **2.6.2. Error Pattern Correlation to Computational Models** We asked which contour properties  
292 contribute to the categorical representations contained in each searchlight location by comparing error  
293 patterns (Walther, Beck, & Fei-Fei, 2012). Previously, we had developed computational descriptions of

294 the same line drawings based on five contour properties – orientation, length, curvature, junction types,  
295 and junction angles (Walther & Shen, 2014). These five image properties were computed directly from the  
296 vectorized line drawings. Separate linear support vector machine classifiers were trained to predict  
297 categories of line drawings of scenes using histograms of each of these five properties in turn. Test errors  
298 from a ten-fold cross-validation analysis were recorded in separate confusion matrices, one for each image  
299 property. Here, we correlated error patterns from decoding scene categories at each searchlight location  
300 with those from the computational analysis based on each of the five properties. We call the property with  
301 the highest correlation at a given searchlight location “max property”. This analysis was performed  
302 separately for each type of line drawing and restricted to voxels with above-chance decoding accuracy.

303

### 304 **3. Results**

#### 305 **3.1. ROI-Based Neural Decoding**

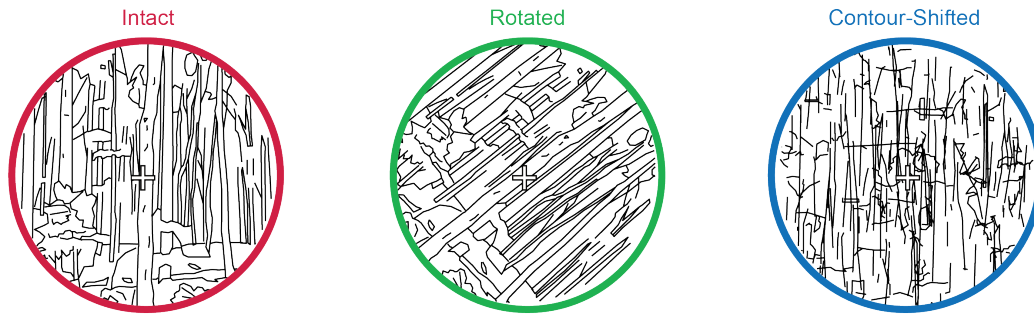
306 **3.1.1. Within-Type Decoding** Separate classifiers were trained to discriminate scene categories  
307 based on neural activity patterns recorded while participants viewed intact, rotated, or contour-shifted line  
308 drawings. The classifiers were then tested on independent data in an LORO cross-validation procedure,  
309 separately for **V1-4, PPA, OPA, RSC, LOC, and FFA**. Consistent with previous findings, one-tailed t-tests  
310 showed that scene categories of intact line drawings were correctly decoded significantly above chance  
311 (1/6) in most of the visually active ROIs: V1-4, PPA, OPA, and LOC. Decoding accuracy for the RSC  
312 was comparable to the previously reported accuracy from the same 6-way category decoding (Walther et  
313 al., 2011), but failed to reach significance. As expected, decoding accuracy from the FFA was not  
314 significantly above chance.

315 How does disruption of orientation or junction statistics affect the neural representation of scene  
316 categories? In **V1-4**, scene categories could be decoded significantly above chance for both rotated and  
317 contour-shifted just as for intact line drawings. In the PPA, OPA, and the LOC, however, scene categories  
318 could only be decoded from rotated line drawings, but not from contour-shifted line drawings. The RSC  
319 showed a similar pattern of results, but category decoding for rotated line drawings was only marginally  
320 above chance (for average decoding accuracies and standard error of means see Figure 1B, for detailed  
321 statistical results see Table 1).

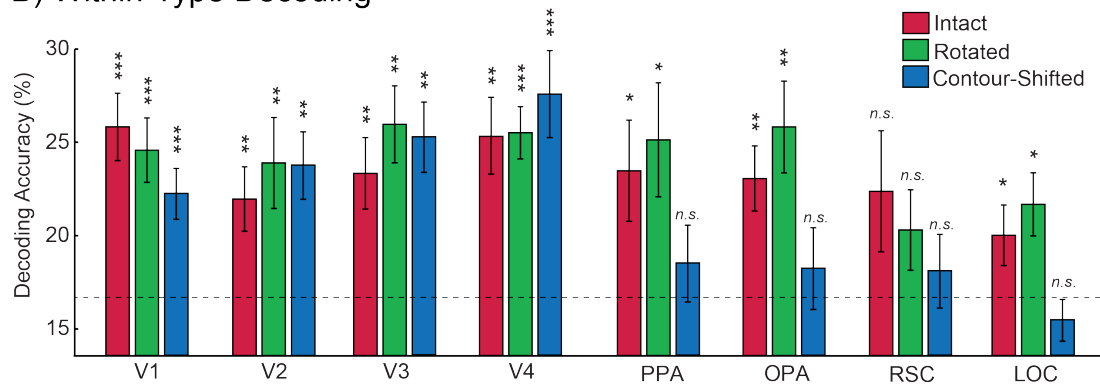
Table 1. Results of one-tailed t-tests for within-type decoding. Significance was adjusted using the false discovery rate for multiple comparisons correction.

ROI	df	Within-Type Decoding					
		Intact		Rotated		Contour-Shifted	
		<i>t</i>	<i>q</i>	<i>t</i>	<i>q</i>	<i>t</i>	<i>q</i>
V1	14	5.047	$2.67 \cdot 10^{-4}$	4.551	$3.40 \cdot 10^{-4}$	4.044	$6.04 \cdot 10^{-4}$
V2	14	3.041	$5.37 \cdot 10^{-3}$	2.941	$5.37 \cdot 10^{-3}$	3.900	$2.40 \cdot 10^{-3}$
V3	14	3.453	$1.94 \cdot 10^{-3}$	4.491	$3.81 \cdot 10^{-3}$	4.544	$3.81 \cdot 10^{-3}$
V4	12	3.892	$1.07 \cdot 10^{-3}$	5.825	$1.22 \cdot 10^{-4}$	4.324	$7.42 \cdot 10^{-4}$
PPA	14	2.492	.0194	2.763	.0194	.871	.199
OPA	11	3.276	$5.54 \cdot 10^{-3}$	3.325	$5.54 \cdot 10^{-3}$	.635	.269
RSC	14	1.746	.0882	1.667	.0882	.698	.248
LOC	14	2.037	.0458	2.944	.0160	-1.108	.857
FFA	14	.575	.431	.000	.500	.636	.431

## A) Examples of Line-drawing Scene Images



## B) Within-Type Decoding



## C) Cross-Type Decoding

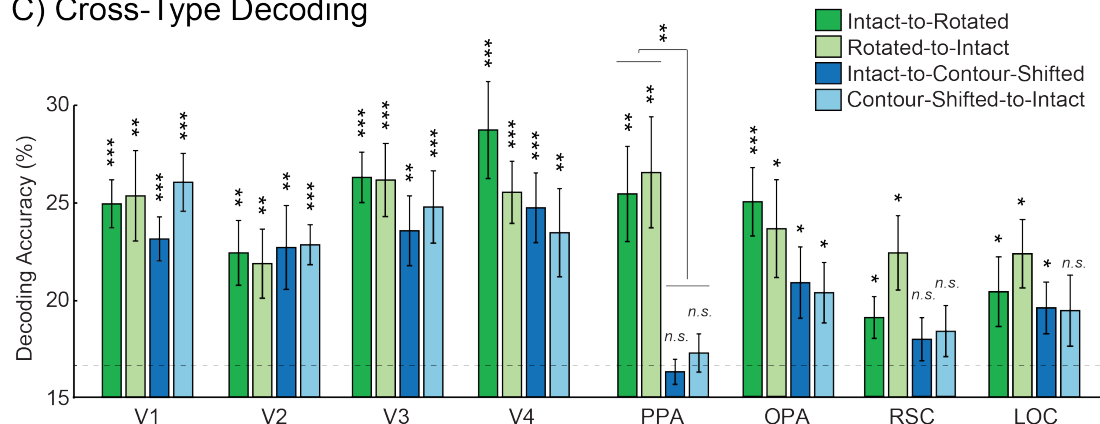


Figure 1. Example stimuli and neural decoding results. A) Examples of intact, rotated, and contour-shifted line drawings were derived from the same color photograph of a forest scene. The set of contours is identical across the three images within a triplet. The circular outline is shown here only for illustration and was absent in the stimuli seen by participants. B) Average accuracy rates of within-type category decoding from ROIs. C) Average accuracy rates of cross-type category decoding from the ROIs. The only significant difference in cross-type decoding accuracy between Intact-to-Rotated/Rotated-to-Intact and Intact-to-Contour-Shifted/Contour-Shifted-to-Intact was found in the PPA, as indicated above the bracket bridging the two bars. Error bars are **standard errors of means**. Dashed lines indicate chance performance (1/6). The significance of the one-sample t-test (one-tailed) **was adjusted for multiple comparisons and marked above each bar, \* $q < .05$ , \*\* $q < .01$ , \*\*\* $q < .001$ .**

322 **3.1.2. Cross-Type Decoding** If the neural representation of scene categories is indeed preserved  
323 under random image rotation but destroyed by shifting contours, then we should expect that a decoder  
324 trained on intact line drawings should be able to predict scene categories for rotated but not contour-  
325 shifted line drawings (and vice versa). We tested these predictions in several visual areas. Using the same  
326 LORO cross-validation procedure, we compared decoding performance across four conditions: intact-to-  
327 rotated (IR), rotated-to-intact (RI), intact-to-contour-shifted (IS), and contour-shifted-to-intact (SI).

328 Figure 1C shows the group average accuracies of the four cross-type conditions. The one-tailed t-  
329 test comparing decoding accuracy to chance showed that cross-type decoding from the early visual areas  
330 was successful for all four cross-type decoding conditions. By contrast, cross-type decoding from the PPA  
331 was significantly more accurate than expected by chance only for IR and RI, but not for IS and SI. The  
332 same pattern was also found in the RSC, although to a reduced extent. Interestingly, in the OPA, cross-  
333 type decoding was possible not only between intact and rotated line drawings but also between intact and  
334 contour-shifted line drawings, although less accurately. Similarly, accuracy from the LOC was  
335 significantly above chance for IR, RI, and IS, also marginally above chance for SI (for details on statistical  
336 tests, see Table 2).

337 To further examine the differences across the brain regions, we conducted a repeated measures  
338 ANOVA by using two factors: (1) Which *type* of disruption was used for the cross-type decoding, rotated  
339 (IR and RI) vs. contour-shifted (IS and SI), and (2) in which *direction* the decoding was conducted,  
340 trained on intact line drawings and tested on disrupted line drawings (IR and IS) vs. trained on disrupted  
341 line drawings and tested on intact line drawings (RI and SI). Consistent with the results from the one-  
342 sample t-tests, the main effect of type was significant in the PPA,  $F(1, 14) = 17.681, p = 8.82 \cdot 10^{-4}, \eta^2 =$   
343  $.558$ . In other scene-sensitive areas, however, it failed to reach significance: in the OPA,  $F(1, 11) = 1.983,$   
344  $p = .188, \eta^2 = .152$ , in the RSC,  $F(1, 14) = 3.278, p = .0917, \eta^2 = .190$ , and in the LOC,  $F < 1$ . As  
345 expected, in the early visual areas the main effect of type was not significant;  $F_s < 1$  in V1-3, and  $F(1, 12)$   
346  $= 1.370, p = .265, \eta^2 = .102$  in V4. Neither the main effect of *direction* nor the interaction between *type*  
347 and *direction* was significant in any of the ROIs.



Table 2. Results of one-tailed one-sample t-tests for cross-type decoding. Significance was adjusted using the false discovery rate for multiple comparisons correction.

Cross-Type Decoding									
ROI	df	Intact to Rotated		Rotated to Intact		Intact to Contour-Shifted		Contour-Shifted to Intact	
		<i>t</i>	<i>q</i>	<i>t</i>	<i>q</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>q</i>
V1	14	6.738	1.86·10 <sup>-5</sup>	3.757	1.06·10 <sup>-3</sup>	5.778	3.18·10 <sup>-5</sup>	6.330	1.86·10 <sup>-5</sup>
V2	14	3.410	4.23·10 <sup>-3</sup>	2.882	7.60·10 <sup>-3</sup>	2.764	7.60·10 <sup>-3</sup>	5.955	7.02·10 <sup>-5</sup>
V3	14	7.378	6.92·10 <sup>-6</sup>	5.013	1.90·10 <sup>-4</sup>	3.780	1.02·10 <sup>-3</sup>	4.326	4.64·10 <sup>-4</sup>
V4	12	4.484	7.48·10 <sup>-4</sup>	5.126	5.02·10 <sup>-4</sup>	4.156	8.90·10 <sup>-4</sup>	2.756	8.70·10 <sup>-3</sup>
PPA	14	3.572	3.89·10 <sup>-3</sup>	3.452	3.89·10 <sup>-3</sup>	-.642	.734	.564	.388
OPA	11	4.231	2.82·10 <sup>-3</sup>	2.469	.0312	2.031	.0336	2.093	.0336
RSC	14	2.200	.0451	2.978	.0120	1.126	.139	1.261	.139
LOC	14	2.099	.0363	3.246	.0112	2.193	.0363	1.512	.0752
FFA	14	-.105	.631	-.341	.631	-.188	.631	1.183	.513

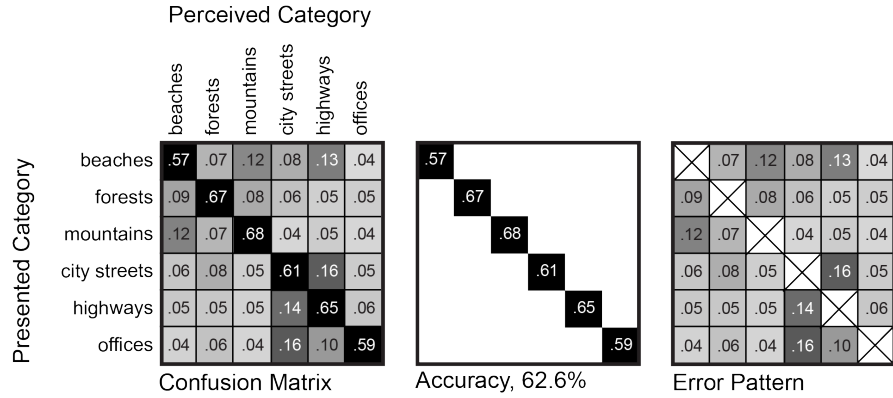
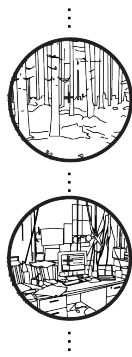
348 **3.1.3. Correlation between Neural and Behavioral Error Pattern** To explore the similarity of the  
349 underlying categorical representations between neural decoding and behavior, we performed a behavioral  
350 categorization experiment with a separate group of 49 participants. Participants were shown a line drawing  
351 of natural scenes, followed by the perceptual mask, and asked to indicate its scene category as either a  
352 beach, a forest, a mountain, a city street, a highway, or an office. Following practice and stair-casing,  
353 participants' performance stabilized at stimulus-onset-asynchronies (SOA) of 13 – 87 ms ( $M = 34$  ms,  $SD$   
354  $= 19$  ms). Average accuracy during the test phase pooled over all line drawing types was 45.5 % (Standard  
355 Errors of Means ( $SEM$ ) = 1.8%). A repeated-effects ANOVA of accuracy showed a significant effect for  
356 type of line drawing,  $F(1.69, 64.06) = 148.476$ ,  $p = 2.22 \cdot 10^{-16}$ ,  $\eta^2 = .796$  (degrees of freedom were adjusted  
357 due to a violation of sphericity). Accuracy was highest for intact line drawings ( $M = 62.3\%$ ,  $SEM = 2.9\%$ ),  
358 followed by rotated line drawings ( $M = 44.1\%$ ,  $SEM = 2.2\%$ ), and lowest for contour-shifted line drawings  
359 ( $M = 30.0\%$ ,  $SEM = 1.2\%$ ). The accuracy of contour-shifted line drawings was still significantly above  
360 chance,  $t(38) = 11.051$ ,  $p = 9.88 \cdot 10^{-14}$ .

361 Responses from the behavioral experiments were recorded in confusion matrices, separately for  
362 the three types of line drawings (see Figure 2A for intact; for behavioral confusion matrices of all three  
363 image types, see Figure S1). Off-diagonal elements of the confusion matrices represent categorization  
364 errors. Errors from the behavioral experiment were correlated with the errors made when decoding scene  
365 categories from brain activity for each of the three types of line drawings, separately per ROI and

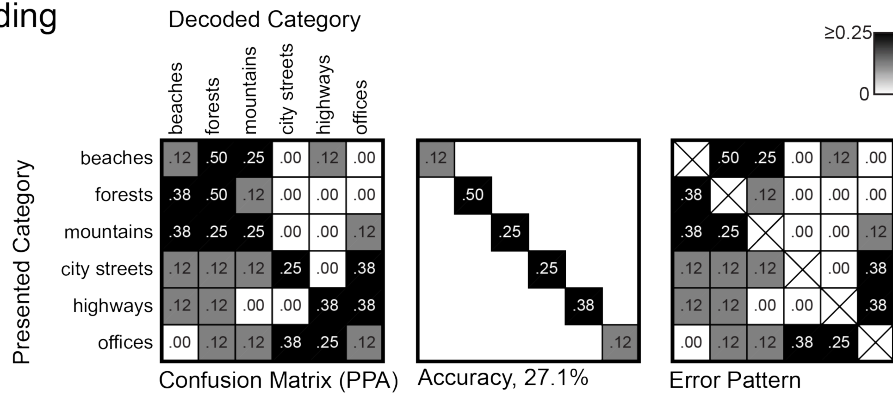
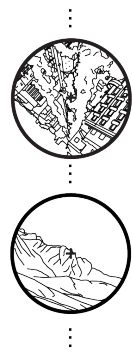
366 participant (see Figure 2B for the confusion matrix for decoding rotated line drawings from the PPA of  
367 one participant; for exclusive confusion matrix for neural decoding see Figure S2-S3). As can be seen in  
368 Figure 3A, error correlation was high when comparing the behavior and neural decoding for the same  
369 types of images (the diagonal of the three-by-three correlation matrices). In the PPA and OPA, the  
370 correlation was also high between intact and rotated, but not between intact and contour-shifted line  
371 drawings. In V1, no particular pattern of error correlations is discernible.

372           We modeled the error correlation patterns as a linear combination of three idealized models. The  
373 same-type model hypothesized that error patterns would match between behavior and decoding only for  
374 the same types of line drawings. The intact–rotated model predicts high brain-behavior error correlations  
375 between intact and rotated line drawings, assuming that disruption of orientation leaves error correlations  
376 largely unaffected. The intact–contour-shifted model, by contrast, predicts high brain-behavior error  
377 correlations between intact and contour-shifted line drawings, assuming that error correlations are  
378 maintained when junctions are disrupted, (Figure 2C).

### A) Behavior



### B) Neural Decoding



### C) Correlation between Neural and Behavioral Error Pattern

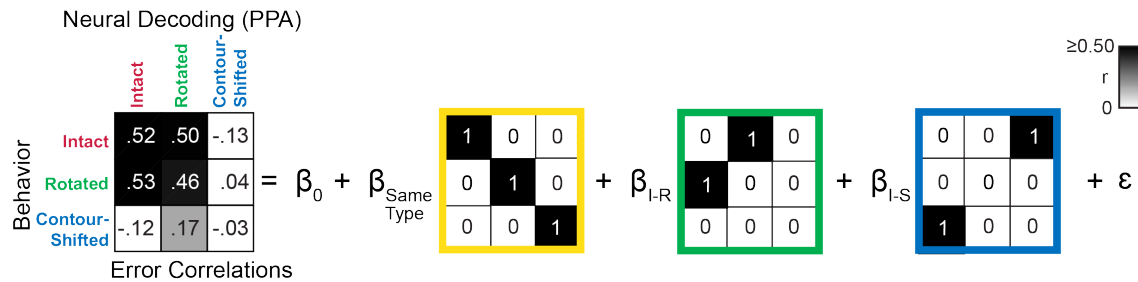


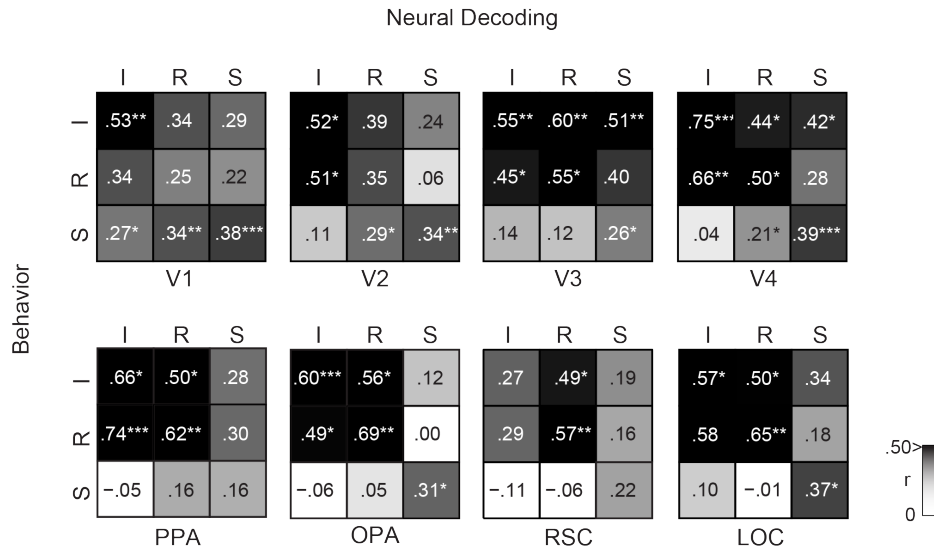
Figure 2. A schematic description correlation between brain and behavior error pattern. A) Group-average confusion matrix for behavioral scene categorization of rapidly presented intact line drawings. B) Confusion matrix obtained from decoding scene categories from rotated line drawings in the PPA for an individual participant. C) Off-diagonal entries of the confusion matrices were correlated for all three image types, resulting in a three-by-three error correlation matrix. The error correlations entered a linear regression analysis to measure how much each of the hypothesized models explains the observed patterns of error correlations.

379 The coefficients corresponding to each of the three idealized models were computed using a  
 380 mixed-effects linear regression model with prediction models as fixed effects and participants as random  
 381 effects. Figure 3B shows the estimated coefficients of the fixed effects for the prediction models for all  
 382 ROIs. In V1-3, none of the three models significantly explained the error correlation patterns between  
 383 neural decoding and behavior. However, further along in the visual processing stream, both the same type  
 384 and the intact-rotated model significantly explained error correlation in V4, PPA, OPA, RSC, and LOC.  
 385 The intact-contour-shifted model, on the other hand, did not contribute to the patterns in any of the ROIs  
 386 (for details on statistical tests see Table 3).

Table 3. Results of two-tailed t-tests for coefficients of the three idealized models for explaining patterns of error correlation between neural decoding and behavior. Significance was adjusted using false discovery rate for multiple comparisons correction.

ROI	df	Same Type		Intact-Rotated		Intact-Contour-Shifted	
		t	q	t	q	t	q
V1	14	1.166	.732	.391	.999	-1.646·10 <sup>-5</sup>	.999
V2	14	1.878	.0718	2.229	.0718	-.103	.918
V3	14	1.786	.111	2.289	.0663	-.912	.362
V4	12	3.376	1.10·10 <sup>-3</sup>	3.267	3.27·10 <sup>-3</sup>	-.103	.918
PPA	14	3.035	3.62·10 <sup>-3</sup>	4.440	2.69·10 <sup>-5</sup>	-.912	.362
OPA	11	3.056	5.61·10 <sup>-3</sup>	2.899	5.61·10 <sup>-3</sup>	-.107	.915
RSC	14	2.439	.0312	2.311	.0312	.238	.812
LOC	14	2.831	8.31·10 <sup>-3</sup>	2.774	8.31·10 <sup>-3</sup>	.795	.427
FFA	14	.568	.570	1.315	.189	-.0324	.974

## A) Error Correlations between Neural Decoding and Behavior



## B) Estimated Coefficients ( $\beta$ ) of the Idealized Models

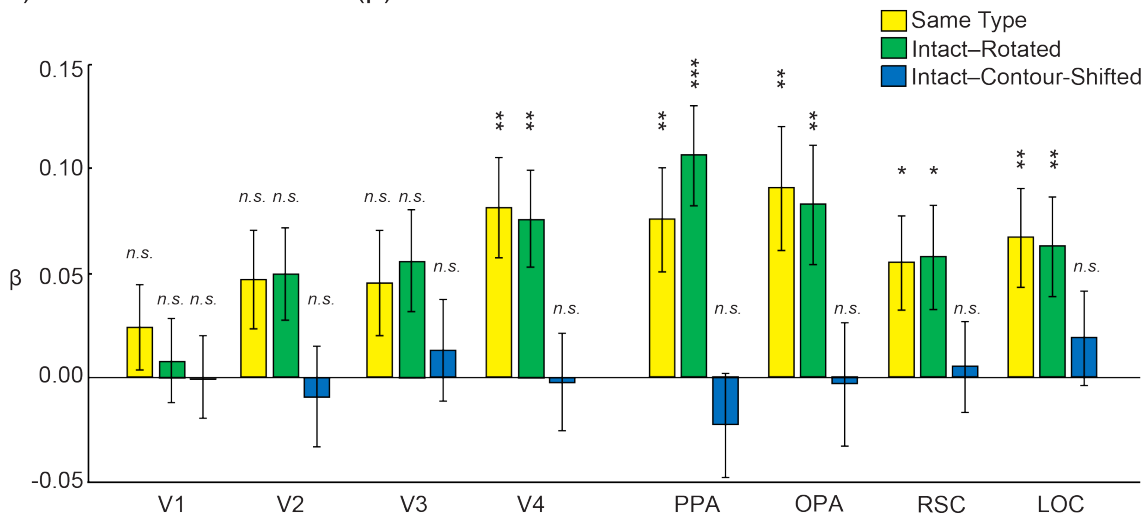


Figure 3. Results of error pattern correlations between neural decoding and behavior. A) Error correlations between neural decoding and behavior in V1-4, PPA, OPA, RSC, and LOC. Three-by-three correlation matrices were created by correlating group average confusion matrices obtained from fMRI and group average confusion matrices from the behavioral experiment. I stands for intact, R for rotated, and S for contour-shifted line drawing conditions for both neural decoding and behavior. The rows represent behavioral conditions, and the columns represent neural decoding conditions. Thus, each entry of a correlation matrix indicates an error pattern correlation value,  $r$ , between neural decoding and behavior. The significance of the correlation was determined non-parametrically using a permutation test, in which we computed correlations for all 720 permutations of the six category labels. B) Estimated coefficients of the three idealized models obtained from the ROIs. Error bars are estimated standard errors of means. The significance of the one-sample t-test (two-tailed) was corrected for multiple comparisons using false discovery rate and marked above each bar, \* $q < .05$ , \*\* $q < .01$ , \*\*\* $q < .001$ .

387 **3.2. Searchlight Analysis**

388 **3.2.1 Decoding Accuracy** To further characterize the representation of scene categories throughout visual  
389 cortex, we performed a searchlight analysis of the posterior 70% of the brain that was included in the  
390 partial acquisition scans. At each searchlight location, we attempted to decode scene categories using  
391 LORO cross-validation, separately for intact, rotated and contour-shifted line drawings. This analysis  
392 resulted in spatial maps of decoding accuracy and confusion matrices for each location. To assess the  
393 agreement between the searchlight analysis and the ROI-based analysis we computed the percentage of  
394 voxels in each of the ROIs that overlapped with the searchlight accuracy maps separately for each  
395 participant. The average of the amount of overlap is shown in Table 4. The searchlight maps for decoding  
396 intact line drawings showed the largest amount of overlap with all ROIs. More importantly, the overlap of  
397 searchlight maps with the PPA and the OPA was larger for rotated than for contour-shifted line drawings.  
398 By contrast, a similar amount of overlap was found for V1-4. In fact, the accuracy map of contour-shifted  
399 line drawings overlapped with slightly more V1 and V2 voxels than the accuracy map of rotated line  
400 drawings.

401

Table 4. Average percentages (%) of overlap of each ROI with searchlight maps computed for the three image types. **Standard errors of means** are shown in parentheses.

ROI	Intact		Rotated		Contour-Shifted	
V1	19.3	(4.1)	13.2	(2.9)	18.8	(3.3)
V2	11.4	(2.7)	9.3	(2.1)	11.4	(1.9)
V3	9.8	(2.0)	8.6	(2.5)	7.8	(2.3)
V4	10.5	(2.5)	11.6	(4.2)	6.8	(2.1)
PPA	29.3	(6.0)	19.3	(5.4)	4.8	(1.6)
OPA	28.8	(6.5)	16.8	(6.8)	10.2	(3.9)
RSC	17.5	(4.9)	8.7	(3.1)	2.8	(1.5)
LOC	8.1	(2.2)	4.5	(1.9)	3.0	(1.1)
FFA	4.3	(1.5)	3.3	(1.5)	2.3	(1.1)

402           The group-mean accuracy map of decoding intact line drawings (Figure 4A) showed a large  
403 cluster of voxels in the posterior visual cortex, including the bilateral parahippocampi, left precuneus, and  
404 the lateral end of the left transverse occipital sulcus, as well as the bilateral fusiform and calcarine gyri.  
405 This large cluster extended to the cerebellum bilaterally, but more to the left than the right cerebellum.  
406 The second large cluster encompassed the right middle occipital gyrus extending to the right transverse  
407 occipital sulcus. Another cluster included the right precuneus and the right posterior cingulate gyrus,  
408 which partially overlapped with the right retrosplenial cortex. The group-mean accuracy map of decoding  
409 rotated line drawings (Figure 4C) showed two clusters that largely overlapped with the accuracy map of  
410 decoding intact line drawings. One large cluster encompassed the left parahippocampal gyrus, the  
411 bilateral fusiform gyri, and bilateral calcarine gyri and extended to the bilateral cerebellum. The other  
412 cluster included the right parahippocampal gyrus. By contrast, the group-mean accuracy map of decoding  
413 contour-shifted line drawings (Figure 4E) revealed only one large cluster, which included bilateral  
414 calcarine gyri, fusiform gyri, lingual gyri, and the cuneus, and extended bilaterally to the cerebellum (for  
415 an exhaustive list of peak coordinates, see Table 5).

Table 5. Clusters identified in the searchlight analysis for within-type decoding of intact, rotated, and contour-shifted line drawings.

Decoding Condition	Peak (MNI coordinates)			Accuracy (%)	Volume ( $\mu$ l)	Description
	x	y	z			
Intact	2.5	90.5	-4.2	30.4	63234	Occipital pole, calcarine gyri, fusiform gyri, parahippocampal gyri, left precuneus, left transvers occipital sulcus, bilateral cerebellum
	-35.0	85.5	25.8	25.4	5219	Right middle occipital gyrus, right transvers occipital sulcus
	-25.0	63.0	18.2	25.6	1922	Right parieto-occipital sulcus, right retrosplenial cortex
	30.0	45.5	68.2	22.0	391	Left superior parietal gyrus
	-32.5	30.5	-26.8	19.6	188	Right lateral occipito-temporal gyrus (fusiform gyrus)
Rotated	10.0	98.0	-14.2	28.0	50938	Occipital pole, calcarine gyri, fusiform gyri, left parahippocampal gyrus, bilateral cerebellum
	-22.5	50.5	-9.2	26.9	4531	Right parahippocampal gyrus
	27.5	85.5	33.2	22.7	375	Left inferior parietal angular gyrus, left middle occipital gyrus
	-47.5	80.5	-1.8	21.8	188	Right middle occipital gyrus
Contour-shifted	-15.0	90.5	-6.8	29.3	54938	Occipital pole, calcarine gyri, precuneuses, fusiform gyri, lingual gyri, bilateral cerebellum
	-30.0	30.5	56.2	22.5	703	Right precentral gyrus
	42.5	33.0	28.2	23.7	563	Left posterior lateral fissure, left inferior parieto-supramarginal gyrus.
	-27.5	53.0	-9.2	21.9	484	Right medial occipito-temporal sulcus, right lingual gyrus
	50.0	25.5	48.2	22.6	453	Left inferior parieto-supramarginal gyrus
	-2.5	20.5	76.2	21.7	422	Right paracentral gyrus, right paracentral sulcus
	-57.5	20.5	78.2	17.8	344	Right inferior parieto-supramarginal gyrus
	10.0	38.0	55.8	21.7	250	Left paracentral gyrus, left paracentral sulcus
	-10.0	43.0	78.2	21.4	234	Right paracentral gyrus, right paracentral sulcus
	35.0	93.0	0.8	21.9	219	Left middle occipital gyrus
-25.0	88.0	25.8	21.0	203	Right superior occipital gyrus.	



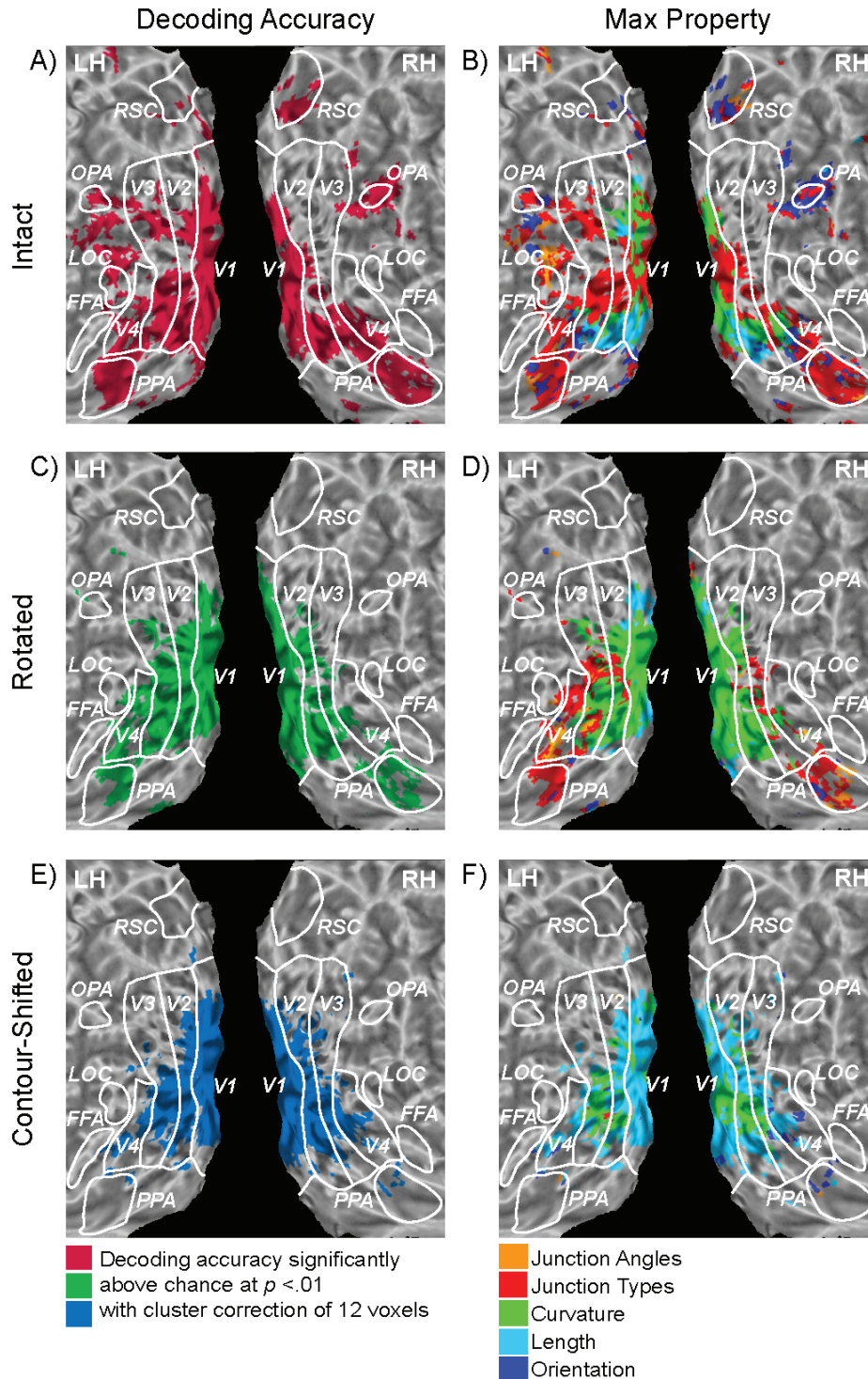


Figure 4. Neural decoding accuracy and max property maps. A) Searchlight locations with above-chance decoding of intact line drawings are highlighted in red. B) Property with the highest error correlation between searchlight decoding and computational feature model (max property) at the searchlight locations with above-chance decoding of intact line drawings. C) Searchlight locations with above-chance decoding of rotated line drawings are highlighted in green. D) Max property at the searchlight locations with above-chance decoding of rotated line drawings. E) Searchlight locations with above-chance decoding of contour-shifted line drawings are highlighted in blue. F) Max property at the searchlight locations with above-chance decoding of contour-shifted line drawings.

416 **3.2.2. Important Image Property for Neural Decoding** To determine the influence of several kinds  
417 of structural properties of line drawings on the neural representation of scene categories, we correlated  
418 error patterns from each searchlight location (that is, the off-diagonal elements of the confusion matrices)  
419 to those from five computational models of scene categorization. Each of these computational models  
420 relies exclusively on one of five structural properties: contour orientation, length, curvature, and types  
421 and angles of contour junctions (Walther & Shen, 2014). Each searchlight location was labeled according  
422 to the contour property with the highest error correlation, the *max property*. Figure 4B, D, and F show  
423 max property maps for the three image types, restricted to the locations that allowed for decoding of scene  
424 categories significantly above chance (see Figures S4-S6 for unrestricted maps of error correlations and of  
425 max property).

426 Representation of scene categories in early visual areas relies most heavily on contour length and  
427 curvature for intact line drawings. Junction types are particularly important in foveal regions of early  
428 visual cortex, and also the PPA, the OPA, and the RSC. For rotated line drawings, curvature dominates  
429 early visual areas, while high-level visual areas continue to rely on junction properties. For contour-shifted  
430 line drawings, early visual areas rely most on contour length and curvature. Only a few searchlight  
431 locations in high-level visual areas allow for decoding of scene categories from contour-shifted line  
432 drawings. These effects are quantified more precisely by assessing overlap of these maps with ROIs  
433 (Figure 5).

434 Voxel statistics for the pre-defined ROIs show several interesting effects. First, representation of  
435 natural scene categories in V1 and V2 relies most heavily on contour length and curvature for intact line  
436 drawings. The importance of junctions increases steadily through V3 and V4, until junctions dominate the  
437 representation of scene categories in the PPA (81.6% of PPA voxels have junction angle or type as max  
438 property), the OPA (66.7%), and the LOC (97.9%). The high reliance on junction properties strongly  
439 manifested even though L-junctions, critical to surface analysis (Biederman, 1987; Guzman, 1968), were  
440 not considered in the computational category prediction (Walther & Shen, 2014). Orientation statistics, by  
441 comparison, play a minor role in the high-level visual areas, except for the RSC (51.8%).

442 Second, the involvement of orientation statistics in the representation of scene categories is  
443 absent for randomly rotated line drawings as **expected**. Whereas contour orientation was the max property

444 for 291 out of 2575 (11.3%) searchlight locations with significant decoding for intact line drawings, this  
 445 number fell to 22 out of 2402 (0.9%) searchlight locations with significant decoding for rotated line  
 446 drawings. Similarly, junction properties ceased to be the max property almost everywhere for contour-  
 447 shifted line drawings. Junction types and angles were the max property for 1469 out of 2575 (57.0%)  
 448 searchlight locations for intact but only 38 out of 2021 (1.9%) searchlight locations for contour-shifted  
 449 line drawings. These findings confirm the effectiveness of our image manipulations as a tool for probing  
 450 the role of structural scene properties in the neural representation of scene categories.

451 Moreover, this shift in the reliance to other structural properties when deprived of contour  
 452 orientation or junctions demonstrates the flexibility of early visual areas to make use of any visual  
 453 regularities present in the images. However, junction properties appear to be critical for the neural  
 454 representation of scene categories in high-level visual areas. Rather than shifting to other structural  
 455 properties, these areas show a dramatic decrease in the number of searchlight locations that allowed for  
 456 the decoding of scene categories when junctions were disrupted.

457

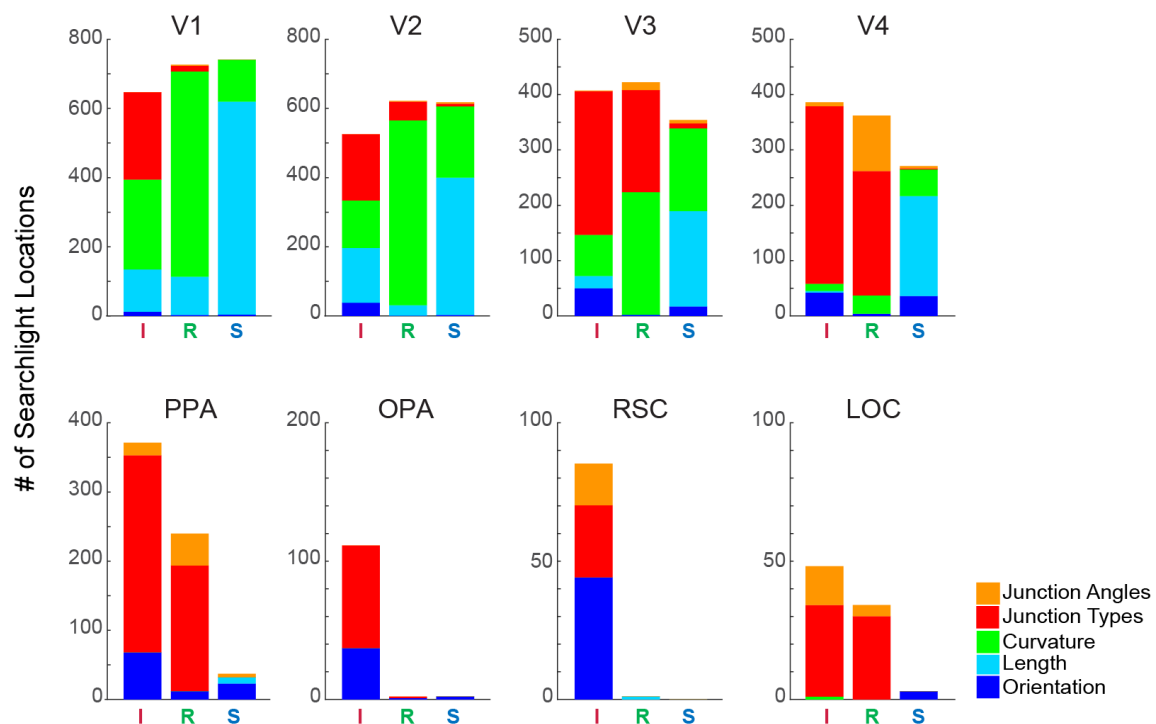


Figure 5. Distribution of max properties within group-level ROIs for decoding of intact (I), rotated (R), and contour-shifted (S) line drawings. Only searchlight locations with above-chance decoding accuracy were counted. Coloring follows the same convention as Figure 4.

## 458 4. Discussion

459 In this study, we used multi-voxel pattern analysis (MVPA) to identify visual properties critical  
460 to the categorical representation of real-world environments in the brain. Consistent with our previous  
461 findings (Walther et al., 2009; 2011), we showed that categorical representations of scenes are distributed  
462 across the human visual cortex. Importantly, they rely on different sets of contour properties along the  
463 course of visual processing. In the PPA, OPA, and LOC, structural relations between contours as  
464 embodied by the distribution of junctions need to be preserved to maintain category-specific brain activity  
465 patterns. Whereas random shifting of contours led to chance-level decoding accuracy, random image  
466 rotation did not undermine decoding accuracy in the PPA, OPA, and LOC.

467 Cross-type decoding was successful between intact and rotated, but not between intact and  
468 contour-shifted line drawings in the PPA. Note that contour-shifted line drawings still preserved the  
469 statistics of contour orientation, length, and curvature. Yet, this information was not sufficient to give rise  
470 to a decodable neural representation of scene categories in the PPA, underscoring the necessity of junction  
471 properties and, thereby, structural relations between contours for the PPA to encode scene categories.

472 In contrast, category decoding was successful with contour-shifted line drawings as well as  
473 rotated line drawings from V1 to V4, suggesting that none of the image manipulations was detrimental to  
474 the category-specificity of neural activity patterns for line drawings in early visual areas. Furthermore,  
475 cross-type decoding was successful not only between intact and rotated but also between intact and  
476 contour-shifted line drawings, indicating that the neural representations of scene categories were  
477 compatible across the three line drawing types. In the early visual areas, category-specific neural activity  
478 patterns are not undermined by disrupting a single type of visual statistics, suggesting that the early visual  
479 areas represent scene categories only implicitly by relying on any available statistics indicative of scene  
480 categories in a parsimonious manner.

481 Analysis of decoding errors confirmed previous findings that error patterns of decoding from the  
482 neural activity in the PPA significantly correlate with those of human scene categorization of intact line  
483 drawings (Walther et al. 2011). Critically, we here show that junction properties but not orientation  
484 statistics of line drawings have to be preserved to maintain such brain-behavior correlation (for searchlight  
485 analysis of error pattern correlation to behavior and its results, see supplementary material; Figure S7).

486 Preservation of junctions starts to be important for brain-behavior error correlation as early as in V2. This  
487 finding agrees well with neurophysiological studies in non-human primates showing that area V2 is  
488 sensitive to changes in types or angles of contour junctions (Pasupathy & Connor, 2002; Peterhans & von  
489 der Heydt, 1989).

490 In a searchlight analysis, we found early visual areas to rely on the distributions of orientation,  
491 length and curvature of contours for the neural representation of scene categories. Junction properties, on  
492 the other hand, became increasingly important in near-foveal regions of V1-4 and peaked in the PPA and  
493 OPA. The importance of junctions for the neural representation of scene categories in the PPA persisted  
494 when scene images were rotated by a random angle but disappeared when junctions were disrupted by  
495 randomly shifting contours.

496 As should be expected, the importance of orientation statistics for the representation of intact line  
497 drawings throughout visual cortex disappeared for rotated images and was supplanted by an increased  
498 reliance on contour length and curvature. Searchlight analysis confirms the hierarchical aspect of neural  
499 representations of scene categories. Once an image of real-world environments proceeds through the  
500 cortical hierarchy, its neural representation relies progressively more on complex visual properties, from  
501 orientation and length extracted in V1 (Hubel & Wiesel, 1962) to curvature and junction properties  
502 extracted in V2 and V4 (Pasupathy & Connor, 2002; Peterhans & von der Heydt, 1989). Recent  
503 computational work showed that junction-like feature representations arise naturally when representations  
504 of complex scenes are learned in simulated multi-layer neural networks (Zeiler & Fergus, 2014).

505 What would be the mechanisms for contour junctions affecting a neural representation of scene  
506 categories in the PPA? While surface features such as color and orientation gradient may initiate contour  
507 detection and surface delineation (by the primal sketch; Marr, 1982), they do not directly give a rise to the  
508 high-level representation for visual recognition. In fact, visual recognition and categorization hardly  
509 benefit from surface features once important contours are analyzed and their relations are determined  
510 (Biederman & Ju, 1988). Instead, junctions present in two-dimensional (2D) images are informative of  
511 three-dimensional (3D) structural information, because their types and angles are indicative of  
512 arrangements and relations of surfaces in 3D space (Biederman, 1987; Guzman, 1968). The 3D structure  
513 of natural environments can be inferred from the distribution of these viewpoint-invariant properties,

514 namely junction types and angles. Furthermore, their invariance to changes in viewpoint makes these  
515 properties particularly useful for visual recognition and categorization

516 In fact, 3D structure of scenes is likely to be related to some global scene properties, such as  
517 whether a scene has an open or closed layout (Harel et al., 2013; Park et al., 2011). For example, scenes  
518 with an open layout are usually clutter-free and contain surfaces not obstructing one another, resulting in a  
519 relatively small number of junctions. In contrast, a closed layout is likely to contain more objects and  
520 multiple surfaces overlaying one another, thus creating proportionally more junctions. Consistent with this  
521 idea, recent neuroimaging evidence showed that the PPA is not only sensitive to changes in the statistics  
522 of simple shapes (Cant & Xu, 2012), but also to junction angles (Nasr, Echavarria, & Tootell, 2014).  
523 When combining these results with the sensitivity of the PPA to global scene properties (Kravitz et al.  
524 2011; Park et al. 2011; Harel et al. 2013), a picture of the PPA arises as a visual area sensitive to several  
525 high-level aspects of scenes, such as semantic category, global layout, or relation of the scene to real-  
526 world locations (Marchette, Vass, Ryan, & Epstein, 2014). A recent study clearly shows that the PPA  
527 indeed can encode several aspects of a scene image such as its entry-level category, spatial layout (open  
528 vs. closed), surface texture property, and content (man-made vs. natural) (Lowe, Gallivan, Ferber, & Cant,  
529 in press). With scenes being spatial arrangements of surfaces and objects, the image properties driving  
530 these diverse visual aspects of scenes are those encoding the relationship between surfaces in 3D space,  
531 namely junction types and their angles.

532 Similar to the PPA, the OPA also relies more heavily on junction properties than contour  
533 orientation. Considering its anatomical proximity to the early visual cortex, the OPA is likely to subserve  
534 relatively primitive scene analysis. In fact, receptive fields in the OPA were reported to be smaller than  
535 those in the PPA (MacEvoy & Epstein, 2007). Furthermore, the OPA preferentially activates to spatial  
536 layouts over collections of multiple objects (Bettencourt & Xu, 2013; MacEvoy & Epstein, 2007). The  
537 OPA has been suggested to contain precursory representations of spatial layout of scenes based on  
538 relatively simple features (Baldassano et al., 2013; Dilks, Julian, Paunov, & Kanwisher, 2013; MacEvoy  
539 & Epstein, 2007). We here propose that contour junctions are critical for these representations.

540 Objects are often diagnostic for scene category (Bar & Aminoff, 2003; Greene, 2013). For  
541 instance, a beach scene is more likely to contain palm trees, beach balls, and umbrellas than desks, swivel

542 chairs, and computer monitors, and *vice versa* in the case of an office scene. It is, therefore, natural that  
543 the object-sensitive LOC contains information about objects, which can be exploited for decoding scene  
544 category or identity (Harel et al., 2013; MacEvoy & Epstein, 2011). The LOC not only activates linearly  
545 to the number of displayed objects (MacEvoy & Epstein, 2011), but it also encodes inter-object  
546 relationships (Kim & Biederman, 2010; 2011). Given recent evidence for a strong functional relationship  
547 between the LOC and PPA, it is highly likely that information about object relations in the LOC is  
548 projected to the PPA (Baldassano et al., 2013), thus contributing the category-specific activity in the PPA.  
549 Our finding that the neural representation of scene categories in the LOC relies almost exclusively on  
550 junction properties is consistent with the reliance of invariant object recognition on non-accidental  
551 properties, most specifically contour junctions (Biederman 1987), which provide cues to the three-  
552 dimensional structure of objects. In addition to their role in defining objects, contour junctions are crucial  
553 to reconstruct spatial arrangements between large-scale surfaces that define the terrain and layout of a  
554 scene, thereby contributing to the neural representations of real-world scene categories.

555 Random contour shifting also affected grouping properties that were not modeled explicitly in  
556 our computation analysis (Walther & Shen, 2014): proximity between parallel contours, colinearity and  
557 curvilinearity of contours indicating parallel surfaces in depth, and to some extent symmetry involving  
558 multiple contours (Biederman, 1987). These non-accidental properties have in common that they are  
559 defined by spatial relations between contours rather than by properties of individual contours. Junction  
560 statistics capture these spatial relations to only a limited extent. It is therefore remarkable that decoding  
561 error patterns in the PPA, OPA and LOC are predicted so well by junction statistics.

562 Unlike the other scene-selective visual regions, the RSC has been found to be critical for  
563 embedding scenes in their real-world context or memory representations rather than for perceptual  
564 analysis of scenes. For instance, the RSC plays a critical role in navigation and route learning (Aguirre &  
565 D'Esposito, 1999; Maguire, 2001), and shows preferential neural activity for landmark buildings  
566 compared to non-landmark buildings (Schinazi & Epstein, 2010). The RSC also mediates between  
567 individual scenes in a broad view rather than representing exact perceptual instances of scenes (Epstein,  
568 Parker, & Feiler, 2007; Epstein, Higgins, Jablonksi, & Feiler, 2007; Park & Chun, 2009). We found  
569 neither significant decoding accuracy nor even significant activation in the RSC for line drawings of

570 scenes (Table S1; for detailed analysis and results see Supplementary materials). We surmise that this may  
571 be because line drawings too dissimilar from actual real-world settings that afford navigation or context-  
572 based memory retrieval.

573

## 574 **5. Conclusion**

575 We have shown that the neural representations of scene categories rely on different image  
576 properties throughout the processing hierarchy in the human visual cortex. In early visual areas, any  
577 statistical regularities available in an image had the potential to elicit category-specific patterns of neural  
578 activity. In the scene-selective high-level visual regions, especially in the PPA, accurate statistics of  
579 junction properties was necessary to generate category-specific activity patterns and, importantly, to  
580 establish high correlation of decoding error patterns with patterns of errors observed in human scene  
581 categorization behavior. We conclude that non-accidental 2D cues to 3D structure, in particular contour  
582 junctions, are causally involved in eliciting a neural representation of scene categories in the PPA and the  
583 OPA by providing a reliable description of the 3D structure of real-world environments. Summary  
584 statistics of orientations, on the other hand, are insufficient to elicit a decodable representation of scene  
585 categories in these brain regions.



586 **Acknowledgments**

587           The authors declare no competing financial interests. All experiments were performed

588 while both authors were at The Ohio State University.

589 **References**

590

591 Aguirre, G. K., & D'Esposito, M. (1999). Topographical disorientation: A synthesis and taxonomy.

592 *Brain*, 122, 1613-1628.

593 Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context, *Neuron*, 38, 347-358.

594 Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the parahippocampal  
595 place area. *NeuroImage*, 75, 236-245.

596 Bates D, Maechler M, Bolker B and Walker S (2014). lme4: Linear mixed-effects models using Eigen and  
597 S4. R package version 1.1-7, <http://CRAN.R-project.org/package=lme4>.

598 Bettencourt, K. C., & Xu, Y. (2013). The role of transverse occipital sulcus in scene perception and its  
599 relationship to object individuation in inferior intraparietal sulcus. *Journal of Cognitive*  
600 *Neuroscience*, 25(10), 1711-1722.

601 Biederman, I. (1987). Recognition-by-components: A theory of human image understanding.  
602 *Psychological Review*, 94(2), 115-147.

603 Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive*  
604 *Psychology*, 20, 38-64.

605 Biederman, I., Mezzanote, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging  
606 objects undergoing relational violation. *Cognitive Psychology*, 14, 143-177.

607 Biederman, I., Teitelbaum, R. C., & Mexxanote, R. J. (1983). Scene perception: A failure to find benefit  
608 from prior expectancy or familiarity. *Journal of Experimental Psychology*, 9(3), 411-429.

609 Cant, J. S. & Xu, Y. (2012). Object ensemble processing in human anterior -medial ventral visual cortex.  
610 *Journal of Neuroscience*, 32, 7885-7700.

611 Chang, C-C., & Lin, C-J. (2001). LIBSVM: A Library for support vector machines. Software available at  
612 <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

613 Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) “brain reading”:  
614 Detecting and classifying distributed patterns of fMRI activity in human visual cortex.

615 *Neuroimage*, 19(2), 261-270.

- 616 Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally  
617 and selectively involved in scene perception. *Journal of Neuroscience*, 33(4), 1331-1336.
- 618 Epstein, R. A., Higgins, J. S., Jablonksi, K., & Feiler, A. M. (2007). Visual scene processing in familiar  
619 and unfamiliar environments. *Journal of Neurophysiology*, 97, 3670-3683.
- 620 Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*  
621 392, 598–601.
- 622 Epstein, R., Parker, W. E., & Feiler, A. M. (2007). Where am I now? Distinct roles for parahippocampal  
623 and retrosplenial cortices in place recognition. *Journal of Neuroscience*, 27, 6141-6149.
- 624 Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world  
625 scene? *Journal of Vision*, 7(1), 1-29.
- 626 Fonov, V., Evans, A. C., Botteron, K., Almlí, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased  
627 average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1), 313-327
- 628 Greene, M. R. (2013). Statistics of high-level scene context. *Frontiers in Perception Science*, 4, 777
- 629 Greene, M. R. & Fei-Fei, L. (2014). Visual Categorization is Automatic and Obligatory: Evidence from a  
630 Stroop-like Paradigm. *Journal of Vision*, 14(1), 1-11.
- 631 Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the  
632 forest without representing the trees. *Cognitive Psychology*, 58(2), 137-176.
- 633 Greene, M. R., & Oliva, A. (2009b). The briefest of glances: The time course of natural scene  
634 understanding. *Psychological Science*, 20(4), 464-472.
- 635 Greene, M. R., & Oliva, A. (2010). High-level aftereffects to global scene properties. *Journal of*  
636 *Experimental Psychology: Human Perception & Performance*, 36(6), 1430-1442.
- 637 Guzmá, A. (1968). Decomposition of a visual scene into three-dimensional bodies. *Proceedings of the*  
638 *Fall Joint Computer Conference*, Part Im 291-304.
- 639 Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed  
640 and overlapping representations of faces and objects in ventral temporal cortex. *Science*,  
641 293(5539), 2425-2430.
- 642 Harel, A., Kravitz, D. J., & Baker, C. I. (2013). Deconstructing visual scenes in cortex: gradients of object  
643 and spatial layout information. *Cerebral Cortex*, 23(4), 947-957.

- 644 Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture  
645 in the cat's visual cortex. *Journal of Physiology*, *160*, 106–154
- 646 Kim, J. G., & Biederman, I. (2010). Where do objects become scenes? *Cerebral Cortex*, *21*, 1738-1746.
- 647 Kim, J. G., & Biederman, I. (2011). The benefit of object interactions arises in the lateral occipital cortex  
648 independent of attentional modulation from the intraparietal sulcus: A transcranial magnetic  
649 stimulation study. *Journal of Neuroscience*, *31*(22), 8320-8324.
- 650 Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-world scene representations in high-level visual  
651 cortex – it's the space more than the places. *Journal of Neuroscience*, *31*(20), 7322-7333.
- 652 Kriegeskorte, N., Goebel R, Bandettini, P. (2006). Information-based functional brain mapping.  
653 *Proceedings of the National Academy of Sciences USA*, *103*, 3863–3868
- 654 Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimarri, T. N. (2010). The role of higher order image  
655 statistics in masking scene gist recognition. *Attention, Perception, & Psychophysics*, *72*(2), 427-  
656 444.
- 657 MacEvoy, S. P., & Epstein, R., A. (2007). Position selectivity in scene- and object-responsive  
658 occipitotemporal regions. *Journal of Neurophysiology*, *98*, 2089-2098.
- 659 MacEvoy, S. P., & Epstein, R., A. (2011). Constructing scenes from objects in human occipitotemporal  
660 cortex. *Nature Neuroscience*, *14*, 1323-1329.
- 661 Maguire, E. A., (2001). The retrosplenial contribution to human navigation: a review of lesion and  
662 neuroimaging findings. *Scandinavian Journal of Psychology*. *42*, 225–238.
- 663 Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A.,(2014). Anchoring the neural compass: coding of  
664 local spatial reference frames in human medial parietal lobe. *Nature Neuroscience*, *17*, 1598–  
665 1606.
- 666 Marr, D. (1982). *Vision: A Computational Approach*. San Francisco, CA: Freeman & Company.
- 667 Nasr, S., Echavarria, C. E., Tootell, R. (2014). Thinking outside the box: rectilinear shapes selectively  
668 activate scene-selective cortex. *Journal of Neuroscience*, *34*(20), 6121-6735.
- 669 Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial  
670 envelope. *International Journal of Computer Vision*, *42*(3), 145-175.

- 671 Park, S., Brady, T. F., Greene, M. R. & Oliva, A. (2011). Disentangling scene content from spatial  
672 boundary: Complementary roles for the PPA and LOC in representing real-world scenes.  
673 *Journal of Neuroscience*, 31(4), 1333-1340.
- 674 Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and  
675 retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage*, 47(4), 1747-1756.
- 676 Park, S., Konkel, T., & Oliva, A. (2015). Parametric coding of scene size and clutter independent of  
677 semantic category. *Cerebral Cortex*, 25(7), 1792-1805. Pasupathy, A., & Connor, C. E. (2002).  
678 Population coding of shape in area V4. *Nature Neuroscience*, 5(12), 1332-1338.
- 679 Pereira, F. & Botvinick, M. (2011). Information mapping with pattern classifiers: a comparative study.  
680 *Neuroimage*, 56, 476-496.
- 681 Peterhans, E., & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. II.  
682 Contours bridging gaps. *Journal of Neuroscience*, 9, 1749–1763.
- 683 Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex  
684 wavelet coefficients. *International Journal of Computer Vision*, 40, 49–71.
- 685 Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of*  
686 *Experimental Psychology*, 81, 10–15.
- 687 Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381,  
688 520–522.
- 689 Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural*  
690 *Systems*, 14, 391-412
- 691 Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology*, 15,  
692 121–149.
- 693 Walther, D. B., Beck, D. M., & Fei-Fei, L. (2012). To err is human: correlating fMRI decoding and  
694 behavioral errors to probe the neural representation of natural scene categories. In: Nikolaus  
695 Kriegeskorte and Gabriel Kreiman (Eds.), *Visual Population Codes – Towards a Common*  
696 *Multivariate Framework for Cell recording and Functional Imaging*. MIT Press: Cambridge,  
697 Massachusetts, pp. 391-415.

- 698 Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in  
699 distributed patterns of activity in the human brain. *Journal of Neuroscience*, *29*, 10573–10581.
- 700 Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice  
701 for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of*  
702 *Sciences USA*, *108*, 9661–9666.
- 703 Walther, D. B., & Shen, D. (2014). Nonaccidental properties underlie human categorization of complex  
704 natural scenes. *Psychological Science*, *25*(4), 851-860.
- 705 Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception &*  
706 *Psychophysics*, *33* (2), 113-120.
- 707 Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In: D. Fleet  
708 et al. (Eds.), *ECCV 2014*, Part I, LNCS 8689, Springer International Publishing: Switzerland, pp.  
709 818–833.

## 710 **Supplementary Materials**

711

### 712 **Localizing regions of interest**

713           Following the main experiment scans, participants viewed blocks of color photographs of faces, scenes,  
714 objects, and grid-scrambled objects as part of a standard face-place-object localizer scan (Epstein & Kanwisher,  
715 1998). Participants were asked to indicate immediate repetitions of images to maintain their attention to images.  
716 Images subtended 640 x 640 pixels (approximately 11 ° of visual angle) and were presented for 500 ms, followed by  
717 a blank screen for 500 ms. Participants saw five blocks of 72 images, sub-divided into four mini-blocks of 18  
718 randomly drawn images for each of the four categories. Including 12 second fixation periods before, between, and  
719 after the blocks, the entire scan lasted for 7 minutes and 12 seconds. Scanning parameters for the face-place-object  
720 localizer differed from those of the main experiment: TR = 3000 ms, TE = 28ms, flip angle = 80°, voxel size = 2.5  
721 x.2.5 x 2.5 mm, matrix size = 90 x 100 x 40 voxels, 40 coronal slices.

722           The fMRI data were motion corrected, registered to the anatomical scans that had been aligned to the  
723 functional volumes of the main experiment, spatially smoothed using a 4 mm FWHM Gaussian filter and converted  
724 to percent signal change with respect to the mean of each run. The pre-processed data entered a GLM analysis with  
725 regressors for all four image types. ROIs were defined as contiguous clusters of voxels with significant contrasts ( $q$   
726  $< 0.05$ ; corrected using false discovery rate) of scenes  $>$  (faces and objects) for PPA, RSC (Epstein & Kanwisher,  
727 1998) and OPA (Dilks et al., 2013); faces  $>$  (scenes and objects) for FFA (Kanwisher et al. 1997); and objects  $>$   
728 (scrambled objects) for LOC (Grill-Spector, Knouf, & Kanwisher, 2004; Grill-Spector, Kourtzi, & Kanwisher  
729 2001). To obtain robust clusters for the RSC, the threshold had to be relaxed to  $p < 0.01$  (uncorrected). We could not  
730 find significant clusters corresponding to the OPA in three participants and used data of the remaining twelve  
731 participants to perform decoding from the OPA.

732           Boundaries of early visual areas were established by stimulating the horizontal and vertical meridians of the  
733 visual field (HV) in alternation (Kastner, Weerd, Desimone, & Ungerleider, 1998). During two HV scans (only one  
734 HV scan for four participants, because we ran out of time), participants viewed flickering checkerboard patterns (2  
735 Hz, a mix of white, red, green, blue, and yellow checkerboards), filling pairs of wedges (width: 10°) aligned with the  
736 horizontal or vertical meridians, respectively. Fixation periods of 20 seconds were included between each alternation

737 as well as at the beginning and the end of the scan (total scan duration: 3 min 20 sec). In order to establish the  
738 anterior boundary of V4, a scan with alternating stimulation of the upper and lower visual field (UL) with similar  
739 flickering checkerboard patterns along the diagonals was included as well, which lasted approximately 3 min. Scan  
740 parameters for the HV and UL scans were as follows; TR = 2000 ms, TE = 28 ms, flip angle = 72°, voxel size =  
741 1.953 x 1.953 x 2 mm, matrix size = 114 x 128 x 30 voxels, 30 coronal slices.

742 Data from the HV and UL scans were motion-corrected, registered to the anatomical scan that had been  
743 aligned to the functional volume of the main experiment, spatially smoothed (4 mm FWHM) and converted to  
744 percent signal change. In separate GLM analyses, data from the HV scans were analyzed for a horizontal-versus-  
745 vertical meridian contrast and data from the UL scans for an upper-versus-lower visual field contrast.

746 Cortical surfaces for each participant's brain were reconstructed from their anatomical scans (MPRAGE)  
747 using Freesurfer. To flatten the cortical surface, each hemisphere was virtually cut along the calcarine fissure and  
748 four additional relaxation cuts. The corpus callosum and mid-brain structures on the medial surface were removed.  
749 Boundaries between early visual areas were identified by projecting the beta-weight maps of the HV and UL  
750 contrasts onto the flattened cortical surfaces using AFNI and SUMA. Following Hansen, Kay, and Gallant (2007)  
751 we identified the V1/V2 border as the first vertical meridian, the V2/V3 border as the second horizontal meridian,  
752 and V3/V4 border as the second vertical meridian. Since V4 represents the entire contralateral hemifield on the  
753 lower bank of the calcarine fissure, we identified the anterior border of V4 as the closest boundary that encompassed  
754 both upper and lower visual field. For two participants, the anterior V4 border could not be clearly delineated. Thus,  
755 we only used data from the remaining thirteen participants to perform decoding from V4. ROIs were drawn  
756 conservatively to minimize the amount of overlap between neighboring areas. Following projection of surface-based  
757 ROIs back into the brain volume of each participant, we excluded voxels that were assigned to more than one ROI.

758

## 759 **Univariate analysis**

760 To explore the effect of the disruption of contour properties on the magnitude of the average neural activity  
761 we conducted a standard general linear model (GLM) analysis using the AFNI software package, and deconvolved  
762 block responses to the three image conditions for each voxel within the ROIs. Nuisance variables were included to  
763 capture variance due to head motion (six affine transform parameters) and scanner drift (4th-degree polynomial).  
764 Beta parameters were extracted for the three contrasts: Intact > Fixation, Rotated > Fixation, and Contour-shifted >



765 Fixation. The beta parameters were averaged over voxels within each ROI. The GLM analysis showed that the line  
 766 drawing images significantly activated all of the ROIs except for the RSC (Table S1). A one-way ANOVA showed  
 767 no significant difference in neural activity between intact, rotated, and contour-shifted line drawings in any of the  
 768 ROIs;  $F_s < 1$  in V1, V2, PPA, RSC, and OPA;  $F(2, 28) = 1.327, p = .281, \eta^2 = .087$  in the V3;  $F(2, 24) = 1.785, p$   
 769  $= .189, \eta^2 = .110$  in V4;  $F(2, 28) = 1.923, p = .165, \eta^2 = .121$  in the LOC; and  $F(2, 28) = 1.727, p = .196, \eta^2 = .110$  in  
 770 the FFA. The mean activity might not be sensitive enough to differentiate between the three types of line drawings  
 771 despite some variety in neural tuning properties in the visual cortex. The line drawing scenes had naturalistic visual  
 772 statistics that often contain various visual features and objects, thus equating the mean activity to an image.  
 773 Moreover, the overall neural activity for any specific line drawings were likely to be smoothed out over the course  
 774 of a block. These findings are consistent to the previous report that the multivariate approach can often recover the  
 775 visual content despite the equivalent univariate results (e.g., Kamitani & Tong, 2005).

Table S1. Results of univariate analysis. One-sample t-tests were conducted to compare average neural activity during presentation of the intact, rotated, and contour-shifted line drawings to average neural activity during fixation presentation. The mean (*M*) and standard errors of means (*SEM*), t-statistics, and the adjusted significance *q* are shown separately for each of the three contrasts: Intact > Fixation, Rotated > Fixation, and Contour-Shifted > Fixation.

ROI	df	Intact > Fixation				Rotated > Fixation				Contour-Shifted > Fixation			
		<i>M</i>	<i>SEM</i>	<i>t</i>	<i>q</i>	<i>M</i>	<i>SEM</i>	<i>t</i>	<i>q</i>	<i>M</i>	<i>SEM</i>	<i>t</i>	<i>q</i>
V1	14	.901	.152	5.923	$5.51 \cdot 10^{-5}$	.872	.146	5.972	$5.51 \cdot 10^{-5}$	.858	.158	5.442	$8.67 \cdot 10^{-5}$
V2	14	.855	.140	6.108	$5.66 \cdot 10^{-5}$	.828	.146	5.683	$5.66 \cdot 10^{-5}$	.847	.148	5.725	$5.66 \cdot 10^{-5}$
V3	14	.823	.104	7.918	$1.85 \cdot 10^{-6}$	.782	.100	7.793	$1.85 \cdot 10^{-6}$	.814	.102	8.019	$1.85 \cdot 10^{-6}$
V4	12	1.15	.127	8.391	$6.90 \cdot 10^{-6}$	1.07	.136	7.329	$9.11 \cdot 10^{-6}$	1.11	.133	7.770	$7.59 \cdot 10^{-6}$
PPA	14	.234	.0629	3.723	$5.18 \cdot 10^{-3}$	.213	.0645	3.308	$5.18 \cdot 10^{-3}$	.228	.0654	3.486	$5.18 \cdot 10^{-3}$
OPA	11	.577	.126	4.105	$1.75 \cdot 10^{-3}$	.554	.116	4.282	$1.75 \cdot 10^{-3}$	.572	.121	4.247	$1.75 \cdot 10^{-3}$
RSC	14	.0225	.134	.427	.938	.0107	.135	.242	.938	.0718	.0966	.762	.938
LOC	14	.572	.113	5.082	$3.86 \cdot 10^{-4}$	.496	.107	4.634	$3.86 \cdot 10^{-4}$	.543	.117	4.640	$3.86 \cdot 10^{-4}$
FFA	14	.371	.0907	4.088	$1.66 \cdot 10^{-3}$	.308	.0747	4.127	$1.66 \cdot 10^{-3}$	.371	.0998	3.716	$2.30 \cdot 10^{-3}$

776 Group-averaged confusion matrices of behavioral categorization and neural decoding

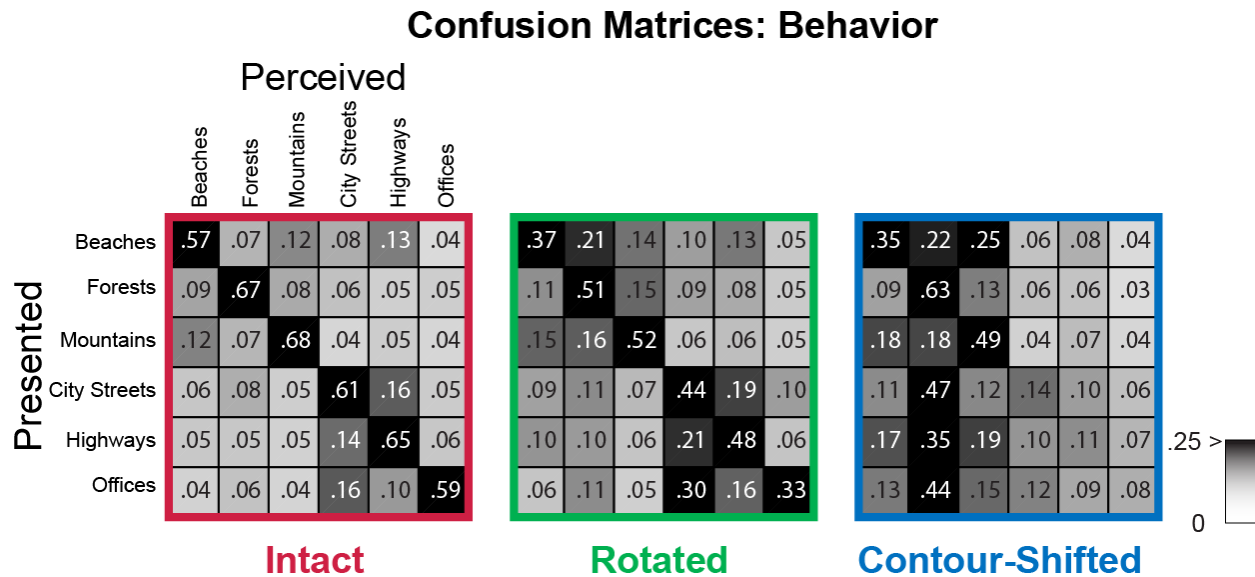


Figure S1. Group average confusion matrices of intact, rotated and contour-shifted line drawings obtained from a separate behavioral scene categorization experiment. The rows indicate true category labels presented to the participants, and columns indicate perceived category labels. In each entry, the probability of a perceived category given a presented category. The entries are also shaded according to the conditional probability: 0 as white, and 0.25 and higher as black.

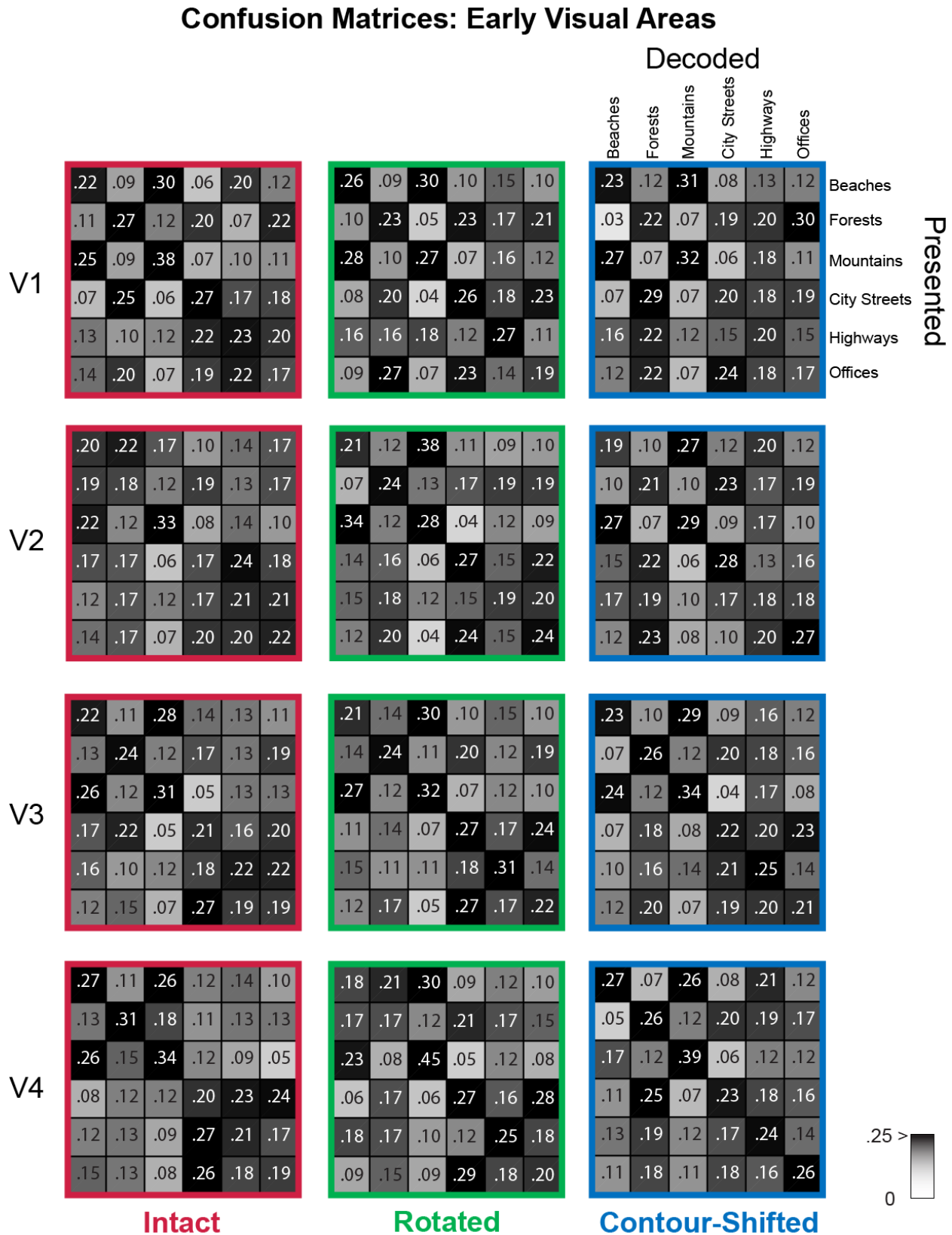


Figure S2. Group average confusion matrices of neural decoding from V1-4.

## Confusion Matrices: Domain-specific Areas

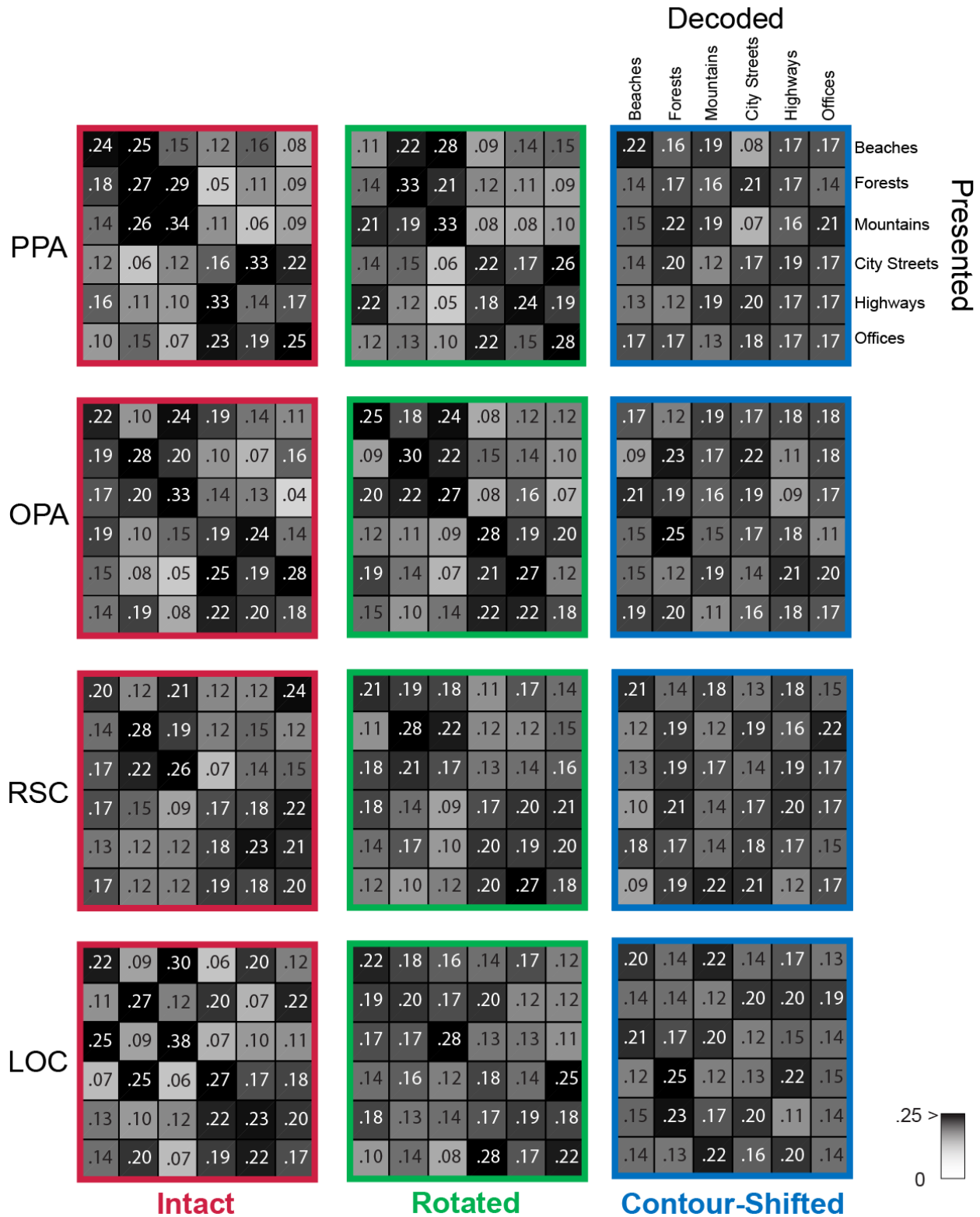


Figure S3. Group average confusion matrices of neural decoding from the PPA, OPA, RSC, and LOC.

777 Un-thresholded error correlation and max property maps

## Intact

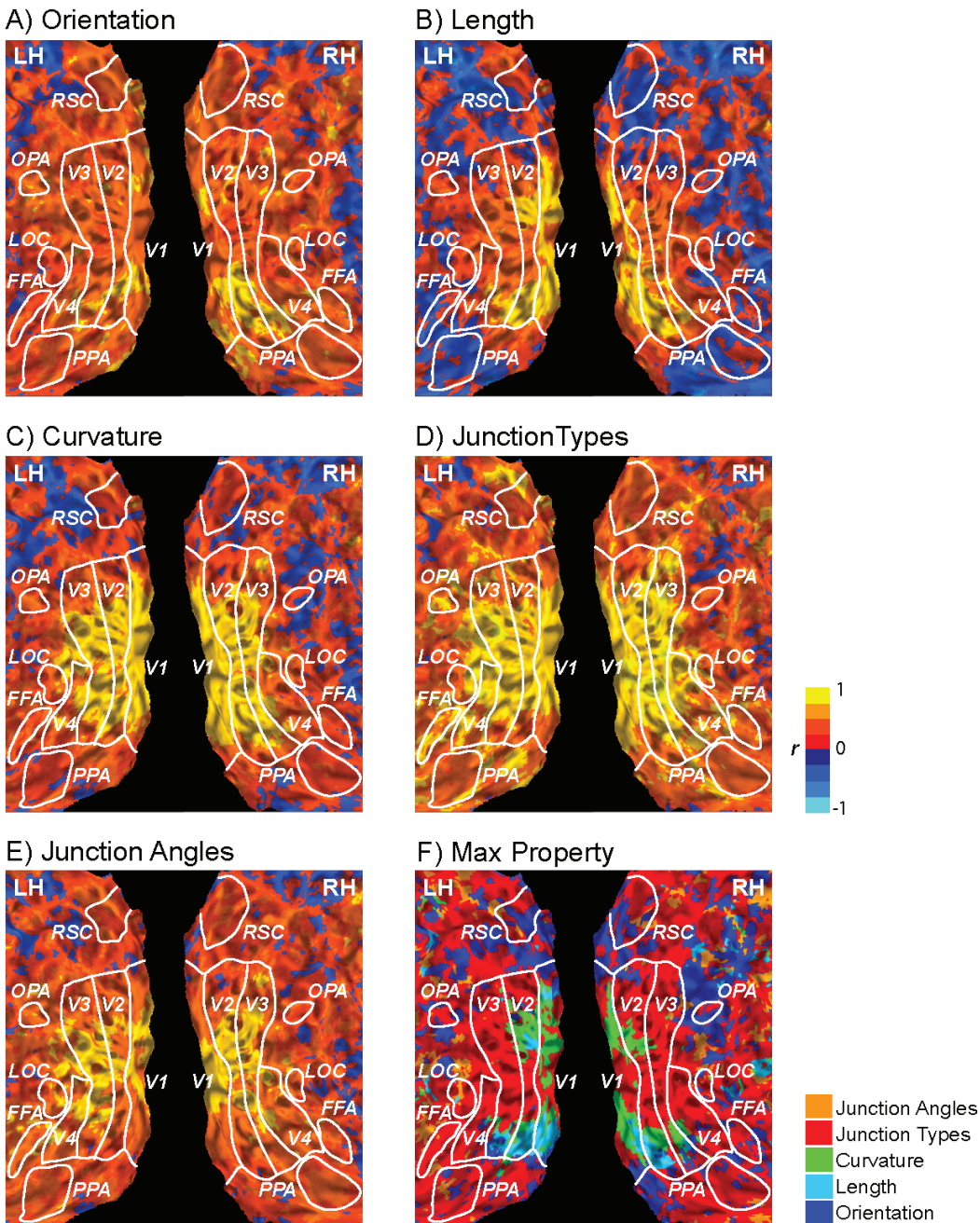
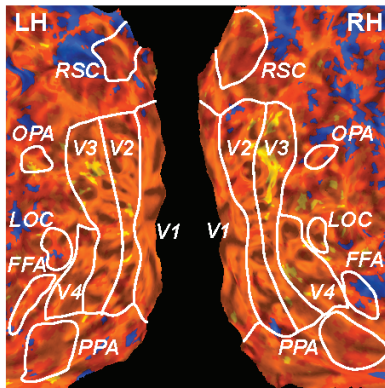


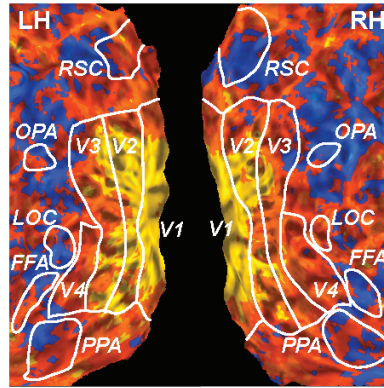
Figure S4. Un-thresholded error correlation maps between searchlight analysis of intact line drawings and computational scene categorization of the same intact line drawings. All searchlight locations were included regardless of their decoding accuracy. A-E) Searchlight locations are colored according to the strength of correlation between their neural decoding error patterns and computational error patterns (warm colors for positive and cold colors for negative correlation). F) Each searchlight location is colored according to the type of contour properties showing the maximum error correlation: orientation in dark blue, length in sky blue, curvature in green, junction types in red, and junction angles in orange.

## Rotated

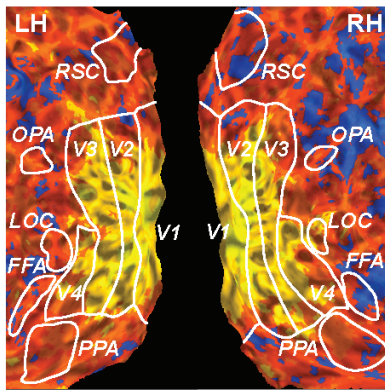
A) Orientation



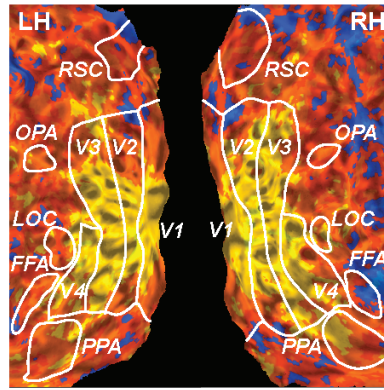
B) Length



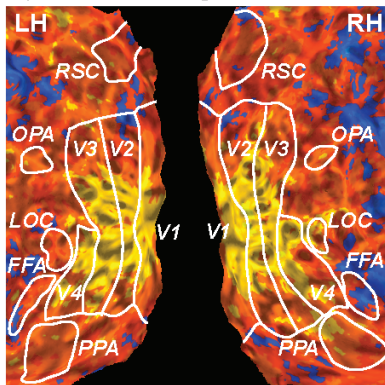
C) Curvature



D) Junction Types



E) Junction Angles



F) Max Property

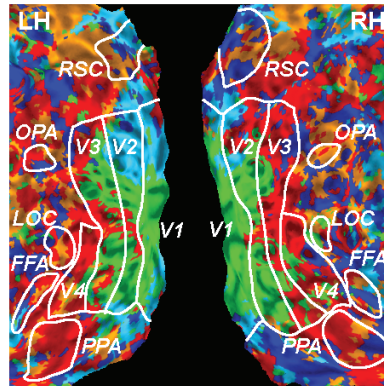
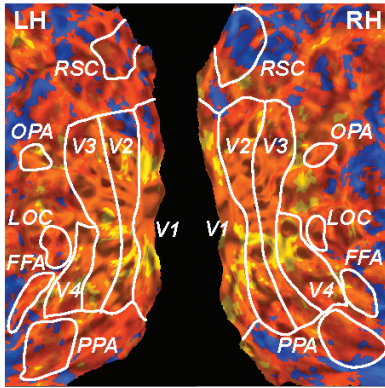


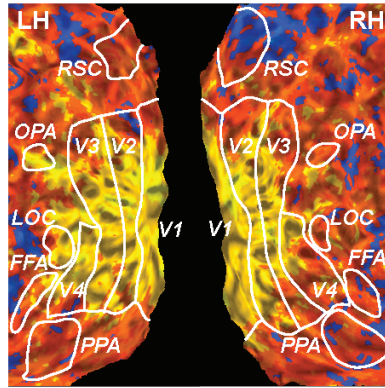
Figure S5. Un-thresholded error correlation and max property maps for rotated line drawings. Coloring follows the same conventions as Figure S4.

## Contour-Shifted

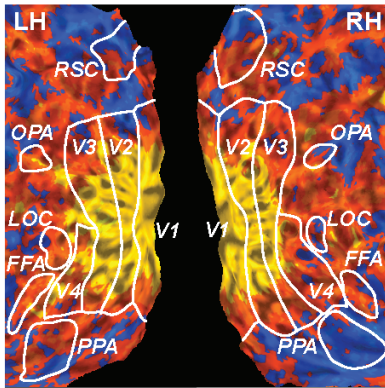
A) Orientation



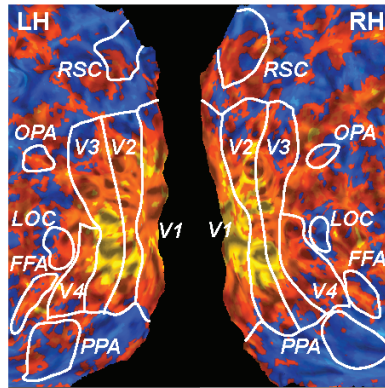
B) Length



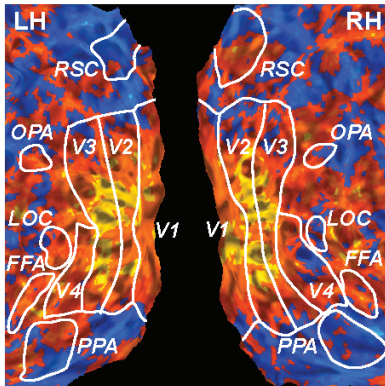
C) Curvature



D) Junction Types



E) Junction Angles



F) Max Property

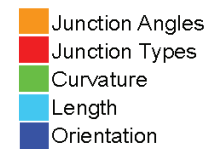
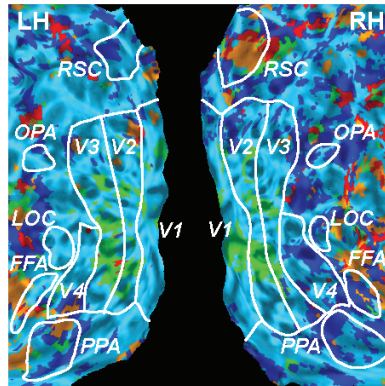


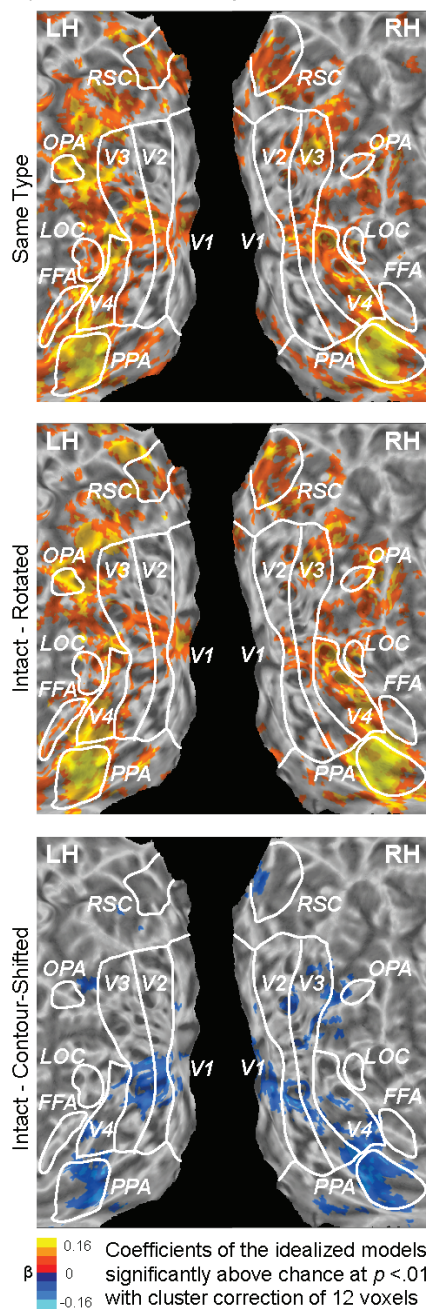
Figure S6. Un-thresholded error correlation and max property maps for contour-shifted line drawings. Coloring follows the same conventions as Figure S4.

778 **Error pattern correlation between brain and behavior: Searchlight analysis**

779 We explored how disruption in contour orientation or junction properties affected the strength of error  
780 pattern correlation between neural decoding and behavior in a voxel-wise manner, using the same linear regression  
781 analysis as used for ROI-based analysis (see Methods in the main text). The linear mixed-effects modeling was  
782 performed with the fitlme function in MATLAB R2014b for faster computation speed compared to R. For each  
783 participant, the pattern of errors from the neural decoding analysis, i.e., the off-diagonal elements of the confusion  
784 matrix, was stored at each voxel location and registered to MNI space. Separately for each of the three line drawing  
785 types, neural decoding error patterns were Pearson-correlated to the group-averaged error patterns obtained from the  
786 separate behavioral experiment, resulting in 3-by-3 error correlation values for each voxel. Error correlations were  
787 transformed using Fisher's z-transform. Using the same linear mixed-effects modeling shown in Figure 4C, we  
788 tested the extent to which the three idealized models can predict the three-by-three error correlation patterns between  
789 neural decoding and behavior. The coefficients were thresholded at  $p < .01$  (two-tailed) with cluster correction of a  
790 minimum cluster size of 12 voxels. The three coefficient maps and their overlap are shown in Figure S7. Significant  
791 contributions from all three models (positive from same-type and intact-rotated, negative from intact-contour-  
792 shifted) are clearly discernible in both PPAs and the right RSC (red in Figure S7B). The overlap between the same-  
793 type and intact-rotated models is wide-spread throughout visual cortex (orange in Figure S7B). Significant  
794 contributions from all three models (positive from same-type and intact-rotated, negative from intact-contour-  
795 shifted) are clearly discernible in both PPAs and the right RSC (red in Figure S7B).



### A) Coefficient Map



### B) Overlap between Maps

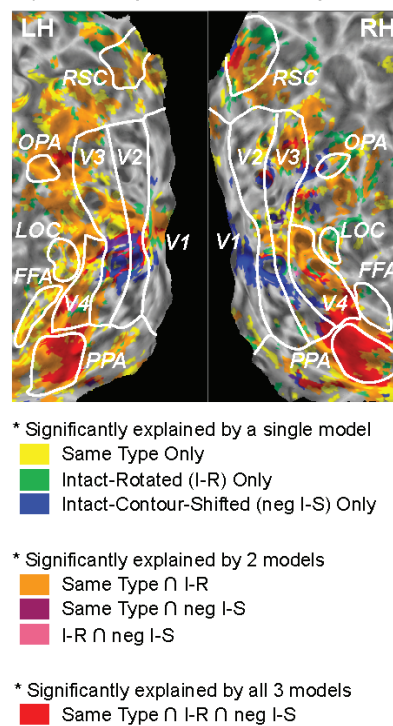
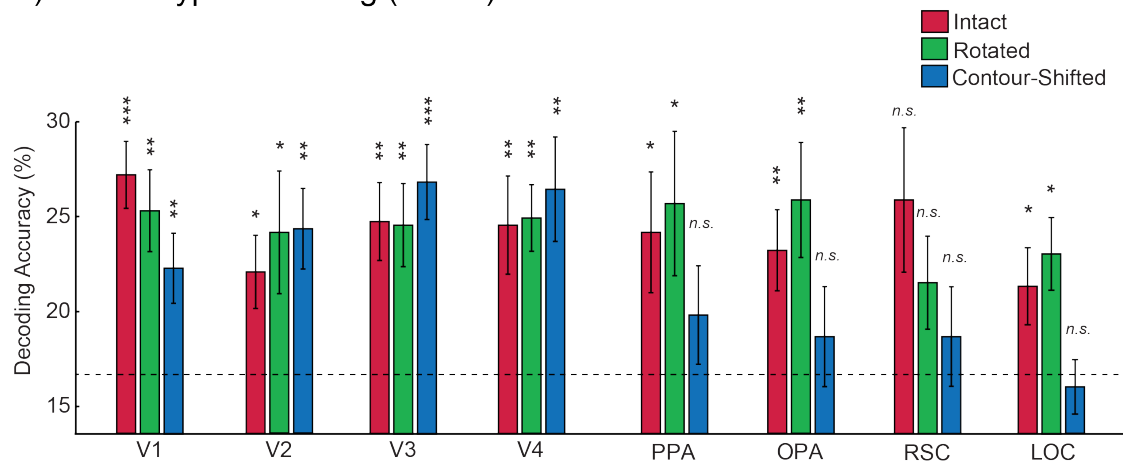


Figure S7. Searchlight analysis of patterns of error correlation between brain and behavior. At each searchlight locations the error correlation analysis were performed using the same three models used in the ROI analysis. A) The coefficient maps of the same type, intact–rotated, and intact–contour-shifted model. Each searchlight location is colored according to the coefficients, warm colors for positive and cold colors for negative. B) Each searchlight location is colored according to the set of models explaining patterns of error correlations at that location: only one model (yellow, green and blue), two of the models (orange, violet, and pink), or all three models (red). Consistent to the ROI results, the three idealized models significantly predicted

796

**Neural decoding using the equal number of participants across all ROIs**

**A) Within-Type Decoding (N = 11)**



**B) Cross-Type Decoding (N = 11)**

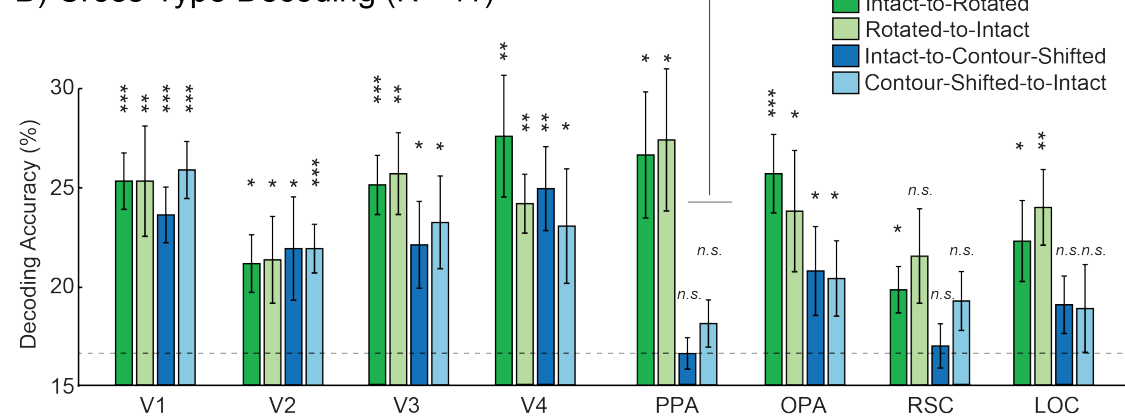
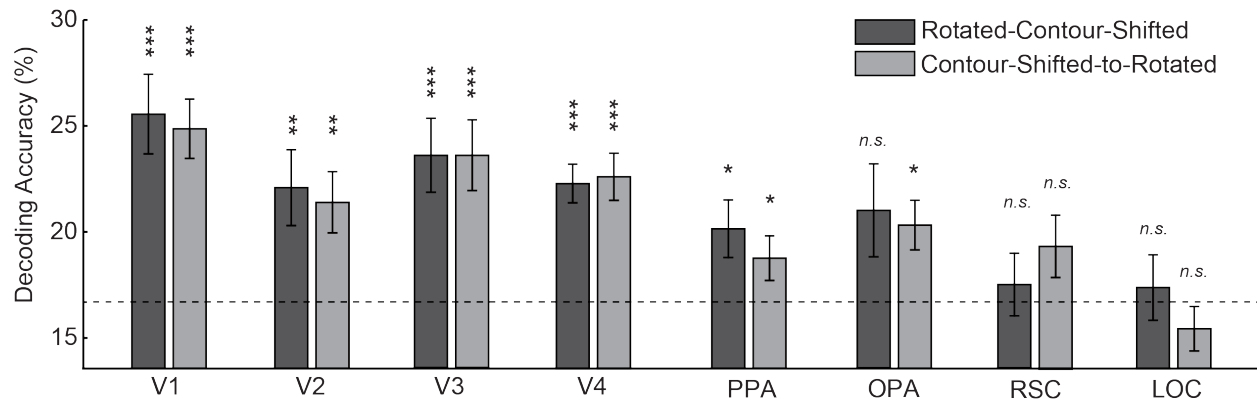


Figure S8. Neural decoding results from 11 participants in which we can delineate all nine ROIs (V1-4, PPA, OPA, RSC, LOC, and FFA). The patterns of results were identical to those from all 15 participants. A) Average accuracy rates of within-type category decoding from the ROIs. C) Average accuracy rates of cross-type category decoding from the ROIs. The only significant difference in cross-type decoding accuracy between Intact-to-Rotated/Rotated-to-Intact and Intact-to-Contour-Shifted/Contour-Shifted-to-Intact was found in the PPA,  $F(1,10) = 12.154$ ,  $p = .006$ ,  $\eta^2 = .549$  as indicated above the bracket bridging the two bars. Error bars are standard errors of means. Dashed lines indicate chance performance (1/6). The significance of the one-sample t-test (one-tailed) was adjusted for multiple comparisons (FDR) and marked above each bar, \* $q < .05$ , \*\* $q < .01$ , \*\*\* $q < .001$ .

797 **Cross-type decoding between rotated and contour-shifted line drawings**



**Figure S9.** Average accuracy rates of cross-type category decoding between rotated and contour-shifted line drawings from the ROIs. Across the two types of line drawings, statistics of contour length and curvature are preserved, which is likely to underlie the robust cross-decoding accuracy in the early visual areas as well as in the OPA. By comparison, cross-decoding in the PPA is relatively weak. Error bars are standard errors of means. Dashed lines indicate chance performance (1/6). False discovery rate correction was employed to adjust significance for multiple comparisons. The significance of the one-sample t-test (one-tailed) is marked above each bar, \* $q < .05$ , \*\* $q < .01$ , \*\*\* $q < .001$ .

798 **Supplementary References**

799

800 Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature* 392, 598 –  
801 601.

802 Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not  
803 generic within-category identification. *Nature Neuroscience*, 7(5) 555-562.

804 Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The Lateral Occipital Complex and its Role in Object  
805 Recognition. *Vision Research*. 41, 1409-1422.

806 Hansen, K. A., Kay, K. N., & Gallant, J. L. (2007). Topographic organization in and near human visual area V4.  
807 *Journal of Neuroscience*, 27, 11896–11911.

808 Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, (1998). Mechanisms of directed attention in the human  
809 extrastriate cortex as revealed by functional MRI. *Science*, 282, 108–111.

810 Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature*  
811 *Neuroscience*, 8, 679-585.