UNIVERSITY *of* York

This is a repository copy of *Patterns of response to scrambled scenes reveal the importance of visual properties in the organization of scene-selective cortex*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/115812/

Version: Accepted Version

Article:

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Accepted Manuscript

Patterns of response to scrambled scenes reveal the importance of visual properties in the organization of scene-selective cortex

David M. Watson, Tom Hartley, Timothy J. Andrews

Please cite this article as: Watson DM, Hartley T, Andrews TJ, Patterns of response to scrambled scenes reveal the importance of visual properties in the organization of scene-selective cortex, *CORTEX* (2017), doi: 10.1016/j.cortex.2017.04.011.

PATTERNS OF RESPONSE TO SCRAMBLED SCENES REVEAL THE IMPORTANCE OF VISUAL PROPERTIES IN THE ORGANIZATION OF SCENE-SELECTIVE CORTEX

David M. Watson, Tom Hartley, Timothy J. Andrews*

Department of Psychology and York Neuroimaging Centre,

University of York, York, YO10 5DD, United Kingdom

* Corresponding author: timothy.andrews@york.ac.uk

Key words: Scene, Cortex, Category, PPA, RSC, OPA, topographic

1 **ABSTRACT**

2 **Neuroimaging studies have found distinct patterns of neural response to different**

3 **categories of scene in scene-selective regions of the human brain. However, it is not clear**

4 **how information about scene category is represented in these regions. Images from**

5 **different categories vary systematically in their visual properties as well as their semantic**

6 **category. So, it is possible that patterns of neural response could reflect variation in visual**

7 **properties. To address this question, we used fMRI to measure patterns of neural**

8 **response to intact and scrambled scene categories. Although scrambling preserved many**

9 **of their visual characteristics, perception of scene categories was severely impaired.**

10 **Nevertheless, we found distinct patterns of response to different scene categories in the**

11 **parahippocampal place area (PPA) and the occipital place area (OPA) for both intact and**

12 **scrambled scenes. Moreover, intact and scrambled scenes produced highly similar**

13 **patterns of response. Our finding that reliable and distinct patterns of response in scene-**

14 **selective regions are still evident when categorical perception is impaired suggests that**

15 **visual properties play an important role in the topographic organization of these regions.**

16

17 **INTRODUCTION**

18 The ability to perceive and recognize the spatial layout of visual scenes is essential for

19 spatial navigation. Neuroimaging studies have identified a number of regions in the human

20 brain that respond selectively to visual scenes (Epstein, 2008). For example, the

21 parahippocampal place area (PPA) is a region on the ventral surface of the temporal lobe

22 that displays preferential activity to images of scenes over and above images of objects and

23 faces (Aguirre, Zarahn, & D'Esposito, 1998; Epstein & Kanwisher, 1998). Other place

24 selective regions include the retrosplenial complex (RSC) located immediately superior to

25 the PPA and the transverse occipital sulcus (TOS) or occipital place area (OPA) on the lateral

26 surface of the occipital lobe (Dilks, Julian, Paunov, & Kanwisher, 2013). Damage to these

27 regions leads to specific impairments in scene perception and spatial navigation (Aguirre &

28 D'Esposito, 1999; Mendez & Cherrier, 2003).

29 Despite the importance of scene-selective regions for spatial navigation, the

30 functional organisation of these regions remains unclear (Lescroart, Stansbury, & Gallant,

31 2015; Groen et al., 2017). For example, although scene-selective regions show distinct

32 patterns of response to images of different scene categories (Walther, Caddigan, Fei-Fei, &

33 Beck, 2009; Watson, Hartley, & Andrews, 2014), the basic organizing principles are

34 unresolved. Some studies have argued that scene-selective regions represent information

35 about 'high-level' semantic properties of natural scenes (Huth, Nishimoto, Vu, & Gallant,

36 2012; Stansbury, Naselaris, & Gallant, 2013; Walther et al., 2009; Walther, Chai, Caddigan,

37 Beck, & Fei-Fei, 2011). This conclusion has, however, been challenged by other studies that

38 have suggested that the patterns of response in scene-selective regions are better explained

39 by properties of the scene, such as openness (Kravitz, Peng, & Baker, 2011; Park, Brady,

40 Greene, & Oliva, 2011) or distance (Amit, Mehoudar, Trope, & Yovel, 2012; Park, Konkle, &

41 Oliva, 2015) rather than by semantic category.

42   Although concepts such as openness or distance provide plausible 'mid-level'

43 dimensions with which to understand the organization of scene-selective regions, it is not

44 clear whether they can be explained at an even more basic level in terms of low-level visual

45 properties that co-vary with these properties (Oliva & Torralba, 2001). In recent studies, we

46 have shown that variance in the patterns of response to different scene categories can be

47 explained by corresponding variance in the image properties of the scenes (Andrews,

48 Watson, Rice, & Hartley, 2015; Watson et al., 2014; Watson, Hymers, Hartley, & Andrews,

49 2016). These findings are consistent with previously reported biases in scene-selective

50 regions for orientation (Nasr, Echavarria, & Tootell, 2014; Nasr & Tootell, 2012), spatial

51 frequency (Musel et al., 2014; Rajimehr, Devaney, Bilenko, Young, & Tootell, 2011) and

52 visual field location (Arcaro, McMains, Singer, & Kastner, 2009; Golomb & Kanwisher, 2012;

53 Levy, Hasson, Avidan, Hendler, & Malach, 2001; Silson, Chan, Reynolds, Kravitz, & Baker,

54 2015) and provide further evidence for the role of image properties in the organization of

55 scene-selective regions. However, a fundamental problem is that images drawn from the

56 same scene category or with the same spatial layout are likely to have similar visual

57 properties (Oliva & Torralba, 2001). So, reliable patterns of response are expected under

58 high-level, mid-level and low-level accounts of scene perception.

59   The aim of this study was to directly determine the extent to which the patterns of

60 neural response across scene-selective regions can be explained by selectivity to more basic

61 properties of the stimulus. To address this question, we measured the neural response

62 across scene-selective regions to intact images of different scene categories, as well as

63 versions of these images that had been phase-scrambled at a global or local level. Our

3

64   rationale for using scrambled images is that they have many of the visual properties found

65   in intact images, but disrupt perception of categorical and semantic information (Andrews,

66   Clarke, Pell, Hartley, 2010; Coggan, Liu, Baker, & Andrews, 2016; Loschky et al., 2007;

67   Loschky, Hansen, Sethi, & Pydimarri, 2010). Applying scrambling both locally and globally

68   allowed us to further investigate the importance of the spatial properties of scenes to the

69   neural response, as local scrambling better preserves the coarse-scale spatial arrangement

70   of visual features in the original image. Our hypothesis was that, if scene-selective regions

71   are sensitive to the visual differences between scene categories, then we would expect to

72   find similar patterns of neural response to these categories even when images are

73   scrambled.

74

75 **METHODS**

76 *Participants*

77 20 participants (5 males; mean age: 25.85; age range: 19-34) took part in the experiment.

78 All participants were neurologically healthy, right-handed, and had normal or corrected-to-

79 normal vision. Written consent was obtained for all participants and the study was approved

80 by the York Neuroimaging Centre Ethics Committee.

81

82 *Stimuli*

83 Participants viewed scene images in two independent runs, one to localize the scene-

84 selective regions, the other to experimentally investigate the effects of local and global

85 scrambling manipulations. Images presented in the experiment runs were taken from the

86 LabelMe database (http://cvcl.mit.edu/database.htm; Oliva & Torralba, 2001). Images for

87 the localiser run were taken from the SUN database

88 (http://groups.csail.mit.edu/vision/SUN/; Xiao, Hays, Ehinger, Oliva, & Torralba, 2010).

89 Stimuli were presented using PsychoPy (Peirce, 2007, 2009) and were back-projected onto a

90 custom in-bore acrylic screen at a distance of approximately 57 cm from the participant,

91 with all images presented at a resolution of 256x256 pixels subtending approximately 10.7°

92 of visual angle.

93　　The image set for the main experiment comprised 180 greyscale images from 5

94 scene categories: city, coast, forest, indoor, and mountain (36 images per category). Each

95 image was shown at 3 levels of image scrambling: intact, locally scrambled, and globally

96 scrambled. Globally scrambled images were created by randomising the phase of the 2D

97 frequency components across the whole image while keeping the magnitude constant.

5

98   Locally scrambled images were created by the same process, except that scrambling was

99   applied independently within each of 64 windows of an 8x8 grid across the image.

100 Luminance histograms across all images in all conditions were normalised using the SHINE

101 toolbox (Willenbockel et al., 2010). Examples of the stimuli used in each condition are

102 shown in Figure 1. Corresponding Fourier amplitude spectra plots are shown in

103 Supplementary Figure 1. In order to assess the impact of the scrambling process on the

104 visual similarity of the scene categories, we assessed the visual statistics of the images using

105 the GIST descriptor (Oliva & Torralba, 2001). This generates a vector for each image

106 describing the spectral energy at assorted spatial frequencies, orientations, and spatial

107 positions within the image. We employed 32 filters spanning 8 orientations and 4 spatial

108 frequencies, within 64 windows of an 8x8 spatial grid, yielding vectors of 2048 values. These

109 vectors were then correlated within- and between-categories using a leave-one-image-out

110 cross-validation procedure for each scrambling condition independently (Supplementary

111 Figure 2a). The resulting similarity matrices are shown in Supplementary Figure 2a. We next

112 tested the ability to distinguish scene categories based on this visual information by

113 contrasting the within- over the between-category correlations (Supplementary Figure 2b).

114 Significantly greater within- than between-category correlations were observed for the

115 intact ($t(35)$ = 29.44, $p < .001$, Cohen's d = 4.91), locally scrambled ($t(35)$ = 25.57, $p < .001$,

116 Cohen's d = 4.26), and globally scrambled scenes ($t(35)$ = 18.69, $p < .001$, Cohen's d = 3.11).

117 Thus, the scene categories remained visually distinct under all conditions of scrambling.

118   The localiser images comprised a separate set of 64 scene images plus their phase

119 scrambled counterparts (128 images total), with all images presented in full colour. Images

120 were chosen in approximately equal number from categories of indoor-manmade, outdoor-

6

121   manmade, and outdoor-natural scenes as these represent the 3 top-level branches of the

122   SUN database hierarchy. Fourier-scrambled images were created by randomising the phase

123   of the 2D frequency components in each colour channel of the original image while keeping

124   the magnitude constant. Mean luminance was then equated across images.

125   [Figure 1 near here]

126   *fMRI Experimental Design*

127   During the experimental runs participants viewed images from the 5 scene categories.

128   Images from each level of image scrambling were presented across separate experiment

129   runs. For all participants, globally scrambled images were presented in the first run, locally

130   scrambled in the second run, and intact images in the third run. This order was chosen as it

131   was crucial to ensure that responses to scrambled scenes could not be primed by earlier

132   viewing of the intact versions.

133   In each run, images from each category were presented in a blocked design. There

134   were 6 images in each block. Each image was presented for 750ms followed by a 250ms

135   grey screen that was equal in mean luminance to the scene images. Each stimulus block

136   was separated by a 9s period in which the same grey screen as used in the inter-stimulus

137   interval was presented. Each condition was repeated 6 times (total 30 blocks) in each run.

138   To maintain attention throughout the experimental runs, participants had to detect the

139   presence of a red dot randomly superimposed on one of the images in each block,

140   responding via a button press.

141   To define scene-selective regions, independent data was collected while participants

142   viewed images from 2 stimulus conditions (intact scenes, scrambled scenes). Images from

143   each condition were presented in a blocked fMRI design, with each block comprising 9

144   images. Each condition was repeated 8 times (16 blocks). In each stimulus block, an image

145   was presented for 750ms followed by a 250ms grey screen. Each stimulus block was

146   separated by a 9s period in which a grey screen was presented. Participants performed a

147   one-back task that involved pressing a button when they detected a repeated image in each

148   block.

149

150   *Imaging Parameters*

151   All scanning was conducted at the York Neuroimaging Centre (YNiC) using a GE 3 Tesla HDx

152   Excite MRI scanner. Images were acquired with an 8-channel phased-array head coil tuned

153   to 127.72MHz. Data were collected from 38 contigual axial slices in an interleaved order via

154   a gradient-echo EPI sequence (TR = 3s, TE = 32.5ms, FOV = 288x288mm, matrix size =

155   128x128, voxel dimensions = 2.25x2.25 mm, slice thickness = 3mm with no inter-slice gap,

156   flip angle = 90°, phase-encoding direction = anterior-posterior, pixel bandwidth = 39.06 kHz).

157   In order to aid co-registration to structural images, T1-weighted in-plane FLAIR images were

158   acquired (TR = 2.5s, TE = 9.98ms, FOV = 288x288mm, matrix size = 512x512, voxel

159   dimensions = 0.56x0.56 mm, slice thickness = 3mm, flip angle = 90°). Finally, high-resolution

160   T1-weighted structural images were acquired (TR = 7.96ms, TE = 3.05ms, FOV =

161   290x290mm, matrix size = 256x256, voxel dimensions = 1.13x1.13 mm, slice thickness =

162   1mm, flip angle = 20°).

163

164   *fMRI Analysis*

165   Univariate analyses of the fMRI data were performed with FEAT v5.98

166   (http://www.fmrib.ox.ac.uk/fsl). In all scans the initial 9s of data were removed to reduce

167 the effects of magnetic stimulation. Motion correction (MCFLIRT, FSL; Jenkinson, Bannister,

168 Brady, & Smith, 2002) was applied followed by temporal high-pass filtering (Gaussian-

169 weighted least-squared straight line fittings, sigma=15s). Spatial smoothing (Gaussian) was

170 applied at 6mm FWHM to both the localiser and experiment runs, in line with previous

171 studies employing smoothing in conjunction with MVPA (Op de Beeck, 2010; Watson et al.,

172 2014).   Parameter estimates were generated for each condition by regressing the

173 hemodynamic response of each voxel against a box-car convolved with a single-gamma HRF.

174 Next, individual participant data were entered into higher-level group analyses using a

175 mixed-effects design (FLAME, FSL). Functional data were first co-registered to an in-plane

176 FLAIR anatomical image then to a high-resolution T1-anatomical image, and finally onto the

177 standard MNI brain (ICBM152).

178   Scene selective regions of interest (ROIs) were defined from the localiser data of

179 both experiments.   ROIs were defined for the parahippocampal place area (PPA),

180 retrosplenial complex (RSC), and occipital place area (OPA) that have been reported in

181 previous fMRI studies (Dilks et al., 2013; Epstein & Kanwisher, 1998; Maguire, 2001). The

182 locations of these ROIs were consistent with those reported in previous literature – see

183 Supplementary Table 1. Within the MNI-2x2x2mm space, seed points were defined at the

184 peak voxels within the intact>scrambled statistical map for each region (PPA, RSC, OPA) in

185 each hemisphere. For a given seed, a flood fill algorithm was used to identify a cluster of

186 spatially contiguous voxels around that seed which exceeded a given threshold.   This

187 threshold was then iteratively adjusted till a cluster size of approximately 500 voxels was

188 achieved (corresponding to a volume of 4000mm^3); actual cluster sizes ranged from 499-502

189 voxels as an optimal solution to the algorithm was not always achievable. This step ensures

190 that estimates of multi-voxel pattern similarity are not biased by the different sizes of ROIs

191 being compared. Clusters were combined across hemispheres to yield 3 ROIs, each

192 comprising approximately 1000 voxels. These regions are shown in Supplementary Figure 3,

193 and MNI co-ordinates of the seeds are given in Table 1. For comparison, we defined two

194 alternative versions of each of the scene ROIs using the same clustering method, based

195 upon independent localiser data from other experiments (not reported here). Specifically,

196 regions were defined using responses from contrasts of 1) Scenes > Faces, and 2) Scenes >

197 Objects. The locations of these regions are shown in Supplementary Figure 8, and MNI co-

198 ordinates of the seeds are given in Supplementary Table 2. In addition, a V1 control ROI was

199 defined from a recent standard atlas of retinotopic regions (Wang, Mruczek, Arcaro, &

200 Kastner, 2015).

201 Next, we measured patterns of response to different stimulus conditions in each

202 ROI. Parameter estimates were generated for each condition in the experimental scans.

203 The reliability of response patterns was tested using a leave-one-participant-out (LOPO)

204 cross-validation paradigm (Poldrack, Halchenko, & Hanson, 2009; Shinkareva et al., 2008) in

205 which parameter estimates were determined using a group analysis of all participants

206 except one. This generated parameter estimates for each scene condition in each voxel.

207 This LOPO process was repeated such that every participant was left out of a group analysis

208 once. These data were then submitted to correlation-based pattern analyses (Haxby et al.,

209 2001; Haxby, Connolly, & Guntupalli, 2014) implemented using the PyMVPA toolbox

210 (http://www.pymvpa.org/; Hanke et al., 2009). Parameter estimates were normalised by

211 subtracting the voxel-wise mean response across all experimental conditions (Haxby et al.,

212 2001). For each iteration of the LOPO cross-validation, the normalized patterns of response

213 to each stimulus condition were correlated between the group and the left-out participant.

214   This allowed us to determine whether there are reliable patterns of response that are

215   consistent across individual participants.

216   *Statistical Analyses*

217   A Fisher's z-transform was applied to the correlation similarity matrices before further

218   statistical analyses. We tested whether scene categories could be distinguished on the basis

219   of the pattern of activity within each region to under each level of image scrambling. For

220   each iteration of the LOPO cross-validation, we calculated an average within-category (on-

221   diagonal) and an average between-category (off-diagonal) value across categories. These

222   values were then entered into a paired-samples t-test.   If scene category can be

223   discriminated based on the pattern of activity it elicits, then significantly greater within-

224   than between-category correlations would be expected.   For the scene regions, a

225   Bonferroni-Holm correction for multiple comparisons was applied across the 3 regions (PPA,

226   RSC, OPA) and 3 scrambling conditions (intact, locally scrambled, globally scrambled). The

227   V1 ROI represents a control analysis and hence was handled separately; here a Bonferroni-

228   Holm correction for multiple comparisons was applied across the 3 scrambling conditions. A

229   possible caveat here is that the leave-one-out procedure means that samples from each

230   iteration are not truly independent, potentially violating the statistical assumptions of the t-

231   test. To address this we repeated these analyses using a sign-flip permutation test on the

232   differences between the scores. The results of these analyses closely followed those of the

233   parametric t-tests – see Supplementary Table 3.

234   Next, we conducted a series of representational similarity analyses (RSAs;

235   Kriegeskorte, Mur, & Bandettini, 2008) to investigate effects of different levels of

236   scrambling. Correlation matrices were averaged across iterations of the cross-validation.

11

237   Representational similarity was assessed by correlating the off-diagonal elements of the

238   averaged similarity matrices between the intact and locally scrambled conditions, and

239   between the intact and globally scrambled conditions. If the scrambling does not abolish

240   the pattern of relative similarity between categories relative to the intact condition, then a

241   significant positive correlation would be expected between the intact and corresponding

242   scrambled matrices. For the scene regions, a Bonferroni-Holm correction was applied

243   across the 3 regions (PPA, RSC, OPA) and 2 analyses (intact versus locally scrambled, intact

244   versus globally scrambled). The V1 ROI represents a control analysis and hence was handled

245   separately; here a Bonferroni-Holm correction for multiple comparisons was applied across

246   the 2 analyses.

247   To test for effects outside our ROIs, we also performed a series of whole-brain

248   searchlight analyses (Kriegeskorte, Goebel, & Bandettini, 2006). A spherical ROI (6mm

249   radius) was iterated over the whole-brain volume, and the MVPA repeated within each

250   sphere. Decoding and representational similarity analyses were conducted in the same

251   manner as for the ROI analyses. For the decoding analysis, for a given sphere an average

252   within- and between-category correlation value was calculated for each LOPO iteration, and

253   then a paired-samples t-test used to test the within > between difference across LOPO

254   iterations. For the representational similarity analyses, for a given sphere the correlation

255   matrices were averaged across LOPO iterations and the off-diagonal elements correlated

256   between the scrambling conditions. In both cases, the p-value of the test was then assigned

257   to the central voxel of the sphere.

258 *Behavioural Experiment*

259 We also tested the ability of participants to recognise the scenes under each level of image

260 scrambling. An independent set of 18 participants naive to the purposes of the study were

261 recruited (6 males; mean age: 21.7; age range: 19-39). Written consent was obtained for all

262 participants and the study was approved by the University of York Psychology Department

263 Ethics Committee. Each participant viewed a subset of 1/6th of the image set comprising 6

264 images from each category. Subsets were counterbalanced across participants. Participants

265 viewed each image under all three levels of scrambling. Crucially, to prevent priming effects,

266 participants viewed globally scrambled images first, followed by locally scrambled images,

267 and finally intact images (as per the fMRI experiment). In each trial participants were shown

268 an image for 750ms, and were then prompted to describe the type of scene they thought

269 was shown, typing their responses. The stimulus duration was chosen to match that of the

270 fMRI experiment. Participants were free to provide any description they wanted, and were

271 also informed that they did not have to give a response if they could not reasonably see

272 what type of scene was depicted. Accuracy was coded manually by two independent raters

273 (both authors of the study). A correct response was defined as any which could reasonably

274 be seen to accurately describe the corresponding intact scene, while an incorrect response

275 was defined as one that did not accurately describe the intact scene or where no response

276 was given. Accuracies were converted to proportions and an arcsine square-root transform

277 was applied prior to further statistical tests. If participants did provide a description, they

278 were next prompted to provide a confidence rating of their decision on a 7 point scale (not

279 at all confident - very confident). No confidence ratings were collected for trials where

280 participants did not provide descriptive responses. Participants were not provided with any

281 information about the scene categories prior to the experiment – this was necessary in

282   order to match the design of the fMRI experiment, where participants were not provided

283   with any information about the structure of the stimulus set beforehand either.

284

285 **RESULTS**

286 *Behavioural Experiment*

287 We tested the effects of the different levels of scrambling on participants' ability to

288 recognise the scenes. Two independent raters (both authors) coded the descriptive

289 responses for accuracy. Inter-rater reliability was high across the subjects (mean Cohen's

290 kappa = .96 ± .01). For all subsequent tests, accuracy values were averaged between the

291 raters. Mean accuracy for each condition is shown in Figure 2a. As expected, accuracy was

292 higher for intact (mean = 98.33 ± 0.80%) compared to locally scrambled (mean = 20.20 ±

293 2.54%) and globally scrambled images (mean = 3.35 ± 0.82%). A one-way repeated

294 measures ANOVA revealed a significant main effect of scrambling ($F(2,34)$ = 374.76, *p* <

295 .001, generalized-η^2 = .95). A series of post-hoc t-tests revealed significantly higher

296 accuracies for intact compared to locally scrambled scenes, intact compared to globally

297 scrambled scenes, and locally scrambled compared to globally scrambled scenes (all *p* <

298 .001). For trials where descriptive responses were given, participants also provided

299 confidence ratings of their descriptions on a scale of 1 (not at all confident) to 7 (very

300 confident). Median ratings for each condition were calculated for each participant and are

301 shown in Figure 3b. One participant's data were excluded from the analysis as they provided

302 no responses, and hence no confidence ratings, for the scrambled images. Similar to

303 accuracy, confidence ratings were higher for intact (median = 7, IQR = 6 - 7) compared to

304 locally scrambled (median = 2, IQR = 2 - 3) and globally scrambled images (median = 2, IQR =

305 1 - 2). A Friedman's ANOVA revealed a significant main effect of scrambling ($\chi^2(2)$ = 31.60, *p*

306 < .001). A series of post-hoc Wilcoxon signed-rank tests revealed significantly higher

307 confidence ratings for intact than locally scrambled scenes (*p* < .001), intact than globally

308 scrambled scenes (*p* < .001), and locally scrambled than globally scrambled scenes (*p* =

309 .004). Thus both types of scrambling significantly impaired participants' recognition and

310 confidence on a scene recognition test.

311 [Figure 2 near here]

312 *Scene Decoding Analysis*

313 Next, we used fMRI to measure the patterns of neural response to each of the conditions.

314 The group normalised responses within the PPA, RSC, and OPA regions are shown in

315 Supplementary Figure 4 (red and blue colours indicate responses above and below the

316 mean respectively). Correlation-based MVPA (Haxby et al., 2001) using a leave-one-

317 participant-out (LOPO) cross-validation scheme was then used to assess the reliability of

318 these responses. Average correlation similarity matrices for each of the ROIs and each of

319 the scrambling types are shown in Figure 3, with symmetrically opposite points averaged

320 across the diagonal to aid visualisation.

321 [Figure 3 near here]

322 We first assessed the ability of the MVPA to decode the scene categories under each

323 of the levels of scrambling. We calculated within- and between-category correlation values

324 averaged across categories for each scrambling type and ROI. These values are shown in

325 Figure 4. Paired-samples t-tests were then used to test for differences between within- and

326 between-category correlations; if categories can be decoded based on patterns of brain

327 activity, then significantly greater within- than between-category correlations would be

328 expected. For the intact scenes, significantly greater within- than between-category

329 correlations were observed in the PPA (t(19) = 10.90, *p* < .001, Cohen's d = 2.44) and OPA

16

330 (t(19) = 9.89, *p* < .001, Cohen's d = 2.21), but not in the RSC (t(19) = 0.17, *p* > .999, Cohen's d

331 = 0.04). In the locally scrambled condition, significantly greater within- than between-

332 category correlations were found in the PPA (t(19) = 5.54, *p* < .001, Cohen's d = 1.24) and

333 OPA (t(19) = 4.57, *p* = .001, Cohen's d = 1.02), but not in the RSC (t(19) = 1.43, *p* = .498,

334 Cohen's d = 0.32). For the globally scrambled scenes, no significant differences were seen

335 for any ROI (PPA: t(19) = 0.43, *p* > .999, Cohen's d = 0.10; RSC: t(19) = 2.20, *p* = .200, Cohen's

336 d = 0.49; OPA: t(19) = 2.14, *p* = .200, Cohen's d = 0.48).

337   A further test of the similarity in response between scrambling conditions is the

338 extent to which neural response patterns generalise across them. This was tested using

339 cross-decoding analyses. MVP analyses were conducted in which the neural response

340 patterns to intact scenes were now correlated with the neural response patterns to 1) the

341 locally scrambled scenes, and 2) the globally scrambled scenes. If response patterns to a

342 given scene category remain similar across the scrambling conditions, then significant

343 decoding of the scene categories from these cross-condition MVP analyses would be

344 expected. The results of these cross-decoding analyses are shown in Supplementary Figure

345 7. The comparison of intact and locally scrambled scenes revealed significant decoding of

346 scene category in the PPA (t(19) = 8.13, *p* < .001, Cohen's d = 1.82) and OPA (t(19) = 7.13, *p* <

347 .001, Cohen's d = 1.59), but not the RSC (t(19) = 1.08, *p* = .583, Cohen's d = 0.24). Similarly,

348 the comparison of intact and globally scrambled scenes also revealed significant decoding of

349 scene category in the PPA (t(19) = 5.62, *p* < .001, Cohen's d = 1.26) and OPA (t(19) = 5.82, *p* <

350 .001, Cohen's d = 1.30), but not the RSC (t(19) = 0.45, *p* = .655, Cohen's d = 0.10). Thus,

351 response patterns in PPA and OPA generalised well between intact and locally scrambled,

352 and intact and globally scrambled conditions.

17

353   [Figure 4 near here]

354   *Representational Similarity Analysis*

355   We next conducted a series of representational similarity analyses (RSAs; Kriegeskorte et al.,

356   2008) to test to what extent the two types of scrambling influence the representational

357   structure of the responses relative to those of the intact scenes. The off-diagonal elements

358   of the group average matrices (20 elements per matrix) were correlated between intact and

359   locally scrambled conditions, and intact and globally scrambled conditions. If the scrambling

360   does not disrupt the representational space, a significant positive correlation would be

361   expected with the intact scenes matrix. A significant positive correlation was observed

362   between intact and locally scrambled scenes in the PPA ($r(18)$ = .66, p = .009), but not in the

363   OPA ($r(18)$ = -.15, p > .999), whilst a significant negative correlation was observed in the RSC

364   ($r(18)$ = -.56, p = .044). A significant positive correlation was observed between intact and

365   globally scrambled conditions in the OPA($r(18)$ = .62, p = .019), but not the PPA ($r(18)$ = .44,

366   p = .160)or RSC($r(18)$ = .02, p > .999). These results are illustrated in Figure 5 (see also

367   Supplementary Figure 6).

368   [Figure 5 near here]

369   To further quantify the degree of preserved pattern similarity under scrambling we

370   undertook an additional analysis of representational similarity, taking into account

371   individual variation and the distribution of correlations this entails (Supplementary Figure

372   7). Such variation leads to a "noise ceiling" (Nili et al., 2014), i.e., an upper bound to the

373   observable correlation between intact and scrambled conditions. By comparing the

374   observed correlations with the noise ceiling, we can determine the degree to which

375   preserved representational structure under scrambled conditions accounts for the

18

376　explicable variance in the data. This approach also permits a more sensitive comparison

377　with a zero correlation, which would be expected if scrambling abolished the

378　representational structure for intact images.

379　　　　The noise ceiling is estimated by correlating each LOPO iteration's intact similarity

380　matrix against the group average intact similarity matrix (calculated across all LOPO

381　iterations for the noise ceiling upper bound, and across all LOPO iterations but the current

382　one for the noise ceiling lower bound), and then averaging these correlations. This reflects

383　the maximum similarity that could be expected for any correlation between the intact and

384　scrambled conditions. Noise ceilings were reasonably high in the PPA and OPA indicating a

385　good degree of reliability in the intact responses across LOPO iterations, but were much

386　closer to zero in the RSC indicating relatively poor reliability in this region.

387　　　　Next, we calculated the correlation between each LOPO iteration's locally- or

388　globally-scrambled similarity matrix and the group average intact similarity matrix. A one-

389　sample t-test was used to contrast each of these correlation distributions against zero. For

390　the comparison of intact and locally scrambled conditions, correlations were significantly

391　greater than zero in the PPA (t(19) = 7.44, p < .001, Cohen' d = 1.66), significantly less than

392　zero in the RSC (t(19) = 3.17, p = .015, Cohen's d = 0.71), and less than zero in the OPA with

393　the difference approaching significance (t(19) = 2.41, p = .053, Cohen' d = 0.54). For the

394　comparison of intact and globally scrambled conditions, correlations were significantly

395　greater than zero in the PPA (t(19) = 5.51, p < .001, Cohen's d = 1.23) and OPA (t(19) = 8.83,

396　p < .001, Cohen's d = 1.97), and did not differ significantly from zero in the RSC (t(19) = 0.09,

397　p = .929, Cohen's d = 0.02). Next, we compared the correlations with the noise ceiling. For

398　the comparison of intact and locally scrambled conditions, correlations were significantly

399   below the lower bound of the noise ceiling in the RSC ($t(19) = 5.00$, $p < .001$, Cohen's d =

400   1.12) and OPA ($t(19) = 14.69$, $p < .001$, Cohen's d = 3.28), but not the PPA ($t(19) = 1.70$, $p =$

401   .211, Cohen's d = 0.38). For the comparison of intact and globally scrambled conditions,

402   correlations were significantly below the lower bound of the noise ceiling in the PPA ($t(19) =$

403   5.89, $p < .001$, Cohen's d = 1.32) and OPA ($t(19) = 2.97$, $p = .023$, Cohen's d = 0.66), but not

404   the RSC ($t(19) = 1.21$, $p = .241$, Cohen's d = 0.27). This shows that in most cases the local

405   and global scrambling conditions ability to predict the intact responses fell significantly

406   below the theoretical maximum of the noise ceiling. Overall, this analysis demonstrates

407   that, for PPA and OPA, significant representational structure is preserved under even global

408   scrambling conditions although it also shows that other sources of variance play a role.

409   We next tested the extent to which the definition of the scene ROIs influenced the

410   MVPA results. The main scene ROIs were defined using a contrast of Scenes > Phase

411   Scrambled versions of those scenes. We defined an alternative set of ROIs for the main

412   scene regions (PPA, RSC, OPA) from independent localiser data of separate experiments (not

413   reported here) using contrasts of Scenes > Faces and Scenes > Objects. The locations of

414   these ROIs are illustrated in Supplementary Figure 8, and co-ordinates of the corresponding

415   peak voxels are given in Supplementary Table 2. Locations of the PPA and RSC regions

416   remained relatively consistent across the definitions (cf. Table 1 and Supplementary Figure

417   3). We next repeated our MVP analyses for these alternative ROIs. Results of the decoding

418   analyses were largely consistent with those for the main ROIs (Supplementary Figure 9; cf.

419   Figure 4). Representational similarity analyses remained broadly consistent between the

420   main and alternative definitions (Supplementary Figure 10; cf. Figure 5).

421   In order to interpret the results of representational similarity analyses within scene

422   selective regions, it is essential to identify any disruption of earlier stages of visual

423   processing. To test whether category specific visual responses in early visual cortex survive

424   scrambling of low-level image properties, we repeated our analyses in a V1 control region

425   defined using a probabilistic atlas (Wang et al., 2015). The results of this analysis are shown

426   in Supplementary Figure 11. Paired-samples t-tests revealed significantly higher within-

427   than between-category correlations for the intact (t(19) = 7.82, p < .001, Cohen's d = 1.75),

428   locally scrambled (t(19) = 4.28, p < .001, Cohen's d = 0.96), and globally scrambled scenes

429   (t(19) = 4.68, p < .001, Cohen's d = 1.05). Representational similarity analyses revealed a

430   significant correlation between the intact and globally scrambled conditions (r(18) = .71, p =

431   .001), but not between intact and locally scrambled conditions (r(18) = .37, p = .112).

432   Overall, these results indicate that preserved low-level features are sufficient to maintain

433   reliable spatial patterns of response in V1 after scrambling.

434   Finally, we repeated our analyses using a whole-brain searchlight approach to

435   identify areas beyond our regions of interest where patterns of response to intact and

436   scrambled images are systematically affected by stimulus category. The results of these

437   analyses are plotted on the cortical surface in Supplementary Figure 12. Spheres showing

438   significant decoding of category for intact scenes were observed throughout occipital and

439   ventro-temporal cortices.   Decoding for the scrambled scene conditions was less

440   widespread; nevertheless, significant spheres were observed in right ventro-temporal

441   cortices and some occipital regions for locally scrambled scenes, and in some occipital

442   regions for globally scrambled scenes. Representational similarity analyses revealed

443   significant spheres in regions including ventro-temporal and occipital cortices, both for the

21

444   comparison of intact and locally scrambled scenes, and intact and globally scrambled

445   scenes.

446

447 **DISCUSSION**

448 The aim of the present study was to directly determine whether category-selective patterns

449 of response in scene-selective regions can be explained by the visual properties of the

450 stimulus. To address this issue, we compared patterns of response to intact and scrambled

451 images. Our hypothesis was that, if category-selective patterns of response purely reflect

452 the semantic content of the images, there should be little similarity between the patterns of

453 response elicited by intact and scrambled images. On the other hand, if category-specific

454 patterns are based on visual properties, similar patterns should be elicited by both intact

455 and scrambled images. Image scrambling significantly impaired the ability to categorize

456 scenes, consistent with previous results showing that local phase information is important

457 for recognition of scene *gist* (Loschky et al., 2007). However, we found distinct and reliable

458 category-selective patterns of response for both the intact and scrambled image conditions

459 in the PPA and OPA scene-selective regions.   Moreover, the patterns of response elicited by

460 intact scenes were similar to the patterns of response to scrambled scenes.

461 Previous studies have identified distinct patterns of neural response to different

462 categories of scene in scene selective regions (Walther et al., 2009, 2011; Watson et al.,

463 2014). Our results show that categorical patterns of response in scene-selective regions are

464 still evident to images with significantly reduced semantic information. These findings are

465 consistent with recent studies in which we have shown that basic image properties of

466 different scene categories can predict patterns of response in scene-selective regions (Rice,

467 Watson, Hartley, & Andrews, 2014; Watson et al., 2014, 2016). However, because images

468 drawn from the same category are likely to have similar visual properties (Oliva & Torralba,

469 2001), it was unclear from this previous work whether or not patterns are determined

470 primarily by categorical or visual properties of the image. The results from the current study

471 provide more direct evidence that lower-level visual properties of the image can account for

472 a substantial proportion of the variance in the patterns of response in scene-selective

473 regions. This does not dispute that there are distinct patterns of response to different scene

474 categories in scene-selective regions, but rather suggests that such effects may be

475 underpinned, at least in part, by sensitivity to the visual properties of scenes.

476 To evaluate the importance of spatial properties in the neural representation of

477 scenes, we compared scrambling across the full global extent of the image, or

478 independently within local windows of the image. The local scrambling thus preserves the

479 coarse-scale global structure of the original image more than the global scrambling, in the

480 sense that the local scrambling leaves the windows of the grid in their original spatial

481 positions (see also Figure 1 & Supplementary Figure 1). In PPA, we found that responses

482 could be discriminated for locally scrambled scenes, but the ability to discriminate globally

483 scrambled images was less reliable. Furthermore, a representational similarity analysis

484 showed that local scrambling preserved the pattern of response to intact images more than

485 globally scrambling. This would suggest that the PPA is sensitive to the coarse-scale spatial

486 organisation of the image, such that responses are disrupted more by global scrambling.

487 Such a conclusion would be consistent with previous studies demonstrating sensitivity of

488 the PPA to the spatial structure of scenes (Epstein, Higgins, Parker, Aguirre, & Cooperman,

489 2006; Kravitz et al., 2011; Park et al., 2011), and displaying visual field biases (Arcaro et al.,

490 2009; Cichy et al., 2013; Silson et al., 2015). Indeed, it has been proposed that the PPA may

491 support extraction of local spatial geometries of the scene (Epstein, 2008; Epstein, Parker, &

492 Feiler, 2007), for which local visual features may be important.

24

493   There was a reduction in the magnitude of the category effect for scrambled scenes

494   relative to intact scenes, suggesting that the scrambling process introduced some disruption

495   to the neural representations. This suggests that patterns of response are dependent on

496   higher-level information about the scene that is only available from the intact images. One

497   possibility is that this higher-level information reflects the semantic or categorical properties

498   conveyed by the image. For example, our noise ceiling analysis suggests that while

499   significant pattern similarity is preserved, a substantial component is disrupted, particularly

500   by global scrambling. However, an alternative possibility is that unexplained variance might

501   reflect image properties that are disrupted by the scrambling process. An important feature

502   of intact images is the strong statistical dependencies between features, such as location-

503   specific combinations of orientation and spatial frequency. Indeed, the behavioural

504   sensitivity to the regularities that occur in intact objects suggests that these properties are

505   critical for differentiating between different classes of images (Loschky et al., 2007, 2010). It

506   is possible that these properties also contribute to the patterns of response in scene-

507   selective regions. When evaluating these possibilities, it is important to recognize that high-

508   and low-level contributions to the observed representational structure are not mutually

509   exclusive. The extraction of any high-level features depends on the availability of relevant

510   low-level features preserved in the scrambled stimuli.

511   We found that category responses in the OPA could be discriminated for intact

512   scenes and locally scrambled scenes, but not globally scrambled scenes. However, in

513   contrast to the PPA the representational structure of the intact scenes was maintained by

514   the global scrambling. Although the OPA has been causally implicated in the perception of

515   scenes (Dilks et al., 2013; Ganaden, Mullin, & Steeves, 2013), its precise functional

516   properties are less well established than other scene regions. The greater similarity

517   between intact and globally scrambled images suggests that the OPA is sensitive to global

518   visual statistics, such as the texture of the image. Interestingly, this implies a possible

519   functional distinction between PPA and OPA, with the PPA more clearly tuned to the local

520   visual features than the OPA. Recent studies have reported a double dissociation in visual

521   field biases between the PPA and OPA (Silson et al., 2015; Silson, Groen, Kravitz, & Baker,

522   2016), suggesting inputs to these regions may at least partially function in parallel rather

523   than in series, and which may therefore support some degree of functional dissociation

524   between them.

525   In contrast to the PPA and OPA, responses in RSC failed to discriminate the scene

526   categories in any of the conditions. The representational similarity analyses showed that

527   neither local nor global scrambling maintained the representational structure relative to the

528   intact scenes. It has been proposed that the RSC may play a role representing the scene as

529   part of the wider spatial environment (Epstein, 2008; Epstein et al., 2007) playing a crucial

530   role in spatial memory, navigation and imagery – for example, translating between ego- and

531   allocentric spatial representations (Byrne, Becker, & Burgess, 2007; Vann, Aggleton, &

532   Maguire, 2009). Such processes suggest a more abstract form of representation, less

533   directly tied to image features.

534   We also examined the response patterns within a V1 control region (Wang et al.,

535   2015). We would expect this region to display sensitivity to the visual features of the

536   scenes, but would be less likely to be modulated by higher-level semantic features of the

537   scene categories. We observed significant decoding of the scene categories under all of the

538   different scrambling conditions, consistent with the reliable differences in visual features

26

539 between different scene categories. We also observed a significant association between the

540 patterns in the intact and globally scrambled conditions, consistent with the presence of the

541 shared global visual features between the intact and globally scrambled conditions.

542 In conclusion, our results demonstrate distinct responses to different categories of

543 scenes even when the perception of scene category is severely impaired by phase

544 scrambling. These results should not be taken to imply that perception of scene category is

545 independent of the neural response in scene-selective regions, but they do suggest that the

546 topographic organization of regions such as the PPA and, to a lesser extent, the OPA can be

547 explained by selectivity for the visual properties of the image.

548 **REFERENCES**

549 Aguirre, G. K., & D'Esposito, M. (1999). Topographical disorientation: a synthesis and
550   taxonomy. *Brain*, *122*(9), 1613–1628. http://doi.org/10.1093/brain/122.9.1613

551 Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex
552   sensitive to "building" stimuli: evidence and implications. *Neuron*, *21*(2), 373–83.
553   Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9728918

554 Amit, E., Mehoudar, E., Trope, Y., & Yovel, G. (2012). Do object-category selective regions in
555   the ventral visual stream represent perceived distance information? *Brain and*
556   *Cognition*, *80*(2), 201–213. http://doi.org/10.1016/j.bandc.2012.06.006

557 Andrews TJ, Clarke A, Pell P & Hartley T (2010) Selectivity for low-level features of objects in
558   the human ventral stream. Neuroimage 49: 703-711.

559 Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of
560   natural images predict topographic patterns of neural response in the ventral visual
561   pathway visual pathway. *Journal of Vision*, *15*(7), 1–12.
562   http://doi.org/10.1167/15.7.3.doi

563 Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic Organization of
564   Human Ventral Visual Cortex. *Journal of Neuroscience*, *29*(34), 10638–10652.
565   http://doi.org/10.1523/jneurosci.2807-09.2009

566 Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future:
567   A neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340–375.
568   http://doi.org/10.1037/0033-295X.114.2.340

569 Cichy, R. M., Sterzer, P., Heinzle, J., Elliott, L. T., Ramirez, F., & Haynes, J.-D. (2013). Probing
570   principles of large-scale object representation: Category preference and location
571   encoding. *Human Brain Mapping*, *34*(7), 1636–1651.
572   http://doi.org/10.1002/hbm.22020

573 Coggan, D. D., Liu, W., Baker, D. H., & Andrews, T. J. (2016). Category-selective patterns of
574   neural response in the ventral visual pathway in the absence of categorical
575   information. *NeuroImage*, *135*, 107–114. http://doi.org/10.1167/15.12.622

576 Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-Image
577   Sensitivity and Invariance in Object and Scene Processing Pathways. *Journal of*
578   *Neuroscience*, *31*(31), 11305–11312. http://doi.org/10.1523/jneurosci.1935-11.2011

579 Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The Occipital Place Area Is
580   Causally and Selectively Involved in Scene Perception. *Journal of Neuroscience*, *33*(4),
581   1331–1336. http://doi.org/10.1523/JNEUROSCI.4081-12.2013

582 Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial
583   navigation. *Trends in Cognitive Sciences*, *12*(10), 388–396.
584   http://doi.org/10.1016/j.tics.2008.07.004

585 Epstein, R. A., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene
586   representations in human parahippocampal cortex. *Neuron*, *37*(5), 865–876.
587   http://doi.org/10.1016/s0896-6273(03)00117-x

588 Epstein, R. A., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place
589   area: Recognition, navigation, or encoding? *Neuron*, *23*(1), 115–125.
590   http://doi.org/10.1016/s0896-6273(00)80758-8

591 Epstein, R. A., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial
592   involvement in three types of visual scene recognition. *Cerebral Cortex*, *17*(7), 1680–
593   1693. http://doi.org/10.1093/cercor/bhl079

594 Epstein, R. A., Higgins, J. S., Parker, W., Aguirre, G. K., & Cooperman, S. (2006). Cortical
595   correlates of face and scene inversion: A comparison. *Neuropsychologia*, *44*(7), 1145–
596   1158. http://doi.org/10.1016/j.neuropsychologia.2005.10.009

597 Epstein, R. A., & Kanwisher, N. (1998). A cortical representation of the local visual
598   environment. *Nature*, *392*(6676), 598–601. http://doi.org/10.1038/33402

599 Epstein, R. A., Parker, W. E., & Feiler, A. M. (2007). Where am I now? Distinct roles for
600   parahippocampal and retrosplenial cortices in place recognition. *Journal of*
601   *Neuroscience*, *27*(23), 6141–6149. http://doi.org/10.1523/jneurosci.0799-07.2007

602 Ganaden, R. E., Mullin, C. R., & Steeves, J. K. E. (2013). Transcranial Magnetic Stimulation to
603   the Transverse Occipital Sulcus Affects Scene but Not Object Processing. *Journal of*
604   *Cognitive Neuroscience*, *25*(6), 961–968. http://doi.org/10.1162/jocn

605 Golomb, J. D., & Kanwisher, N. (2012). Higher Level Visual Cortex Represents Retinotopic,
606   Not Spatiotopic, Object Location. *Cerebral Cortex*, *22*(12), 2794–2810.
607   http://doi.org/10.1093/cercor/bhr357

608 Groen, I. I. A., Silson, E. H. & Baker, C. I. (2017) Contributions of low- and high-level
609   properties to neural processing of visual scenes in the human brain. *Philos. Trans. R.*
610   *Soc. B* 1–22. doi:10.1098/rstb.2016.0102

611 Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V, & Pollmann, S.
612   (2009). PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data.
613   *Neuroinformatics*, *7*(1), 37–53. http://doi.org/10.1007/s12021-008-9041-y

614 Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry
615   organization of human occipito-temporal object areas. *Neuron*, *37*(6), 1027–1041.
616   http://doi.org/10.1016/s0896-6273(03)00144-2

617 Haxby, J. V, Connolly, A. C., & Guntupalli, J. S. (2014). Decoding Neural Representational
618   Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience*, *37*, 435–
619   456. http://doi.org/10.1146/annurev-neuro-062012-170325

620 Haxby, J. V, Gobbini, M., Furey, M., Ishai, A., Schouten, J., & Pietrini, P. (2001). Distributed
621   and overlapping representations of faces and objects in ventral temporal cortex.
622   *Science*, *293*(5539), 2425–2430. http://doi.org/10.1126/science.1063736

623 Henderson, J. M., Zhu, D. C., & Larson, C. L. (2011). Functions of parahippocampal place area
624   and retrosplenial cortex in real-world scene analysis: An fMRI study. *Visual Cognition*,
625   *19*(7), 910–927. http://doi.org/10.1080/13506285.2011.596852

626 Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space
627   Describes the Representation of Thousands of Object and Action Categories across the
628   Human Brain. *Neuron*, *76*(6), 1210–1224. http://doi.org/10.1016/j.neuron.2012.10.014

629 Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the
630   Robust and Accurate Linear Registration and Motion Correction of Brain Images.
631   *NeuroImage*, *17*(2), 825–841. http://doi.org/10.1006/nimg.2002.1132

632 Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). NeuroImage An algorithmic
633   method for functionally defining regions of interest in the ventral visual pathway.
634   *NeuroImage*, *60*(4), 2357–2364. http://doi.org/10.1016/j.neuroimage.2012.02.055

635 Köhler, S., Crane, J., & Milner, B. (2002). Differential contributions of the parahippocampal
636   place area and the anterior hippocampus to human memory for scenes. *Hippocampus*,
637   *12*(6), 718–723. http://doi.org/10.1002/hipo.10077

638 Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-World Scene Representations in High-
639   Level Visual Cortex: It's the Spaces More Than the Places. *Journal of Neuroscience*,
640   *31*(20), 7322–7333. http://doi.org/10.1523/jneurosci.4588-10.2011

641 Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain
642   mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868.
643   http://doi.org/10.1073/pnas.0600244103

644 Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis -
645   connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*,
646   *2*(4), 1–28. http://doi.org/10.3389/neuro.06.004.2008

647 Lescroart, M. D., Stansbury, D. E., & Gallant, J. L. (2015). Fourier power, subjective distance,
648   and object categories all provide plausible models of BOLD responses in scene-selective
649   visual areas. *Frontiers in Computational Neuroscience*, *9*.
650   http://doi.org/10.3389/fncom.2015.00135

651 Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center – periphery
652   organization of human object areas. *Nature Neuroscience*, *4*(5), 533–539.

653 Levy, I., Hasson, U., Harel, M., & Malach, R. (2004). Functional analysis of the periphery
654   effect in human building-related areas. *Human Brain Mapping*, *22*(1), 15–26.
655   http://doi.org/10.1002/hbm.20010

656 Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimarri, T. N. (2010). The role of higher order
657   image statistics in masking scene gist recognition. *Attention, Perception, &
658   Psychophysics*, *72*(2), 427–444. http://doi.org/10.3758/APP.72.2.427

659 Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The
660   importance of information localization in scene gist recognition. *Journal of
661   Experimental Psychology: Human Perception and Performance*, *33*(6), 1431–1450.
662   http://doi.org/10.1037/0096-1523.33.6.1431

663 Maguire, E. (2001). The retrosplenial contribution to human navigation: A review of lesion
664   and neuroimaging findings. *Scandinavian Journal of Psychology*, *42*(3), 225–238.
665   http://doi.org/10.1111/1467-9450.00233

666 Mendez, M. F., & Cherrier, M. M. (2003). Agnosia for scenes in topographagnosia.
667   *Neuropsychologia*, *41*(10), 1387–1395. http://doi.org/10.1016/s0028-3932(03)00041-1

668 Mullally, S. L., & Maguire, E. A. (2011). A new role for the parahippocampal cortex in
669   representing space. *The Journal of Neuroscience*, *31*(20), 7441–9.

670   http://doi.org/10.1523/JNEUROSCI.0267-11.2011

671   Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C.
672   (2014). Coarse-to-fine Categorization of Visual Scenes in Scene-selective Cortex. *Journal*
673   *of Cognitive Neuroscience*, *26*(10), 2287–2297. http://doi.org/10.1162/jocn

674   Nasr, S., Echavarria, C. E., & Tootell, R. B. H. (2014). Thinking Outside the Box: Rectilinear
675   Shapes Selectively Activate Scene-Selective Cortex. *Journal of Neuroscience*, *34*(20),
676   6721–6735. http://doi.org/10.1523/JNEUROSCI.4802-13.2014

677   Nasr, S., & Tootell, R. B. H. (2012). A cardinal orientation bias in scene-selective visual
678   cortex. *The Journal of Neuroscience*, *32*(43), 14921–6.
679   http://doi.org/10.1523/JNEUROSCI.2036-12.2012

680   O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates
681   corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6),
682   1013–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11177421

683   Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation
684   of the Spatial Envelope. *International Journal of Computer Vision*, *42*(3), 145–175.
685   http://doi.org/10.1023/a:1011139631724

686   Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: spatial smoothing does not
687   hurt multivariate fMRI analyses? *NeuroImage*, *49*(3), 1943–8.
688   http://doi.org/10.1016/j.neuroimage.2009.02.047

689   Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling Scene Content from
690   Spatial Boundary: Complementary Roles for the Parahippocampal Place Area and
691   Lateral Occipital Complex in Representing Real-World Scenes. *Journal of Neuroscience*,
692   *31*(4), 1333–1340. http://doi.org/10.1523/jneurosci.3885-10.2011

693   Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M. M. (2007). Beyond the Edges of a
694   View: Boundary Extension in Human Scene-Selective Visual Cortex. *Neuron*, *54*(2), 335–
695   342. http://doi.org/10.1016/j.neuron.2007.04.006

696   Park, S., Konkle, T., & Oliva, A. (2015). Parametric Coding of the Size and Clutter of Natural
697   Scenes in the Human Brain. *Cerebral Cortex*, *25*, 1792–1805.
698   http://doi.org/10.1093/cercor/bht418

699   Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience*
700   *Methods*, *162*, 8–13. http://doi.org/10.1016/j.jneumeth.2006.11.017

701   Peirce, J. W. (2009). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in*
702   *Neuroinformatics*, *2*(10), 1–8. http://doi.org/10.3389/neuro.11.010.2008

703   Poldrack, R. A., Halchenko, Y. O., & Hanson, S. J. (2009). Decoding the large-scale structure
704   of brain function by classifying mental states across individuals. *Psychological Science*,
705   *20*(11), 1364–72. http://doi.org/10.1111/j.1467-9280.2009.02460.x

706   Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The
707   "Parahippocampal Place Area" Responds Preferentially to High Spatial Frequencies in
708   Humans and Monkeys. *PLoS Biol*, *9*(4), e1000608.
709   http://doi.org/10.1371/journal.pbio.1000608

710 Rice, G. E., Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Low-level image properties of
711   visual objects predict patterns of neural response across category-selective regions of
712   the ventral visual pathway. *Journal of Neuroscience*, *34*(26), 8837–8844. Retrieved from
713   http://www.jneurosci.org/content/34/26/8837.short

714 Schinazi, V. R., & Epstein, R. A. (2010). Neural correlates of real-world route learning.
715   *Neuroimage*, *53*(2), 725–735. http://doi.org/10.1016/j.neuroimage.2010.06.065

716 Shinkareva, S. V, Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., & Just, M. A. (2008).
717   Using FMRI brain activation to identify cognitive states associated with perception of
718   tools and dwellings. *PloS One*, *3*(1), e1394.
719   http://doi.org/10.1371/journal.pone.0001394

720 Silson, E. H., Chan, a. W.-Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A Retinotopic
721   Basis for the Division of High-Level Scene Processing between Lateral and Ventral
722   Human Occipitotemporal Cortex. *Journal of Neuroscience*, *35*(34), 11921–11935.
723   http://doi.org/10.1523/JNEUROSCI.0137-15.2015

724 Silson, E. H., Groen, I. I. A., Kravitz, D. J., & Baker, C. I. (2016). Evaluating the correspondence
725   between face-, scene-, and object-selectivity and retinotopic organization within lateral
726   occipitotemporal cortex. *Journal of Vision*, *16*(6), 14. http://doi.org/10.1167/16.6.14

727 Stansbury, D. E., Naselaris, T., & Gallant, J. L. (2013). Natural Scene Statistics Account for the
728   Representation of Scene Categories in Human Visual Cortex. *Neuron*, *79*(5), 1025–1034.
729   http://doi.org/10.1016/j.neuron.2013.06.034

730 Vann, S. D., Aggleton, J. P., & Maguire, E. a. (2009). What does the retrosplenial cortex do?
731   *Nature Reviews Neuroscience*, *10*(11), 792–802. http://doi.org/10.1038/nrn2733

732 Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural Scene Categories
733   Revealed in Distributed Patterns of Activity in the Human Brain. *Journal of
734   Neuroscience*, *29*(34), 10573–10581. http://doi.org/10.1523/jneurosci.0559-09.2009

735 Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings
736   suffice for functional MRI decoding of natural scene categories. *Proceedings of the
737   National Academy of Sciences*, *108*(23), 9661–9666.
738   http://doi.org/10.1073/pnas.1015666108

739 Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic Maps of Visual
740   Topography in Human Cortex. *Cerebral Cortex*, *25*(October), 3911–3931.
741   http://doi.org/10.1093/cercor/bhu277

742 Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are
743   linked to the low-level properties of the image. *NeuroImage*, *99*, 402–410.
744   http://doi.org/10.1016/j.neuroimage.2014.05.045

745 Watson, D. M., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response
746   in scene-selective regions of the human brain are affected by low-level manipulations
747   of spatial frequency. *NeuroImage*, *124*, 107–117.
748   http://doi.org/10.1016/j.neuroimage.2015.08.058

749 Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010).
750   Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*,

751   *42*(3), 671–684. http://doi.org/10.3758/brm.42.3.671

752 Xiao, J. X., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN Database: Large-scale
753   Scene Recognition from Abbey to Zoo. In *IEEE Conference on Computer Vision and*
754   *Pattern Recognition* (pp. 3485–3492). Los Alamitos: IEEE Computer Soc.
755   http://doi.org/10.1109/cvpr.2010.5539970

756

757 **TABLES**

758 **Table 1.** Peak MNI mm co-ordinates, voxel counts, and thresholds of standard scene
759 selective clusters (PPA, RSC, OPA).

| Region	Hemisphere	x	y	z	Voxel count	Threshold (Z)
PPA	L	-34	-46	-22	500	5.06
	R	26	-50	-18	500	5.59
RSC	L	-18	-52	-2	500	4.63
	R	16	-58	6	502	4.79
OPA	L	-36	-90	2	500	5.14
	R	38	-82	4	499	5.03

760

34

761 **FIGURES**
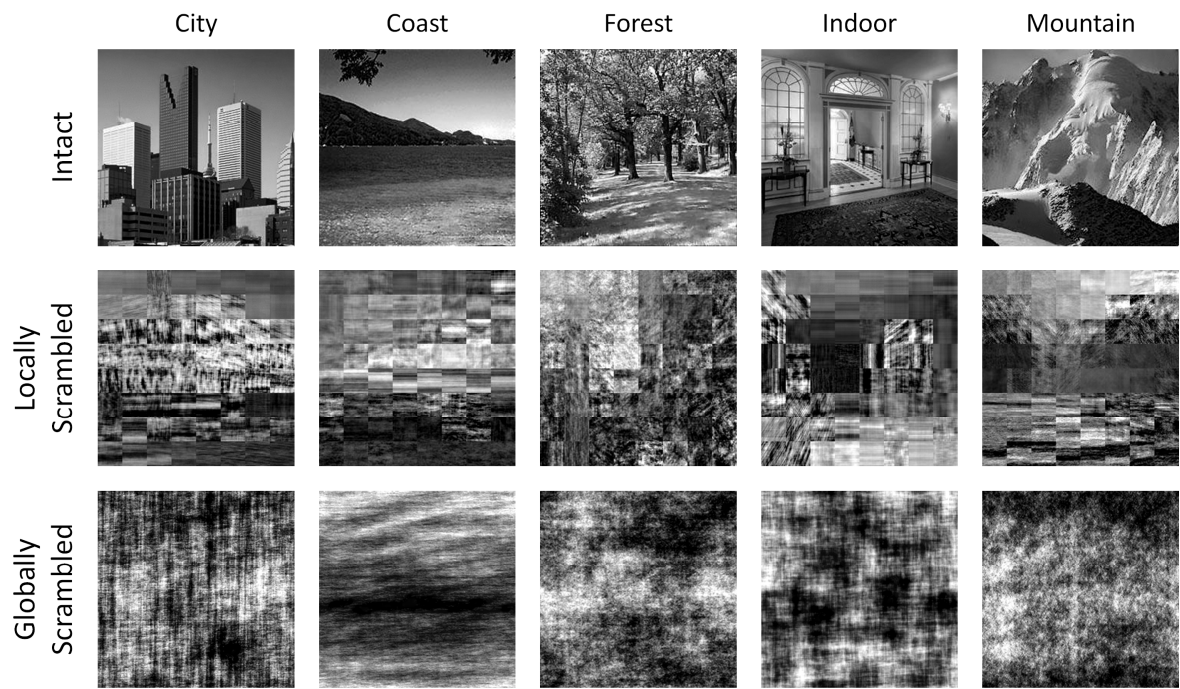
763 **Figure 1.** Examples of the scene images used in each condition.

764
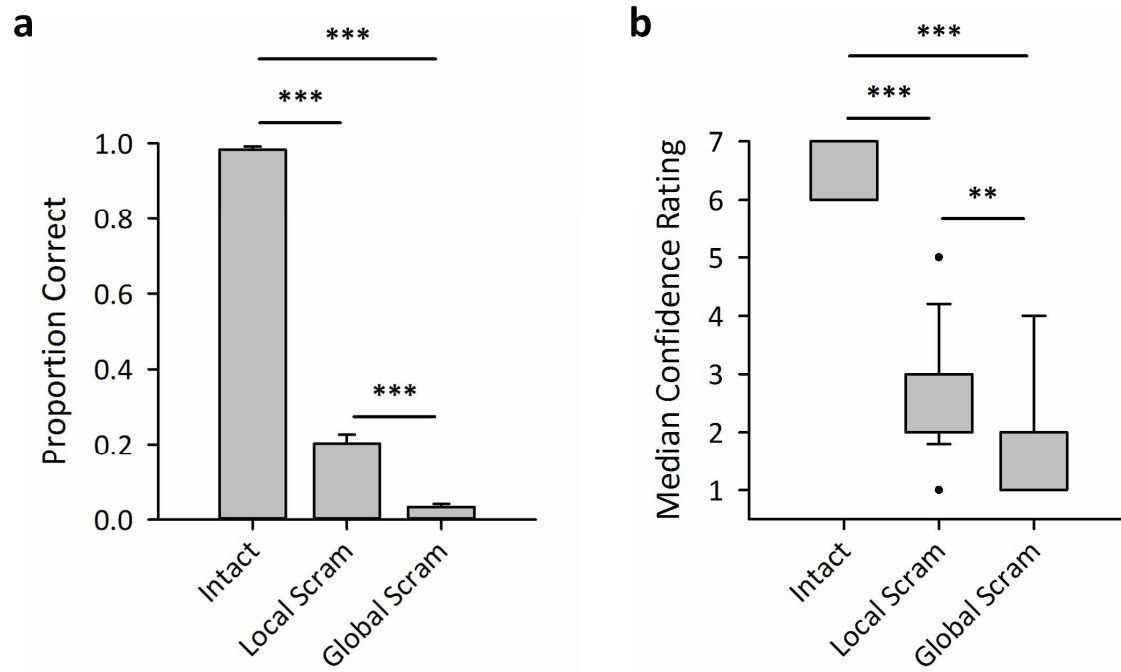
765 **Figure 2.** Results of the behavioural experiment. (a) Mean scene identification accuracies

766 for each level of scrambling. Error bars represents 1 SEM. (b) Box-plots of median

767 confidence ratings for each level of scrambling. (*** $p < .001$, ** $p < .01$, * $p < .05$).

768
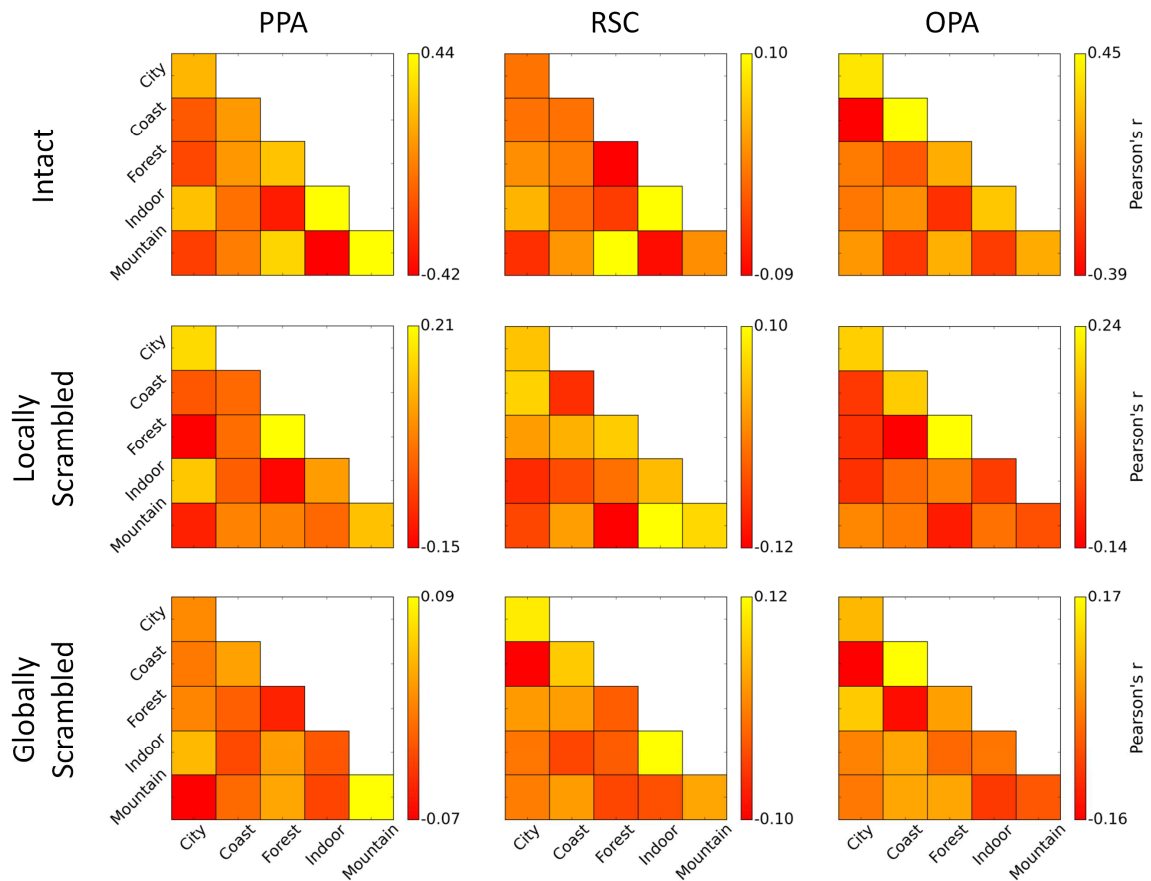
Figure 3. MVPA results: correlation similarity matrices for each level of scrambling in each region of interest. To aid visualisation, symmetrically opposite points across the diagonal have been averaged and displayed within the lower-triangle portion of the matrix only.

772
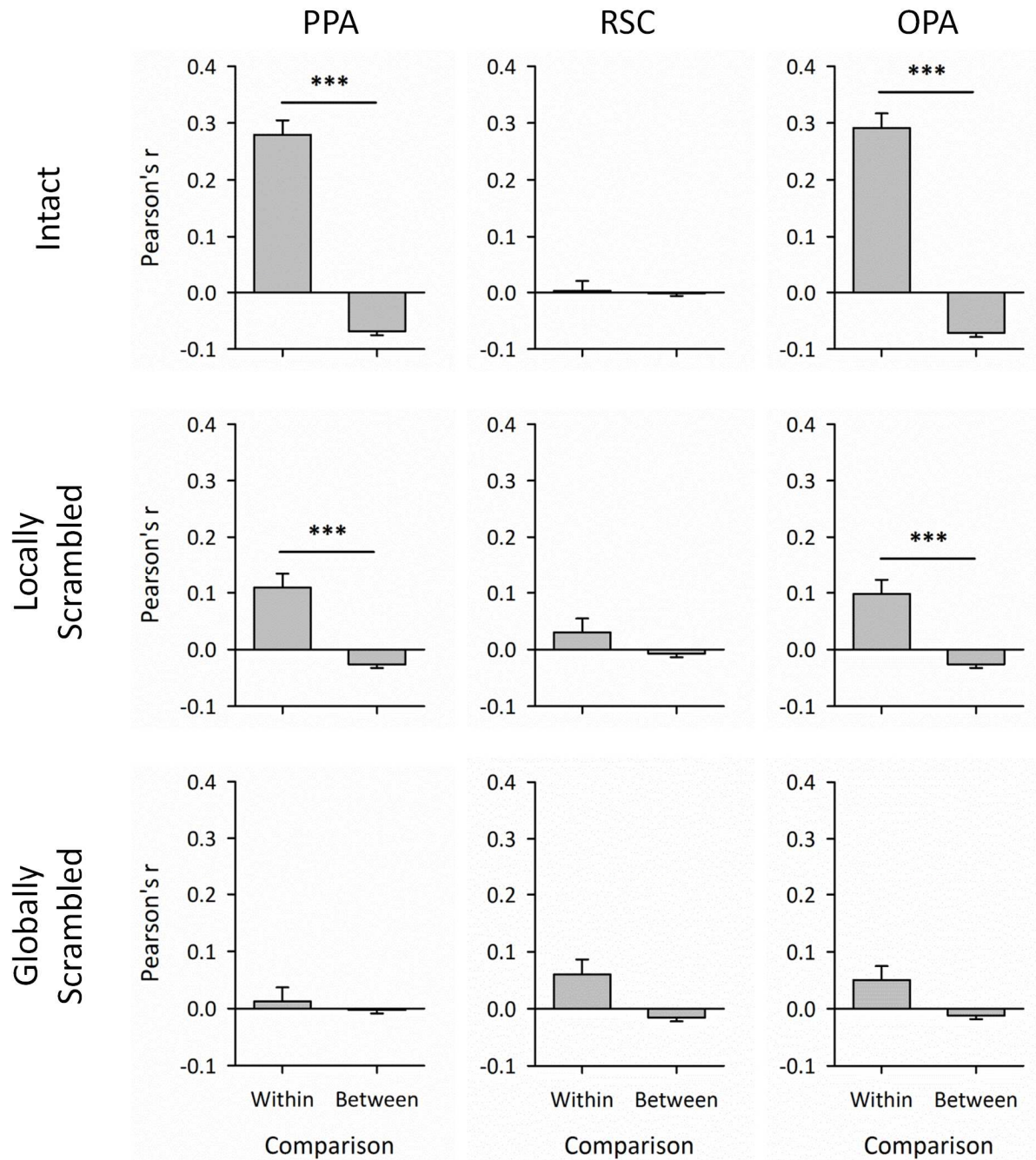
773 **Figure 4.** Decoding of categories from MVPA. Average within-category (on-diagonal) and

774 between-category (off-diagonal) values were calculated from the MVPA correlation

775 matrices. Significantly greater within- than between-category correlations indicate

776 categories can be successfully decoded. Error bar represent 1 SEM. (*** $p < .001$, ** $p <$

777 $.01$, * $p < .05$).

778
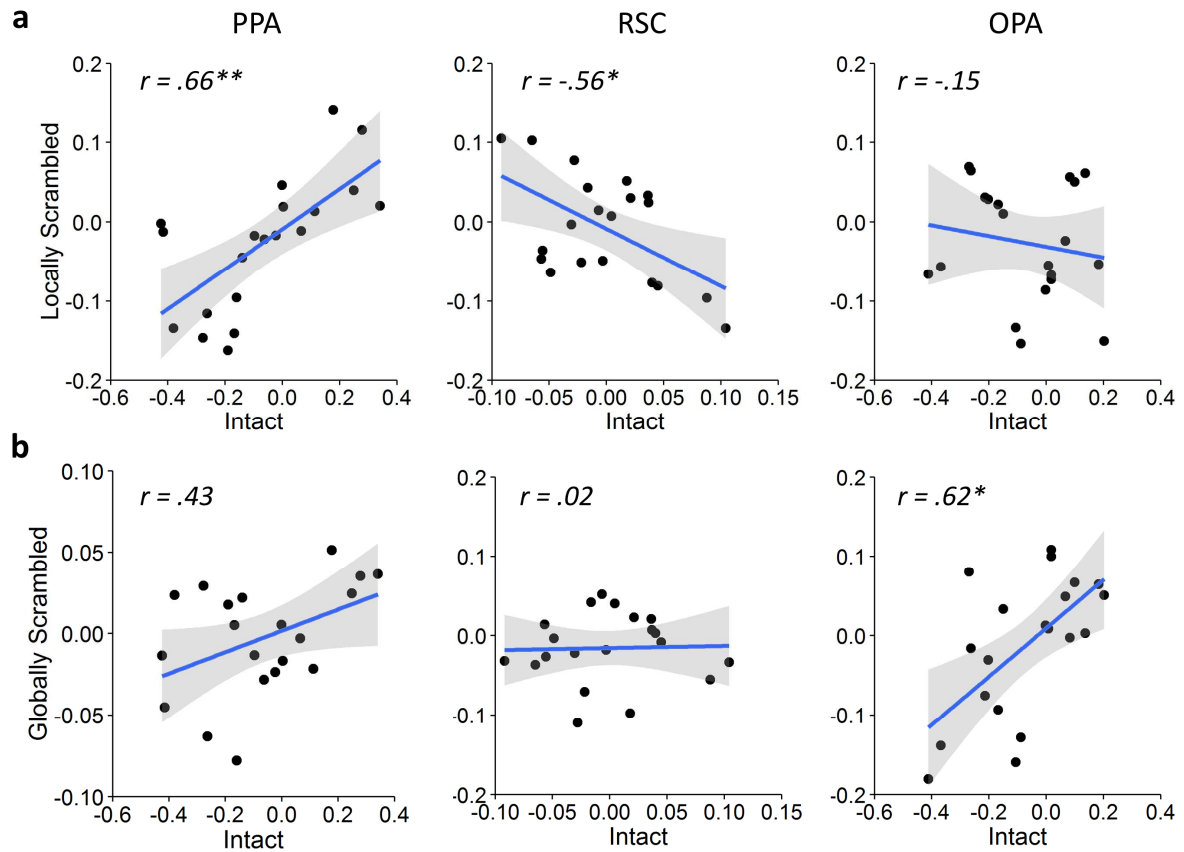
Figure 5. Representational similarity analyses. Off-diagonal elements of group average
MVPA correlation matrices (Figure 3) are correlated between (a) intact and locally-
scrambled conditions, and (b) intact and globally-scrambled conditions. Shaded regions
represent 95% confidence intervals. (*** $p < .001$, ** $p < .01$, * $p < .05$).

783

784