

## Where Do Objects Become Scenes?

Jiye G. Kim<sup>1</sup> and Irving Biederman<sup>1,2</sup>

<sup>1</sup>Department of Psychology and <sup>2</sup>Neuroscience Program, University of Southern California, Los Angeles, CA 90089-1061, USA

Address correspondence to Jiye G. Kim, Department of Psychology, University of Southern California, 3620 McClintock Avenue, Seely G. Mudd Building, Room 501, Los Angeles, CA 90089-1061, USA. Email: jiyekim@usc.edu.

**Regions tuned to individual visual categories, such as faces and objects, have been discovered in the later stages of the ventral visual pathway in the cortex. But most visual experience is composed of scenes, where multiple objects are interacting. Such interactions are readily described by prepositions or verb forms, for example, a bird perched on a birdhouse. At what stage in the pathway does sensitivity to such interactions arise? Here we report that object pairs shown as interacting, compared with their side-by-side depiction (e.g., a bird besides a birdhouse), elicit greater activity in the lateral occipital complex, the earliest cortical region where shape is distinguished from texture. Novelty of the interactions magnified this gain, an effect that was absent in the side-by-side depictions. Scene-like relations are thus likely achieved simultaneously with the specification of object shape.**

**Keywords:** LOC, object recognition, object relations, scene perception, ventral pathway

### Introduction

The coding of relations among entities allows us to represent complex structures beyond what could be derived from a mere “bag” of items (Pinker 2007). In language, relations play the core role not only in syntax but also in minimal phrases as illustrated by the contrast of a “Venetian blind” with a “blind Venetian” (Miller 1965). Visually, a pair of objects can be depicted side by side or as interacting, for example, a cup “on” a chair. Such interactions markedly facilitate cued recall (chair-?) compared with the side-by-side depictions (Epstein et al. 1960). Object interactions, readily described by prepositions or gerunds, for example, a hand “holding” a pen, are fundamental to visual cognition and appear to be immediately appreciated (Biederman et al. 1974, 1982; Biederman 1981; Green and Hummel 2006).

Interacting stimuli that have a meaningful functional relationship (e.g., a pitcher pouring water into a cup) enhance object identification (Green and Hummel 2004). In a visual search task with displays composed of pairs of objects, targets were detected more accurately when shown as an appropriate interaction than not (e.g., the pitcher pouring away from the cup). The cost of the number of display distractors could be reduced when nontarget pairs were interacting, as if each nontarget pair then acted as a single object rather than 2.

Green and Hummel (2006) further demonstrated that near simultaneous presentation of an object pair is required for their relations to facilitate object recognition. When 2 objects were presented sequentially with a short (100-ms) stimulus onset asynchrony (SOA) so that the 2 could be grouped into a single percept, object identification was facilitated when they were depicted as interacting compared with when they were not. When SOA was increased to 250 ms, the 2 objects were

perceptually segregated, and the facilitation of object identification was lost. If the relations between objects were encoded post-perceptually, the longer SOA should not have eliminated the benefit of the interactions.

Where in the brain are these visual relations registered? The ventral visual cortical pathway, generally regarded as the system mediating shape recognition, consists of a series of stages that can be distinguished through contrasting functional magnetic resonance imaging (fMRI) blood oxygenation level-dependent (BOLD) activation to different classes of stimuli, such as objects, faces, and houses (Malach et al. 1995; Haxby et al. 1996; Kanwisher et al. 1997; Aguirre et al. 1998; Epstein and Kanwisher 1998). One candidate for the cortical locus for the coding of scene interactions is the parahippocampal place area (PPA). PPA is suggestive of scene processing in that it manifests a stronger response to images depicting spatial layouts than it does to individual objects (Epstein and Kanwisher 1998; Epstein et al. 1999; Epstein et al. 2003).

Recently Bar and colleagues have proposed that the parahippocampal cortex (PHC) that includes the PPA, processes contextual information such that objects (or faces) that have strong associations to a given scene context (e.g., a shopping cart for a supermarket), evokes greater activity than objects that are not associated with a particular setting, for example, a basket (Bar and Aminoff 2003; Bar 2004; Bar et al. 2008). Further, they have proposed that the posterior PHC (i.e., the PPA) encodes spatial contextual information. In this region, greater activity was observed for object triplets that were previously studied in the same spatial relations to each other than those that were not (Aminoff et al. 2006).

Lateral occipital complex (LOC), composed of the lateral occipital cortex (LO) and the posterior fusiform gyrus (pFs), is defined as the region that manifests a greater BOLD response to intact pictures of individual objects compared with their scrambled counterparts (Malach et al. 1995; Grill-Spector et al. 1999). The scrambled images resemble texture, and recent work documents that LOC is, indeed, the earliest stage in the ventral pathway where shape is distinguished from texture (Grill-Spector et al. 2001; Cant and Goodale 2007). LOC has been implicated as critical for shape recognition in that bilateral lesion to this area renders an individual deeply agnostic for shape, particularly for line drawings where inferences cannot be made on the bases of color or material properties (James et al. 2003). Insofar as LOC is posterior to PPA, it might be assumed to be earlier in the ventral pathway, although the precise time course of these areas and their sequential dependency, if any, have yet to be firmly determined. A simple hypothesis, consistent with the finding that activity in LOC to 2 simultaneously presented noninteracting objects can be well predicted by the average responses to the constituents presented singly (MacEvoy and Epstein 2009), is that LOC defines critical shapes that are then fed forward for scene-like

specification in PPA. This suggests that LOC would not differentiate between 2 objects presented as interacting or side by side.

An alternative hypothesis is that scene-like relations among the objects are specified in LOC, an account consistent with the subjective impression that objects and their relations are perceived simultaneously and relations among objects can facilitate or interfere (when incongruous) with the identification of the objects themselves (Biederman 1972; Biederman et al. 1974, 1982; Green and Hummel 2004, 2006). Biederman (1981) demonstrated that the identification of a scene could readily be achieved by the structure of the objects, even when the contours of the objects comprising that scene were so degraded that they could not be identified when presented in isolation. LOC has also recently been shown to be particularly sensitive to the relative positions of 2 objects (Hayworth et al. forthcoming) and important in coding the relative relations between multiple parts within an object (Behrmann et al. 2006).

We examined activity in functionally defined regions of interest (ROIs), PPA, and LOC as well as control ROIs such as early visual retinotopic areas and potentially attention-sensitive regions—the intraparietal sulcus (IPS) (Wojciulik and Kanwisher 1999; Kanwisher and Wojciulik 2000; Riddoch et al. 2003) and dorsolateral prefrontal cortex (DLPFC)—as subjects viewed a series of scenes. Each scene consisted of a pair of objects presented either side by side (e.g., a bird and a birdhouse), or interacting with each other (e.g., the bird perched on the birdhouse), similar to the stimuli of Epstein et al. (1960) (Fig. 1*a*). In half of the scenes, the stimuli were paired in either familiar or novel depictions (e.g., a bird perched on an ear).

In all experiments, differences in the effects of the variables—larger BOLD responses to interacting than side-by-side depictions and to novel over familiar interactions (but not for side-by-side conditions)—were consistently evident in LOC but not in PPA. A supporting electroencephalography (EEG) study that sought to examine the time course of the scene-like interaction effects (see Supplementary Table 2 and Supplementary Fig. 2) also showed that the sensitivity to object interactions was not apparent in any area prior to that of the occipital-temporal area, roughly corresponding to LOC.

## Materials and Methods

### Subjects

In the fMRI experiments, 25 subjects (all right handed, 12 males, mean age of 25.7 years, range 19–30,  $n = 9$  for Experiment 1,  $n = 6$  for Experiment 2, and  $n = 10$  for Experiment 3) were scanned at the Dana and David Dornsife Cognitive Neuroscience Imaging Center at the University of Southern California (USC). All subjects were screened for safety and gave informed consent in accordance with the procedures approved by USC's Institutional Review Board.

### Stimuli

Stimuli were selected from a set of 96 line drawings of individual objects. These were combined, pairwise, to make 48 different 2-object scenes. Each scene depicted the 2 objects either as interacting (Inter) in half of the blocks or side by side (Side) in the other half. The objects could be a familiar (Fam) pair, in that they could be expected to co-occur, or novel (Novel), with a low probability of co-occurrence (for a listing of the pairs, see Supplementary Table 3). The interacting scenes subtended an average of  $5.1^\circ \times 4.9^\circ$  (width  $\times$  height) presented centrally for Experiments 1 and 2 and presented  $3^\circ$  to the left or right

side of fixation for Experiment 3. The 2 side-by-side objects were depicted at their appropriate real-world relative sizes, which were the same as their sizes when interacting, subtending on average of  $3.6^\circ \times 3.1^\circ$  in Experiments 1 and 2 and in approximately equal sizes (typically by enlarging the smaller of the 2 objects) subtending on average of  $4.9^\circ \times 4.4^\circ$  for Experiment 3 (see Fig. 3*b*). This manipulation allowed an assessment as to whether the BOLD response to the side-by-side depictions might be increased if the objects were presented at equivalent sizes. Each of the side-by-side objects in all experiments were positioned such that their center was  $3^\circ$  left and right of fixation chosen randomly with equal probability on both sides. Stimuli were presented as black line drawings on a white background.

### Behavioral Task in the Scanner

In Experiments 1 and 2, subjects detected if there was a repeat of a scene within a block. At the end of each block, when prompted with a question mark, the subject responded with 1 of 2 buttons indicating whether there was or was not a repeat. A repeat occurred in 33% of the blocks counterbalanced across the 4 conditions. In Experiment 3, subjects performed an odd-man out task where they counted how many instances in each block only a single object appeared. At the end of the block, subjects responded with a button press if they saw 0, 1, or 2 instances. Presentation of only single objects occurred in 11% of all trials, equally distributed across the 4 conditions. The task was changed in Experiment 3 to test whether the pattern of responses in the ROIs would differ with an alternative task.

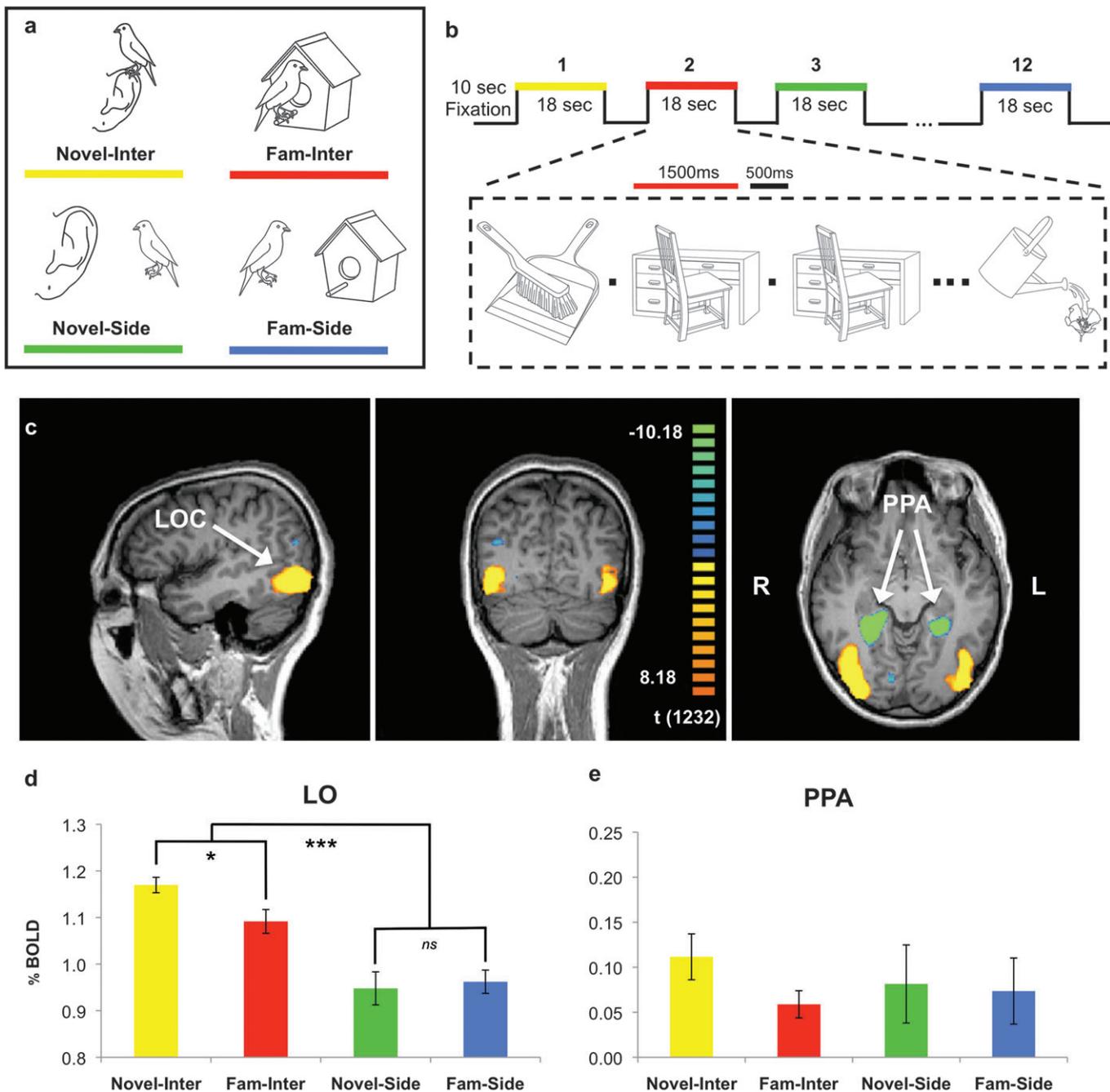
Both the one-back and odd-man-out tasks were orthogonal to the major variables in that subjects did not need to pay attention to the relationship between the 2 objects or the familiarity of their pairing to perform the task accurately. Subjects were instructed (and frequently reminded) to maintain central fixation throughout the experiment.

### Imaging Parameters and fMRI Analyses

Scanning was performed with a Siemens MAGNETOM Trio 3-T scanner with a standard 12-channel coil. Responses were recorded with an MRI compatible button box. One anatomical and 6 functional scans were run for each subject. High-resolution  $T_1$ -weighted structural scans with the magnetization-prepared rapid gradient-echo sequence were acquired with time repetition (TR) = 1950 ms, time echo (TE) = 2.26 ms,  $1 \times 1 \times 1$  mm voxels, and 160 sagittal slices. The functional images were collected using a  $T_2^*$ -weighted echo planar sequence with the following parameters: TR = 2000 ms, TE = 30 ms, field of view =  $224$ , flip angle =  $90^\circ$ , voxel size =  $3.5 \times 3.5 \times 3$  mm, and 32 transversal slices.

Each experimental run was composed of 12 blocks lasting approximately 6.0 min for Experiment 1 and 2 and 5.6 min for Experiment 3. Each block consisted of 10 trials (stimuli) for Experiment 1 and 2 and 9 trials for Experiment 3. Within a run, an equal number of the 4 conditions were presented in a Latin-square block design: Novel-Inter, Fam-Inter, Novel-Side, and Fam-Side. This resulted in a total of 18 blocks per condition. Each run started and ended with 6 s of fixation dot, and each block was separated by 10 s of a fixation dot to allow the hemodynamic response to return to baseline before the start of the next block. Within a block, each scene was presented for 1500 ms, followed by 500 ms of the fixation dot in Experiments 1 and 3. In Experiment 2, each scene was presented for 200 ms followed by 1800 ms of fixation dot. BOLD responses for the 4 main conditions were examined in functionally and anatomically defined ROIs, LOC (composed of LO and pFs), PPA, IPS, DLPFC, and early visual areas (V1, V2, and V4).

All functional imaging runs were preprocessed using the BrainVoyager software package (Brain Innovation BV) including slice scan time correction with sinc interpolation, 3-dimensional motion correction with trilinear interpolation, spatial smoothing with 4-mm full-width at half-max Gaussian filter, and temporal filtering using a high-pass filter of 3 cycles over the run's length for linear trend removal. All functional images were coregistered to each individual subject's anatomical scan. The anatomical scans were transformed into Talairach coordinates (Talairach and Tournoux 1988), on which all the statistical analyses were performed.



**Figure 1.** Stimuli, design, and results for Experiment 1. (a) Sample stimuli. (b) Presentation sequence of a single run with an illustration of a Fam-Inter block containing a "repeat" of a scene. (c) Activation map of LOC (in yellow) and PPA (in green) on a representative subject's brain. Results in LO (d) and PPA (e) for Experiment 1. The error bars (here and for all subsequent figures) represent the standard errors computed from the deviation scores around each subject's overall mean. \*\*\* $P < 0.001$ , \* $P < 0.05$ .

The average response for each condition was computed as a percent signal change by using the averaged value of the BOLD signal at seconds  $-4$  to  $0$  preceding the start of each block as the baseline. For a given ROI for each subject, all voxels' responses were averaged, and the average peak hemodynamic responses were calculated from time points  $6$ – $24$  s for Experiments 1 and 2 and  $6$ – $22$  s for Experiment 3 (corresponding to the number of trials for a given experiment) after the start of each block. With the peak responses for each condition, a repeated-measures  $2 \times 2$  analysis of variance with Relations (Inter vs. Side) and Novelty (Novel vs. Fam) were performed in all ROIs with  $\alpha = 0.05$ .

#### ROI Procedure

Two additional block-designed localizer scans (approximately 3.5 min each) were run to define LOC, PPA, and IPS for each subject. Each run

was composed of sixteen 12-s blocks with alternating blocks between intact objects, places, faces, and scrambled images. Each image subtended a visual angle of  $6^\circ \times 6^\circ$  presented centrally. Subjects were asked to passively view the stimuli.

For each subject, LOC was defined by comparing the BOLD activations of the conjunction contrast of object minus places and object minus scrambled with a  $t$ -map threshold of  $P < 0.05$ , Bonferroni corrected. These LOC voxels were then divided into 2 subregions, LO and pFs based on anatomical location as described previously (Grill-Spector et al. 1999; Kourtzi et al. 2003; Haushofer et al. 2008). LO, the posterior part of LOC, was defined as those voxels in the most dorsal-caudal part of LOC. PFs, the anterior part of LOC, was defined as those nonoverlapping voxels anterior and ventromedial to LO. PPA was

defined by the conjunction contrast of places minus objects and places minus scrambled with the same  $t$ -map threshold (Fig. 1c).

The sizes and locations of these ROIs varied across subjects. The mean volume for each ROI and the mean peak activation coordinates in Talairach space (Talairach and Tournoux 1988) are shown in Supplementary Table 1.

#### Functional and Anatomical Localizer for IPS and DLPFC

Using the same functional localizer scans for defining LOC and PPA, IPS was defined functionally and anatomically as those voxels in the intraparietal region showing greater activity to objects minus scrambled blocks with a  $t$ -map threshold of  $P < 0.05$ , Bonferroni corrected. This method of defining IPS was similar to that used by Xu and colleagues (Xu and Chun 2006; Xu 2008). IPS was defined in 23 of 25 subjects.

Using 1 of the 6 main experimental functional scans, DLPFC was functionally and anatomically defined as those voxels showing greater activation to all of the main experimental conditions minus fixation in the dorsolateral region of the frontal cortex. The same  $t$ -map threshold was used. The functional run used to define DLPFC was excluded in examining the differences in BOLD responses to the experimental conditions to avoid potential circular analysis giving rise to invalid statistical results, (Kriegeskorte et al. 2009). DLPFC was defined in all subjects.

#### Retinotopic Areas

Standard retinotopic maps were acquired for 10 of the subjects run in the fMRI experiments (2 from Experiment 1, 3 from Experiment 2, and 5 from Experiment 3) in separate scan sessions. Two functional runs with flashing checkerboard wedge stimuli were run to define V1, V2, and V4 boundaries by utilizing a standard voxel-wise correlation method (Serenio et al. 1995; Engel et al. 1997). Details of imaging parameters are described in the Supplementary Material.

## Results

For each experiment, results in LOC and PPA, the 2 main ROIs, are reported first. In subsequent sections, results in control ROIs including early visual areas, IPS, and DLPFC are reported followed by results from a whole-brain contrast analysis.

### Experiment 1

All of our experiments showed virtually identical patterns of responses to the experimental conditions in LO and pFs, and only the peak responses for LO are shown in Figure 1d (and in subsequent figures). Although there was a slight reduction in the number of contrasting pixels in the interacting images (when contours overlapped), the interacting objects nonetheless elicited a larger BOLD signal than their side-by-side versions, in LO,  $F_{1,8} = 41.35$ ,  $P < 0.001$ , and in pFs,  $F_{1,8} = 5.36$ ,  $P < 0.05$ . The novel interactions elicited greater activation than

the familiar interactions ( $t_8 = 2.72$ ,  $P = 0.03$  for LO and  $t_8 = 2.45$ ,  $P = 0.04$  for pFs) but not for the side-by-side depictions (both  $t_8 < 1.00$ ), producing a significant Relation  $\times$  Novelty interaction in LO,  $F_{1,8} = 8.75$ ,  $P = 0.02$ , and in pFs,  $F_{1,9} = 8.47$ ,  $P < 0.02$  (see Fig. 1d).

Unlike LOC, PPA (Fig. 1e) did not show a greater response to the interacting than side-by-side conditions,  $F_{1,8} < 1.00$ , and did not show a statistical interaction between Relations and Novelty,  $F_{1,8} = 1.68$ ,  $P = 0.23$ . Although typically PPA has been shown to be activated by scenes (or places), spatial structure, or contextual processing of objects (Epstein and Kanwisher 1998; Epstein et al. 1999, 2003; Bar 2004; Aminoff et al. 2006), the scene-like interaction effects were only evident in LOC.

Accuracy on the one-back task was high (96.4%) and did not differ across conditions ( $F$ s for all comparisons  $< 1.00$ ).

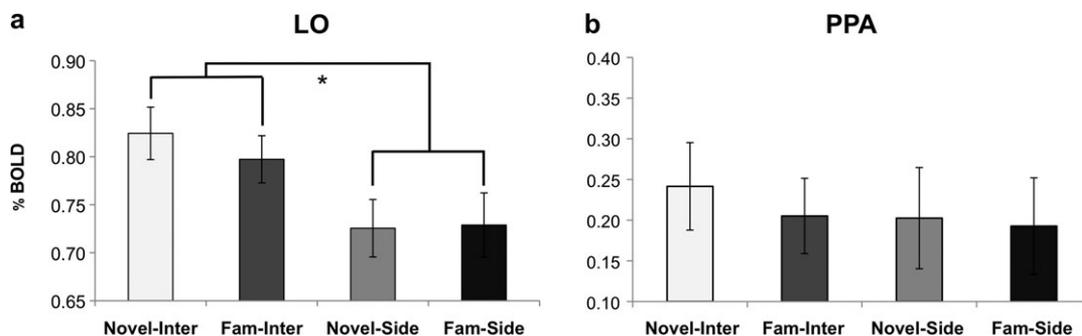
### Experiment 2: Brief Exposure Duration

Although subjects were instructed to maintain central fixation at all times, it is possible that some subjects on some trials did not. To control for potential eye movements as an account of the activity pattern in LOC, an experiment was run with 6 subjects with a presentation duration of only 200 ms (vs. 1500 ms in Experiment 1), followed by a 1800-ms fixation dot. All other parameters were the same as before.

As in Experiment 1, subjects were highly accurate in the behavioral task in the scanner (92.4%), and none of the comparisons between conditions showed a difference in performance ( $F$ s for all comparisons  $\leq 1.00$ ).

fMRI results are shown in Figure 2. Consistent with the previous experiment, the Inter conditions still showed a significantly greater BOLD response than the Side conditions in both LO,  $F_{1,5} = 9.26$ ,  $P = 0.03$ , and pFs,  $F_{1,5} = 6.71$ ,  $P < 0.05$ . There was a trend toward the novel stimuli producing greater activity than the familiar ones in the interacting but not the side-by-side conditions, but this interaction fell short of significance in LO,  $F_{1,5} = 4.81$ ,  $P = 0.08$ , and pFs,  $F_{1,5} = 4.30$ ,  $P = 0.09$ . In PPA, the Inter conditions did not produce a greater BOLD responses than the Side conditions,  $F_{1,5} = 1.84$ ,  $P = 0.23$ . The statistical interaction of Relations and Novelty also was not reliable in PPA,  $F_{1,5} = 1.67$ ,  $P = 0.26$ .

Even with a brief presentation duration, rendering it unlikely that subjects were making eye movements during the stimulus presentation, the interacting conditions still produced greater activity in LOC than the side-by-side conditions. It is thus highly unlikely that the pattern of responses in LOC is due to differential eye movements across conditions.



**Figure 2.** Experiment 2: Brief exposure duration. Peak BOLD responses are shown for LO (a) and PPA (b) for Experiment 2 where each scene was presented for 200 ms to reduce eye movements during stimulus presentations. LO showed a pattern of responses similar to that of Experiment 1. \* $P < 0.05$ .

### Experiment 3: Control for Foveal Overrepresentation

In Experiments 1 and 2, interacting objects were presented centrally while the side-by-side objects flanked central fixation. Could the greater BOLD activation of interacting objects be due to foveal overrepresentation in the cortex? In Experiment 3, interacting objects were presented off-center, either to the left or to the right of fixation chosen randomly, matching one of the positions of the side-by-side objects (Fig. 3*a*). If foveal overrepresentation was affecting the BOLD results, LOC should now be expected to show greater BOLD responses to the side by side than the interacting conditions, since the interacting stimuli were presented off-center and reduced in size by 50% compared with the side-by-side images. Experiment 3 also provided a test as to whether the performance of a different task (odd-man out), and presenting the objects in the Side condition at equal (rather than real-world relative sizes), would alter the pattern of results in LOC.

As in previous experiments, LO and pFs (Fig. 3*c*) showed significantly greater BOLD responses to the Inter than the Side conditions in LO,  $F_{1,9} = 67.14$ ,  $P < 0.001$ , and pFs,  $F_{1,9} = 52.44$ ,  $P < 0.001$ . Again, the greater BOLD responses to the novel than familiar were present only for the interacting objects (LO:  $t_8 = 2.72$ ,  $P = 0.03$  and pFs:  $t_8 = 3.72$ ,  $P < 0.01$ ) and not for the side-by-side depictions (both  $t_8 < 1.00$ ), producing a significant Relation  $\times$  Novelty interaction in LO,  $F_{1,9} = 6.01$ ,  $P < 0.05$ , and pFs,  $F_{1,9} = 5.35$ ,  $P < 0.05$ . This time, PPA also showed a greater response to the interacting than side-by-side conditions,  $F_{1,9} = 12.43$ ,  $P < 0.01$ , but not a reliable statistical interaction of Relations and Novelty,  $F_{1,9} = 2.69$ ,  $P = 0.14$ . The magnitude of the Inter minus Side conditions in PPA, however, was markedly smaller than that of LO,  $t_8 = 3.90$ ,  $P < 0.01$ , and pFs,  $t_8 = 3.65$ ,  $P < 0.01$ .

Subjects' performance on the odd-man out task was high, 92.5%, and did not differ across conditions ( $F$ s for all comparisons  $< 1.00$ ).

### Early Visual Areas, IPS, and DLPFC

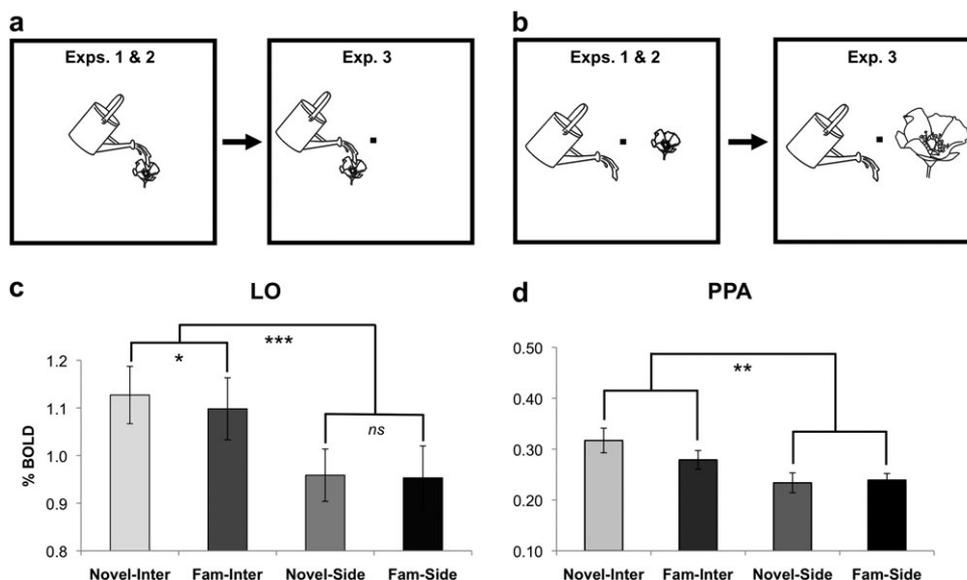
To address whether the effects witnessed in LOC could be fed forward from earlier visual areas, we examined activity in V1,

V2, and V4 (Fig. 4*a*). Unlike LOC, the earliest visual areas did not show a reliable difference for Inter versus Side conditions (V1:  $F_{1,9} < 1.00$ ; V2:  $F_{1,9} = 1.69$ ,  $P = 0.23$ ). Although consistent with prior findings (e.g., Kobatake and Tanaka 1994; Grill-Spector et al. 1998), V4, which is adjacent to LO, did show a trend partially reflecting LOC tuning for the Inter versus Side comparison ( $F_{1,9} = 2.80$ ,  $P = 0.13$ ). There was no main effect of Novelty or a reliable Relations  $\times$  Novelty interaction in any of the regions ( $0.39 < P$ s  $< 0.87$ ). LOC for these subjects evidenced the same pattern of results as the other subjects. The pattern of results in LOC was not generated by feed-forward activity from early visual areas.

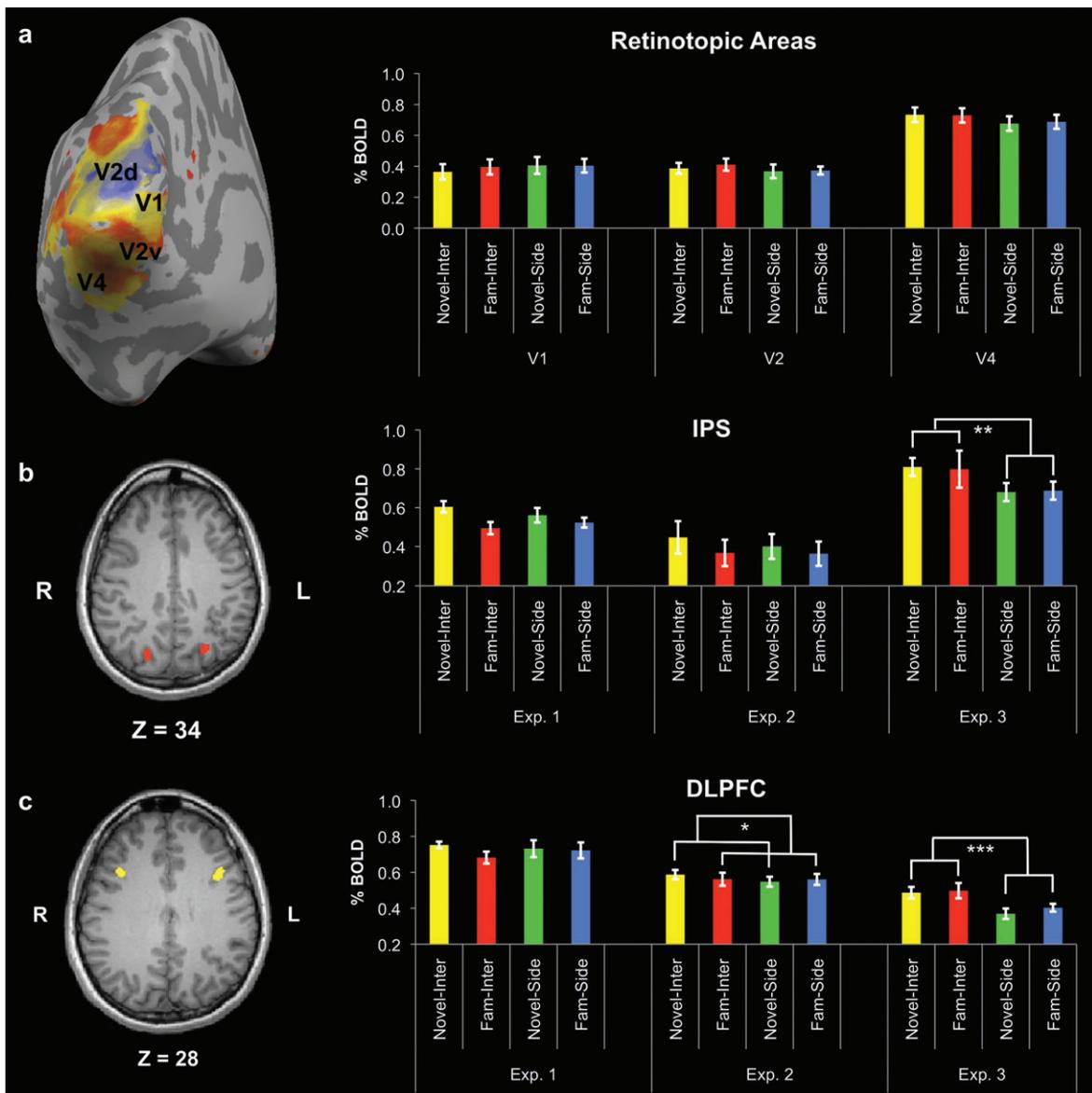
IPS, an area associated with attentional processes (Wojciulik and Kanwisher 1999; Kanwisher and Wojciulik 2000), did not show a consistent pattern of responses across experiments. The main effect of Relations (Fig. 4*b*) was not evident in Experiments 1 and 2 ( $F_{1,7} < 1.00$  and  $F_{1,5} = 2.17$ ,  $P = 0.20$ , respectively) but was reliable in Experiment 3 ( $F_{1,8} = 19.96$ ,  $P < 0.01$ ). Unlike that of LOC, IPS did not show a reliable interaction between Relations and Novelty in any of the experiments (Experiment 1:  $F_{1,7} = 3.01$ ,  $P = 0.12$ ; Experiment 2:  $F_{1,5} = 2.26$ ,  $P = 0.19$ ; and Experiment 3:  $F_{1,8} < 1.00$ ). The pattern of responses in LOC was not evident in IPS across the experiments, rendering it implausible that feedback from this area generated the pattern in LOC.

The novel interactions, which might have been expected to attract attention in IPS, did not reliably produce greater activation than the familiar ones in IPS in any of the experiments (Experiment 1:  $F_{1,7} = 3.30$ ,  $P = 0.12$ ; Experiment 2:  $F_{1,5} < 1.00$ ; and Experiment 3:  $F_{1,8} = 2.57$ ,  $P = 0.15$ ).

Recent studies have shown that activity in the inferior IPS, functionally defined in a manner similar to our definition, increased with the number of objects (or grouped objects) present, regardless of the complexity of the stimuli (Xu and Chun 2006; Xu 2008). Given this result, if anything, IPS should have shown greater responses to the side by side than interacting stimuli, in that the side by side displays presented 2 positionally distinct objects whereas the interacting stimuli



**Figure 3.** Experiment 3: Control for foveal overrepresentation. (a) The interacting conditions (Novel-Inter and Fam-Inter) in Experiment 3 were presented off-center, matching one of the positions of the side-by-side objects. (b) The objects in the side-by-side conditions which were matched in relative sizes in Experiments 1 and 2, were presented in equivalent absolute sizes in Experiment 3. The average peak hemodynamic responses in (c) LO and (d) PPA. \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ .



**Figure 4.** BOLD responses for retinotopic areas, IPS, and DLPFC. Retinotopic early visual areas V1, V2, and V4 are represented in one subject's inflated left hemisphere (a). These regions did not reliably manifest any of the effects seen in LOC. BOLD responses in IPS (b) and DLPFC (c) were inconsistent across experiments. \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ .

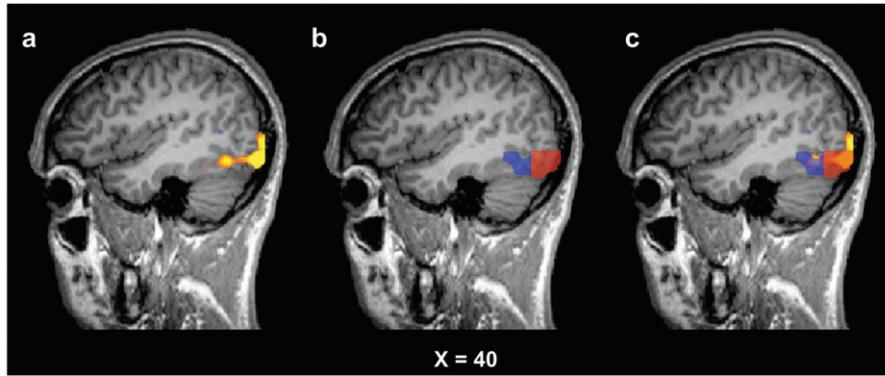
could be regarded as 1 (complex) object. That we did not observe this pattern also suggests that IPS was influenced by the pattern of activity in LOC rather than the reverse.

Like IPS, DLPFC's responses were highly inconsistent across experiments (Fig. 4c) and did not reflect the pattern of results in LOC. In Experiments 1 and 2, DLPFC did not show a greater BOLD response to the interacting than side-by-side conditions, Experiment 1:  $F_{1,8} < 1.00$  and Experiment 2:  $F_{1,5} = 2.21$ ,  $P = 0.20$ . In Experiment 3, however, there was a significant main effect of Relations,  $F_{1,9} = 45.34$ ,  $P < 0.001$ . DLPFC did show a somewhat greater activity to novel than familiar conditions in Experiment 1,  $F_{1,8} = 3.39$ ,  $P = 0.10$ , and Experiment 2,  $F_{1,5} = 11.61$ ,  $P = 0.02$ , but not in Experiment 3,  $F_{1,9} < 1.00$ . Unlike LOC, the Novelty effect was the same for Inter and Side conditions so that in none of the experiments was the interaction of Relations and Novelty in DLPFC reliable;  $F_{1,8} =$

1.86,  $P = 0.21$  for Experiment 1;  $F_{1,5} = 1.49$ ,  $P = 0.28$  for Experiment 2; and  $F_{1,9} < 1.00$  for Experiment 3.

#### Whole-Brain Contrast Analysis

A whole-brain analysis was done to test if any areas either overlapping with the independently defined ROIs or regions possibly missed by the ROI analyses were differentially activated by the conditions. The functional data from the 15 subjects in Experiments 1 and 2 (that had the same behavioral task and stimulus display size and position) were concatenated to increase power, and the contrast of Inter minus Side was done with a threshold of  $P < 0.05$ , Bonferroni corrected. This contrast resulted in only one bilateral region in the ventral pathway (peak activation Talairach coordinates: 31, -83, -4 for the right and -38, -81, -5 for the left patch) located near LOC (as referenced by the peak Talairach coordinates of LO and pFs in Supplementary Table 1). In Figure 5, the right activation



**Figure 5.** Whole-brain contrast analysis. The contrast of Inter minus Side conditions is shown on one representative subject's brain in the right hemisphere (a). LO (in red) and pFs (in blue), defined with independent functional localizers, are shown on the same subject's brain in (b). The activation patch revealed by the contrast analysis overlaps well with the independently defined LOC (c).

patch is shown on one representative subject's brain that overlaps almost completely with independently defined LOC. That the contrast of Inter minus Side revealed only an area overlapping with independently defined LOC confirms the inference that the interaction effect arises at LOC and that we are not missing additional regions that could be sensitive to the Inter versus Side conditions.

The whole-brain contrast of Novel minus Familiar did not reveal any regions of activation.

## Discussion

In all the fMRI experiments, we witnessed the identical signatures of scene-like relations in LOC: significantly greater activation from interacting compared with side-by-side depictions and an additional boost from the novelty of the interaction. The pattern was clearly due to neither feed-forward activity from early visual areas nor a function of eye movements (given that it was also manifested with a brief exposure duration), foveal overrepresentation, or difficulty in object identification. With respect to the last factor, an experiment (see Supplementary Material) in which subjects named both objects in a briefly presented display showed that naming accuracy of the 2 objects in the Inter conditions was not more difficult than in the Side conditions.

PPA showed an inconsistent and markedly weaker pattern of BOLD responses than LOC, suggesting that PPA is not the locus of the interaction effect. That there might be a dependency of PPA on LOC is suggested by an fMRI finding with D.F., a deeply shape agnostic individual with bilateral LOC lesions. It is clear that without LOC, there is no shape—object or scene—perception, even with PPA intact (James et al. 2003; Steeves et al. 2004). D.F. is not only deeply shape agnostic with no capacity for recognition of line drawings of objects (such as those presented in the present experiments) but she is also critically impaired in scene classification when the scenes are devoid of color information, that is, allowing D.F. to completely rely on shape or spatial relations between shapes within the scene (Steeves et al. 2004). PPA has recently been shown to be sensitive to texture or high spatial frequency information that characterize object details (Rajimehr et al. 2008). By using line drawings that, nonetheless, clearly depicted the relations among objects, our study likely reduced that portion of PPA activity driven by texture.

What about feedback from the parietal cortex? Activity patterns in IPS, an area associated with attentional processes (Wojciulik and Kanwisher 1999; Kanwisher and Wojciulik 2000; Riddoch et al. 2003), were not consistent with that of LOC, rendering it implausible that feedback from IPS generated the pattern of responses in LOC. Moreover, the novel interactions did not produce greater activation than the familiar ones in IPS in any of the experiments. A supporting EEG experiment (in Supplementary Material) showed that the parietal region did not manifest an earlier divergence of the interacting versus side-by-side conditions than the LOC region, further eliminating the possibility that effects seen in LOC are a direct consequence of parietal cortex activity.

Consistent with the interpretation that the scene effects do not arise in the parietal cortex is the finding that patients with parietal lesions who show extinction, an attentional phenomenon where they fail to report the stimulus presented on the contralesional side due to competition with a stimulus shown on the ipsilesional side, demonstrate significant recovery when an object pair is depicted as interacting (e.g., a wine bottle pouring into a wineglass) than presented side by side (Riddoch et al. 2003, 2006). It is likely that these patients are utilizing their intact ventral streams to organize the 2 stimuli into a single entity by encoding the relations between the 2 objects. This grouping is visual rather than a post-visual semantic effect (Green and Hummel 2006): there was no recovery of extinction to pairs of objects with similar semantic information but not interacting, for example, a pen and a pencil (Riddoch et al. 2003).

Like IPS, DLPFC's activity was inconsistent across experiments and did not show the pattern of responses that characterized LOC. Lesions to prefrontal areas do not result in deficits in scene perception, or in vision, for that matter. Rather the deficits they present are in planning, problem solving, and reasoning (e.g., Milner and Petrides 1984; Delis et al. 1992; Bechara et al. 1994). Whatever the feedback from prefrontal cortex to LOC, it would seem implausible that such feedback was causal to the pattern seen in LOC.

Unlike Experiments 1 and 2, where LOC was the only area that evidenced a larger BOLD response to the Inter than Side conditions, in Experiment 3 PPA, IPS, and DLPFC joined LOC in evidencing that effect. Whether sensitivity to object interactions in regions other than LOC is dependent on stimulus position (for PPA, Levy et al. 2001, 2004) or task remains to be

determined. That these regions did not show an interaction of Relations and Novelty in any of the experiments suggests that this neural signature seen in LOC is specific to this region. The consistency of the effects of object relations and novel interactions in LOC across all experiments also shows that LOC activity is robust across subtle differences in stimulus size, presentation, and tasks.

The greater response to interacting pairs is inconsistent with the characterization of activity in LOC as a simple average of the BOLD responses produced by the individual objects (MacEvoy and Epstein 2009). Instead, LOC is clearly sensitive to the relations among objects. One possibility is that the distinctive relation of each interacting pair (e.g., perching) activated additional associations, and it is this activity that magnified the BOLD response. Consistent with this view is that LOC has been shown to encode relative positions among objects (Hayworth et al. forthcoming) and that its responses are greater for structured versus random sequences of shapes (Turk-Browne et al. 2008).

Similarly, it is possible that the familiar (vs. novel) relations have already undergone competitive learning, which would be expected to diminish neural activity. That is, initially strongly responsive neurons may inhibit ones more weakly activated, an explanation that has been advanced to account for the reduction of the BOLD signal observed in longer term priming experiments (Schacter and Buckner 1998). When objects are presented side by side, objects may be encoded as a list or a simple conjunction (i.e., “and”), both already well-adapted relations. That none of the experiments showed a difference in BOLD responses to the novel versus familiar pairings in the side-by-side condition suggests that the effects shown in LOC are not driven by the semantics of the objects as would be expected from the finding that adaptation in LOC is not sensitive to the semantic category of individual objects (Kim et al. 2009). However, DLPFC, which showed a main effect of novelty (in Experiments 1–2), may have been driven by the semantics of the object pairs. This is consistent with the findings that the prefrontal cortex is sensitive to semantic categories of objects (Freedman et al. 2001, 2003).

The interactions define the functional relations between objects even for the novel pairs where the interactions were plausible (e.g., a bird perched on an ear) as in the stimuli of Epstein et al. (1960). Recent studies have indicated that LOC shows sensitivity to action relatedness of objects (Roberts and Humphreys 2010; Valyear and Culham 2010). The stimulus set used in the current experiments (see Supplementary Table 3) consisted of both action-related pairs, for example, a butterfly being caught by a net or a dust brush dusting an oven, and non-action-related pairs, for example, a cup on a saucer and a plant on a chair. Given the recent fMRI findings, it is possible that action relatedness played a role in the scene-like interaction effect for both the familiar and novel pairs. This is consistent with Roberts and Humphreys’ (2010) results that both familiar and unfamiliar object pairs showed greater activity in LOC when positioned to interact than when they were not interacting and regardless of whether the subjects were paying attention to the stimuli. Given that our pairs included both types of items (in approximately equal proportions) mixed together, additional research will be necessary to determine the extent to which action relatedness may have played a role.

Could difficulty of object identification produce the scene-like interaction effects? This seems unlikely as previous

behavioral experiments showed the opposite pattern; subjects show greater efficiency of object recognition/encoding when presented as interacting compared with not interacting (Epstein et al. 1960; Green and Hummel 2004, 2006). As noted previously, a behavioral experiment (see Supplementary Material) showed that the accuracy of object identification did not differ between the Inter and Side conditions, suggesting that the BOLD response to the 2 conditions in LOC cannot be explained by differences in difficulty of object naming.

## Conclusions

Neuroimaging studies examining LOC’s properties have largely been based on displays of single objects (e.g., Grill-Spector et al. 1999; Vuilleumier et al. 2002). The present results provide clear evidence that there is a distinct, robust signature in LOC of the interactions between objects. The processing of these relations appears to be obligatory in that none of the tasks required their processing. The picture that emerges is one in which scene-like relations are not inferred at some stage following object identification but are likely achieved simultaneously with the specification of object shape.

## Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

## Funding

National Science Foundation (04-20794, 05-31177, 06-17699); National Institute of Health (BRP EY016093) to I.B.

## Notes

The authors thank Kenneth J. Hayworth, Mark D. Lescoart, Xiaokun Xu, and Ori Amir for helpful discussions, Jiancheng Zhuang for his role in maintaining the MRI scanner, and Xiangrui Li for assistance with the EEG recording. *Conflict of Interest*: None declared.

## References

- Aguirre GK, Zarahn E, D’Esposito M. 1998. An area within human ventral cortex sensitive to “building” stimuli: evidence and implications. *Neuron*. 21:373–383.
- Aminoff E, Gronau N, Bar M. 2006. The parahippocampal cortex mediates spatial and nonspatial associations. *Cereb Cortex*. 17:1493–1503.
- Bar M. 2004. Visual objects in context. *Nat Rev Neurosci*. 5:617–629.
- Bar M, Aminoff E. 2003. Cortical analysis of visual context. *Neuron*. 38:347–358.
- Bar M, Aminoff E, Ishai A. 2008. Famous faces activate contextual associations in the parahippocampal cortex. *Cereb Cortex*. 18:1233–1238.
- Bechara A, Damasio AR, Damasio H. 1994. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*. 50:7–15.
- Behrmann M, Peterson MA, Moscovitch M, Suzuki S. 2006. Independent representation of parts and the relations between them: evidence from integrative agnosia. *J Exp Psychol*. 32:1169–1184.
- Biederman I. 1972. Perceiving real-world scenes. *Science*. 177:77–80.
- Biederman I. 1981. On the semantics of a glance at a scene. In: Kubovy M, Pomerantz JR, editors. *Perceptual organization*. Hillsdale (NJ): Lawrence Erlbaum 213–253.
- Biederman I, Mezzanotte RJ, Rabinowitz JC. 1982. Scene perception: detecting and judging objects undergoing relational violations. *Cognit Psychol*. 14:143–177.

- Biederman I, Rabinowitz J, Glass AL, Stacy EW, Jr. 1974. On the information extracted from a glance at a scene. *J Exp Psychol.* 103:597-600.
- Cant JS, Goodale MA. 2007. Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cereb Cortex.* 17:713-731.
- Delis DC, Squire LR, Bihrlle A, Massman P. 1992. Componential analysis of problem-solving ability: performance of patients with frontal lobe damage and amnesic patients on a new sorting test. *Neuropsychologia.* 30:683-697.
- Engel SA, Glover GH, Wandell BA. 1997. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex.* 7:181-192.
- Epstein R, Grahn KS, Downing PE. 2003. Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron.* 37:865-876.
- Epstein R, Harris A, Stanley D, Kanwisher N. 1999. The parahippocampal place area: recognition, navigation, or encoding? *Neuron.* 23:115-125.
- Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature.* 392:598-601.
- Epstein W, Rock I, Zuckerman CB. 1960. Meaning and familiarity in associative learning. *Psychol Monogr Gen Appl.* 74:1-22.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK. 2001. Categorical representation of visual stimuli in the primate prefrontal cortex. *Science.* 291:312-316.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK. 2003. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *J Neurosci.* 23:5235-5246.
- Green C, Hummel JE. 2004. Functional interactions affect object detection in non-scene displays. In: Forbus K, Gentner D, Reiger T, editors. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society.* Mahwah (NJ): Erlbaum. p. 488-493.
- Green C, Hummel JE. 2006. Familiar interacting object pairs are perceptually grouped. *J Exp Psychol.* 32:1107-1119.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R. 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron.* 24:187-203.
- Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzhak Y, Malach R. 1998. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp.* 6:316-328.
- Grill-Spector K, Kourtzi Z, Kanwisher N. 2001. The lateral occipital complex and its role in object recognition. *Vision Res.* 41:1409-1422.
- Haushofer J, Livingstone MS, Kanwisher N. 2008. Multivariate patterns in object selective cortex dissociate perceptual and physical shape similarity. *PLoS Biol.* 6:1459-1467.
- Haxby JV, Ungerleider LG, Howitz B, Maisog JM, Rapoport SI, Grady CL. 1996. Face encoding and recognition in the human brain. *Proc Natl Acad Sci U S A.* 93:922-927.
- Hayworth K, Lescroart MD, Biederman I. Forthcoming. Visual relation encoding in anterior LOC. *J Exp Psychol Hum Percept Perform.*
- James TW, Culham J, Humphrey GK, Milner AD, Goodale MA. 2003. Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. *Brain.* 126:2463-2475.
- Kanwisher N, Wojciulik E. 2000. Visual attention: insights from brain imaging. *Nat Rev Neurosci.* 1:91-100.
- Kanwisher N, Woods RP, Iacoboni M, Mazziotta JC. 1997. A locus in human extrastriate cortex for visual shape analysis. *J Cogn Neurosci.* 9:133-142.
- Kim JG, Biederman I, Lescroart MD, Hayworth KJ. 2009. Adaptation to objects in the lateral occipital complex (LOC): shape or semantics? *Vision Res.* 49:2297-2305.
- Kobatake E, Tanaka K. 1994. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol.* 71:856-867.
- Kourtzi Z, Erb M, Grodd W, Bulthoff HH. 2003. Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cereb Cortex.* 13:911-920.
- Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. 2009. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci.* 12:535-540.
- Levy I, Hasson U, Avidan G, Hendler T, Malach R. 2001. Center-periphery organization of human object areas. *Nat Neurosci.* 4:533-539.
- Levy I, Hasson U, Harel M, Malach R. 2004. Functional analysis of the peripheral effect in human building related areas. *Hum Brain Mapp.* 22:15-26.
- MacEvoy SP, Epstein RA. 2009. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr Biol.* 19:1-5.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady BR, Tootell RBH. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A.* 92:8135-8139.
- Miller GA. 1965. Some preliminaries to psycholinguistics. *Am Psychol.* 20:15-20.
- Milner B, Petrides M. 1984. Behavioural effects of frontal-lobe lesions in man. *Trends Neurosci.* 7:403-407.
- Pinker S. 2007. *The stuff of thought: language as a window onto human nature.* New York: Viking Press.
- Rajimehr R, Devaney K, Young J, Postelnicu G, Tootell RB. 2008. The 'parahippocampal place area' responds selectively to high spatial frequencies in humans and monkeys. *J Vis.* 8:85a.
- Riddoch MJ, Humphreys GW, Edwards S, Baker T, Willson K. 2003. Seeing the action: neuropsychological evidence for action-based effects on object selection. *Nature.* 6:82-89.
- Riddoch MJ, Humphreys GW, Hickman M, Clift J, Daly A, Colin J. 2006. I can see what you are doing: action familiarity and affordance promote recovery from extinction. *Cogn Neuropsychol.* 23:583-605.
- Roberts KL, Humphreys GW. 2010. Action relationships concatenate representations of separate objects in the ventral visual system. *Neuroimage.* 54:1541-1548.
- Schacter DL, Buckner RL. 1998. Priming and the brain. *Neuron.* 20:185-195.
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen RB, Tootell RBH. 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science.* 268:889-893.
- Steeves JKE, Humphrey GK, Culham JC, Menon RS, Milner AD, Goodale MA. 2004. Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patient with visual form agnosia. *J Cogn Neurosci.* 16:955-965.
- Talairach J, Tournoux P. 1988. *Co-planar stereotaxic atlas of the human brain.* Stuttgart (Germany): Thieme.
- Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. 2008. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. *J Cogn Neurosci.* 21:1934-1945.
- Valyear KF, Culham JC. 2010. Observing learned object-specific functional grasps preferentially activates the ventral stream. *J Cogn Neurosci.* 22:970-984.
- Vuilleumier P, Henson RN, Driver J, Dolan RJ. 2002. Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci.* 5:491-499.
- Wojciulik E, Kanwisher N. 1999. The generality of parietal involvement in visual attention. *Neuron.* 23:747-764.
- Xu Y. 2008. Representing connected and disconnected shapes in human inferior intraparietal sulcus. *Neuroimage.* 40:1849-1856.
- Xu Y, Chun MM. 2006. Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature.* 440:91-95.