

# What Is Actually Affected by the Scrambling of Objects When Localizing the Lateral Occipital Complex?

Eshed Margalit, Irving Biederman, Bosco S. Tjan, and Manan P. Shah

## Abstract

■ The lateral occipital complex (LOC), the cortical region critical for shape perception, is localized with fMRI by its greater BOLD activity when viewing intact objects compared with their scrambled versions (resembling texture). Despite hundreds of studies investigating LOC, what the LOC localizer accomplishes—beyond distinguishing shape from texture—has never been resolved. By independently scattering the intact parts of objects, the axis structure defining the relations between parts was no longer defined. This led to a diminished BOLD response, despite the increase in the number of inde-

pendent entities (the parts) produced by the scattering, thus indicating that LOC specifies interpart relations, in addition to specifying the shape of the parts themselves. LOC's sensitivity to relations is not confined to those between parts but is also readily apparent between objects, rendering it—and not subsequent “place” areas—as the critical region for the representation of scenes. Moreover, that these effects are witnessed with novel as well as familiar intact objects and scenes suggests that the relations are computed on the fly, rather than being retrieved from memory. ■

## INTRODUCTION

The lateral occipital complex (LOC), composed of the lateral occipital cortex (LO) and the posterior fusiform gyrus (pFs), is defined as the cortical region that shows a greater fMRI BOLD response when viewing images of intact objects compared with the pixel-scrambled versions of these images, which resemble texture. The intact images do not have to be familiar to yield a robust advantage over their scrambled counterparts (Margalit et al., 2016; Malach et al., 1995), suggesting that the greater BOLD response in LOC for the intact objects is not a consequence of greater semantic associations. James, Culham, Humphrey, Milner, and Goodale (2003) showed that this region was critical for shape recognition in that its bilateral lesioning in an individual (D. F.) rendered that person severely shape agnostic but with spared perception of color and material properties, such that the individual could readily identify surfaces as, for example, aluminum or red plastic. That it is shape that is critical for LOC activation—rather than surface properties—is documented by the equivalence in the maintenance of adaptation when line drawings depicting only the orientation and depth discontinuities of photographed objects are viewed compared with a re-presentation of the original photograph (Grill-Spector, Kourtzi, & Kanwisher, 2001). D. F.'s deficit was specific to shape recognition in that her grasping of unfamiliarly shaped objects was normal (Goodale et al., 1994).

The operation of scrambling an image of an object is a drastic one, disrupting just about everything that could be employed in determining its shape. Despite hundreds of studies exploring LOC, it remains unclear just what aspects of the scrambling of shape results in the observed reduction in LOC BOLD response. It would seem implausible that the only function of LOC is to distinguish shape from texture rather than (also) coding specific aspects of shape. Here we assess whether the reduced BOLD response to scrambling as compared with intact objects (both familiar and novel) is due to the loss of the shape of the parts or the loss of the relations between parts or both. The coding of relations revealed in this investigation may not be restricted to the parts of individual objects but are apparent in the relations between objects that define the representation of scenes.

## Response of LOC to Object Parts and Relations

The neural representation of objects in LOC has been shown to be based primarily on the objects' parts and relations (Lescroart & Biederman, 2012; Hayworth & Biederman, 2006), as evidenced in fMRI adaptation and multivoxel pattern analysis paradigms. fMRI adaptation takes advantage of neuronal adaptation—the tendency for the BOLD response to be diminished when the same stimulus is repeated in succession compared with when the second stimulus differs from the first. Although there is uncertainty as to the underlying cause of repetition suppression in fMRI adaptation, we note that fMRI adaptation often reflects psychological similarity. Hayworth

and Biederman (2006) created complementary pairs of line drawings of objects in which each member of a complementary pair had half of the contour deleted so that if the two images were superimposed they would compose the original intact objects without any overlap in contour. With local feature-deleted (LFD) images, every other line and vertex from each part was deleted but the parts remained largely identifiable. Viewing a sequence of two images consisting of the two members of a LFD complementary pair produced no release from adaptation in LOC (compared with presentation of the identical image), indicating that the representation was not sensitive to the change in local features, as long as the same parts could be recovered. The insensitivity to the particular contours that comprise an object's shape has also been documented by Kourtzi and Kanwisher (2001). That it was the parts that were critical was shown by Hayworth and Biederman with part-deleted (PD) complementary pairs of images in which whole parts were deleted from each member of a complementary pair with each image containing approximately half of the (intact) parts (and half of the contour). Viewing a sequence of PD complements led to a complete release from adaptation—equivalent to viewing a sequence of same-name, different-shaped exemplars of that class (e.g., a grand piano followed by an upright piano)—indicating that the representation of objects in LOC, as assessed by the adaptation paradigm, is completely specified by the representation of the parts (in their appropriate relations to other parts) with no contribution of local features, lexical (i.e., name), or subordinate level concept priming. This result is consistent with earlier behavioral priming studies of Biederman and Cooper (1991). Visual priming, as measured by the reduction in naming RTs and error rates from the first to the second block (10 min apart), was equivalent between LFD complements and the repetition of an identical image. However, there was no visual priming between PD complements in that their second block naming perfor-

mance was equivalent to naming a same-name, different-shaped exemplar, indicating that behavioral priming, as well, is dependent on a reinstatement in the image of the object's parts.

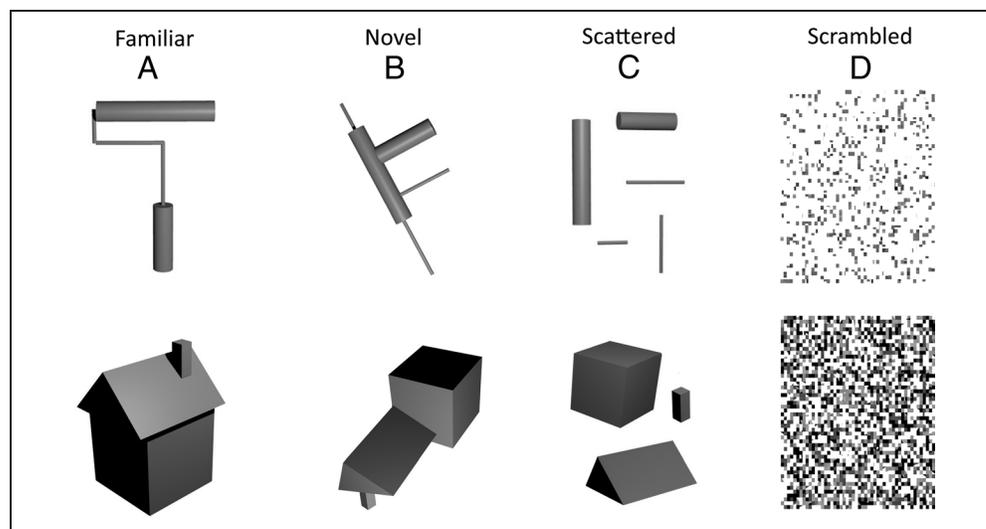
LOC is not only sensitive to the presence of object parts, it is sensitive to the relations between object parts. Lescroart and Biederman (2012) showed that multivoxel pattern analysis could reliably distinguish among different medial axes configurations created by rearrangements of the same geons, despite variations in the overall orientations of the objects and the participants' task of judging which parts were present rather than their configuration. The role of LOC in specifying relations may not be limited to the relations among the parts of an object. There is now considerable evidence that LOC is sensitive to the relations between objects, which is the essence of the representation of the semantics of a scene (e.g., Kim & Biederman, 2011, 2012; Hayworth, Lescroart, & Biederman, 2011; Kim, Biederman, & Juan, 2011; Biederman, 1981).

### Approach

To investigate the selectivity of LOC to part shape and interpart relations, we conducted an fMRI study in which participants were shown 3-D-rendered computer graphic images of familiar and novel objects composed of the same parts, scattered parts of these objects, and pixel-scrambled images of the objects (Figure 1).

To determine whether the greater BOLD response of LOC to intact objects (Figure 1A, B) than to scrambled objects (Figure 1D) was due to the disruption of the shape of the individual parts or the relations between parts, we created a stimulus set, which “scattered” an object's parts by separating, translating, and rotating the parts (Figure 1C). If viewing the scattered objects yielded activation equivalent to that of the corresponding intact

**Figure 1.** Sample stimuli from the experiment. (A) Two examples of Familiar stimuli, a paint roller and a house. (B) Novel, or unfamiliar, versions of the Familiar stimuli created by a rearrangement of the parts that defined a different axis structure from the Familiar versions. (C) Scattered parts. (D) Scrambled images. All the stimuli could be enclosed in a  $3^\circ \times 3^\circ$  bounding box in the center of the screen.



images (familiar or novel), then the loss of part shape would be implicated as the factor responsible for the effects of scrambling. If, however, activation to scattered objects yields a BOLD response equivalent to that of scrambled objects, then part relations would be implicated as the critical variable. If the magnitude of the BOLD response to the scattered images was greater than that to the scrambled images but less than that to the intact images, then the loss of both part shape and the disruption of the axis structure defining the relations between the parts would be implicated as the consequence of scrambling the images.

## METHODS

### Participants

Thirteen right-handed university students (6 women, mean age = 20.7 years, age range = 18–27 years) participated in the study. All participants were screened for MRI safety and provided informed consent in accordance with the University of Southern California's institutional review board guidelines. Data from 12 participants were used in the following analyses, as functionally defined ROIs could not be reliably localized in one female participant.

### Stimuli

In the “Familiar” condition, stimuli were 3-D-rendered objects, composed of geons, which participants in another experiment (Margalit et al., 2016) had quickly and reliably identified as familiar and nameable (Figure 1A). The stimulus set was generated from a set of images of 51 different objects, all exemplars of familiar basic level categories. In the “Novel” condition, the parts composing the Familiar objects were rearranged such that they formed an unfamiliar object (Figure 1B). In the “Scattered” condition, the object parts were separated in space in a different arrangement from the original, breaking the contiguity of the object's parts (Figure 1C); that is, object parts appeared dissociated from one another, thus leaving undefined a medial axis structure that could specify the relations among the parts. All 3-D-rendered stimuli had no apparent texture but were designed to display realistic lighting and shadows. They were presented in the center of the display such that they could, in principle, be enclosed within a  $3^\circ \times 3^\circ$  bounding box with no apparent differences in eccentricity between the conditions. For some of the stimuli in the Scattered condition, the geons were moderately reduced in size (e.g., as for the lower object in Figure 1) but they all remained readily identifiable. Scrambled stimuli were generated by randomly permuting chunks of  $3 \times 3$  pixels from the familiar object images (Figure 1D). All nonscrambled stimuli were created in Blender ([www.blender.org](http://www.blender.org)), and the scrambled versions of the intact objects were created in MATLAB (The MathWorks, Natick, MA).

## Procedure and Experimental Design

### Structural Scans

A high-resolution T1-weighted structural scan was acquired for each participant for the purpose of within- and between-subject registration during fMRI analysis.

### Functional Scans

Each individual participated in four functional runs: The first two runs followed a block design (340 sec each), and the last two runs followed a rapid event-related design (560 sec each). Participants were encouraged to take 1-min breaks between each run but had the option to begin each run at their convenience. Each run in the block design experiment consisted of 17 blocks. The order of conditions was counterbalanced across the four conditions—familiar, novel, scattered, and scrambled—such that each condition preceded every other condition with uniform frequency. The first block of each run, which had no preceding block, was excluded from the statistical analyses. Within each block, 10 images were presented for 1667 msec each, followed by a fixation cross presented for 333 msec (ISI), resulting in a block duration of 20 sec.

For the two event-related runs, stimuli from five conditions—familiar, novel, scattered, scrambled, and null—were presented in pseudorandomized and counterbalanced order with a balancing look-back of two trials. In the null condition, only the fixation cross was presented. Each stimulus was shown for 500 msec and followed by a fixation cross that remained on screen for 2, 3, or 4 sec (ISI and condition order was pseudorandomized and counterbalanced).

For all functional runs, participants were instructed to maintain focus on the stimuli, which were presented in the center of the screen and surrounded by a rectangular colored border. As an orthogonal task designed to maintain attentiveness, participants were instructed to respond as quickly and accurately as possible by button press when the surrounding border changed color, without moving their eyes away from central fixation. The experimenter monitored eye movements to ensure stability and centrality of gaze. In the block design runs, the border changed color once per block, with the change occurring at a random point within each block. In the event-related runs, the border changed color with 25% probability on any given trial. This task was designed to maintain attentiveness while viewing the stimuli without any explicit stimulus-dependent task, thus relying on the automaticity of object processing (Smith & Magee, 1980). Stimuli were presented in MATLAB using the PsychToolbox package (Brainard, 1997; Pelli, 1997).

### Data Acquisition and Imaging Parameters

Data were collected at the Dana and David Dornsife Cognitive Neuroscience Imaging Center at the University of

Southern California using a Siemens 3T Magnetom Prisma (Siemens, Erlangen, Germany) with a 32-channel head coil. Responses were collected using an MRI-compatible button box. High-resolution ( $0.8 \times 0.8 \times 0.8$  mm) T1-weighted images were collected using an MPRAGE sequence with repetition time = 2400 msec, echo time = 2.24 msec, inversion time = 1060 msec, 192 sagittal slices, and flip angle =  $8^\circ$ . Functional T2\*-weighted images ( $2.0 \times 2.0 \times 2.0$  mm) were collected using an EPI sequence with repetition time = 1000 msec, echo time = 37 msec, field of view (FOV) = 192 mm, flip angle =  $52^\circ$ , and multiband acceleration of 8 to obtain full-brain coverage with 72 axial slices.

## Data Analysis

### Whole-brain Voxel-wise Analysis

The functional imaging data processing was carried out using FEAT (fMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, [www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl)). 3-D motion was corrected using MCFLIRT. No slice timing correction was performed. A space-domain 3-D spatial smoothing was performed using a 5-mm FWHM Gaussian filter on all volumes. Lastly, each volume sequence was filtered using a high-pass filter set to 80 sec (the length of four blocks).  $Z$  (Gaussianized  $T/F$ ) statistic images were thresholded using clusters determined by  $Z > 2.3$  and a (corrected) cluster significance threshold of  $p = .05$  (Worsley, 2001). FILM (FMRIB's Improved Linear Model) prewhitening was used to provide a robust and accurate nonparametric estimation of time series autocorrelation on each voxel's time series (Woolrich, Ripley, Brady, & Smith, 2001). Registration to high-resolution

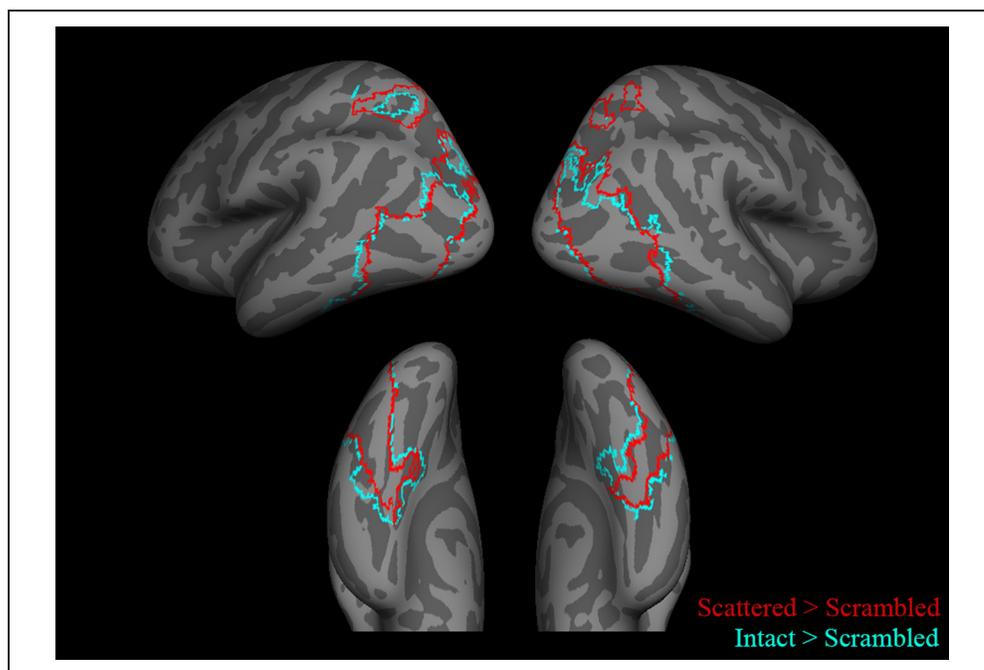
structural and MNI standard space images was carried out using FLIRT (Jenkinson, Bannister, Brady, & Smith, 2002).

For each run, activations were calculated in all acquired voxels from the block design runs using the general linear model. Three predictor variables were explicitly modeled, one for each of the Familiar, Novel, and Scattered conditions. The Scrambled condition was treated as the baseline and not explicitly modeled. Bared parameter estimates are therefore contrasts relative to the Scrambled condition (i.e., "Familiar" is essentially "Familiar" minus "Scrambled"). For each predictor variable, a regressor was formulated as the convolution of a boxcar spanning the duration of the block with a double-gamma canonical hemodynamic response function. Temporal derivatives of each original regressor and motion correction parameters (6 degrees of freedom as computed by MCFLIRT) were also added to the model as regressors of no interest. The two runs of each participant were combined via fixed effects analysis using FEAT. Mixed-effects analysis across participants utilized FEAT FLAME 1 and 2 (FMRIB's Local Analysis of Mixed Effects) to model and estimate the random-effects components of the measured inter-session mixed-effects variance.

### ROI Analysis

Functional ROIs were defined separately for each participant. For each of the two runs, bilateral ROIs LO and pFs were each defined by the union of Familiar > Scrambled with Novel > Scrambled based on the block design runs. For these contrasts, clusters often extended into retinotopic areas V2, V3, and V4, as approximately defined by the PALS visuotopic annotations (Van Essen, 2005) in

**Figure 2.** Outlines of statistical activation maps for the Scattered > Scrambled contrast (red). Activation appears similar in location to the Familiar > Scrambled contrast and Novel > Scrambled contrast (jointly represented as Intact > Scrambled in cyan).



Freesurfer. Thus, to isolate the LOC ROI, clusters were rethresholded beyond the initial  $Z = 2.3$  cutoff, so that clusters extending into areas V2, V3, and V4 could be distinguished from clusters near areas LO and pFs. This process was conducted iteratively by inspection for each participant. The resulting maps contained a group of clusters that included areas LO and pFs but did not extend into earlier retinotopic areas, as well as a separate group of clusters that included retinotopic areas but did not include LO or pFs. Figure 2 shows a loose threshold, such that the two ROIs bleed into a contiguous mass. The ROI definition process, however, involved rethresholding each participant's individual statistical maps to the point that both ROIs were cleanly separated and could be unambiguously defined. This process was conducted manually for each participant. For ROI statistical analyses, mean activation values in terms of percent BOLD signal change from Run 1 were extracted from the ROI defined by Run 2 and vice versa. The values from each run were then averaged for each participant.

#### Peristimulus Time Course Extraction

The hemodynamic response function (percent BOLD signal change as a function of time) in LOC, relative to the null trials, was estimated at 26 time points (2 before stimulus onset and 24 including and following stimulus onset) in 1-sec intervals using a finite impulse response model. Activation values, in terms of percent BOLD signal change, were extracted from the ROIs defined above. At each time point, the magnitude of BOLD signal change for a given condition is taken as the difference in BOLD response between that condition and fixation. Error bars are calculated as the standard error of the BOLD signal change estimates across participants.

## RESULTS

### Whole-brain Voxel-wise Analyses

We observed robust activation in areas LO and pFs for both the Familiar > Scrambled and Novel > Scrambled contrasts. The Scattered > Scrambled contrast also

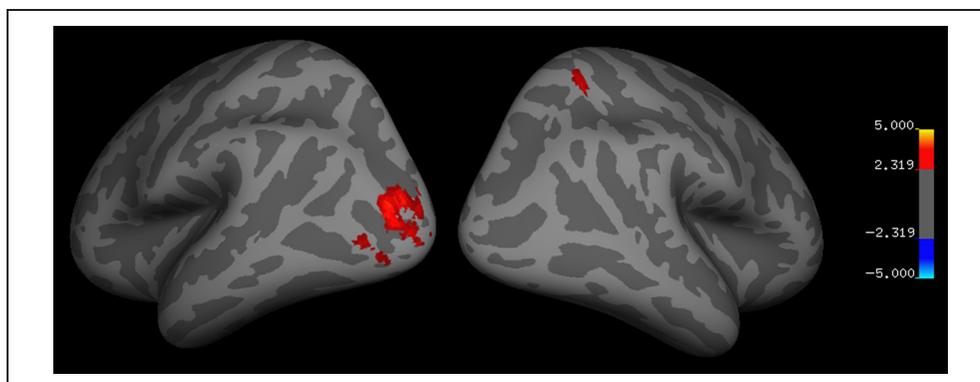
yielded reliable activation in these regions (Figure 2). The Familiar > Scrambled and Novel > Scrambled conditions are collapsed into an "Intact > Scrambled" condition in Figure 2 to more clearly distinguish between the areas defined by traditional localizers (Intact > Scrambled) and the Scattered > Scrambled condition.

The numbers of significant voxels in the Intact > Scrambled and Scattered > Scrambled contrasts were not significantly different from one another, as determined by a paired  $t$  test,  $t(11) \leq 1.00$ ,  $ns$ . Dice similarity coefficients were computed to assess the degree of overlap of the most strongly activated voxels in each contrast map. Given voxels in two maps, the Dice coefficient (Dice, 1945) equals twice the intersection volume divided by the sum of the voxels in each of the maps. Mean Dice coefficients were calculated for the 12 participants over six percentile thresholds between the 50th and 99th percentiles. At the 95th percentile of each contrast, the mean Dice coefficient across participants was 0.50, with a range of 0.34–0.65. With one exception, the 72 Dice coefficients (12 participants  $\times$  6 six percentile threshold values) ranged from 0.34 to 0.72. A single coefficient (from one participant) was at 0.14 at the 99th percentile. There is thus considerable overlap in the peaks of the Intact > Scrambled and Scattered > Scrambled maps, suggesting that the same voxels that are sensitive to the integrity of the parts are also sensitive to the integrity of the relations. Accordingly, the definition of LO and pFs based on the Intact > Scrambled contrast is highly unlikely to bias mean values within those ROIs to be higher for Intact > Scrambled than for Scattered > Scrambled.

### Effects of Familiarity

No voxel clusters showed significantly greater activation to Familiar than to Novel objects, suggesting no preferential effect of familiarity on net BOLD response magnitudes. Novel objects, however, did elicit greater BOLD responses in the left occipital and the right parietal lobes (Figure 3), effects that are consistent with those recently reported by Margalit et al. (2016).

**Figure 3.** Statistical activation maps for the Novel > Familiar contrast. Color scale bar indicates  $Z$  value.



## ROI Analyses

As in Figure 2, the Familiar minus Scrambled and Novel minus Scrambled conditions are combined (averaged) for the following analyses into an Intact minus Scrambled condition. In both ROIs (LO and pFs), a one-sample  $t$  test relative to 0 revealed that the BOLD signal was significantly greater for Intact images than for Scrambled images in LO ( $t(11) = 2.51, p < .05$ , Cohen's  $d = 0.73$ ) and pFs ( $t(11) = 3.48, p < .01$ , Cohen's  $d = 1.0$ ), indicating preferential activation to shapes compared with texture in both regions (Figure 4). Response magnitude was also significantly higher for Scattered images than to Scrambled images in pFs ( $t(11) = 2.73, p < .05$ , Cohen's  $d = 0.79$ ) but was not significantly higher in LO ( $t(11) = 1.13, p = .28$ , Cohen's  $d = 0.33$ ), suggesting preferential activation to intact part shape than to pixel-scrambled texture in pFs (Figure 4). The BOLD signal was significantly higher for Intact images than for Scattered images in both LO and pFs (Figure 4), indicating preferential activation to interpart relations than to isolated object parts in both regions: for LO ( $t(11) = 4.45, p < .001$ , Cohen's  $d = 1.28$ ) and for pFs ( $t(11) = 4.32, p < .01$ , Cohen's  $d = 1.25$ ).

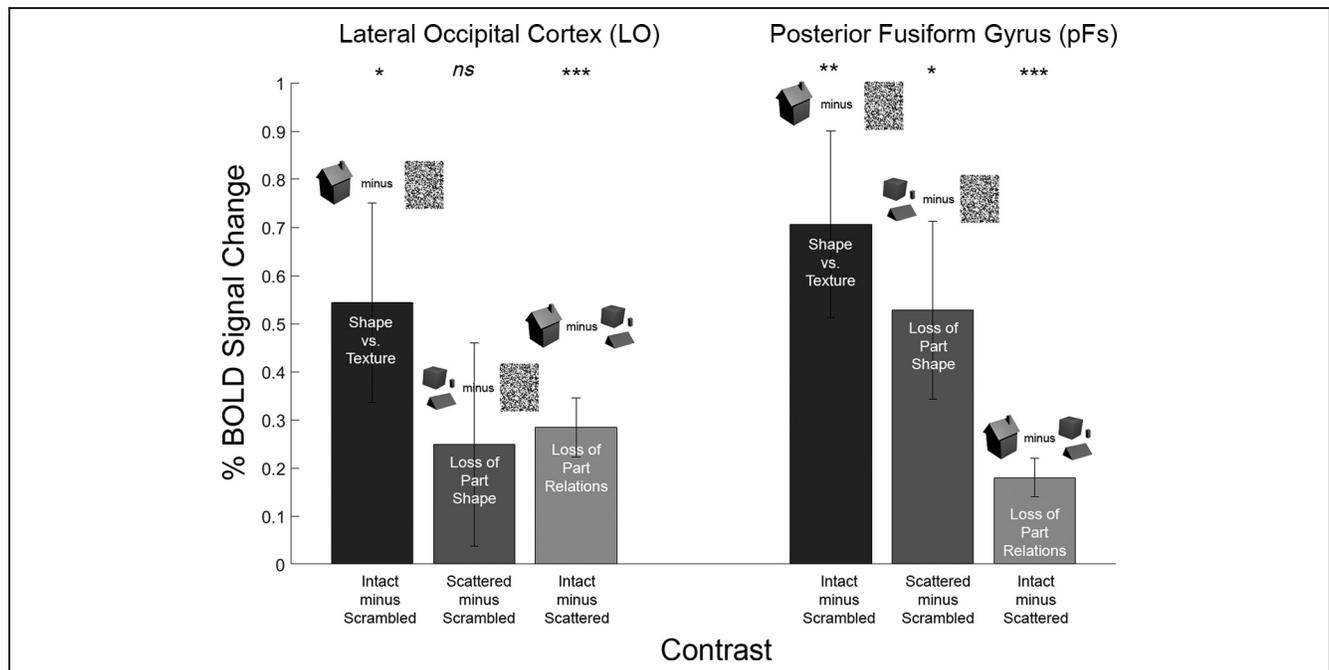
Statistically, the difference between the contrast of Scattered minus Scrambled and that of Intact minus Scattered was not significant in LO ( $t(11) = 0.15, p > .05$ ) or pFs ( $t(11) = 1.76, p > .05$ ). Moreover, such a comparison of the relative size of the effects with different baselines would be dependent on an assumption of linearity, which would be difficult to justify (Bao, Purington, &

Tjan, 2015). We thus cannot conclude that the loss of part shape has a larger impact than the loss of interpart relations in the reduction of the BOLD signal (Figure 4).

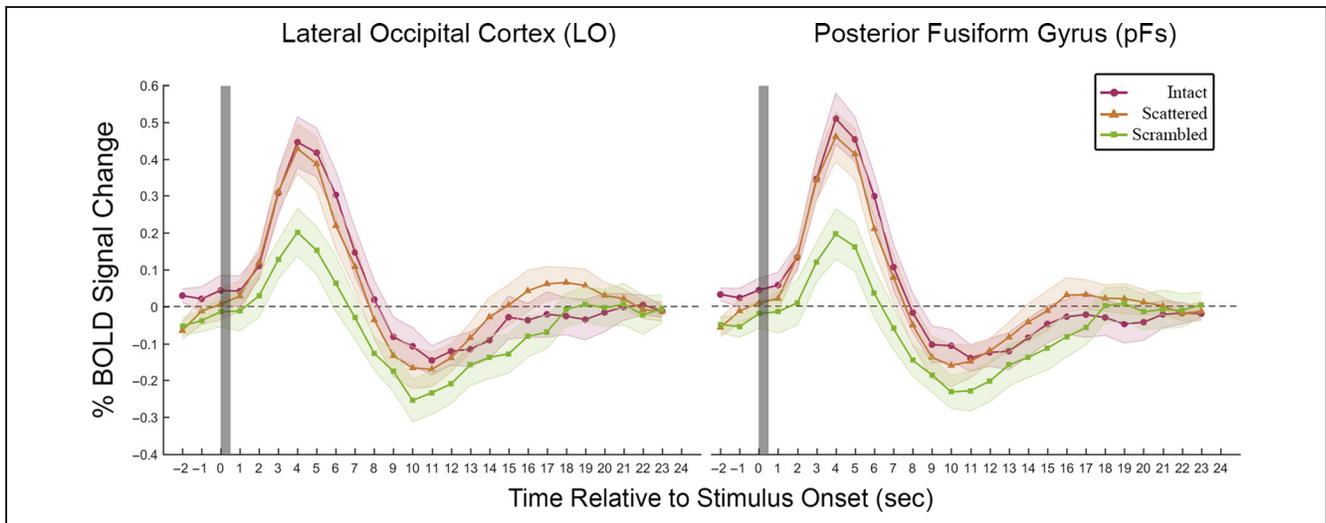
Although Margalit et al. (2016) did not find a net difference in the BOLD response to Novel and Familiar objects, in this study, the response to Novel objects was greater than that to Familiar objects both in LO ( $t(11) = 3.75, p < .01$ , Cohen's  $d = 1.08$ ) and in pFs ( $t(11) = 2.77, p < .05$ , Cohen's  $d = 0.8$ ). As suggested by subjective reports and the greater activation in parietal and early retinotopic areas to Novel than Familiar objects, Novel objects appear to be more likely to elicit attentional exploration and implicit motor activation. A two-way ANOVA revealed that this novelty advantage, however, did not interact with the calculation of increased activity to intact objects than to scrambled objects in LOC or the calculation of the effect of Loss of Part Relations (defined as Intact minus Scattered) in either LO ( $F(2, 71) = 1.77, p > .05$ ) or pFs ( $F(2, 71) = 1.00, p > .05$ ), indicating that these effects would have been present whether the Intact objects were confined solely to the Familiar objects or solely to the Novel objects.

## Peristimulus Time Course Evaluation

The relationship between conditions with respect to the magnitude of the percent BOLD response observed in the block design ROI analysis (Intact > Scattered > Scrambled) was generally replicated in the data from the event-related design runs (Figure 5). The peak BOLD



**Figure 4.** Percent BOLD signal change in LO (left) and pFs (right) for the Intact minus Scrambled contrast, the Scattered minus Scrambled contrast, and the Intact minus Scattered contrast. The Intact condition is for the combined responses to Familiar and Novel objects. Error bars depict the SEM. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ .



**Figure 5.** Peristimulus time course of the hemodynamic response in areas LO (left) and pFs (right). Peak BOLD increase values were higher for Intact objects than for Scattered objects, which was higher than for Scrambled objects. The shaded areas surrounding the data points depict the SEM. The gray vertical bar marks the stimulus duration (500 msec).

signal change was highest for the Intact minus Fixation contrast ( $0.45\% \pm 0.07\%$  increase in LO;  $0.51\% \pm 0.07\%$  in pFs), followed closely by the Scattered minus Fixation contrast ( $0.43\% \pm 0.07\%$  increase in LO;  $0.46\% \pm 0.07\%$  in pFs). The BOLD response to the Scrambled images relative to fixation was the lowest of the three ( $0.20\% \pm 0.06\%$  increase in LO;  $0.20\% \pm 0.07\%$  in pFs). A two-way repeated-measures ANOVA (Stimulus condition  $\times$  ROI) revealed a significant difference between the peak values of the Intact, Scattered, and Scrambled conditions ( $F(2, 71) = 22.4, p < .001$ ), where the peak is defined as the time interval between 4 and 6 sec after stimulus onset. Post hoc *t* tests (Tukey's honest significance difference) revealed that the difference between the peak values of all conditions were significantly different from one another (all pairwise *ps*  $< .05$ ).

### Behavioral Results

Participants were near ceiling in the accuracy of their detection of changes in border color (mean proportion of border changes identified =  $99.3\% \pm 0.02\%$ ). Responses were made well within the allotted window of 2000 msec (median RT =  $526 \pm 114$  msec), indicating that participants remained alert and attentive during the scan.

### DISCUSSION

LOC is localized by determining those regions where the BOLD response to intact images of objects is decreased by the scrambling of those images. The scrambling disrupts all aspects of shape including the shape of the parts and the interpart relations, so what had been an intact object becomes an array of texture.

To assess whether a portion of the reduction in the BOLD response could be independently attributed to the disruption of the relations among the parts, we created a condition where the parts of the original object were scattered, but the parts themselves remained intact. Prior studies of short-term visual memory have documented that an increase in the number of independent visual entities (at about the same complexity of the geons in this study) result in an increase in the BOLD response in LOC (Xu & Chun, 2006), but we observed a decrease in the BOLD response. Whereas there was no instruction for our participants to match the stimuli to a subsequent probe, although they certainly did attend to the stimuli, we cite this result to note that a reduction in the BOLD response in LOC from an increase in the number of identifiable, independent stimuli is not at all an obvious result.

In this study, the same stimuli were used to define the ROIs and to evaluate the magnitude of responses within those ROIs. Although the use of Intact  $>$  Scrambled to define LOC is standard and direct circularity was avoided by using independent runs to define and evaluate ROIs, we cannot exclude the possibility that the BOLD responses within the ROIs may have been biased toward the Intact stimuli. We note, however, that the number of voxels and peak activity in the Scattered  $>$  Scrambled contrast map are statistically indistinguishable from those in the Intact  $>$  Scrambled map. Thus, we conclude that our method of ROI definition is unlikely to explain the preference in LO and pFs for Intact over Scattered stimuli.

As reviewed in the Introduction, there is independent evidence for LOC representing objects in terms of parts (e.g., Hayworth & Biederman, 2006) and relations (Lescroart & Biederman, 2012; Hayworth et al., 2011). A major contribution of this study is in showing that the defining

operation of the localization of LOC—the reduction in the BOLD response from scrambling object images—has, at its core not merely the distinction between object shape and texture and the loss of part shape but the disruption of interpart relations.

The one apparent inconsistency in the overall pattern of results was that the contrast of (Scattered minus Scrambled), presumably reflecting the cost of the loss of part shape, fell short of significance in LO in the block design trials.

### **Possible Role of Low-level Effects in the Scattered Condition**

The separation of the geons in the Scattered condition necessarily involved several low-level image variations compared with the Intact condition. As the stimuli were constrained to be clearly separated but remain enclosed within a  $3^\circ \times 3^\circ$  bounding box in the center of the display, there was a slight to moderate reduction in the size of the geons to stay within the confines of the bounding box producing more white pixels (the color of the background). This effect is partially offset by the partial occlusion of the geons in the Intact condition. Overall, 21 of the 51 images had approximately 2% more whitespace in the Scattered condition than in the Intact condition, a nonsignificant difference,  $t(100) < 1$ , *ns*.

Several findings argue against the likelihood that these low-level feature differences accounted for the reduced BOLD response in the Scattered condition. First, there was no difference between Intact and Scattered conditions in early visual areas, for example, V1, V2, and V4. Second, the geons remained readily identifiable in both conditions and there is strong evidence that LOC activation reflects the stimulus as perceived rather than low-level feature differences (Kourtzi & Kanwisher, 2001). In the Kourtzi and Kanwisher study, novel 2-D shapes were displayed either behind partially occluding vertical bars or in front of the bars where none of the contours would be occluded. The adaptation effects in LOC were independent of the occlusion/depth conditions but were affected by relatively modest differences in the perceived shape of the stimuli. Indeed, much more drastic variations in image appearance—that of photographs versus line drawings—have no effect on adaptation in LOC (Grill-Spector et al., 2001). Thus, the BOLD response to viewing a photograph of a cup is maintained as strongly when viewing a line drawing of that object as the original photograph. Hayworth and Biederman (2006) showed that complementary versions of line drawings of familiar objects, in which every other extended contour and vertex was deleted from each of an object's geons—so that members of a complementary pair had no contours or vertices in common—primed each other as much as they primed themselves. Hayworth and Biederman also studied adaptation between complementary pairs of object images in which half of the geons were deleted from

each member of a complementary pair with its complementary mate composed of the remaining geons. These stimuli thus had different parts. They found that there was no priming (maintenance of adaptation) between such images, suggesting that the priming could not be attributable to a common basic or subordinate level concept or name.

There is also ample evidence that the invariance of shape tuning to modest differences in the size and appearance of objects is a general property in anterior, shape-selective areas in the ventral pathway, not limited to priming/adaptation studies. Thus, the selectivity of macaque IT cells is only minimally affected by whether an image is a rendered 3-D appearing object or a silhouette (Kayaert, Biederman, & Vogels, 2005) or, even more striking, whether the shape is conveyed by luminance differences or disparities of texture or motion (Sary, Vogels, & Orban, 1993). Intracranial single unit recordings in humans document the invariance of the activity of single IT neurons to variations in the size and orientation in depth of images of individual chairs (Liu, Agam, Madsen, & Kreiman, 2009).

We found a preference for Novel images in LO, which is consistent with our whole-brain voxel-wise finding that the Novel minus Familiar contrast yielded clusters of activation in or near early visual cortex, an effect that was also apparent in Margalit et al. (2016). As discussed previously, the greater lateral occipital activation for Novel over Familiar objects may be a result of the Novel objects inviting greater inspection, whereas the parietal activation apparent in the whole-brain results may reflect greater implicit (i.e., imagined) motor interactions elicited by the novel 3-D objects. Specifically, participants might have pondered the possible functionality of the novel objects and how they might be manipulated (yielding parietal activation).

The absence of any boost to the LOC BOLD response from familiar (vs. novel) objects suggests that LOC is a stage where the perceived physical shape is represented, with no appreciable contribution from object semantics. The lack of an effect of object semantics in LOC has also been seen in two fast event-related adaptation studies where the viewing of a sequence of two object images from different basic level categories did not result in a greater BOLD response (which would presumably reflect a release from adaptation) than when the objects were from the same basic level category (Kim, Biederman, Lescroart, & Hayworth, 2009; Chouinard, Morrissey, Kohler, & Goodale, 2008).

The coding of between-part relations within LOC, as evidenced by the greater BOLD response to the Intact objects compared with the separated geons in the Scattered condition, bears a parallel to LOC's coding of the relations between objects in simple scenes. Kim and Biederman (2011) and Kim et al. (2011) compared LOC activation to pairs of objects that either were depicted as interacting to form a single scene, for example, a watering can depicted as watering a plant, or noninteracting,

for example, mirror reversing the watering can so that it was now depicted as oriented away from the plant. In the latter case, the array would be interpreted as two separate, noninteracting objects, verbally describable with a conjunction “and” or a simple list. The interacting objects could be in a familiar relation, for example, a bird perched on a birdhouse, or novel, for example, the bird perched on an ear. LOC was more strongly activated by the interacting objects than by the noninteracting objects with an additional boost from the novelty of the interaction (perhaps akin to the attentional boost from the novel objects in this study). None of these effects were observed in the parahippocampal place area (PPA). In Kim et al.’s (2011) fast event-related design, the enhanced BOLD signal from the interaction appeared to arise in LO in that TMS applied to LO abolished the advantage of the between-object interactions but TMS applied to the intraparietal sulcus, which would presumably affect visual attention, had no effect. Further evidence that LOC codes between-object relations was reported by Hayworth et al. (2011), who showed that, whereas LOC was insensitive to the translation of a pair of objects, it was highly sensitive to a change in relative position. Thus, shifting an entire scene of a turtle above a bus right or up yielded no release from adaptation but depicting the bus above the turtle, with the same change in visual angle of the individual objects as when the whole array was translated, resulted in a sizeable release from adaptation. Thus, it appears that the coding of relations in LOC is not restricted to the coding of between-part relations in single objects but is also engaged in coding the between-object relations that define scenes. That these effects are witnessed with novel as well as familiar intact objects and scenes suggests that the relations are computed on the fly, rather than being retrieved from memory.

We are here distinguishing scenes, based on object interactions, from settings, for example, a forest, roadway, or beach, which can be specified, in part, by the statistical (spectral) properties of the contours of the array, decoded not in LOC but in the PPA or the retrosplenial cortex (Walther, Chai, Caddigan, Beck, & Li, 2011). Rather than specifying the relations between objects, PPA and retrosplenial cortex may be more involved in defining the physical layout of a scene, perhaps as a guide for navigation (Epstein, Parker, & Feiler, 2007). It is of some interest that D. F., who would be absolutely unable to interpret the scenes depicted as line drawings in the Kim and Biederman (2011) study, is able to, nonetheless, competently navigate through her environment without bumping into or tripping over objects (Goodale & Milner, 2013).

The specification of part shape and the specification of the relations between parts (defining objects) and between objects (defining scenes) thus endows LOC as the critical region where visual shape descriptions of objects and scenes are achieved that serve as the input to the subsequent associative processes that yield visual cognition.

## Acknowledgments

This study was supported by NIH/NEI grants R01-EY017707 to B. S. T. and NSF BCS 0617699 and BCS 0420794 to I. B.

Reprint requests should be sent to Irving Biederman, Department of Psychology/Neuroscience, University of Southern California, Room 316, 3641 Watt Way, Hedco Neurosciences Bldg., Los Angeles, CA 90089-2520, or via e-mail: [bieder@usc.edu](mailto:bieder@usc.edu).

## REFERENCES

- Bao, P., Purington, C. J., & Tjan, B. S. (2015). Using an achiasmic human visual system to quantify the relationship between the fMRI BOLD signal and neural response. *eLife*, *4*, e09600.
- Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (Chap. 8, pp. 213–263). Hillsdale, NJ: Lawrence Erlbaum.
- Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology*, *23*, 393–419.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Chouinard, P. A., Morrissey, B. F., Kohler, S., & Goodale, M. (2008). Repetition suppression in occipital-temporal visual areas is modulated by physical rather than semantic features of objects. *Neuroimage*, *41*, 130–144.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, *26*, 297–302.
- Epstein, R. E., Parker, W. E., & Feiler, A. M. (2007). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *Journal of Neuroscience*, *27*, 6141–6149.
- Goodale, M. A., Meenan, J. P., Bulthoff, H. H., Nicolle, D. A., Murphy, K. J., & Racicot, C. I. (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology*, *4*, 604–610.
- Goodale, M. A., & Milner, A. D. (2013). *Sight unseen: An exploration of conscious and unconscious vision* (2nd ed.). Oxford: Oxford University Press.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*, 1409–1422.
- Hayworth, K. J., & Biederman, I. (2006). Neural evidence for intermediate representations in object recognition. *Vision Research*, *46*, 4024–4031.
- Hayworth, K. J., Lescroart, M. D., & Biederman, I. (2011). Neural coding of relative position. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 1032–1050.
- James, T. W., Culham, J., Humphrey, K. G., Milner, D. A., & Goodale, M. A. (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: An fMRI study. *Brain*, *126*, 2463–2475.
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*, 825–841.
- Kayaert, G., Biederman, I., & Vogels, R. (2005). Representation of regular and irregular shapes in macaque inferotemporal cortex. *Cerebral Cortex*, *15*, 1308–1321.
- Kim, J., Biederman, I., Lescroart, M. D., & Hayworth, K. J. (2009). Adaptation to objects in the lateral occipital complex (LOC): Shape or semantics? *Vision Research*, *49*, 2297–2305.
- Kim, J. G., & Biederman, I. (2011). Where do objects become scenes? *Cerebral Cortex*, *21*, 1738–1746.

- Kim, J. G., & Biederman, I. (2012). Greater sensitivity to nonaccidental than metric changes in the relations between simple shapes in the lateral occipital cortex. *Neuroimage*, *63*, 1818–1826.
- Kim, J. G., Biederman, I., & Juan, C.-H. (2011). The benefit of object interactions arises in the lateral occipital cortex independent of attentional modulation from the intraparietal sulcus: A TMS study. *Journal of Neuroscience*, *31*, 8320–8324.
- Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital cortex. *Science*, *293*, 1506–1509.
- Lescroart, M. D., & Biederman, I. (2012). Cortical representation of medial axis structure. *Cerebral Cortex*, *23*, 629–637.
- Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neurons*, *62*, 281–290.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *92*, 8135–8139.
- Margalit, E., Shah, M. P., Tjan, B. S., Biederman, I., Keller, B., & Brenner, R. (2016). The lateral occipital complex shows no net response to object familiarity. *Journal of Vision*, *16*, 3.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Sáry, G., Vogels, R., & Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, *260*, 995–997.
- Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General*, *109*, 373–392.
- Van Essen, D. C. (2005). A Population-average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *Neuroimage*, *28*, 635–662.
- Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Li, F.-F. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, *108*, 9661–9666.
- Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage*, *14*, 1370–1386.
- Worsley, K. J. (2001). Statistical analysis of activation images. In P. Jezzard, P. M. Matthews, & S. M. Smith (Eds.), *Functional MRI: An introduction to methods* (Chap. 14, pp. 251–270). Oxford: Oxford University Press.
- Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*, 91–95.