Multisensory Part-based Representations of Objects in Human Lateral Occipital Cortex

Goker Erdogan, Quanjing Chen, Frank E. Garcea, Bradford Z. Mahon, and Robert A. Jacobs

Abstract

■ The format of high-level object representations in temporal-occipital cortex is a fundamental and as yet unresolved issue. Here we use fMRI to show that human lateral occipital cortex (LOC) encodes novel 3-D objects in a multisensory and part-based format. We show that visual and haptic exploration of objects leads to similar patterns of neural activity in human LOC and that the shared variance between visually and haptically induced patterns of BOLD contrast in LOC reflects the part structure of the objects. We also show that linear classifiers trained on neural data from LOC on a subset of the objects successfully predict a novel object based on its component part structure. These data demonstrate a multisensory code for object representations in LOC that specifies the part structure of objects. ■

INTRODUCTION

While eating breakfast, the object shape you perceive when viewing your coffee mug is the same as the shape you perceive when grasping your mug. This phenomenon illustrates modality invariance, an important type of perceptual constancy. Modality invariance suggests that people have representations of objects that are multisensory (i.e., with a significant degree of modality independence).

From behavioral studies, we know that participants trained in the visual modality to recognize novel objects show partial or near-complete transfer to the haptic modality, and vice versa (Lawson, 2009; Lacey, Peters, & Sathian, 2007; Norman, Norman, Clayton, Lianekhammy, & Zielke, 2004), and that object similarity is judged in similar ways across modalities (Gaissert & Wallraven, 2012; Gaissert, Bülthoff, & Wallraven, 2011; Gaissert, Wallraven, & Bülthoff, 2010; Cooke, Jäkel, Wallraven, & Bülthoff, 2007; Cooke, Kannengiesser, Wallraven, & Bülthoff, 2006). Those findings suggest that participants base their similarity judgments on a multisensory representation. Where is the neural substrate for these representations and how are the representations structured?

Prior brain imaging work suggests that human lateral occipital cortex (LOC) is one seat of multisensory representations of object shape, at least across the visual and haptic modalities. Previous research shows that LOC represents visual information about object shape (Grill-Spector, Kourtzi, & Kanwisher, 2001; Kourtzi & Kanwisher, 2001) and responds to haptic exploration of objects in sighted and congenitally blind individuals (Naumer et al., 2010; Amedi, Jacobson, Hendler, Malach, & Zohary, 2002; James et al., 2002; Amedi, Malach, Hendler, Peled, & Zohary, 2001). Furthermore, neural shape similarity matrices from blind participants are correlated with neural shape similarity matrices from sighted individuals (Peelen, He, Han, Caramazza, & Bi, 2014), suggesting that LOC is biased to represent object shape even if the principal modality of input is not vision.

To date, researchers have relied mainly on two measures—amount of neural "activation" (e.g., BOLD contrast) and correlations between neural similarity matrices—to argue for the multisensory nature of representations in LOC. Most studies compared the amount of BOLD contrast in LOC in response to visually and haptically presented stimuli. For example, James et al. (2002) showed that both visual and haptic exploration of objects led to neural activity in LOC. Similarly, Amedi and colleagues (2001, 2002) argued for multisensory shape representations in LOC on the basis of increased neural activity in response to objects compared with textures for visual and haptic stimuli. In a more recent study, Naumer and colleagues (2010) showed that the amount of neural activation when stimuli are presented through both visual and haptic modalities is higher than the amount of neural activation when stimuli are presented through a single modality.

Importantly, comparing the amount of activation in response to visual and haptic presentation of objects is an indirect test of multimodality of neural representations. It is quite possible that LOC carries distinct modality-specific representations for both visual and haptic object shape. A stricter test is possible by measuring the similarity in patterns of neural activity. Recently, Peelen and colleagues (2014) calculated neural similarity matrices for a set of

University of Rochester

Journal of Cognitive Neuroscience 28:6, pp. 869–881
doi:10.1162/jocn_a_00937

objects presented visually and verbally to blind and sighted individuals. By measuring the correlations between these neural similarity matrices, Peelen and colleagues (2014) argued that LOC carries a cross-modal shape representation. With respect to our current goals, there are two limitations associated with this study. First, Peelen and colleagues (2014) did not measure neural activity in response to haptic stimuli. Second, the correlation between two neural similarity matrices is a measure of second-order relations between two representations. It is possible for visual and haptic neural similarity matrices to be highly correlated even though the visual and haptic representations themselves are not. Here, we present a stricter test of the multisensory nature of object representations in LOC by correlating activations from different modalities directly to form cross-modal neural similarity matrices. Our analyses show that cross-modal correlation of an object with itself is larger than the cross-modal correlations among different objects and that objects can be decoded cross-modally from neural activations in LOC.

The second question we focus on is concerned with the structure of multisensory shape representations in LOC. Two competing theories emerge from previous research on object shape representations. First, view-based theories argue that the representation for an object is a collection of 2-D images of the object from different views (Peissig & Tarr, 2007). View dependency of object recognition is usually advanced as the main evidence for the view-based hypothesis. For example, a previous study (Bülthoff & Edelman, 1992) showed that the recognition performance for previously seen views of an object is better than the performance for views of the same object not previously seen. However, the view-based hypothesis is difficult to reconcile with the hypothesis that LOC encodes multisensory object representations, because the view-based hypothesis presumes a strictly visual nature of object representations.

Alternatives to the view-based hypothesis are part-based or structural description theories (e.g., Peissig & Tarr, 2007; Riddoch & Humphreys, 1987). These theories assume that objects are represented as collections of parts and the spatial relations among these parts. There is behavioral and neural evidence for both aspects of the part-based theory: representation of parts and spatial relations among those parts. An influential study by Biederman (1987) showed that priming is principally mediated by parts, and recognition suffers dramatically when part-related information is removed. Later studies also investigated whether spatial relations are explicitly represented. For example, Hayworth, Lescroart, and Biederman (2011) found that it was impossible for participants to ignore relations between objects in a scene even when that information was irrelevant. Importantly for our current study, previous work has found evidence that LOC encodes object parts and spatial relations explicitly. Using fMRI adaptation, Hayworth and Biederman (2006) found that, when part-related information was removed from an image, there was a release from adaptation in LOC, suggesting that different parts involve different LOC representations. A separate study (Hayworth et al., 2011) showed that a comparable amount of release from adaptation in LOC is observed when the spatial relation between two objects is changed as when one of the objects is replaced with a new object. This suggests that spatial relations are encoded explicitly by this region. More recently, Guggenmos and colleagues (2015) tested whether LOC encodes objects in a part-based or holistic manner by measuring decoding accuracy for split and intact objects. They showed that a classifier trained on neural activations for intact objects can successfully discriminate between activations for split objects (e.g., a camera with its lens and body separate) and vice versa. These studies suggest that LOC represents objects in a part-based format. Here, we provide further evidence for this hypothesis by showing that a novel object can be decoded from the neural activations in LOC based on part-based representations.

METHODS

Participants

Twelve (six in Experiment 1 and six in Experiment 2) University of Rochester students (mean age = 21.5 years, *SD* = 1.57 years, five men) participated in the study in exchange for payment. All participants were right-handed (assessed with the Edinburgh Handedness Questionnaire), had normal or corrected-to normal vision, and had no history of neurological disorders. All participants gave written informed consent in accordance with the University of Rochester research subjects review board.

General Procedure

Stimulus presentation was controlled with "A Simple Framework" (Schwarzbach, 2011) written in MATLAB Psychtoolbox (Brainard, 1997; Pelli, 1997) or E-Prime Professional Software 2.0 (Psychology Software Tools, Inc., Sharpsburg, PA). For all fMRI experiments with visual presentation of stimuli, participants viewed stimuli binocularly through a mirror attached to the head coil adjusted to allow foveal viewing of a back-projected monitor (temporal resolution = 120 Hz). Each participant completed four 1-hr sessions: one session for retinotopic mapping and somatosensory and motor cortex mapping (data not analyzed herein), one session for an object-responsive cortex localizer, and two sessions for the experiment proper (visual and haptic exploration of objects).

Object-responsive Cortex Localizer (LOC Localizer)

The session began with (i) one 6-min run of resting state fMRI, (ii) eight 3-min runs of the object-responsive cortex localizer experiment, and (iii) one 6-min run of resting

state fMRI. The resting state fMRI data are not analyzed herein.

To localize object-responsive areas in the brain, participants viewed scrambled and intact images of tools, animals, famous faces, and famous places (see Chen, Garcea, & Mahon, in press, for all details on stimuli and design; also see Fintzi & Mahon, 2013). For each of four categories (tools, animals, faces, and places) 12 items were selected (e.g., hammer, Bill Clinton, etc.), and for each item, eight exemplars (gray-scale photographs) were selected (e.g., eight different hammers, eight different pictures of Bill Clinton, etc.). This resulted in a total of 96 images per category and 384 total images. Phase-scrambled versions of the stimuli were created to serve as a baseline condition. Participants viewed the images in a miniblock design. Within each 6-sec miniblock, 12 stimuli from the same category were presented, each for 500 msec (0 msec ISI), and 6-sec fixation periods were presented between miniblocks. Within each run, eight miniblocks of intact images and four miniblocks of phase-scrambled versions of the stimuli were presented with the constraint that a category of objects did not repeat during two successive miniblock presentations. All participants completed eight runs of the object-responsive cortex localizer experiment (91 volumes per run).

Experimental Materials

The stimuli used in Experiment 1 were taken from the set of objects known as Fribbles (Tarr, 2003). We picked 12 Fribbles (four objects from three categories) for Experiment 1. For the stimuli used in Experiment 2, we created a new set of objects by taking parts from Fribbles and combining them in the following way. Each object is made up of five components where the body (one component) is common to all objects. The remaining four components are located at four fixed locations on the body. For each location, there are two possible parts or values that the component can take (i.e., 2^4, hence 16 objects). Figures 1 and 2A show the entire set of objects used in Experiments 1 and 2, respectively. Figure 7 shows how we constructed the set of objects for Experiment 2 from the parts and how these were combined to create an example object. For the haptic stimuli, we used 3-D-printed plastic models of the objects. The physical objects were approximately 11.5 cm long, 6.0 cm wide, and 7.5 cm high.

To summarize, the stimuli used in Experiment 1 were drawn from three "categories" of objects (four items per category) but the part structure was not explicitly (i.e., factorially) manipulated across the stimulus set. In contrast, in Experiment 2, the materials were created by creating all possible combinations of part values (two values) at each of four possible locations, leading to a factorial stimulus space defined by part structure.

Visual and Haptic Exploration of Novel Objects (Two Sessions)

Each participant completed two 1-hr sessions of the experiment proper. Each session was composed of four runs, two runs dedicated to visual exploration of objects and two runs dedicated to haptic exploration of objects. In the first experiment, the participants

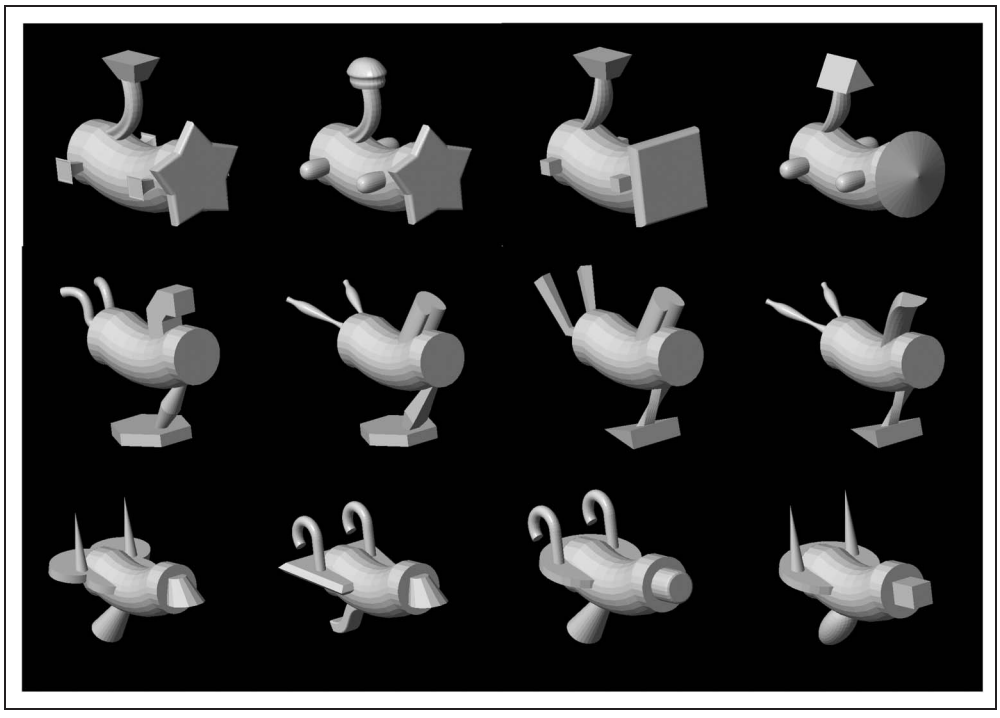

Figure 1. Experimental stimuli used in Experiment 1. The stimuli are taken from the set of novel objects known as Fribbles (Tarr, 2003).

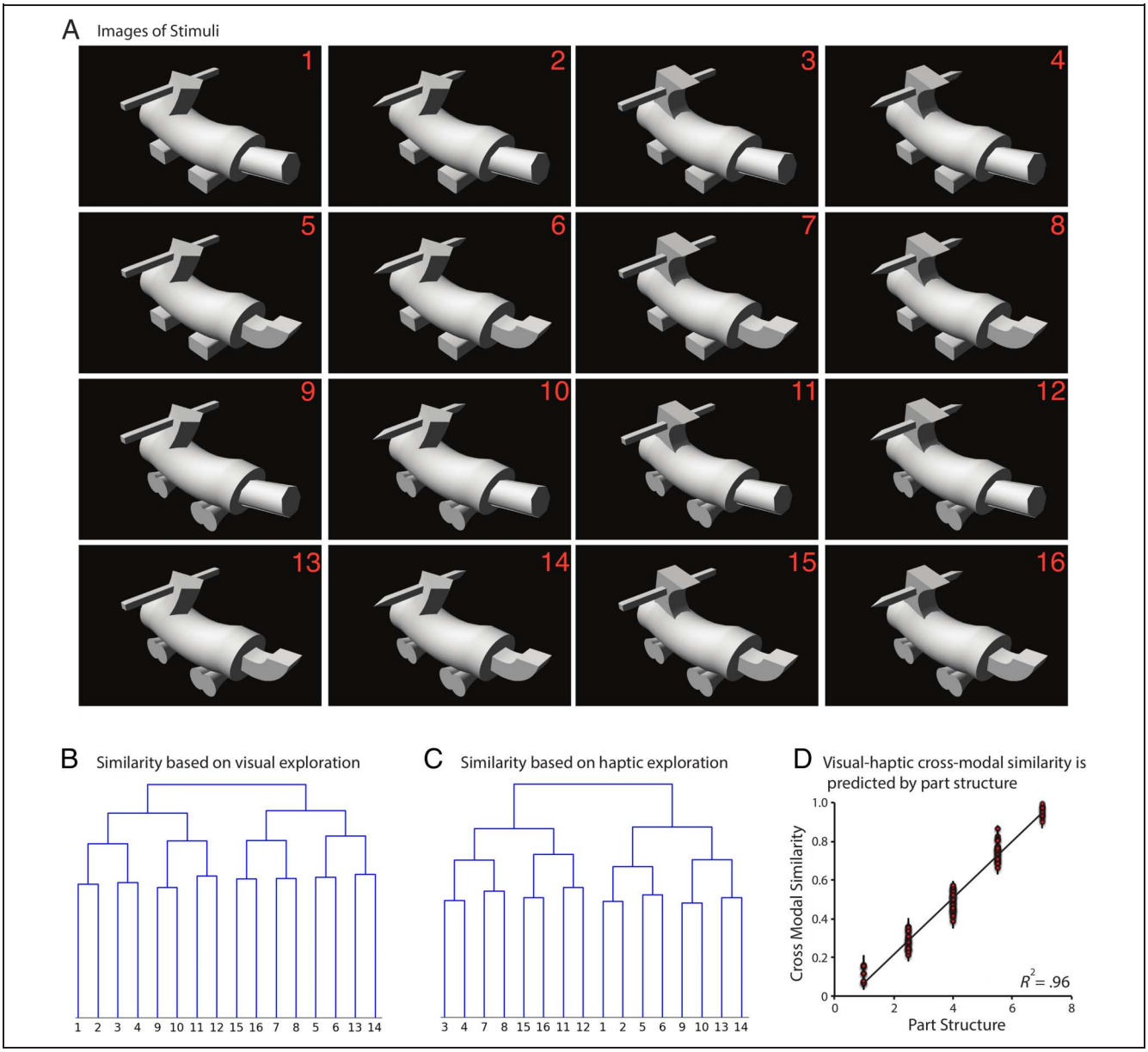

Figure 2. (A) Experimental stimuli used in Experiment 2. The stimuli are based on Fribbles (Tarr, 2003). Each object is made up of four components at four fixed locations. For each location, there are two possible values or parts (i.e., 2^4; hence 16 objects). (B) Results of agglomerative clustering applied to behavioral similarity data from the visual condition. In the behavioral experiment (Erdogan, Yildirim, & Jacobs 2015), participants either viewed or haptically explored a pair of objects and provided similarity ratings on a scale of 1–7. Similarity judgments are averaged across participants to get a condition level similarity matrix. (C) Results of agglomerative clustering applied to haptic behavioral similarity data. (D) Scatter plot of cross-modal behavioral similarity judgments versus similarities calculated from part structure. In the cross-modal condition, participants viewed one of the objects and haptically explored the other object. Similarities based on part structure are calculated by counting the number of shared parts between pairs of objects.

observed each novel object stimulus in the visual and haptic conditions; that is, all 12 objects were presented in each run. In the second experiment, the novel object stimuli were divided (arbitrarily) into two sets, A and B. Within a given scanning session, a participant was presented (for instance) Set A for haptic exploration and Set B for visual exploration; that is, in each run, participants saw eight objects. In their second session for the experiment proper, that same participant was presented Set B for haptic exploration and Set A for visual exploration. This ensured that participants only viewed or only haptically explored a given object in a given scanning session. The order of a given item set (Set A first, Set B first) by modality (visual, haptic) was also counterbalanced across participants. For both Experiments 1 and 2, visual and haptic exploration was blocked by run, organized in an ABBA/BAAB fashion, and counterbalanced evenly across participants.

While laying supine in the scanner, participants were visually presented with the objects or were required to

keep their eyes closed while haptically exploring the objects. In the haptic condition, the objects were handed to the participant by the experimenter. For runs in which items were visually presented, participants were instructed to deploy their attention to the features of the object.

In the visual condition in Experiment 1, the objects were presented in the center of the screen for the participants to fixate upon. Miniblocks were 4 sec long and were interspersed by 8-sec fixation periods. Each object was presented in four miniblocks per run, with the constraint that the same object did not repeat on two successive miniblocks. This meant that there were a total of 48 (12 × 4) object presentations in each run. In Experiment 2, the objects were presented centrally and rotated 40 degrees per second along the vertical axis (i.e., the objects revolved in the depth plane). Miniblocks in the visual condition were 9 sec long and were interspersed by 9-sec fixation periods. Each object was presented in four miniblocks per run, in a similar manner to Experiment 1. Therefore, there were in total 32 (8 × 4) object presentations in each run. In the haptic condition, participants were instructed to form a mental image of the plastic object while haptically exploring the object with their hands. In Experiment 1, miniblocks were 12 sec long and were interspersed by 9-sec periods in which their hands were unoccupied. Each plastic object was presented in four miniblocks per run, with the constraint that the same item did not repeat across two successive miniblock presentations. Miniblocks in Experiment 2 were 16 sec long and were interspersed by 16-sec periods in which their hands were unoccupied. Each plastic object was presented in four miniblocks per run, in a similar manner to Experiment 1.

In our experiments, participants performed no explicit task other than visually or haptically exploring the presented objects. We believe such a design enables us to investigate visual-haptic processing without any potential task-related effects. Previous research shows that, even in the absence of any explicit task, visual and haptic processing converges in LOC (Naumer et al., 2010). Although our participants did not perform an explicit task, we asked them to mentally picture the object they were exploring in the haptic condition. This might raise suspicions about whether the activation in LOC was due to mental imagery rather than haptic processing. However, previous research suggests that LOC is minimally activated by mental imagery (James et al., 2002; Amedi et al., 2001).

Before the experiment began, participants were introduced to comparable plastic objects outside the scanner. These objects were not used in the experiment proper and were dissimilar to the experimental stimuli. Visual analogs of the objects were also presented to the participants to inform them of the format of the visual experiment and to practice the implicit task that they were required to carry out while in the scanner.

MR Acquisition and Analysis

MRI Parameters

Whole-brain BOLD imaging was conducted on a 3-T Siemens (Amsterdam, The Nertherlands) MAGNETOM Trio scanner with a 32-channel head coil located at the Rochester Center for Brain Imaging. High-resolution structural T1 contrast images were acquired using a magnetization prepared rapid gradient-echo pulse sequence at the start of each participant's first scanning session (repetition time = 2530, echo time = 3.44 msec, flip angle = 7°, field of view = 256 mm, matrix = 256 × 256, 1 × 1 × 1 mm sagittal left-to-right slices). An EPI pulse sequence was used for T2* contrast (repetition time = 2000 msec, echo time = 30 msec, flip angle = 90°, field of view = 256 × 256 mm, matrix = 64 × 64, 30 sagittal left-to-right slices, voxel size = 4 × 4 × 4 mm). The first six volumes of each run were discarded to allow for signal equilibration (four at acquisition and two at analysis).

fMRI Data Analysis

fMRI data were analyzed with the BrainVoyager software package (Version 2.8) and in-house scripts drawing on the BVQX toolbox written in MATLAB (wiki2.brainvoyager.com/bvqxtools). Preprocessing of the functional data included, in the following order, slice scan time correction (sinc interpolation), motion correction with respect to the first volume of the first functional run, and linear trend removal in the temporal domain (cutoff: two cycles within the run). Functional data were registered (after contrast inversion of the first volume) to high-resolution deskulled anatomy on a participant-by-participant basis in native space. For each participant, echo-planar and anatomical volumes were transformed into standardized space (Talairach & Tournoux, 1988). Functional data for the localizer experiment (object-responsive cortex localizer) were smoothed at 6 mm FWHM (1.5 mm voxels) and interpolated to 3 mm^3 voxels; functional data for the experiment proper (visual and haptic exploration of objects) were interpolated to 3 mm^3 but were not spatially smoothed.

For all experiments, the general linear model was used to fit beta estimates to the experimental events of interest. Experimental events were convolved with a standard 2-gamma hemodynamic response function. The first derivatives of 3-D motion correction from each run were added to all models as regressors of no interest to attract variance attributable to head movement. Thus, all multivoxel pattern analyses were performed over beta estimates.

In all multivoxel analyses, we normalized individual voxel activations within a run to remove baseline differences across runs. In other words, for each voxel, we subtracted the mean activation for that voxel over all objects in the run and divided it by the standard deviation of that voxel's activation across objects. Additionally, for linear correlation multivoxel analyses, activations for all eight repeats of a single item (in a given modality, i.e., visual/haptic) were

averaged to obtain a single activation vector for each item. In our correlation analyses, we transformed correlation values using Fisher's z transformation and ran all statistical tests on those transformed values. When calculating correlations between correlation matrices, we used only the upper triangles of matrices. All statistical tests were two-tailed. For training the support vector machine (SVM) for decoding, we used the library libsvm (www.csie.ntu.edu.tw/~cjlin/libsvm/). We used linear kernels with cost parameter set to 1.

Whole-brain pattern analyses were performed using a searchlight approach (Kriegeskorte, Goebel, & Bandettini, 2006). Whole-brain searchlight maps were computed with a mask fit to the deskulled Talairach anatomy of individual participants. The "searchlight" passes over each voxel (in each participant) and extracts the beta estimates (for 16 items) for the cube of voxels ($n = 125$) that surround the voxel. The analysis was carried out based on the pattern of responses across the 125 voxels, and the results were assigned to the center voxel of that cube. All whole-brain analyses were thresholded at $p < .005$ (corrected), cluster threshold for nine contiguous voxels. If no regions were observed at that threshold, a more lenient threshold was used ($p < .05$, uncorrected, nine voxels).

Definition of ROIs (LOC)

Left and right LOC were identified at the group level using the object-responsive localizer experiment with the contrast of [intact images] > [scrambled images]. The result used cluster size corrected alpha levels by thresholding individual voxels at $p < .05$ (uncorrected) and applying a subsequent cluster size threshold generated with a Monte Carlo style permutation test (1000 iterations) on cluster size to determine the appropriate alpha level that maintains Type I error at 1% (using AlphaSim as implemented in Brain Voyager). The Talairach coordinates were as follows: left LOC: $x = -40, y = -71, z = -9$; right LOC: $x = 38, y = -65, z = -12$. We note as well that none of the results in this study change qualitatively if LOC is defined individually for each participant, rather than at the group level.

RESULTS

Our study consisted of two experiments. In both experiments, participants either viewed or haptically explored a set of objects during fMRI. The stimuli for Experiment 1 consisted of 12 objects (four objects from three categories; see Figure 1) picked from the set of objects known as Fribbles (Tarr, 2003). For Experiment 2, we created a novel set of objects based on Fribbles. Each object in this set was composed of one component that was common to all objects and four components that varied across objects. The variable components were located at four fixed locations (Figure 2A), and there were two possible parts (or values) that each component could take (i.e., $2^4 = 16$ objects in total).

Cross-modal Decoding of Novel Objects in LOC

If object representations in LOC are multisensory across haptic and visual modalities, it should be possible to decode object identity using cross-modal representational similarity analyses. To that end, we correlated the voxel patterns in LOC elicited when a participant was viewing objects with the voxel patterns elicited when the same participant haptically explored the objects. The resulting representational similarity analysis quantifies the similarity of voxel patterns across modalities, comparing every object to every other object as well as to itself. Previous studies have calculated neural similarity matrices separately for each modality and then subsequently correlated

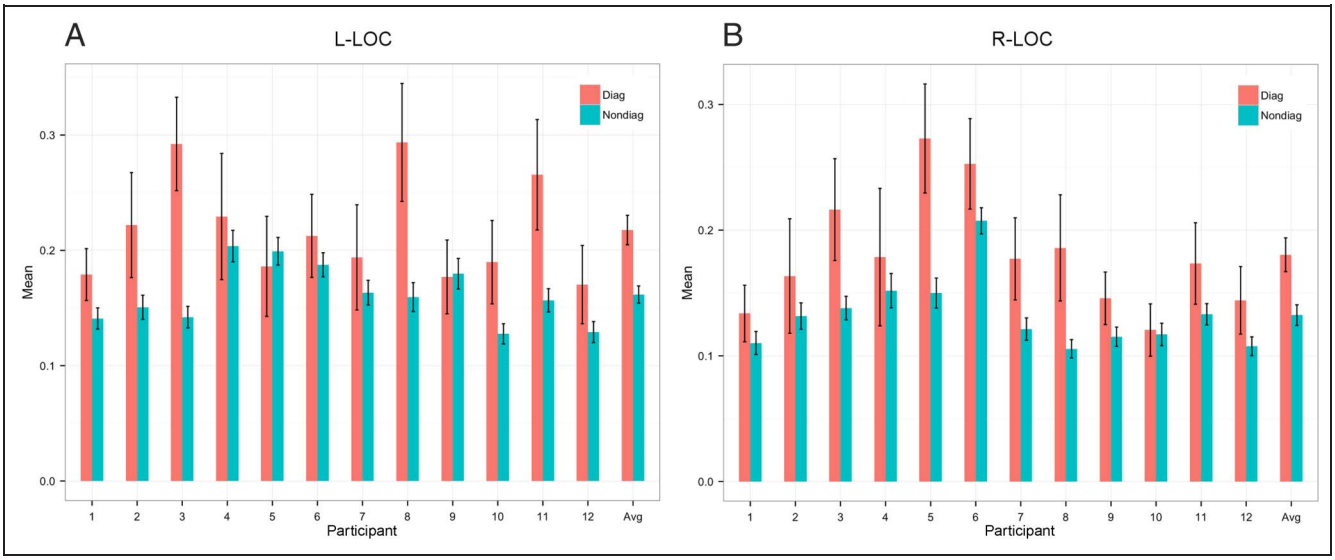

Figure 3. Comparison between diagonals and nondiagonals of cross-modal similarity matrices for both experiments. Participants 1–6 are in Experiment 1, and participants 7–12 are in Experiment 2. Avg = average of all 12 participants. (A) Results for left LOC. (B) Results for right LOC.

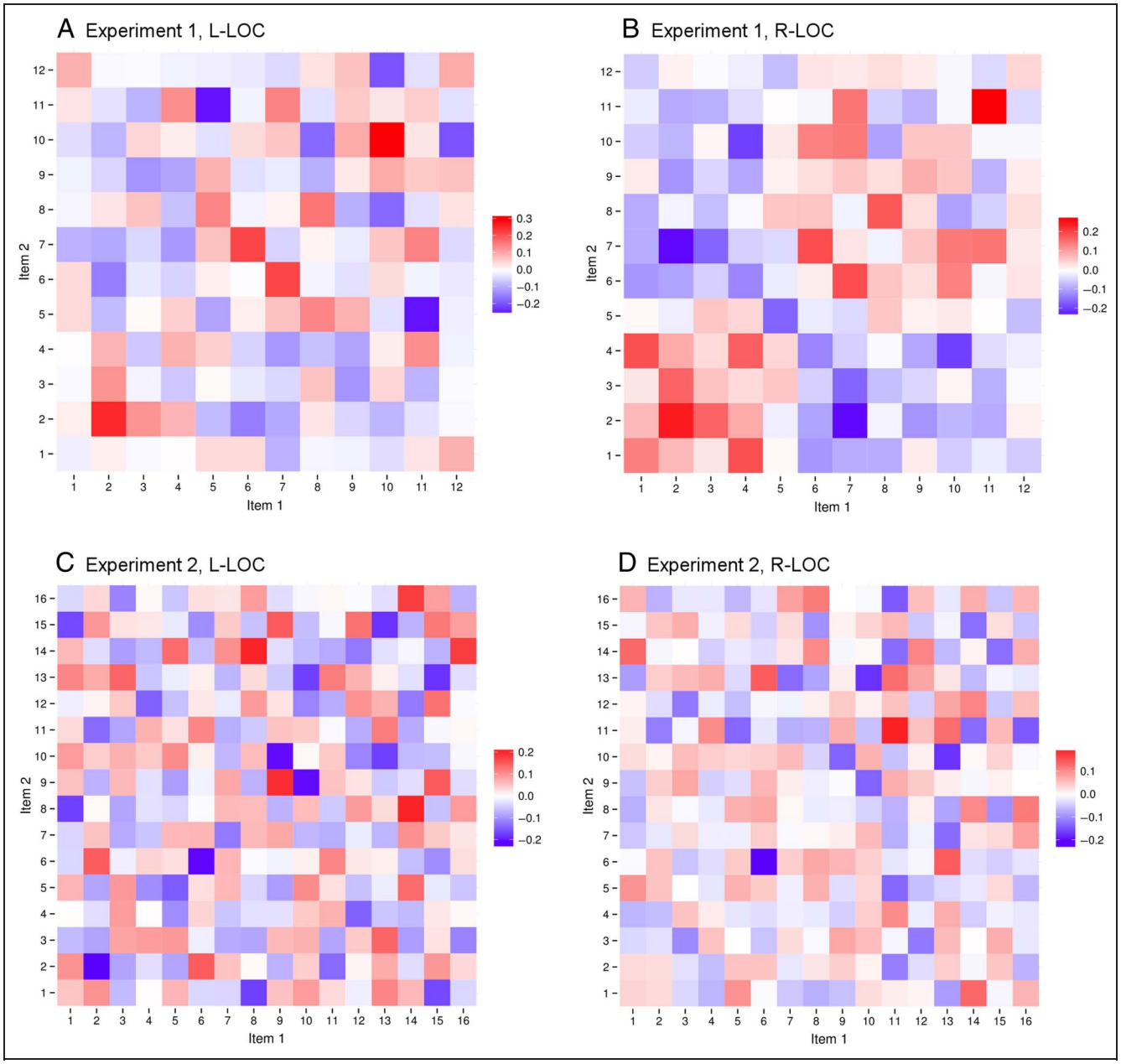

Figure 4. Cross-modal similarity matrices for both experiments. (A, B) Cross-modal similarity matrices calculated from left (A) and right (B) LOC activations from Experiment 1. (C, D) Cross-modal similarity matrices calculated from left (C) and right (D) LOC activations from Experiment 2.

those matrices (e.g., Peelen et al., 2014). Such an approach amounts to showing that neural correlations among objects in one modality correlate with the neural correlations among objects in another modality. The goal of the current analysis is to run a stricter test of the hypothesis that LOC encodes objects in a multisensory manner by correlating voxel patterns from different modalities directly to form a cross-modal neural similarity matrix.

Two predictions are made by the hypothesis that object representations in LOC are multisensory. First, cross-modal correlations between the visual and haptic voxel patterns for the same object will be higher than cross-modal correlations among the voxel patterns for different objects (i.e., the diagonal values will be greater than the nondiagonal values in the cross-modal representational similarity matrix). The results of this analysis for each participant in Experiments 1 and 2 can be seen in Figure 3. For every participant in right LOC and for 10 of 12 participants in left LOC, cross-modal correlations were in fact higher for identical objects than they were for different objects (see Figure 4 for average cross-modal correlation matrices). Because an initial ANOVA analysis found no effect of Experiment (L-LOC, $F = 0.17$, $p = .69$; R-LOC, $F = 0.48$, $p = .50$), we combined the results from both experiments. Diagonal versus nondiagonal differences reached statistical significance in both L-LOC and R-LOC (L-LOC, difference = 0.06; $t = 3.86$, $p < .004$; R-LOC, difference = 0.05; $t = 5.08$, $p < .001$),

indicating that LOC contains multisensory representations of objects. A second and stricter prediction is that it should be possible to decode object identity using the representational similarity matrix by testing whether each object is more correlated with itself (across modalities) than it is with each of the other objects in the set (also across modalities). We calculated the decoding accuracies for each participant and compared these to the chance decoding accuracy (1/12 for Experiment 1 and 1/16 for Experiment 2). Again, because an initial ANOVA analysis found no effect of Experiment (L-LOC, $F = 0.67$, $p = .43$; R-LOC, $F = 0.82$, $p = .39$), we combined the results from both experiments. Our results showed that it is possible to decode object identity cross-modally in both L-LOC and R-LOC (L-LOC, difference from chance accuracy = 0.09, $t = 2.48$, $p < .04$; R-LOC, difference = 0.10, $t = 3.48$, $p < .006$). These data indicate that LOC contains multisensory representations of objects.

We then tested whether multisensory coding of novel objects was specific to LOC or was a property observed throughout the brain. To that end, a whole brain searchlight analysis was conducted in which each voxel was coded according to whether it (and its immediate neighbors) showed higher pattern similarity for an object correlated with itself (across modalities) than with other objects (also cross modality). Converging with the ROI analyses, the

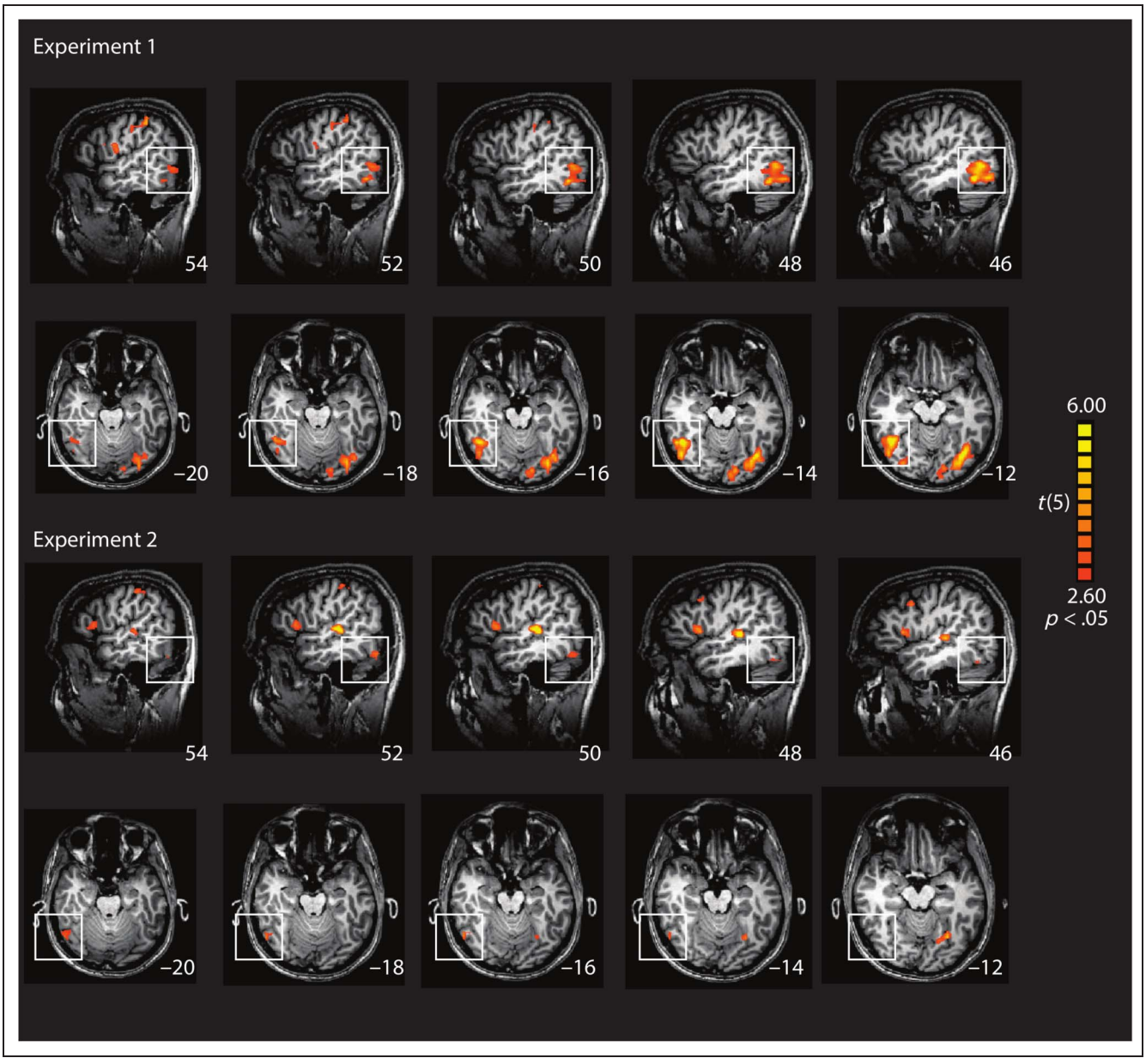

Figure 5. Whole searchlight analysis of brain regions in which the diagonal of the cross-modal neural similarity matrix is greater than the off-diagonal values. The cross-modal similarity matrix was created by correlating the voxel patterns elicited when visually exploring objects with the voxel patterns elicited when haptically exploring objects. If the diagonal of the matrix is greater than the off-diagonal values, that means that the pattern of voxel activations elicited by an object (across modalities) is more similar than the patterns elicited by two different objects.

Table 1. Talairach Coordinates, Cluster Sizes, Significance Levels, and Anatomical Regions for the Searchlight Results

Region	Talairach Coordinates			Cluster Size (mm^2)	t	p
	x	y	z			
Exp1: Diagonal of the Cross-modal Neural Similarity Matrix > Off-diagonal Values ($p < .05$, Cluster > 9 Voxels)						
Precentral gyrus LH	−51	−13	34	6790	8.55	<.001
Middle occipital gyrus LH	−24	−88	19	25007	12.28	<.001
Lateral occipital cortex LH	−39	−67	−14		8.40	<.001
Precentral gyrus RH	57	−1	19	1469	7.06	<.001
Postcentral gyrus RH	63	−25	38	3175	7.51	<.001
Lateral occipital cortex RH	39	−55	−5	23568	13.83	<.001
Exp2: Diagonal of the Cross-modal Neural Similarity Matrix > Off-diagonal Values ($p < .05$, Cluster > 9 Voxels)						
Inferior frontal gyrus LH	−39	20	10	1100	5.47	<.01
Precentral gyrus LH	−30	−16	52	1514	7.30	<.001
Superior parietal lobule LH	−21	−58	58	4417	12.27	<.001
Inferior frontal gyrus RH	39	17	16	1450	6.19	<.002
Superior temporal gyrus RH	51	−25	7	877	10.04	<.001
Lateral occipital cortex RH	50	−62	−18	257	3.88	<.01
Exp2: Correlation between Neural and Behavioral Similarity for Visual Exploration of Objects ($p < .05$, Cluster > 9 Voxels)						
Parietal lobe LH	−18	−58	46	539	8.67	<.001
Lateral occipital cortex RH	42	−70	1	742	9.84	<.001
Lingual gyrus RH	0	−73	−11	2607	6.07	<.002
Exp2: Correlation between Neural and Behavioral Similarity for Haptic Exploration of Objects ($p < .005$, Cluster > 9 Voxels)						
Lateral occipital cortex LH	−39	−67	−14	1230	9.99	<.001
Precentral gyrus RH	42	−13	34	1577	10.78	<.001
Postcentral gyrus RH	51	20	34	2323	19.19	<.001
Parietal lobe RH	9	−37	61	3143	12.86	<.001
Superior temporal gyrus RH	42	−49	19	2110	13.79	<.001
Lateral occipital cortex RH	33	−73	−8	2190	11.84	<.001

LH = left hemisphere; RH = right hemisphere.

results (Figure 5) identified the right LOC in both experiments (see Table 1 for coordinates). The left posterior temporal-occipital cortex was also identified in the searchlight analyses from both experiments.

A Common Similarity Space of Novel Objects as Derived from Neural and Behavioral Metrics

The stimuli used in Experiment 2 were designed to have a clear part-based structure for the purpose of testing the part-based hypothesis through representational similarity and neural decoding analyses. In a prior study (Erdogan, Yildirim, & Jacobs, 2015), we collected behavioral similarity judgments for these stimuli while participants viewed or haptically explored the objects. Similarity ratings consisted of Likert similarity ratings (range 1:7) for each pair of objects. We evaluated how well participants' judgments of the similarity among the objects were explained by the part-based structure of the objects. As shown in Figure 2D, the agreement was extremely good ($R^2 = .96$). This indicates that participants perceive the similarity among these object stimuli in terms of their part structure. Therefore, a significant agreement

between the neural similarity matrices and behavioral similarity judgments will lend support to both the hypothesis that LOC representations are multisensory and to the hypothesis that they are part based. We tested this prediction by calculating correlations between behavioral similarity judgments and measures of object similarity derived from neural data. A visual similarity matrix was formed by correlating voxel patterns when participants viewed the objects during fMRI, and a haptic similarity matrix was formed when participants haptically explored the objects during fMRI. As predicted by the hypothesis that LOC encodes multisensory, part-based representations of objects, the neural similarity matrices obtained from R-LOC for both modalities were correlated with the behavioral similarity matrices (neural similarity measures based on visual exploration: L-LOC: $r = .02$, $t = 1.10$, $p = .33$, R-LOC: $r = .08$, $t = 4.21$, $p < .009$; Haptic condition, L-LOC: $r = .08$, $t = 1.52$, $p = .187$, R-LOC: $r = .14$, $t = 3.28$, $p < .03$).

To evaluate the degree to which the observed relationship between behavioral and neural similarity measures was specific to LOC, we again carried out a whole-brain searchlight analysis that maps how similar the neural similarity matrices were to the behavioral similarity matrices. The most stringent test of whether LOC encodes multisensory representations of novel objects is to test whether LOC is identified by two independent searchlight analyses: The first analysis relates neural and behavioral similarity data for visual exploration of objects, and the second analysis relates neural and behavioral similarity data for haptic exploration of objects. Thus, the key test is whether these two independent searchlight analyses overlap in LOC. The results indicate overlap in right LOC (see Table 1 for Talairach coordinates). As can be seen in Figure 6, there is good overlap (35 voxels, 958 mm^3, across the maps in Figure 6A, B, and C) between the independent functional definition of right LOC (objects > scrambled images) and right LOC as identified by the two independent multivoxel pattern searchlight analyses. Interestingly, the whole-brain searchlight analysis over haptic data also identified several other regions in the temporal and frontal lobes involved in sensory processing (see Table 1 for coordinates).

Object Category Representations in LOC

Stimuli in Experiment 1 formed three families or categories of objects (Figure 1). This raises the possibility of evaluating whether LOC object representations encode category structure. Using analyses of the LOC cross-modal similarity matrix, we found that neural activations were more similar when considering two objects belonging to the same category than when considering two objects belonging to different categories. Using decoding analyses, we found that we can decode the category to which an object belongs at above-chance levels. However, because we are uncertain about the proper interpretation of these results, we do not study LOC object category representations here. One possibility is that LOC encodes the category structure of objects. Another possibility is that LOC encodes object shape and that the results regarding category structure are due to the fact that objects belonging to the same category have similar shapes in our experiment and objects belonging to different categories have dissimilar shapes. Because we cannot distinguish these two possibilities

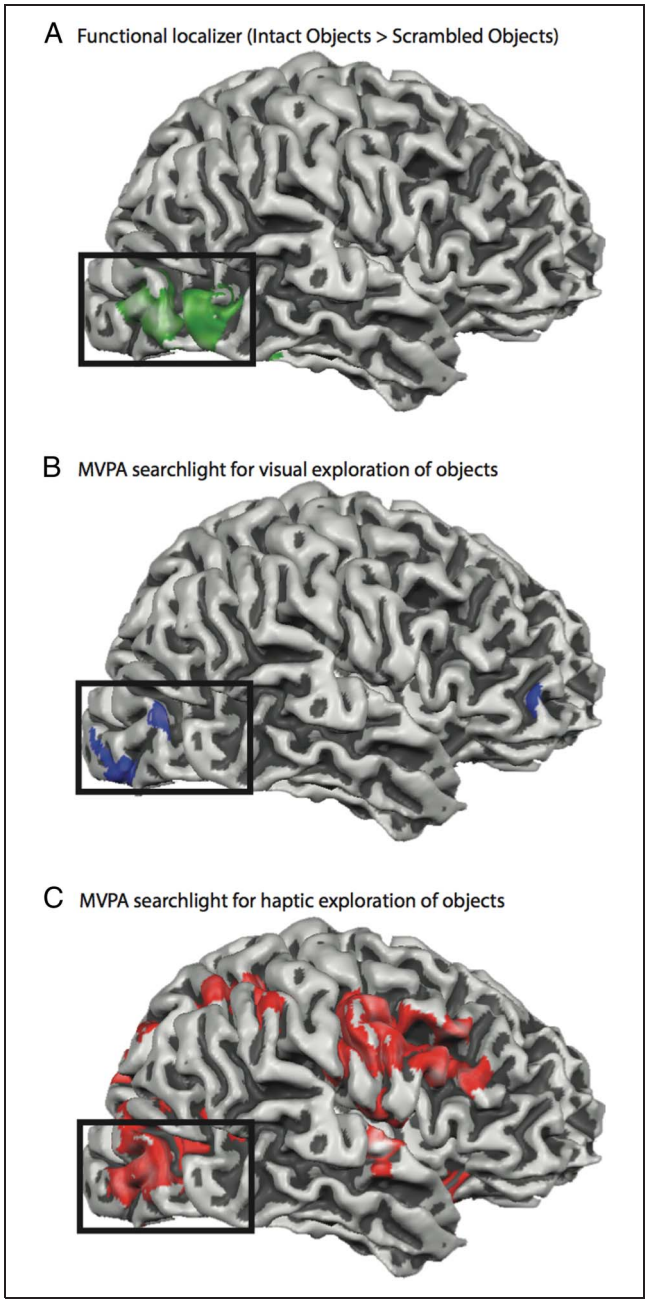

Figure 6. Overlap in right LOC for the (A) functional localizer (i.e., objects > scrambled objects), (B) a whole brain searchlight analysis of the correlation between neural similarity matrices and behavioral similarity for visual exploration of objects, and (C) a whole brain searchlight analysis of the correlation between neural similarity matrices and behavioral similarity for haptic exploration of objects.

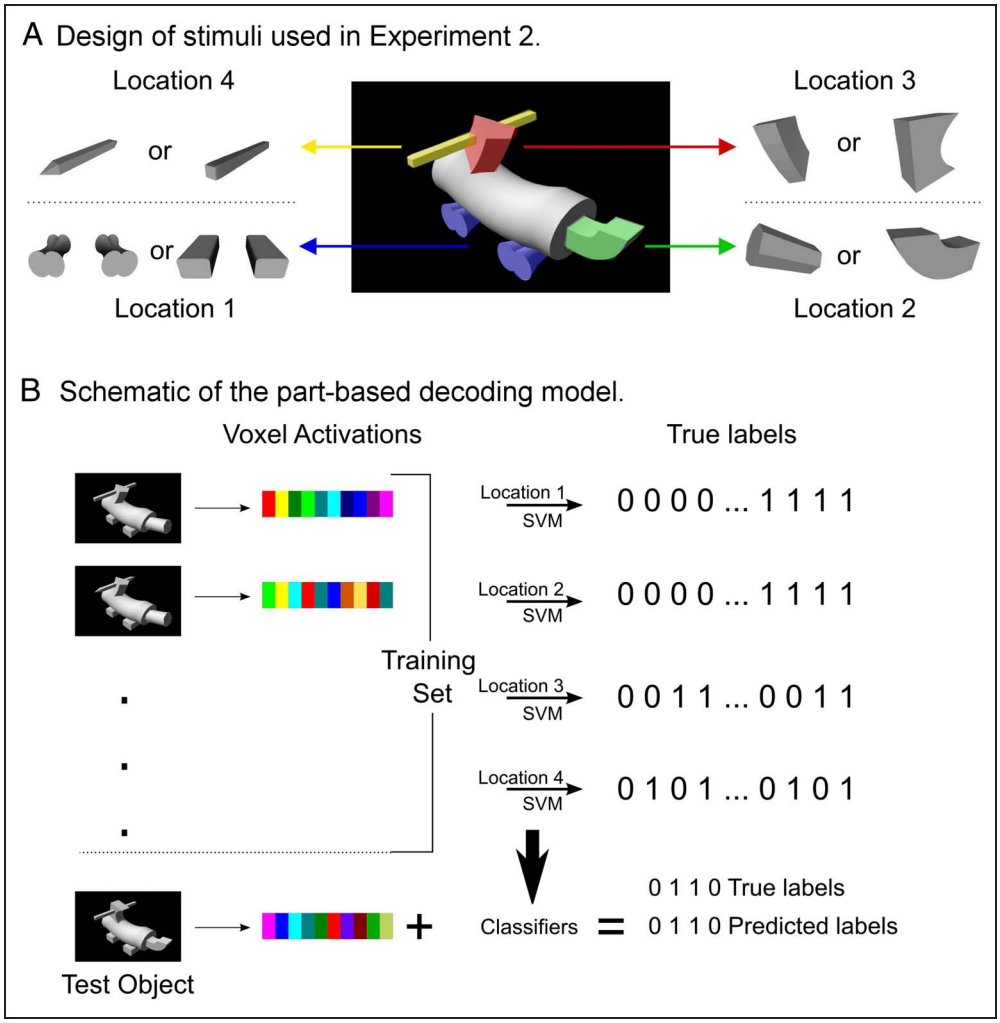

Figure 7. (A) Design of stimuli. Each object is composed of four components at four fixed locations. (Parts are colored for illustration purposes. All images were grayscale in the experiment.) (B) Schematic of the decoding model. Neural activations for 15 of the objects are used as the training set to train four linear SVMs to predict parts at each location. Then, the trained classifiers are used to predict the parts of the left-out test object, and these predictions are compared with the true parts of the object.

based on the stimuli used here and because there is substantial evidence indicating that LOC represents object shape, a stronger test of the nature of LOC object representations is provided by fine-grained analysis of the part structure within the materials from Experiment 2.

Part-based Object Representations in LOC

Finally, we sought to directly test the hypothesis that LOC encodes objects in a part-based manner. If the shape representations in LOC are encoding object parts, we should be able to decode the parts that make up an object from neural activations. We focused these analyses only on our second experiment because the stimuli in our first experiment are not suited to testing the part-based hypothesis. Although all objects used in Experiment 1 have a clear part-based structure, each part is at most shared by two objects, which drastically limits the amount of data available for decoding part identities. However, the stimuli in our second experiment were designed specifically to test the part-based hypothesis, with each part being shared by 8 of 16 objects in the stimulus set. The objects in our second experiment can be represented as four binary digits with each digit coding which one of the two possible parts for each of the part locations is present (see Figure 7 for a schematic of this analysis approach). In our decoding analyses, we thus sought to predict the four-digit binary representation of each object using neural activity patterns. We trained four separate linear SVMs, one for each location. Each SVM model was trained to predict which of the two possible part values for that location was present in an object. Each of the four classifiers was trained on 15 of the 16 objects, and the classifiers were tested by having them jointly predict the four-digit binary representation for the 16th object. If all four of the predictions (one for each location) were correct, we counted that as a successful decoding of the object (see Figure 7B). Thus, chance for this classification test was $0.5^4 = 0.0625$. This analysis approach was performed using 16-fold leave-one-out cross-validation, each time leaving one object out (for test) and training the classifiers on the remaining 15 objects. We then averaged the classification accuracies over folds to obtain an estimate of the classification accuracy across all objects for each participant. Statistical analysis was then performed over subject means. The

results of this analysis indicated that it was possible to decode novel objects in LOC, both for fMRI data obtained during visual and during haptic exploration of the objects (visual condition, L-LOC: classification accuracy = 0.198, $t = 3.61, p < .016$, R-LOC: classification accuracy = 0.250, $t = 5.81, p < .003$; haptic condition, L-LOC: classification accuracy = 0.167, $t = 2.50, p = .055$, R-LOC: classification accuracy = 0.302, $t = 5.86, p < .003$).

DISCUSSION

We have shown that it is possible to decode object identity from a cross-modal similarity matrix created by correlating LOC voxel patterns during visual and haptic exploration of the same set of objects. This suggests that there is a unique neural code generated during perceptual exploration of each of the novel objects that is similar regardless of whether the sensory modality is vision or touch. We also found that linear classifiers successfully predict a novel object based on its part structure. Thus, the fundamental units of object representation in LOC are expressed in terms of an object's composite parts. These findings provide further evidence for part-based visual representations of objects in LOC and multisensory representations of whole objects, at least across the haptic and visual modalities (Peelen et al., 2014; Naumer et al., 2010; Amedi et al., 2001, 2002; James et al., 2002). Crucially, our cross-modal decoding analyses relied on a direct comparison between activations from different modalities, representing a more direct test of the multisensory nature of object representations in LOC than was present in prior studies. Additionally, we believe our part-based decoding of novel objects presents a significant step towards understanding the nature of object representations in LOC. The only previous study that used a similar decoding analysis (Guggenmos et al., 2015) employed simpler stimuli (two-part objects) and presented objects only visually. Our study used a richer set of stimuli and showed that decoding of a novel object is possible from both visual and haptic activation in LOC. We believe that the findings we have reported strongly suggest that object representations in LOC are multisensory and part based.

Our results show an interesting hemispheric asymmetry; in most of our analyses, the findings are stronger in R-LOC. We do not have a clear understanding of why this is the case. A recent study suggests that haptic processing is stronger in LOC for the nondominant hand (Yalachkov, Kaiser, Doehrmann, & Naumer, 2015). However, it is important to note that participants in our experiment used both of their hands to explore objects. Additionally, these hemispheric differences are seen in the visual condition as well, making an explanation based on haptic processing unlikely. Future research should investigate whether this hemispheric asymmetry is a consistent characteristic of object shape processing or merely an artifact of our particular sample.

Although we have referred to the object representations in LOC as multisensory, it is worth pointing out that our study focused on visual and haptic processing, simply because shape information is conveyed mainly through these two modalities. For example, as previous research (Naumer et al., 2010; Amedi et al., 2002) shows, LOC does not respond to auditory stimulation. Similarly, our study says little about the representation of objects that lack a clear part-based structure, for example, bell peppers, or that are processed holistically, for example, faces. The question of how an object without a clear part-based structure is represented lies at a finer level than that on which our study focused; we did not investigate how an individual part might be neurally represented but whether parts are explicitly represented in the first place. Future research should focus on this more difficult question of how individual parts are represented.

In this study, we focused mainly on LOC and the nature of object representations in this region. However, looking at Table 1, we see that our searchlight results identified other regions, for instance, the precentral gyrus and the left posterior temporal-occipital cortex. Although none of those regions show the consistent activity that LOC shows across various analyses, it is possible that multisensory object representations reside in a larger network of brain regions and likely that multisensory object representations in LOC are embedded in a broader network of regions that support multisensory processing. This is an empirical question that needs to be addressed by future research.

A key claim of the part-based hypothesis is that objects are represented as a combination of shape primitives from a finite set. Although our data cannot speak to the inventory of shape-based primitives that the brain may encode, further research using the methods we have developed may be able to describe that inventory. A second key aspect of part-based theories of object representation is that spatial relations among parts are directly represented. The findings we have reported motivate a new approach to test whether the spatial arrangement among an object's parts are encoded in the same region (LOC) that encodes the part information. Alternatively, information about the spatial arrangement of parts may be stored elsewhere in the brain.

Our findings also bear on the principal alternative theoretical model to part-based object representations: image- or view-based models. View-based theories argue that the representation of an object is a concatenation of 2-D images of the object from different views (for discussion, see Peissig & Tarr, 2007). View dependency in object recognition is advanced as the main evidence for the view-based hypothesis. However, view-based models have difficulty accounting for our finding that there is a high degree of similarity in the voxel patterns elicited by haptic and visual exploration of objects and that the shared variance in voxel pattern maps onto the part structure of the stimuli.

In this study, we have presented evidence that LOC carries multisensory and part-based representations of objects. In addition to the empirical evidence presented here and in earlier studies, we believe this hypothesis is also appealing from a theoretical perspective as it elegantly captures how information can be transferred across modalities, how inputs from multiple modalities can be combined, and more generally, how we cope with a world that is in its essence multisensory.

Acknowledgments

We thank Elon Gaffin-Cahn for assistance with fMRI data collection. Preparation of this paper was supported by NIH R01 NS089609 to B. Z. M. and by AFOSR FA9550-12-1-0303 and NSF BCS-1400784 to R. A. J. F. E. G. was supported by a University of Rochester Center for Visual Science predoctoral training fellowship (NIH Training Grant 5T32EY007125-24).

Reprint requests should be sent to Robert A. Jacobs, Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627-0268, or via e-mail: robbie@bcs.rochester.edu.

REFERENCES

Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex, 12,* 1202–1212.

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience, 4,* 324–330.

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

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, U.S.A., 89,* 60–64.

Chen, Q., Garcea, F. E., & Mahon, B. Z. (in press). The representation of object-directed action and function knowledge in the human brain. *Cerebral Cortex*. doi: 10.1093/cercor/bhu328.

Cooke, T., Jäkel, F., Wallraven, C., & Bülthoff, H. H. (2007). Multimodal similarity and categorization of novel, three-dimensional objects. *Neuropsychologia, 45,* 484–495.

Cooke, T., Kannengiesser, S., Wallraven, C., & Bülthoff, H. H. (2006). Object feature validation using visual and haptic similarity ratings. *ACM Transactions on Applied Perception, 3,* 239–261.

Erdogan, G., Yildirim, I., & Jacobs, R. A. (2015). From sensory signals to modality-independent conceptual representations: A probabilistic language of thought approach. *PLoS Computational Biology, 11,* e1004610.

Fintzi, A. R., & Mahon, B. Z. (2013). A bimodal tuning curve for spatial frequency across left and right human orbital frontal cortex during object recognition. *Cerebral Cortex, 24,* 1311–1318.

Gaissert, N., Bülthoff, H. H., & Wallraven, C. (2011). Similarity and categorization: From vision to touch. *Acta Psychologica, 138,* 219–230.

Gaissert, N., & Wallraven, C. (2012). Categorizing natural objects: A comparison of the visual and the haptic modalities. *Experimental Brain Research, 216,* 123–134.

Gaissert, N., Wallraven, C., & Bülthoff, H. H. (2010). Visual and haptic perceptual spaces show high similarity in humans. *Journal of Vision, 10,* 1–20.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41,* 1409–1422.

Guggenmos, M., Thoma, V., Cichy, R. M., Haynes, J. D., Sterzer, P., & Richardson-Klavehn, A. (2015). Non-holistic coding of objects in lateral occipital complex with and without attention. *Neuroimage, 107,* 356–363.

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 encoding of relative position. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1032–1050.

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S., & Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia, 40,* 1706–1714.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital cortex. *Science, 293,* 1506–1509.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 3863–3868.

Lacey, S., Peters, A., & Sathian, K. (2007). Cross-modal object recognition is viewpoint-independent. *PLoS One, 2,* e890.

Lawson, R. (2009). A comparison of the effects of depth rotation on visual and haptic three-dimensional object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 911–930.

Naumer, M. J., Ratz, L., Yalachkov, Y., Polony, A., Doehrmann, O., Van De Ven, V., et al. (2010). Visuohaptic convergence in a corticocerebellar network. *European Journal of Neuroscience, 31,* 1730–1736.

Norman, J. F., Norman, H. F., Clayton, A. M., Lianekhammy, J., & Zielke, G. (2004). The visual and haptic perception of natural object shape. *Perception & Psychophysics, 66,* 342–351.

Peelen, M., He, C., Han, Z., Caramazza, A., & Bi, Y. (2014). Nonvisual and visual object shape representations in occipitotemporal cortex: Evidence from congenitally blind and sighted adults. *Journal of Neuroscience, 34,* 163–170.

Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology, 58,* 75–96.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Riddoch, M. J., & Humphreys, G. W. (1987). A case of integrative visual agnosia. *Brain, 110,* 1431–1462.

Schwarzbach, J. (2011). A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behavior Research Methods, 43,* 1194–1201.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.

Tarr, M. J. (2003). Visual object recognition: Can a single mechanism suffice? In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes* (pp. 177–207). New York: Oxford University Press.

Yalachkov, Y., Kaiser, J., Doehrmann, O., & Naumer, M. J. (2015). Enhanced visuo-haptic integration for the non-dominant hand. *Brain Research, 1614,* 75–85.