“What?” and “where?” versus “what is where?”: The impact of task on coding of object form and position in the lateral occipital complex

Sean P. MacEvoy

Fast and accurate recognition of both the identities and positions of objects in visual space is critical to deciphering visual environments. Studies in both humans and nonhuman primates have demonstrated that neural populations in ventral temporal visual areas are jointly tuned to both the form and position of objects, allowing information about the identities of objects to be “tagged” with their positions. Because not all behaviors demand that the identities of objects be associated with position information with equal precision, however, the present study asked whether the spatial tuning of form-encoding populations in the human lateral occipital complex (LOC) is sculpted by task demands. Subjects were scanned using functional magnetic resonance imaging while viewing matches of the game Rock, Paper, Scissors played with exemplar pairs from those categories. Subjects first performed a repetition-detection task that depended on object form but not position; subsequently, subjects viewed the same stimuli while determining the position of each pair’s “winner,” a task that depended upon the conjunction of object form and position. Compared to data from the initial scan, multivoxel activity patterns evoked in the lateral occipital (LO) subdivision of LOC while subjects judged winners showed enhanced sensitivity to the relative positions of objects in pairs. Although superficially consistent with dynamic position tuning, this effect appears to be attributable to an accompanying task-dependent improvement in the sensitivity of LO populations to object form. The results thus suggest that the spatial tuning of form-encoding populations in LO does not depend upon the precision of spatial information demanded by a task.

Introduction

A growing body of literature demonstrates that temporal-lobe visual areas thought to be critical to object recognition are considerably more sensitive to the locations of objects in the visual field than the traditional ventral/dorsal, “what”/“where” heuristic might have predicted (Mishkin, Ungerleider, & Macko, 1983). In the human visual system, activity patterns in the lateral occipital complex (LOC), a region regarded as central to object recognition, can differ as significantly between two different positions of a single object as between objects from different categories at the same retinal location (Carlson, Hogendoorn, Kanai, Mesik, & Turret, 2011; Cichy, Chen, & Haynes, 2011; Kravitz, Kriegeskorte, & Baker, 2010; MacEvoy & Epstein, 2009; Sayres & Grill-Spector, 2008; Schwarzlose, Swisher, Dang, & Kanwisher, 2008). Moreover, LOC patterns differ significantly between the relative positions of objects within pairs (Hayworth, Lescroart, & Biederman, 2011; MacEvoy & Yang, 2012), which is consistent with the presence in LOC of neural populations that are jointly tuned for both object form and position, similar to tuning profiles observed in single-unit recordings from macaque inferotemporal cortex (Aggelopoulos & Rolls, 2005; DiCarlo & Maunsell, 2003; Op de Beeck & Vogels, 2000).

Joint form/position tuning presents a straightforward solution to the problem of binding the identities of objects to their locations in space, which human observers appear to automatically do (Kravitz et al., 2010; Lin & He, 2012). Not all behaviors require equally tight coupling between these variables, however. To an observer attempting to decide quickly whether a room in an unfamiliar house is a kitchen or an office, the identities of objects are of great importance (“Is there a computer or a refrigerator somewhere?”) but their positions are unlikely to matter much. In other circumstances, however, behavior may require tight binding between the identities of objects and their positions. For instance, rapid identification of a specific office, or of the viewer’s location within that office, may require an assessment of not just what
objects are present but where in space each is located. In theory, simple detection of object presence would be best served by relatively relaxed spatial-tuning profiles among neurons encoding object “form” (i.e., visual features linked to identity), since this would enhance sensitivity by maximizing the number of receptive fields overlapping each point in space. In contrast, a task that required knowledge of each object’s identity and position would benefit from comparatively narrow spatial tuning among the same population of neurons. The present study was designed to discern whether the visual system reconciles these competing demands by adjusting the spatial tuning of form-encoding populations in LOC in response to task demands. Using functional magnetic resonance imaging (fMRI), multivoxel activity patterns evoked by object pairs were recorded first while participants viewed pairs of objects and performed a task that demanded attention solely to the identities of without without regard to their positions. These patterns were compared to those recorded during a second session, in which participants viewed the same object pairs while performing a task that required linking object identity and position. The hypothesis that the spatial tuning of form-encoding neural populations in LOC varies to meet task demands was tested by comparing the amount of information about the spatial configurations of pairs that was contained in multivoxel activity patterns evoked during the two sessions. Any improvement in spatial tuning of form-encoding populations resulting from task demands was expected to be evident as an improvement in the sensitivity of patterns to pair configuration, above and beyond improvements to either the identities of objects within pairs or their absolute positions.

**Methods**

**Participants**

Nine participants (six female, three male, all aged 19–21 years) with normal or corrected-to-normal visual acuity gave written informed consent in compliance with procedures approved by the Institutional Review Board of Boston College in accordance with the Declaration of Helsinki. Neuroimaging data from two participants who failed to perform behavioral tasks at above-chance levels were excluded from analysis.

**MRI acquisition**

All scan sessions were conducted at the Brown University MRI Research Facility using a 3T Siemens Trio scanner and a 32-channel head coil. Structural T1*-weighted images for anatomical localization were acquired using 3-D MPRAGE pulse sequences (TR = 1620 ms, TE = 3 ms, TI = 950 ms, voxel size = 0.9766 × 0.9766 × 1 mm, matrix size = 192 × 256 × 160). T2*-weighted scans sensitive to blood-oxygenation-level-dependent (BOLD) contrasts were acquired using a gradient-echo echo-planar pulse sequence (TR = 3000 ms, TE = 30 ms, voxel size = 3 × 3 × 3 mm, matrix size = 64 × 64 × 45). Visual stimuli were rear projected onto a screen at the head end of the scanner bore and viewed through a mirror affixed to the head coil. The entire projected field subtended 24° × 18° at a pixel resolution of 1024 × 768.

**Experimental procedure**

Scan sessions encompassed six experimental scans followed by two functional localizer scans. Experimental stimuli were 16 color photographs from each of the categories of rocks/stones, scissors, and sheets of paper. Object images were edited to remove any background elements and rescaled and cropped so that the longer of either the horizontal or vertical dimension of each object spanned 3°.

Pairs of objects from different categories were presented horizontally at either 2° or 3° below screen center (Figure 1). At each of these elevations, objects within pairs were positioned 1.7° to the left and right of the vertical meridian. Pairs were only presented in the lower visual field to take advantage of the lower-field preference of LOC (Niemeier, Goltz, Kuchinad, Tweed, & Vilis, 2005; Sayres & Grill-Spector, 2008). Pairs were presented for each three-way cross of three types of object content (rock/scissors, rock/paper, and paper/scissors), two screen elevations, and two left-right configurations (e.g., left-rock/right-scissors and left-scissors/right-rock), resulting in a total of 12 stimulus conditions. Each trial consisted of two object pairs with the same content, elevation, and configuration presented sequentially for 140 ms each, separated by a 40-ms fixation-only gap and followed by an interstimulus interval of 1140 ms (Figure 1). In one half of trials the two pairs contained different exemplars, while in the remaining half one exemplar was repeated. Repetition trials were equally distributed across the 12 condition types, with an equal number of repetitions by the left and right objects in each condition type. Pair trials along with 1.5-s and 3-s null events were sequenced according to a type 1-index 1 sequence (Aguirre, 2007); the 1.5-s null appeared 3 times as often as any one of the 12 pair types, and the 3-s null appeared with the same frequency as each pair type. Each scan run contained 16 repetitions of each pair type, and each object exemplar appeared approxi-
mately six times per scan. Unique type 1-index 1 sequences were constructed for each scan and each subject. Each scan included 15-s fixation-only intervals at its beginning and end, and lasted for a total of 7 min 48 s.

Each subject participated in two scan sessions, separated by at least 2 days. Both sessions for each subject used the same six stimulus sequences as described previously but differed in the subject’s task. In the first session, subjects were instructed to respond by button press whenever they detected a trial containing an object repetition. In the second session, they were told to ignore repetitions and instead evaluate each trial according to the rules of the game of Rock, Paper, Scissors (RPS), indicating by button press the side of the screen where the “winner” of each object pair appeared. The goal of this task was to force subjects to attend to the conjunction of object category and position, with the circular RPS hierarchy (rock defeats scissors, scissors defeats paper, paper defeats rock) demanding that subjects attend to both objects in each pair. The RPS task was always introduced in the second scan session, to prevent its covert performance during repetition detection.

Both scan sessions also included two functional localizer scans lasting 6 min 15 s each, during which subjects viewed blocks of color photographs of scenes, faces, common objects, and scrambled objects presented at a rate of 1.33 pictures per second (Epstein & Higgins, 2006). Localizer stimuli occupied the central 15° of visual space.

**fMRI analysis**

Functional images were corrected for differences in slice timing by resampling slices in time to match the first slice of each volume, realigned with respect to the first image of the scan, and spatially normalized to the Montreal Neurological Institute template. Brain volumes from localizer scans were spatially smoothed with a 9-mm FWHM Gaussian filter and analyzed using a general linear model (GLM) as implemented in VoxBo, including an empirically derived 1/f noise model, filters that removed high and low temporal frequencies, and nuisance regressors to account for global signal variations and between-scan signal differences. Localizer models were also executed on unsmoothed volumes to construct voxel maps of object selectivity used in feature selection during multivariate analyses of data from experimental scans.

Volumes from experimental scans were preprocessed as described and passed to custom routines in MATLAB for multivariate analysis. Voxel time-series data were detrended and up-sampled to 500-ms resolution with a linear-phase finite-impulse-response filter interpolating symmetrically from the nearest eight values of the original data (MATLAB interp() function). Voxel signals evoked by individual stimulus trials were calculated as the average of the first three 500-ms data points aligned with and following stimulus onset, after shifting stimulus time stamps by 4.5 s to account for hemodynamic delays. Signals from all voxels within each region of interest (ROI; see the “Regions of interest” subsection for definitions) were concatenated to form a response vector that was stored for each stimulus event. Voxel signals were z-scored across the full library of single-trial response vectors within each scan.

To measure the sensitivity of each ROI to pair differences, voxel patterns for each stimulus type were averaged within complementary data halves (e.g., even scans versus odd scans, etc., exhaustively), with sensitivity to each stimulus dimension measured by subtracting the median between-half correlation among all pairs of patterns evoked by stimuli differing along...
that dimension from the median correlation among all pairs of patterns evoked by stimuli sharing a position along that dimension; the resulting values are referred to as "correlation differentials" hereafter. Typically, these differentials would be calculated from correlations between patterns evoked by individual stimulus types; for example, the correlation between the patterns evoked by left-rock/right-scissors and by left-scissors/right-rock would be pooled with the correlations from all the other between-configuration stimulus pairings, and the median value from that pool subtracted from the median within-configuration correlation, such as between patterns evoked by left-rock/right-scissors in the two data halves. This simple approach was not suited for the present study, owing to the correlation between behavioral responses and pair configuration during the RPS task. Because opposite configurations of each pair were associated with different behavioral responses, it was possible that between-configuration correlations that were significantly less positive than within-configuration correlations (i.e., resulting in significantly positive correlation differentials) might result from signals related to differing behavioral responses rather than to differing visual stimuli.

To avoid this confound, correlation differentials for each dimension were computed from correlations among "composite" patterns that were each the average of the patterns evoked by the two object pairs at the same locus along that dimension. For instance, to assess pattern sensitivity to pair content, composite patterns were constructed by averaging together the two patterns at each elevation that had the same content but different left/right configurations, and hence opposite behavioral responses. Correlation differentials for content were computed from a similarity matrix computed from the six (three pair-content types by two screen elevations) resulting patterns (Figure 2A, B). The same matrix was used to compute correlation differentials for pairs at the same or different elevations, which were taken as a measure of pattern sensitivity to the absolute positions of pairs.

In the configuration dimension, composite patterns were the averages of the patterns evoked by the two pairs containing one object category at a given left/right position (Figure 2C). The resulting 12 patterns (three object types by two left/right positions, each at two elevations) were used to generate a similarity matrix from which correlation differentials for configuration were computed by subtracting the correlations between composite patterns corresponding to opposite left/right positions of one object from the correlation between patterns corresponding to the same left/right position of that same object (Figure 2D). This procedure essentially recasts the measure of configuration sensitivity from the correlation differential between right-rock/left-paper and right-paper/left-rock to the differential between right-rock/left-"everything else" and right-"everything else"/left-rock. Because the circular RPS hierarchy dictated that the two pairs contributing to each composite were associated with opposite behavioral responses, this new configuration dimension was orthogonal to the dimension that determined behavioral responses, leaving composite patterns free from behavioral contamination. Note that this averaging scheme necessarily destroyed any information about specific dyadic object relationships contained in patterns. However, such relationships were not the target of this study; relationships were only important insofar as they formed a basis for a task requiring participants to link the identity of each object to its position on the screen. It should also be noted that configuration sensitivity measured via composite pattern provides as direct a measure of relative position tuning as configuration sensitivity measured by direct comparison of patterns evoked by different configurations of single pairs. In both cases, complete insensitivity to object position among form-encoding neurons would produce correlation differentials of zero; further, significantly positive correlation differentials could only occur with either analysis approach if form-encoding populations were also tuned for position, with the sizes of differentials necessarily dependent upon tuning precision. Separate correlation differentials for content and configuration were computed for each pair elevation; however, because there were no differences between differentials at the two elevations, all differentials presented here are averaged across elevations.

For scans in which subjects performed the RPS task, only patterns from trials in which subjects responded correctly contributed to analysis. Mean accuracy on the RPS task among participants whose functional data were analyzed was 93.2% (SD = 5%, range = 80.9%–96.8%). To equate the number of trials contributing to analysis of the results of the two task types, analysis of each subject's data from repetition-detection scans drew upon a similarly circumscribed library of responses; for example, if a subject only delivered 10 correct RPS choices with a left-rock/right-scissors pair in the second of the six RPS scans, then only 10 left-rock/right-scissors response patterns acquired during the second of six repetition-detection scans were eligible for analysis. The latter were drawn randomly, and similarity scores were averaged across 1,000 draws.

**Statistical tests**

Single-sample t tests were used to assess the significance of positive correlation differentials for each dimension, and paired t tests were used to assess the significance of increases in correlation differentials during the RPS task versus the repetition-detection task.
task. Because the study posed clear hypotheses about
the sign of effects in all cases, all tests were one-tailed
unless otherwise noted.

Searchlight analysis

Whole-brain maps of local pattern sensitivity to
pattern dimensions were generated using a searchlight
procedure (Kriegeskorte, Goebel, & Bandettini, 2006).
A 5-mm-radius mask was defined for each voxel in
each participant’s brain, and the pattern-analysis
procedures described previously were performed for
the patterns defined by each of these 19-voxel regions.

Searchlight masks near brain margins were truncated
as needed to include only voxels inside the brain;
 masks containing fewer than three voxels were
discarded. Correlation differentials for each search-
light mask were assigned to the voxel at its center,
generating separate differential maps for pair content,
elevation, and configuration.

Single-subject correlation-differential maps were
passed to a second-level group analysis to identify
brain regions demonstrating significant sensitivity to
each dimension. Using a permutation-based approach
(Nichols & Holmes, 2002), group average maps were
generated for each of the possible sign permutations of
the set of seven single-subject maps. The resulting 27-
element distributions of permuted values for each voxel were used to identify all voxels within each permutation map with values exceeding their respective 95th percentiles, and the size of each permutation’s largest six-connected cluster of such voxels was recorded. Suprathreshold voxels were similarly identified in actual (i.e., nonpermuted) group maps, within which regions of significantly positive correlation differentials were identified as clusters with sizes exceeding the 95th percentile of the distribution of maximum cluster sizes from permuted maps, providing control of Type I error. The same overall approach was used to search for regions displaying task-related enhancement in sensitivity, except that input maps were the difference between each subject’s correlation-differential maps from the repetition detection and RPS tasks.

Regions of interest

LOC was defined from spatially smoothed localizer scans as the set of voxels in the lateral/ventral occipitotemporal region exhibiting stronger responses ($p < 0.001$, uncorrected) to objects than to scrambled objects. LOC was divided into lateral occipital (LO) and posterior fusiform (pF) areas on the basis of anatomy and activity clusters (Figure 3). The inferior intraparietal sulcus (IPS) was defined from the same contrast as a separate posterior parietal cluster (Xu, 2008). Early visual cortex (EVC) was defined by significantly higher responses to scrambled objects than intact objects in the posterior occipital lobe. A second GLM was run on unsmoothed localizer volumes to derive voxelwise $t$-value maps from the contrast between objects and scrambled objects, which were used to generate finer-grained ROIs (see the “Selective voxel analysis” subsection of Results).

Whole-ROI analysis

Dynamic optimization of spatial tuning of LOC to meet task demands predicts that multivoxel activity patterns should be more sensitive to the configurations of object pairs during behaviors to which the conjunction of object identity and position is relevant than during behaviors to which it is not. This prediction was tested first in the LO subdivision of LOC, which has been shown in previous studies to be sensitive to both the positions and identities of single objects (Grill-Spector et al., 1999; Schwarzlose et al., 2008). Consistent with these results, the subject-averaged LO pattern-similarity matrix constructed with respect to object category and derived from data collected during the repetition-detection task showed a ridge of high correlations along the main diagonal and generally higher values in in the upper-right and lower-left quadrants compared to the other two quadrants (Figure 4A; correlations are from response patterns concatenated across hemispheres), indicating sensitivity to pair content and elevation, respectively. Average correlation differentials (Figure 4E, light bars) computed from single-subject matrices were significantly greater than zero for both of these dimensions when patterns were concatenated across hemispheres (content: $t(6) = 4.4$, $p = 0.002$; position: $t(6) = 3.9$, $p = 0.004$). Hemisphere-level analysis (Figure 4F, G, light bars) showed significant sensitivity to both form and elevation in both the left (content: $t(5) = 4.7$, $p = 0.002$; elevation: $t(5) = 3.2$, $p = 0.01$) and right (content: $t(5) = 3.5$, $p = 0.009$; elevation: $t(5) = 3.7$, $p = 0.007$) LO.

Because sensitivity to elevation could be at least in part conferred by neural populations which are selective for position but not object category, the spatial sensitivity of those populations which also carry information about object form (i.e., which are jointly tuned for form and position) must be estimated from pattern sensitivity to pair configuration. Again consistent with previous work demonstrating LO sensitivity to relative position (Hayworth et al., 2011; MacEvoy & Yang, 2012), the subject-averaged LO similarity matrix derived from configuration-specific patterns showed a ridge of high values along the main diagonal (Figure 4C), and the average correlation differential for configuration was significantly greater than zero, $t(6) = 4.1$, $p = 0.003$. Configuration sensitivity was significant in both hemispheres, left: $t(5) = 2.7$, $p = 0.02$; right: $t(5) = 2.6$, $p = 0.02$.

To test the hypothesis that a task requiring subjects to link object identity and position improves the spatial tuning of form-encoding neural populations in LO,
each participant was rescanned on a different day while he or she viewed the same sequence of stimuli but was asked to ignore repetitions and instead determine the “winner” of each two-pair event according to the rules of RPS. This task required participants to attend not only to the form of both objects in a pair, in order to ascertain the correct decision rule to apply, but also to the positions of each object in order to determine the location of the winner. Any concomitant sharpening of spatial tuning among form-encoding LO populations should have been evident as a significant increase in sensitivity of LO patterns to pair configuration relative to sensitivity to pair content. This final proviso is critical given that the more engaging and/or demanding RPS task was expected to produce an overall improvement in pattern sensitivity along all stimulus dimensions.

The content-similarity matrix for the RPS task shows a clear enhancement along the main diagonal relative to the matrix for the repetition-detection task (Figure 4B vs. 4A), indicative of an improvement in pattern sensitivity to the identities of objects within pairs. This is confirmed by a significantly greater average correlation differential for content in LO patterns concatenated across hemispheres (Figure 4E, comparison between brown bars; \( t(6) = 3.42, p = 0.007 \)). Although the configuration-similarity matrix also shows higher values along the main diagonal for data acquired during the RPS task versus the repetition-detection task (Figure 4D vs. 4C), the increase in correlation differential for configuration did not achieve significance (Figure 4E, comparison between green bars; \( t(6) = 1.5, p = 0.09 \)). Patterns analyzed...
separately for each hemisphere show that task-related effects principally arose in left LO (Figure 4F, G). However, even in left LO, improvements in configuration sensitivity, although significant, were no greater than improvements in content sensitivity and in fact were exceeded by them. This combination of results suggests that whatever LO improvements in configuration sensitivity the RPS task produced appear to have resulted from enhancements in the precision with which LO populations encoded the identities of objects rather than from enhancements to their spatial tuning.

Consistent with this interpretation, there was no significant improvement in sensitivity to pair elevation in either LO as a whole, \(t(6) = 0.3, p = 0.4\), or in the left hemisphere, \(t(5) = 1.4, p = 0.1\). The absence of an effect on elevation sensitivity also suggests that the improvement in content sensitivity did not result from a generalized task-related increase in the magnitude or reliability of BOLD signals, nor from task order.

In contrast to LO, the content similarity matrix for the pF subdivision of LOC indicates sensitivity to object identity during the repetition-detection task but not to elevation (Figure 5A). Consistent with this and previous demonstrations of comparatively weak sensitivity in pF to position (Schwarzlose et al., 2008), correlation differentials during repetition detection (Figure 5E, light bars) were significantly greater than zero for pair content, \(t(6) = 2.1, p = 0.04\), but not for pair elevation, \(t(6) = 0.28, p = 0.4\), or configuration, \(t(6) = 1.0, p = 0.17\). Although correlation values along the main diagonal appear even more prominent in the content-similarity matrix derived from patterns evoked by the RPS task (Figure 5B), there was no significant effect of task on correlation differentials for object content (Figure 5E, brown bars; \(t(6) = 1.3, p = 0.12\)) or for elevation, \(t(6) = 1.24, p = 0.13\). Similarly, apparent differences between configuration-similarity matrices under the two task types (Figure 4C, D) did not translate into significantly different configuration correlation differentials (Figure 5E, green bars; \(t(6) = 1.22, p = 0.13\)). Similar patterns were evident in each hemisphere (Figure 5F, G), with the exception that the improvement in content sensitivity in right pF crossed the threshold of statistical significance, \(t(6) = 2.07, p = 0.04\). It is important to point out that no inference should be drawn from the failure of these improve-
mments to reach formal significance; indeed, they may have done so if the sample had been larger or if stimuli had been positioned to take advantage of the reported preference of the pF for stimuli in the upper visual field (Kravitz et al., 2010). What is important is that these data provide absolutely no indication of even a trend towards greater improvements in configuration versus content sensitivity that would be expected from a selective improvement in the position sensitivity of form-encoding populations.

To provide context for results in LO and pF, the sensitivity of EVC to each dimension was also analyzed. In addition to the expected sensitivity to pair elevation, $t(6) = 5.2, p = 0.001$, activity patterns in EVC showed significant although comparatively weak sensitivity to pair content, $t(6) = 2.1, p = 0.04$, and somewhat greater sensitivity to configuration, $t(6) = 6.2, p = 0.0004$, during the repetition-detection task (Figure 6). This combination of results is counterintuitive, since sensitivity to configuration presumably depends upon sensitivity to object form. In this case, however, sensitivity to configuration likely reflects the pattern-averaging scheme that was employed to avoid contamination of activity patterns during the RPS task with signals related to behavioral responses. As described in the Methods section, content sensitivity was measured using comparisons among patterns that were the averages of the patterns evoked by the two configurations of each object pair, while configuration sensitivity was measured using comparisons between patterns that were the averages of those evoked by the two patterns that included a given object at a single position. Because the same averaging scheme applied to the stimuli themselves (rather than the patterns they evoked) would likely produce retinal images that differed more between configurations than between object contents, greater sensitivity to configuration than content should be expected in regions characterized by high-resolution retinotopy. (Image analysis confirming this assertion can be found in Supplemental Figure 3.) During the RPS task, EVC showed a nearly significant improvement in sensitivity to pair content, $t(6) = 1.93, p = 0.051$, a nonsignificant decrease in sensitivity to configuration (two-tailed $t$ test; $t(6) = -1.8, p = 0.13$), and a nonsignificant improvement in elevation sensitivity, $t(6) = 1.5, p = 0.09$. Correlation differentials for each hemisphere followed an almost identical profile, except that the improvement in content sensitivity reached formal significance in left EVC, $t(6) = 2.04, p = 0.043$, but not in right EVC, $t(6) = 1.77, p = 0.063$; both hemispheres showed no significant effect of task on correlation differentials for either configuration or elevation.

Because the repetition-detection and RPS tasks potentially differed in the degree of attention they demanded from participants, the impact of task on pair-evoked responses in IPS was also assessed. Average correlation differentials in IPS did not differ significantly from zero for any pair dimension during either task type in this sample, nor was there any significant effect of task on sensitivity to any dimension ($p > 0.07$ for all cases).

Selective voxel analysis

The foregoing analyses are based on the responses of all voxels within each ROI. To focus analyses more precisely on neural populations contributing most to codes for object form, voxels were ranked according to their $t$ values from the contrast between objects and scrambled objects in localizer scans. These rankings were used to construct a hierarchical family of “child” ROIs ranging in content from the 10 highest ranked voxels within their parent ROI to the 150 highest ranked voxels, in 10-voxel increments. The sensitivity to each dimension along which object pairs varied was then calculated for each of these 15 child ROIs.

For child ROIs generated from voxels pooled across hemispheres, sensitivity to pair content during the repetition-detection task was uniformly higher than for LO as a whole during the same task (Figure 7A, dotted brown line; compare to values in Figure 4E). The average correlation differential for content peaked for the child ROI composed of just the top 30 voxels. Sensitivity to pair elevation showed no similar peak within child ROIs ranging in size up to 150 voxels (Figure 7A, dashed line), indicating that the ranking procedure succeeded in identifying voxels containing information about object identity and not just those with generally high signal-to-noise ratios. It is also suggestive of some systematic trade-off between neural tuning for object form and position in LO. Sensitivity to configuration was also greater for each child ROI than for LO as a whole (Figure 7C).
As in LO as a whole, the RPS task increased correlation differentials across all child ROIs for both pair content and configuration (Figure 7A, C, solid lines). However, while 10 of the 15 tested child ROIs showed significant increases in content differentials \((p < 0.05, \text{Bonferroni corrected})\), none showed a significant improvement in sensitivity to configuration. Although child ROIs were not independent voxel samples and the statistical correction was highly conservative, this stark difference in the frequency of significant improvements nonetheless shows that the effect of task was largely manifested as an improvement in content sensitivity, even among voxels that were already more sensitive to content than average. When child ROIs were generated within each hemisphere (data not shown), significant effects of task were limited to left LO, similar to the results of the whole-LO analysis. Even in the left hemisphere, however, significant improvements were only evident for content sensitivity, for which 11 of 15 tested child ROIs exceeded a corrected threshold of \(p = 0.05\). A similar profile emerged in pF (Figure 7B, D). While 4 of 15 child ROIs showed significant increases in content differentials \((p < 0.05, \text{corrected})\), none showed a significant increase in configuration differential, even at an uncorrected statistical threshold. Essentially identical results were obtained for pF child ROIs assembled within hemispheres.

**Searchlight analysis**

A searchlight pattern-analysis approach was used to identify any brain areas in which the RPS task produced an enhancement in sensitivity to any pair dimension. Enhanced sensitivity to content was evident in a single significant cluster of voxels (permutation test, \(p < 0.002, \text{corrected} \)) centered in left LO (Figure 8). No significant clusters of enhancement were evident for either configuration or elevation sensitivity.

**Discussion**

Previous studies have shown that activity patterns evoked by single objects in LOC differ significantly with object identity at either the category or individual level, as well as with objects’ retinal positions (Carlson et al., 2011; Cichy et al., 2011; Kravitz et al., 2010; MacEvoy & Epstein, 2009; Sayres & Grill-Spector, 2008; Schwarzlose et al., 2008). Although these findings are consistent with underlying populations of neurons jointly tuned for object form and position, they could also reflect the combined activity of two complementary neural populations: one encoding object form but insensitive to object position and another sensitive to object position but not to form. This possibility makes it difficult to draw inferences about the spatial tuning of LOC populations encoding object form from studies using single objects as stimuli. Measuring the sensitivity of patterns to the configurations of paired objects
avoids this problem because any such sensitivity necessarily requires joint tuning to both object form and position (MacEvoy & Yang, 2012). In the present study, sensitivity to pair configuration was used to selectively probe the impact of task upon spatial tuning of form-encoding LOC populations, under the hypothesis that a task requiring attention to the conjunction of object identity and position should enhance their spatial sensitivity relative to a task which required attention solely to object form. Ultimately, the results of the present study offer no evidence for any such enhancement in either the LO or pF subdivisions of LOC. Although the RPS task was associated with a marginal improvement in the average sensitivity of LO patterns to pair configuration, it was not significantly different from the improvement in sensitivity to pair content. Inasmuch as sensitivity to configuration is logically a product of sensitivity to both object identity and position, the observed improvement in configuration sensitivity can therefore be fully accounted for by the simultaneous improvement in content sensitivity. This suggests that the position sensitivity of form-encoding LOC populations was unaltered by task demands.

It warrants emphasis that this study was aimed at understanding task-dependent flexibility in the spatial specificity with which individual objects in pairs were encoded, not at the ways that relationships between those objects—spatial or otherwise—were encoded. Although the RPS task did draw upon relationships among objects, in this experiment the task was only important insofar as its dependence on relative object position required, as a prerequisite, that participants link the left/right position of each object to its identity. (It would be possible to design a task performed on single objects that similarly required joint encoding of object form and position; however, as outlined earlier, an experiment using single objects would be unable to target the tuning of form-encoding populations as precisely as one using object pairs.) Furthermore, the pattern-averaging technique used to analyze pattern sensitivity to configuration precludes any statements about neural correlates of spatial or functional relationships within specific object dyads. As explained in the Methods section, the configuration of specific pairs was confounded with behavioral response during the RPS task. Therefore, configuration correlation differentials were based on correlations among patterns that were each the average of the patterns evoked by the two pairs with a single object at each position. While this served to decorrelate behavioral responses from configuration polarity, it also destroyed any information pertaining to specific dyads. Thus the analyses presented here do not foreclose the possibility that one or both LOC subunits may encode information about object relationships abstracted from retinal position, as has been demonstrated previously in humans using both neuroimaging and behavioral approaches (Hayworth et al., 2011; Lin & He, 2012) and has been observed among single units in macaque visual areas (Olson & Gettner, 1995; Pasupathy & Connor, 2001). Completion of the RPS task can be formalized as a two-step process (albeit with numerous subordinate steps): First the identity of the winner in each pair is determined, and then its location is ascertained. In theory, the second step could be completed via search, eliminating the need for any improvement in spatial tuning over the repetition task (and indeed for any joint encoding of object position at all, although position tuning in LO is not in doubt). However, the short presentation time of pairs in the present study (both pairs in each trial visible for a total of 320 ms) makes it unlikely that participants could have completed the task in this way. Given that response latencies of temporal-lobe visual neurons in awake macaques have been reported to be on the order of 75–160 ms (Perrett, Rolls, & Caan, 1982; Richmond, Optican, Podell, & Spitzer, 1987; Sugase, Yamane, Ueno, & Kawano, 1999; Tamura & Tanaka, 2001), participants using a search strategy would have had at most less than 250 ms to (1) identify both objects in each pair, (2) recall the correct rule for that pair, and (3) direct attention to one of the objects and determine whether it matched the winner. Thus, it would appear more likely that participants were forced to link each object’s position to its identity before a decision was made.

Although the present results do not support the existence of task-dependent flexibility in the precision of position tuning in LO, the enhancement in pattern sensitivity to pair configuration and content does suggest that the RPS task specifically altered either the quantity or quality of information about object identity encoded by LO. Two features of the data support this interpretation over the alternatives that either (a) the RPS task, by demanding greater attention, merely produced a general improvement in BOLD signal-to-noise ratios by way of elevated neural responses (Smith, Kosillo, & Williams, 2011; Tong, Harrison, Dewey, & Kamitani, 2012) or (b) that greater sensitivity to content and configuration was a consequence of the fixed task order used to avoid participants’ implicitly performing the RPS task during repetition-detection scans. The first feature is the absence of any discernible improvement in the sensitivity of LO patterns to pair elevation during the RPS task. If the improvement in LO pattern sensitivity to pair content and configuration simply reflected a general attention- or order-related increase in overall LO responsiveness or other improvement in signal quality, there should have been an improvement in sensitivity to pair elevation in addition to content and configuration. This was not observed.
The absence of such an improvement is unlikely to be a ceiling effect (i.e., saturation of population sensitivity to position), since a previous study using almost identical stimuli observed graded sensitivity to elevation changes greater than that used here (MacEvoy & Yang, 2012). The second feature arguing against a nonspecific attention- or order-based explanation is the laterality of effects. Any general attention or order effects should have been similar across hemispheres. Instead, significant improvements in sensitivity to both content and configuration were limited to the left hemisphere, which has previously been implicated in categorical processing (Grossman et al., 2002; Laeng, Zarrinpar, & Kosslyn, 2003; Li, Ostwald, Giese, & Kourtzi, 2007; Seger et al., 2000).

Somewhat surprisingly, evidence in the literature for an impact of behavioral object categorization upon the quality of categorical information in LOC is scant. Although recognition of the identities of degraded objects (Hsieh, Vul, & Kanwisher, 2010) and training on categorization tasks can enhance the degree to which human temporal-lobe visual areas distinguish between stimuli (Folstein, Palmeri, & Gauthier, 2013; Gillebert, Op de Beeck, Panis, & Wagemans, 2008; Jiang et al., 2007; Li, Mayhew, & Kourtzi, 2009; Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006), an effect of a categorization task on neural representations of very familiar objects at suprathreshold visibility has not previously been reported. Closest to the results of the present study are those of Li et al. (2007), who showed that performance of a pattern classifier applied to LOC varied dynamically with the dimensions along which subjects were instructed to categorize stimuli. However, the design of that study made it difficult to exclude a role for signals stemming from subjects’ decisions in mediating this effect, a confound that likely explains the general paucity of studies targeting the effects of categorization tasks on LOC category information. The combination of task and stimuli used in the present experiment avoids this problem, since although the RPS task required extraction of the categories of objects in each pair, this information remained orthogonal to participants’ decisions.

Given the shortcomings of attention- or order-based explanations for the impact of the RPS task, and the exclusion of behavioral confounds by the design of the experiment, the present results suggest that the RPS task, by behaviorally prioritizing information about the category identity of objects relative to the repetition-detection task, induced a parallel shift towards categorical object representations in LOC. Although this shift could conceivably reflect the recruitment of neural populations encoding semantic category per se, a more plausible explanation is that attention to category altered the way LO populations encoded the visual form of objects. This could have occurred in two ways. First, attention to category could have sharpened the tuning profiles of individual neurons in LO to visual features, thereby making population responses to individual objects sparser. Inasmuch as exemplars from the same category generally share more features than do objects from different categories, this sharpening would have tended to make patterns evoked by different categories of objects more distinctive. Similar attention-dependent sharpening of neural tuning in LOC has been reported previously (Murray & Wojciulik, 2003). Second, attention to category might have driven neural representations of objects towards that of the prototype for each category. In this view, attention to object category triggered LO to encode either (a) a minimal set of features defining each category or (b) the set of features most commonly associated with each category; either of these would presumably make the average pattern evoked by each category less susceptible to noise resulting from visual idiosyncrasies among the inherently limited exemplar set used in this, or any, experiment.

A final possible explanation for enhanced sensitivity to pair content in LOC during the RPS task is explicit coding of interobject associations. A recent study (Kim & Biederman, 2011) found that BOLD responses in LO to pairs of objects that commonly interact with each other (e.g., a bird and birdhouse) are greater when the objects are shown actually interacting than when they are shown side by side with empty space between them. Although the objects used in the present study are not commonly associated with each other outside the context of the RPS game, the rules of the game are of course rooted in a set of distinct interactions: a rock smashes scissors, scissors cut paper, and paper covers a rock. Even if category-related information were encoded identically during the repetition-detection and RPS tasks, activation by the latter of additional neurons encoding the interactions themselves could plausibly enhance the discriminability of pair-evoked patterns.

If the improvement in sensitivity to pair content is attributable to the relevance of object identity to the RPS task, as opposed to nonspecific arousal effects, why then was sensitivity to object position not similarly enhanced despite its similar importance to the task? The 3.4° separation used in this study was unlikely to have simply saturated LO sensitivity to configural changes, based on a previous related study, which found increasing pattern discriminability for configuration changes for pairs with up to 6° between objects (MacEvoy & Yang, 2012). One possibility therefore is that the RPS task did, in fact, improve LO position tuning, but only along the task-relevant horizontal dimension, leaving sensitivity to the irrelevant elevation dimension unaltered. While such task-dependent anisotropy in spatial tuning is an intriguing possibility,
the fact that the RPS task produced an almost identical improvement in sensitivity to content (an equal determinant of configuration sensitivity) suggests that the more parsimonious explanation for the configuration effect lies with changes in codes for object form. Thus the most likely explanation for the absence of an improvement in position sensitivity may simply be that the RPS task did not require it. The “default” position tuning of LO that was already present during the repetition-detection task may have been sufficient to complete the RPS task, which indeed only required participants to know whether objects were in the left or right hemifields. Note that this possibility does not trivialize the observation that the RPS task enhanced sensitivity to pair content; rather, it instead suggests flexibility in the apportionment of coding efficiency that favors dimensions for which tuning improvements benefit a task, while having no effect on the representations of those dimensions for which tuning is already sufficient. Moreover, it leaves open the possibility that a task requiring greater precision in the discernment of the positions of specific objects might yet reveal flexibility in position tuning as well.

Conclusions

In summary, although a task requiring attention to the relative positions of objects did produce a marginal improvement in the sensitivity of LO to the spatial configurations of object pairs, this effect did not appear to reflect an increase in the precision of spatial tuning of form-encoding populations. Instead, it appears to have arisen from a specific improvement in the sensitivity of LO to the category membership of pair contents, a result which has not been previously reported. Further experiments will be needed to assess whether this stems from a shift towards categorical neural representations of visual features or activation of populations encoding implicit interactions between objects.

Keywords: object recognition, lateral occipital complex, fMRI, attention

Acknowledgments

This research was funded by Boston College.

Commercial relationships: none.
Corresponding author: Sean P. MacEvoy.
Email: sean.macevoy.1@bc.edu.
Address: Department of Psychology, Boston College, Chestnut Hill, MA, USA.

References


Hsieh, P.-J., Vul, E., & Kanwisher, N. (2010). Recognition alters the spatial pattern of fMRI


