Please cite this article in press as: MacEvoy and Epstein, Decoding the Representation of Multiple Simultaneous Objects in Human Occipitotemporal Cortex, Current Biology (2009), doi:10.1016/j.cub.2009.04.020

Current Biology *19*, 1–5, June 9, 2009 ©2009 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2009.04.020

Report

Decoding the Representation of Multiple Simultaneous Objects in Human Occipitotemporal Cortex

Sean P. MacEvoy[1,2,*] and Russell A. Epstein[1,2]
[1]Department of Psychology
[2]Center for Cognitive Neuroscience
University of Pennsylvania
Philadelphia, PA 19104
USA

Summary

Previous work using functional magnetic resonance imaging has shown that the identities of isolated objects viewed by human subjects can be extracted from distributed patterns of brain activity [1]. Outside the laboratory, however, objects almost never appear in isolation; thus it is important to understand how multiple simultaneously occurring objects are encoded by the visual system. We used multivoxel pattern analysis to examine this issue, testing whether activity patterns in the lateral occipital complex (LOC) evoked by object pairs showed an ordered relationship to patterns evoked by their constituent objects. Applying a searchlight analysis [2] to identify voxels with the highest signal-to-noise ratios, we found that responses to object pairs within these informative voxels were well predicted by the averages of responses to their constituent objects. Consistent with this result, we were able to classify object pairs by using synthetic patterns created by averaging single-object patterns. These results indicate that the representation of multiple objects in LOC is governed by a response normalization mechanism similar to that reported in visual areas of several nonhuman species [3–6]. They also suggest a population coding scheme that preserves information about multiple objects under conditions of distributed attention, facilitating fast object and scene recognition during natural vision.

Results

Single- and Paired-Object Classification

In a block design, subjects viewed single objects in four categories (shoes, chairs, cars, or brushes), as well as object pairs containing objects from two of these categories. Several previous studies have shown that information about the category of viewed objects is present in distributed patterns of activity measured with functional magnetic resonance imaging (fMRI) [1, 7] and we first wished to replicate this finding as a means of validating the quality of our data. Figure 1A shows classification performance for single objects within standard functionally defined regions of interest (ROIs) (see Supplemental Experimental Procedures available online). Consistent with previous work, classification accuracy was significantly above chance in lateral occipital complex [LOC; two-tailed t test, $t(11) = 4.95$, $p = 0.0004$]. Classification accuracy was also above chance in the parahippocampal place area [PPA; $t(11) = 2.77$, $p = 0.018$] but not in the fusiform face area [FFA; $t(11) = 1.56$, $p = 0.15$] or a nonbrain ROI [$t(11) = 0.13$, $p = 0.89$]. (See Supplemental Results for additional classification analyses, including the impact of changes in stimulus position upon accuracy.)

We next assessed the accuracy of the classifier in distinguishing among object pairs (Figure 1B). For classification purposes, each unique object pair was treated as a distinct stimulus (e.g., chair+brush and car+brush were treated as different stimulus categories), producing six pairs from the pool of four object categories. Classification accuracy for pairs was significantly above chance in LOC [$t(11) = 4.68$, $p = 0.0007$] but not in the PPA [$t(11) = 1.04$, $p = 0.31$], FFA [$t(11) = 1.68$, $p = 0.12$], or the nonbrain ROI [$t(11) = 0.86$, $p = 0.40$]. These results, along with whole-brain maps of local classification accuracy (Figure S1), indicate that activity patterns in LOC reliably discriminate among object pairs as well as among single objects.

Relationship of Paired-Object to Single-Object Responses

Do LOC patterns evoked by pairs bear any relationship to patterns evoked by their constituent objects? To answer this question, we first assessed the ability of a linear model to explain responses to object pairs [3, 5]. For each voxel, we performed a linear regression of the responses to pairs against the sum of responses to their constituent objects. This procedure is illustrated in Figure 2A for a voxel with a strong linear relationship between responses to pairs and to single objects ($R^2 = 0.96$).

Many voxels had much lower R^2 values, which could have reflected the impact of noise on a linear relationship, a nonlinear relationship, or no relationship at all. To differentiate between these possibilities, we used a searchlight classification technique to identify local voxel clusters that carried information about stimulus identity (see Experimental Procedures). We reasoned that searchlight clusters that most accurately differentiated among object pairs would contain voxels that were the most instructive of the "true" relationship between responses to pairs and constituent single objects. Therefore, if a linear model provides a good description of this relationship, we would expect to see R^2 increase as a function of searchlight classification accuracy. (See Supplemental Results and Discussion for a detailed treatment of this approach.)

Figure 2B plots median R^2 within each LOC searchlight cluster as a function of cluster classification rank for one subject. (We used classification rank, rather than raw classification accuracy, as the independent variable in order to facilitate averaging data across subjects, among whom overall classification accuracy varied.) For this subject, there was a clear trend toward higher R^2 values as classification accuracy improved. This relationship was also apparent in R^2 averaged across subjects (Figure 2C). To quantify this trend, we computed correlation coefficients between R^2 and classification rank within LOC for each subject. All subjects had positive correlation coefficients and all but two were significantly greater than zero at a $p < 0.05$ threshold. Across subjects, mean correlation coefficients were significantly above zero [mean = 0.33, $t(11) = 6.32$, $p = 0.00006$]. From the positive relationship between R^2 and classification rank, we infer that

*Correspondence: macevoy@gmail.com

Please cite this article in press as: MacEvoy and Epstein, Decoding the Representation of Multiple Simultaneous Objects in Human Occipitotemporal Cortex, Current Biology (2009), doi:10.1016/j.cub.2009.04.020

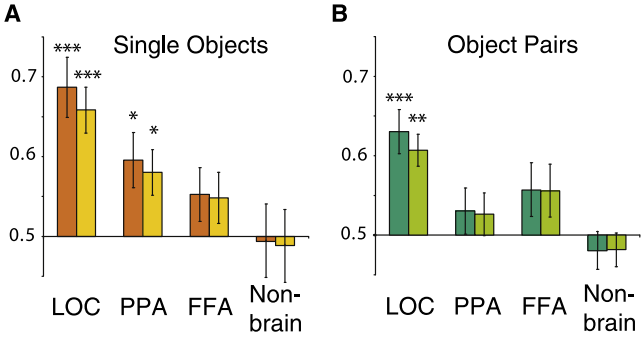

Figure 1. Multivoxel Pattern Classification Accuracies for Single Objects and Object Pairs

(A) Classification accuracy was significantly above chance (0.5) for the four single-object categories in both LOC and the PPA.

(B) Classification accuracy for the six possible category pairs was significantly above chance only in LOC.

Patterns were averaged across stimulus position (singles) or configuration (pairs) prior to classification. Dark-hued bars represent classification accuracy based on all voxels within each ROI, and lighter-hued bars represent accuracy for ROIs matched in voxel count to the smallest ROI for each subject. Asterisks denote significance of difference from chance performance (*$p < 0.05$; ***$p < 0.001$). Data are for 12 subjects. Error bars represent standard error of the mean (SEM).

responses to object pairs are well approximated by a linear combination of responses to single objects. (Similar analyses for the PPA, FFA, and retinotopic cortex can be found in the Supplemental Results.) A permutation-based control analysis demonstrated that this relationship was not a trivial outcome of voxel selection (i.e., "peeking;" see Supplemental Results).

Although R^2 captures the quality of a linear relationship, it does not specify its parameters. To understand whether voxels in LOC obeyed any specific linear relationship between pair and single-object responses, we examined the slope terms returned by the linear regressions described above. Figure 2D illustrates the relationship between classification rank and median slope for each searchlight position for one subject, and Figure 2E plots the same relationship averaged over all subjects. As with R^2, median searchlight slopes increased as classification accuracy improved. More importantly, slope values among high performing clusters fell close to 0.5, indicating that pair responses were approximately the average of responses to their constituent single objects. This result echoes a previous finding by Zoccolan et al. [5] that neuronal responses in macaque inferotemporal (IT) cortex to pairs of objects are well predicted by the average of the single-object responses. Although the terminal slope value in Figure 2E was 0.62, this value was not significantly different from 0.5. Terminal slope values for LOC were fairly consistent across subjects, with 8 of 12 subjects' values falling between 0.35 and 0.65. Furthermore, analysis of the distribution of residual error between actual pair responses and regression lines indicated that these results are more consistent pair responses that were simple, rather than weighted, averages of responses to constituent single objects (see Supplemental Results).

Linear regression returns an intercept term in addition to slope, which was not significantly different from zero in the top 30 LOC searchlight clusters [$t(11) = 0.76$, $p = 0.45$]. Thus, the responses to object pairs were truly the averages of responses to single objects, without any additional offset reflecting systematic differences in overall activity evoked by pairs and single objects.

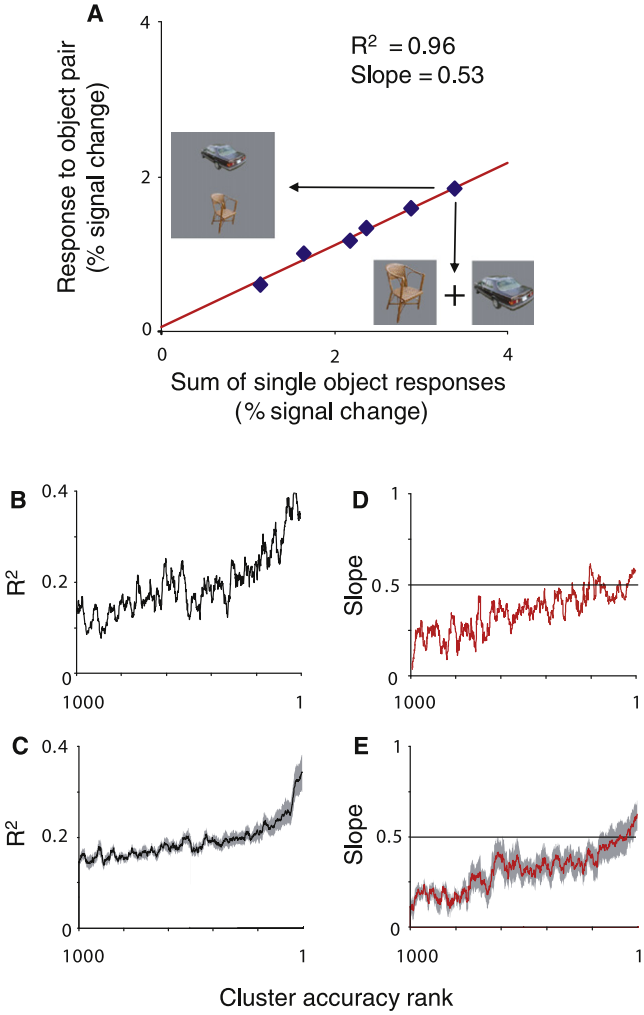

Figure 2. Relationship between Single- and Paired-Object Responses in LOC

(A) Responses of a single voxel to each of the six object pairs, plotted against the responses to the sum of responses evoked by each pair's constituent objects. For this voxel, pair responses showed a tight linear relationship ($R^2 = 0.96$) to single-object responses.

(B) Median R^2 within LOC searchlight clusters plotted as a function of each cluster's rank in classifying object pairs, for one subject. Data were smoothed with a 20-bin mean filter. R^2 increased with classification rank, suggesting that a linear model provides a good prediction of pair responses as noise is reduced.

(C) Same as (B), averaged across all subjects. Missing data for subjects with fewer than 1000 searchlights were ignored when the computing the average curve. Shaded regions fall within SEM.

(D) Median slope within searchlight clusters for one subject plotted as function of searchlight accuracy rank, smoothed with a 20 bin mean filter. Regression slopes fell close to 0.5 for the highest ranked searchlight clusters.

(E) Same data as in (D), averaged across subjects, with the same conventions as in (C).

Classification Using Synthetic Patterns

The preceding analyses suggest that we may approximate the responses of LOC voxels to object pairs as the averages of responses evoked by their constituent objects. To test this assertion, we repeated the pair pattern classification procedure but replaced pair patterns in one half of the data with "synthetic" patterns that were the averages of patterns

Please cite this article in press as: MacEvoy and Epstein, Decoding the Representation of Multiple Simultaneous Objects in Human Occipitotemporal Cortex, Current Biology (2009), doi:10.1016/j.cub.2009.04.020

evoked by the corresponding single objects. Patterns were limited to voxels that fell within the 30 highest performing searchlights in terms of pair classification, which typically afforded the highest average classification performance across subjects (Figure S4). It is critical to note that although these voxels were selected on the basis of high pair classification in their searchlights, this criterion was completely independent of the responses to single objects that were used to construct synthetic patterns.

Classification accuracies using synthetic patterns are shown in Figure 3A. At a rate well above chance [$t(11) = 8.54$, $p < 0.00001$], the classifier was able to correctly identify patterns evoked by object pairs based on comparison to synthetic response patterns derived by averaging the single-object responses within each voxel. Although classification based on these synthetic patterns was not as accurate as classification based on actual pair patterns, it was significantly more accurate [$t(11) = 5.45$, $p = 0.0002$] than classification based on a set of "MAX" function synthetic patterns generated by taking the higher of each voxel's responses to the two single objects comprising each pair [8]. This is consistent with the idea that pair responses reflect linear rather than nonlinear combinations of single-object responses.

Our ability to classify pairs from single-object patterns suggests that inverting the operation should allow us to decode the identities of single objects from the pattern evoked by a pair. Reddy and Kanwisher [7] found that classification accuracy for single objects was markedly degraded in LOC when a second object was present. The origin and nature of this "clutter cost" was unclear, however. Was information about the identity of objects actually lost? Or did the "cost" simply reflect the joint representation of both objects? Under the second scenario, we should be able to recoup clutter costs through appropriate decoding of patterns evoked by pairs.

We first assessed the impact of clutter in our own data by measuring classification accuracy for single objects within pairs. A correct classification decision was recorded when the Euclidean distance between the pattern evoked by a pair and the pattern evoked by one of its component objects (the "target" object for the purposes of classification only) was less than the distance between the pair pattern and the pattern for a comparison object not in the pair. Consistent with Reddy and Kanwisher [7], accuracy for single objects in pairs was significantly lower than accuracy for single objects by themselves (Figure 3B) in LOC [$t(11) = 4.27$, $p = 0.0013$], reflecting a substantial clutter cost.

To recover this clutter cost, we assumed that patterns evoked by object pairs were the averages of patterns evoked by their constituent objects. Accordingly, to extract the pattern evoked by a target object from a pair response, we subtracted a half-scaled version of the pattern evoked by the nontarget object and multiplied the resulting pattern by two. Applying this treatment produced a significant improvement in classification [Figure 3B; $t(11) = 4.02$, $p = 0.002$]. Linear decoding via this approach recovered an average of 48% of clutter costs associated with the presence of a second object. This result confirms the claim that information about each individual object is embedded in patterns evoked by object pairs.

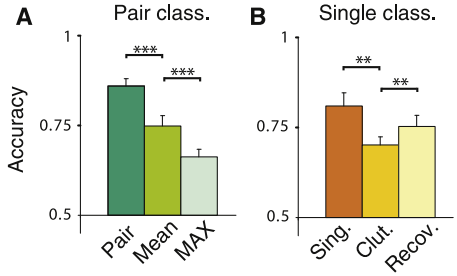

Figure 3. Synthetic Patterns Decode Patterns Evoked by Object Pairs in LOC

(A) Performance in classifying patterns evoked by pairs using actual pair patterns (Pair), synthetic patterns derived the means of single-objects patterns (Mean), or a MAX-function combination of single-object patterns (Max). Although classification accuracy for the mean predictor was not as high as for actual pairs, it was significantly higher than accuracy for the nonlinear MAX predictor.

(B) Performance for classifying single-object patterns (Sing.) was significantly reduced when a second, cluttering object appeared simultaneously (Clut.). A large portion of this clutter cost was recovered by assuming that pair patterns were the average of their component object patterns and by linearly decomposing responses to pairs accordingly (Recov.). Patterns included all voxels that fell within the top 30 clusters ranked in terms of pair classification rank. Across subjects, this corresponded to an average of 154 voxels. Error bars represent SEM.

Discussion

The principal finding of this study is that under conditions of distributed attention, voxelwise patterns of activity in object-selective cortex evoked by pairs of objects are the average of the patterns evoked by the individual component objects. Consistent with this result, pair patterns could be decoded with high accuracy by reference to synthetic patterns generated by averaging the single-object responses. Conversely, subtraction of an appropriately-scaled version of the voxel pattern evoked by one object of a pair recovered the pattern evoked by the second object.

This work builds on and extends two previous findings. First, Zoccolan et al. [5] demonstrated that responses of object-selective neurons of macaque area IT to pairs of objects were precisely predicted by the average of responses to their constituent objects. Our results demonstrate that a similar averaging rule applies to human LOC. Second, Reddy and Kanwisher [7] demonstrated a clutter cost for classification of single, focally attended objects when a second, unattended stimulus was present. Here we demonstrate that when the two objects are equally attended, a substantial portion of this cost for one object can be recouped if the response pattern to the second object is known.

These results potentially provide important insights into how visual recognition might proceed in the real world. The fact that objects in natural scenes almost always appear amidst other objects presents both a challenge and an opportunity for the visual system. The challenge is to identify single objects even when they are surrounded by the clutter of other stimuli. Attentional mechanisms might help solve this problem by boosting up the neural response to attended objects while suppressing the neural response to unattended objects [9–11]. However, this suppression of unattended object response can potentially negate an important informational opportunity. Specifically, the multiple objects within the scene might, if considered together, convey information about the "gist" or "context" of the scene [12–15]. Behavioral studies indicate that humans can indeed extract this gist information very rapidly [12, 16]; furthermore, observers can report the identities of objects within a scene even after very brief

presentation times that are unlikely to permit attention to be moved serially from object to object [17]. Our results suggest a way in which the visual system might accomplish this feat. In particular, if the pattern evoked by a multiple-object scene is linearly related to the patterns evoked by its constituent objects, then the gist might correspond simply to an initial hypothesis about the set of objects contributing to this overall pattern and a judgment about the category of scene that is most likely to contain such objects. Indeed, if there is a lawful relationship between the representations of a whole scene and of its component objects, then the same neural system can be use to represent both.

This reasoning explains why it would be advantageous for the visual system to maintain a linear relationship between single- and multiple-object responses, but it does not explain why the voxel patterns evoked by pairs resemble the average of single-object patterns. In its adherence to the mean, LOC appears to obey rules similar to those that have been described previously in a variety of visual areas in nonhuman species [3, 4, 6, 18–20] and which have traditionally been explained as an outcome of competition between stimuli for limited neural bandwidth [10, 11, 19, 21]. Our results suggest an alternative framing of this phenomenon in which response averaging reflects a normalization process that actively supports the coding of multiple simultaneous objects by avoiding the problems presented by saturation of neural responses [22]. Because individual neurons have finite firing rates, pure summation of responses to multiple objects runs the risk of driving some neurons to saturation, particularly those that respond vigorously to both objects. Once this happens, the population response to a pair of objects is no longer a linear combination of the patterns evoked by each object by itself and information about the identity of each object is lost. By scaling population responses by the number of stimuli present, normalization helps avoid this problem by ensuring that response saturation cannot be reached.

The presence of multistimulus normalization in LOC might also provide a window into its functional organization. Whenever response normalization has been found in nonhuman visual areas, such as with multiple oriented contours in V1, V2, and V4 [3, 4, 18, 19, 23], directions of motion in middle temporal and medial superior temporal cortex [6, 20], or shape in IT [5], the simultaneously presented stimuli have differed along some dimension that is "mapped" across the surface of the area under study (i.e., the individual stimuli presented by themselves activate spatially distinct clusters of neurons [24–27]). We speculate that this sort of mosaic-like organization might be a prerequisite to multistimulus normalization. If so, our results provide additional evidence that LOC neurons are clustered according to shape or category. (Indeed, such functional clustering might be necessary for multivoxel pattern analyses to work in the first place [28, 29].)

Finally, our data revealed two additional novel and somewhat surprising phenomena. First, the LOC territory that best encoded object pairs was largely identical to the LOC territory that best encoded single objects (Figure S1). In contrast, the PPA did not encode object pairs as reliably as LOC even though it did encode information about single objects. These findings are consistent with previous claims that LOC rather than the PPA is the primary region involved in encoding object identity information, even when more than one object is present in a scene [30]. Second, LOC response patterns did not distinguish between different spatial configurations of a pair (i.e., shoe over brush was indistinguishable from brush over shoe). This suggests that when attention is distributed evenly across a scene, object identity is encoded independently of object location in the ventral stream [31].

Experimental Procedures

Stimulus and Task

Stimuli were 60 photographic images (1.7° square) of common objects from four categories (brushes, cars, chairs, and shoes) with all background elements removed. Stimuli were presented in 15 s blocks (see Supplemental Experimental Procedures). In single-object blocks, 15 exemplars from the same object category were presented one at a time at a single screen position that was centered either 1.7° above or below the fixation point. In paired-object blocks, 15 exemplars from two categories (30 in total) were presented two at a time, with exemplars from one category appearing in the top screen position and exemplars from the other category appearing in the bottom screen position. Within each scan run, each object category was presented twice in the single-object condition (once in the upper screen position and once in the lower screen position) and each category pairing was shown twice (corresponding to the two possible spatial configurations; e.g., top brush/bottom chair and top chair/bottom brush).

To ensure that attention was paid equally to all objects, we required subjects (n = 12) to perform a one-back repetition detection task while maintaining central fixation. In paired-object blocks the repetition could occur at either stimulus location, forcing subjects to attend to both.

Data Analysis

Following standard preprocessing, fMRI data were passed to a general linear model implemented in VoxBo, from which voxelwise beta values associated with each stimulus condition were extracted (see Supplemental Experimental Procedures). Multivoxel pattern classification was implemented with custom code written in MATLAB and using an algorithm similar to Haxby et al. [1]. In brief, response patterns were extracted for each ROI from each of the six experimental scans. Data were then divided into halves (e.g., even runs versus odd runs) and the patterns within each half were averaged. A "cocktail" mean pattern (consisting of the average pattern across all stimuli) was calculated separately for each half of the data and then subtracted from each of the individual patterns before classification. Separate cocktails were computed for single objects and paired objects. No pattern normalization was applied at any point.

Pattern classification proceeded through a series of pairwise comparisons between stimulus conditions. Correct classification decisions were recorded when the Euclidean distance between the patterns evoked by condition A in opposite halves of the data was shorter than between condition A and condition B in opposite halves of the data. This procedure was repeated for every possible stimulus pairing, and correct decisions were accumulated across every possible binary split of the six scan runs. Additional analysis showed that the Euclidean distance metric produced classification accuracies similar to a correlation-based classifier [1].

Searchlight voxel selection [2] was implemented with custom MATLAB code. For each voxel, we defined a spherical mask that included all other voxels within a 5 mm radius. Searchlight clusters near the cortical surface were truncated to ensure that only voxels within the brain were included. Similarly, when searchlights were used on predefined ROIs, searchlight masks were truncated where necessary so that only voxels within the ROI were included.

Supplemental Data

Supplemental Data include Supplemental Results, Supplemental Discussion, Supplemental Experimental Procedures, and four figures and can be found with this article online at http://www.cell.com/current-biology/supplemental/S0960-9822(09)00980-4.

Acknowledgments

This work was supported by National Institutes of Health grant EY-016464 to R.A.E.

Received: December 17, 2008
Revised: April 6, 2009
Accepted: April 7, 2009
Published online: May 14, 2009

Please cite this article in press as: MacEvoy and Epstein, Decoding the Representation of Multiple Simultaneous Objects in Human Occipitotemporal Cortex, Current Biology (2009), doi:10.1016/j.cub.2009.04.020

References

1. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425–2430.
2. Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. Proc. Natl. Acad. Sci. USA *103*, 3863–3868.
3. MacEvoy, S.P., Tucker, T.R., and Fitzpatrick, D. (2009). A precise form of divisive suppression supports population coding in primary visual cortex. Nat. Neurosci. *12*, 637–645.
4. Luck, S.J., Chelazzi, L., Hillyard, S.A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J. Neurophysiol. *77*, 24–42.
5. Zoccolan, D., Cox, D.D., and DiCarlo, J.J. (2005). Multiple object response normalization in monkey inferotemporal cortex. J. Neurosci. *25*, 8150–8164.
6. Recanzone, G.H., Wurtz, R.H., and Schwarz, U. (1997). Responses of MT and MST neurons to one and two moving objects in the receptive field. J. Neurophysiol. *78*, 2904–2915.
7. Reddy, L., and Kanwisher, N. (2007). Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr. Biol. *17*, 2067–2072.
8. Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nat. Neurosci. *2*, 1019–1025.
9. Posner, M.I. (1980). Orienting of attention. Q. J. Exp. Psychol. *32*, 3–25.
10. Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. *18*, 193–222.
11. Kastner, S., De Weerd, P., Desimone, R., and Ungerleider, L.C. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science *282*, 108–111.
12. Potter, M.C. (1975). Meaning in visual search. Science *187*, 965–966.
13. Treisman, A. (2006). How the deployment of attention determines what we see. Vis. Cogn. *14*, 411–443.
14. Oliva, A., and Torralba, A. (2007). The role of context in object recognition. Trends Cogn. Sci. *11*, 520–527.
15. Schyns, P.G., and Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. Psychol. Sci. *5*, 195–200.
16. Biederman, I. (1972). Perceiving real-world scenes. Science *177*, 77–80.
17. Fei-Fei, L., Iyer, A., Koch, C., and Perona, P. (2007). What do we perceive in a glance of a real-world scene? J. Vis. *7*, 10.
18. Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science *229*, 782–784.
19. Reynolds, J.H., Chelazzi, L., and Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. J. Neurosci. *19*, 1736–1753.
20. Britten, K.H., and Heuer, H.W. (1999). Spatial summation in the receptive fields of MT neurons. J. Neurosci. *19*, 5074–5084.
21. Miller, E.K., Gochin, P.M., and Gross, C.G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. Brain Res. *616*, 25–29.
22. Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. Vis. Neurosci. *9*, 181–197.
23. Carandini, M., Heeger, D.J., and Movshon, J.A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. J. Neurosci *17*, 8621–8644.
24. Ghose, G.M., and Ts'o, D.Y. (1997). Form processing modules in primate area V4. J. Neurophysiol. *77*, 2191–2196.
25. Fujita, I., Tanaka, K., Ito, M., and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. Nature *360*, 343–346.
26. Albright, T.D., Desimone, R., and Gross, C.G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. J. Neurophysiol. *51*, 16–31.
27. Hubel, D.H., and Livingstone, M.S. (1987). Segregation of form, color, and stereopsis in primate area 18. J. Neurosci. *7*, 3378–3415.
28. Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. *8*, 679–685.
29. Haynes, J.D., and Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. *8*, 686–691.
30. Epstein, R.A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn. Sci. *12*, 388–396.
31. Evans, K.K., and Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? J. Exp. Psychol. Hum. Percept. Perform. *31*, 1476–1492.

Supplemental Data

Decoding the Representation of Multiple Simultaneous Objects in Human Occipitotemporal Cortex

Sean P. MacEvoy and Russell A. Epstein

Department of Psychology and Center for Cognitive Neuroscience, University of Pennsylvania, Philadelphia PA 19104 USA

Supplemental Material

Supplemental Experimental Procedures

Subjects

Twelve subjects (5 female, aged 19 to 30 years) with normal or corrected-to-normal vision were recruited from the University of Pennsylvania community and gave written informed consent in compliance with procedures approved by the University of Pennsylvania Institutional Review Board. Subjects were paid for their participation. An additional subject was also scanned, but was excluded from the study prior to data analysis because of excessive head motion.

MRI acquisition

Scans were performed at the Center for Functional Neuroimaging at the University of Pennsylvania on a 3T Siemens Trio scanner equipped with a Siemens body coil and an eight-channel multiple-array Nova Medical head coil. Structural T1* weighted images for anatomical localization were acquired using a 3D MPRAGE pulse sequences (TR = 1620 ms, TE = 3 ms, TI = 950 ms, voxel size = 0.9766 x 0.9766 x 1mm, matrix size = 192 x 256 x 160). T2* weighted scans sensitive to blood oxygenation level-dependent (BOLD) contrasts were acquired using a gradient-echo echo-planar pulse sequence (TR = 3000ms, TE = 30ms, voxel size = 3x3x3mm, matrix size = 64 x 64 x 45). Visual stimuli were rear projected onto a mylar screen at the head end of the scanner bore with an Epson 8100 3-LCD projector equipped with a Buhl long-throw lens and viewed through a mirror affixed to the head coil. The entire projected field subtended 22.9 x 17.4° and was viewed at 1024 x 768 pixel resolution.

The scanning session for each subject consisted of six experimental scans and two functional localizer scans. Experimental scans were 6 minutes 30 seconds in length,

and were divided into 20 18-second stimulus blocks, with additional 15-second fixation periods at the beginning of each scan and after every fourth stimulus block. Localizer scans were 6 minutes 15 seconds long and were divided into blocks during which subjects viewed color photographs of scenes, faces, common objects, and scrambled objects presented at a rate of 1.33 pictures per second as described previously [S1].

MRI 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, spatially normalized to the Montreal Neurological Institute (MNI) template. Data for localizer scans were spatially smoothed with an 8 mm FWHM Gaussian filter; all other data were left unsmoothed. Data were analyzed using the general linear model as implemented in VoxBo (www.voxbo.org) including an empirically-derived 1/f noise model, filters that removed high and low temporal frequencies, regressors to account for global signal variations, and nuisance regressors to account for between-scan differences.

Regions of Interest

The lateral occipital complex (LOC) was defined by stronger responses ($t > 3.5$) to objects than to scrambled objects during the functional localizer scans. LOC was trimmed to exclude any voxels that had 1) significantly greater responses to scenes than objects or 2) had significantly greater responses to faces than to objects. This was the same procedure used to define object selective cortex by Reddy and Kanwisher [S2], to whose results we wished to compare our own.

The parahippocampal place area (PPA) was defined as voxels in the posterior parahippocampal/collateral sulcus region that responded more strongly ($t > 3.5$) to scenes than to common objects. To focus on voxels characterized by selectivity for scenes, the PPA was further refined to exclude any voxels with significantly higher responses to intact objects than to scrambled objects. (Voxels that had significantly higher responses to scenes than to objects *and* significantly higher responses to objects than to scrambled objects were not included in any ROI.) The fusiform face area (FFA) was defined by voxels responding more strongly to faces than to objects. We did not exclude voxels from the FFA that also met the criteria for inclusion in LOC since this often produced exceedingly small FFA voxel counts.

To provide a baseline from which to judge multi-voxel pattern classification accuracy, we also defined a non-brain ROI for each subject by selecting 100 contiguous voxels within a supraorbital portion of the skull.

Supplemental Results

Additional classification analyses

We performed several additional analyses to validate our classification results. First, to ensure that differences in classification accuracy between ROIs did not result

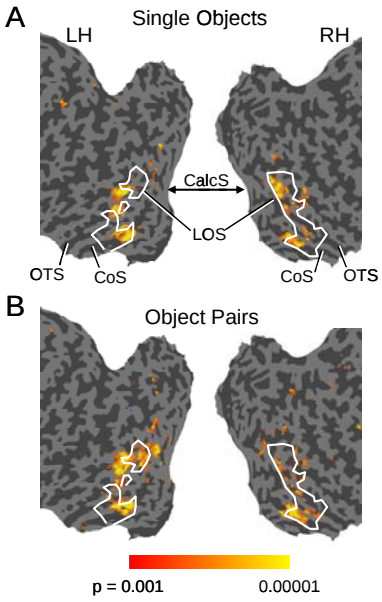

Figure S1. Group analysis of searchlight classification accuracy for single objects (A) and object pairs (B). Colored voxels are those whose associated searchlight clusters had classification accuracies above chance ($p < 0.001$, uncorrected), determined from random-effects analysis across searchlight volumes from all subjects. White contour indicates boundaries of LOC defined from a random-effects group analysis of functional localizer data, thresholded at $p < 0.01$, uncorrected. Maps for single objects and pairs are highly overlapping, indicating that the same clusters encoded information about both single objects and object pairs. CoS: collateral sulcus; OTS: occipitotemporal sulcus; LOS: lateral occipital sulcus CalcS: apex of calcarine sulcus.

from differences in ROI size, we repeated pattern classification for randomly-drawn voxel subsets matched in size to the smallest ROI for each subject, usually the FFA. The median voxel count across subjects for the smallest ROI was 53. Average classification accuracy was computed over 200 draws (light-shaded bars in Figure 1A). Accuracy remained significantly above chance in LOC ($t(11) = 5.54$, $p = 0.00006$) and the PPA ($t(11) = 2.80$, $p = 0.016$), indicating that the superior classification accuracy for single objects in LOC and the PPA relative to the FFA was not simply a result of greater ROI size. Similarly, pair classification accuracy in LOC ($t(11) = 5.25$, $p = 0.0002$) for voxel subsets was significantly above chance.

Second, we adopted a "searchlight" analysis approach to assess local classification accuracy for both single objects and object pairs throughout the whole brain [S3]. This analysis was particularly important given that LOC was defined based on responses to single objects in the localizer, leaving open the possibility that our ROI-based analysis may have overlooked voxels outside of LOC that preferentially carried information about object pairs. Supplemental Figure 1 shows the results of group random-effects analysis of local classification accuracy for both single objects and pairs. The maps show considerable overlap between regions exhibiting above-chance accuracy for pair and single object classification. Importantly, we observe no regions of high classification performance for pairs that are outside the boundaries of LOC *and* are absent from the map for singles. These results, together with our ROI analyses, indicate that LOC is the primary region of the brain involved in representing object pairs.

Position-specificity of classification

Our main interest in conducting these experiments was to quantify the relationship between responses evoked by object pairs and by their constituent objects. A potential

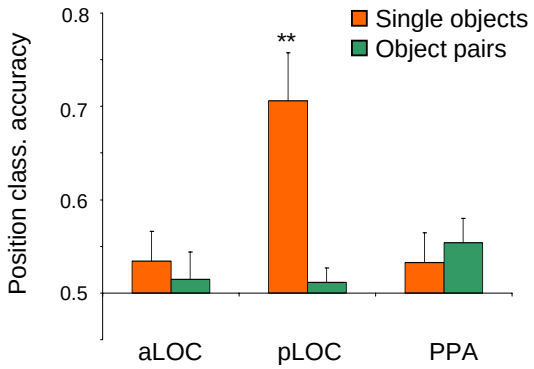

Figure S2. Spatial configuration classification accuracy. Patterns evoked by each single object and object pair were split by stimulus position (singles) or spatial configuration (pairs). Data are accuracy in classifying position/configuration in one half of the data, based on patterns for the same object or pair in the opposite half of the data. ROIs are defined in Supplemental Results. Patterns in aLOC did not discriminate between different positions for single objects or different configurations for pairs. Although pLOC patterns distinguished between single object positions, they did not discriminate between different spatial configurations of a pair, suggesting functional segregation of category and position information. Position/Configuration classification accuracy in the PPA was marginally above chance for object pairs, but not for single objects. Error bars are s.e.m.

complicating factor in any such analysis was the fact that object pairs necessarily contained objects at two positions in the visual field. If category-selective neurons were also position-selective, the relationship between the pattern evoked by a given object pair and the patterns evoked by its constituent objects might vary depending on whether the single objects were presented at the same locations as they appeared in the pair or at opposite locations. It was therefore important for us to understand the degree to which multi-voxel patterns varied with object position.

To assess the position-specificity of activity patterns in LOC, we measured our ability to classify the positions of single objects based on multi-voxel patterns. In other words, we assessed whether (for example) brush-on-top/nothing-on-bottom could be distinguished from nothing-on-top/brush-on-bottom. This was done by examining the patterns evoked by each kind of object (brushes, cars, chairs, and shoes) at each screen position (top, bottom) and assessing the extent to which the patterns elicited by an object of a given category in the two halves of the data were more similar when stimuli were presented at the same position than when they were presented in opposite positions. For these analyses, we divided LOC in each subject into anterior (aLOC) and posterior (pLOC) regions which corresponded to regions that have been previously termed posterior fusiform and LO, respectively [S4]. A recent study by Schwarzlose et al. [S5] showed that activity patterns evoked by single objects in pLOC carried significant information about object position, while activity patterns in aLOC were considerably less dependent upon objects' screen positions. In keeping with these results, we were able to classify single object positions in pLOC ($t(11) = 3.95$, $p = 0.002$) but not in aLOC ($t(11) = 1.04$, $p = 0.32$), nor in the PPA ($t(11) = 1.04$, $p = 0.23$), as shown in Supplemental Figure 2 (orange bars).

These data suggest a convergence between information about object category and position in pLOC. However, a more nuanced picture emerged when we examined the spatial-specificity of patterns evoked by pairs. Classification accuracy was not significantly above chance in either aLOC ($t(11) = 0.58$, $p= 0.62$) or pLOC ($t(11) = 0.75$, $p = 0.47$) when we attempted to distinguish between the two spatial configurations of each pair, for example between brush-on-top/chair-on-bottom and

chair-on-top/brush-on bottom (Supplemental Figure 2, green bars). These data indicate that, to the extent that neurons in pLOC encode information about object position, this information is carried by a population of neurons different from that which carries information about category. That is, while the responses of some neurons may discriminate between a stimulus at the top or bottom stimulus position, and other neurons may discriminate between a chair and shoe, neurons tuned jointly for position and object category do not appear to compose a significant fraction of the pLOC population.

In the PPA, we did observe a trend towards above-chance classification accuracy for spatial configurations of object pairs ($t(11) = 2.02$, $p = 0.065$). Although falling short of significance, this result is particularly interesting insofar as it suggests the possibility that some neurons in the PPA may be tuned to the spatial relationship between objects.

We also used the results of the searchlight classification analysis to identify any regions in the posterior half of the brain that reliably differentiated between spatial configurations of object pairs. Searchlight positions matching this criterion were relatively rare: while subjects had an average of 169 searchlights that could classify pair identity above an arbitrary threshold of 75%, they only had an average of 62 searchlights that could classify pair configuration at the same threshold. By comparison, an average of 2,356 searchlights in each subject could correctly identify the position of a single object at a rate greater than 75%. Group random-effects analysis showed no notable regions of high pair configuration accuracy outside LOC.

Relationship between pair and single object responses in other ROIs

We analyzed the relationship between responses to single objects and pairs for each voxel in the PPA, the FFA, and the non-brain ROI, using the same regression analysis detailed for LOC in the main text. In contrast to the findings in LOC, we saw no positive relationship between searchlight classification accuracy and R^2 in the FFA (mean correlation = 0.06, $t(11) = 0.77$, $p = 0.45$) or non-brain region (mean correlation = 0.095, $t(11) = 1.23$, $p = 0.24$). Taking note of the fact that some non-brain searchlights had, by chance, classification accuracies above 50%, the absence of any significant positive correlation demonstrates that higher R^2 values are not a trivial accompaniment to high classification scores, providing a control against which to judge the positive correlations we observed in LOC.

Interestingly, we did observe a significant positive correlation between classification rank and R^2 in the PPA (mean = 0.21, $t(11) = 3.25$, $p = 0.0077$). This result may seem surprising given the fact that whole-PPA patterns did not discriminate reliably among object pairs (Figure 1B). However, patterns derived from voxels in the top 30 searchlight positions in the PPA did exhibit pair classification accuracies significantly above chance ($t(11) = 5.17$, $p = 0.0002$). As in LOC, the slope of linear regressions between pair and summed single-object responses for the top 30 searchlight clusters in the PPA was 0.51, with upper and lower 95% confidence limits of 0.73 and 0.27, respectively. As in LOC, the average regression intercept term in the PPA did not differ significantly from zero ($t(11) = 1.59$, $p = 0.14$). By analogy to LOC, these results suggest that at least a subset of PPA voxels do contain information about object pairs,

and that responses of these voxels to pairs are linearly related to their responses to constituent objects.

With this in mind, we tested whether patterns evoked by object pairs in the PPA could be classified based on synthetic patterns created by averaging the patterns evoked by the constituent single objects. As in LOC, the classifier was able to correctly identify patterns evoked by object pairs based on these synthetic patterns at rates well above chance (PPA, $t(9) = 4.85$, $p = 0.0007$) when PPA data were restricted to the top 30 searchlight clusters.

We also performed regression analysis on retinotopic cortex, which was defined from functional localizer scans as regions with significantly higher responses to scrambled objects than intact objects. Classification accuracy in retinotopic cortex was not above chance for either single objects ($t(11) = 1.31$, $p = 0.22$) or pairs ($t(11) = 0.005$, $p = 0.99$). This indicates that the responses of voxels in early visual cortex did not vary reliably across single objects or pairs. Consistent with this, when we calculated regressions between voxel responses to single objects and object pairs there was no trend toward higher R^2 as a function of searchlight accuracy; across subjects, the mean correlation between R^2 and pair accuracy rank did not differ significantly from zero ($t(11) = 0.8693$, $p(11) = 0.40$).

Validation of Regression Analysis

Our regression analysis was based on the responses of each voxel to stimuli averaged across all six scans in our experiment. We used data from these same six scans to measure searchlight classification accuracy. In standard pattern classification studies, using the same data for both voxel selection and pattern classification ("peeking") can produce artificially elevated performance. It was therefore necessary to ensure that the positive relationship we observed between accuracy and regression R^2 (and slope) was not a trivial outcome of voxel selection.

Such a confound was unlikely in our analysis because voxels were selected entirely on the basis of searchlight classification performance for pairs without any consideration whatsoever of their classification performance for single objects. Because single and pair responses constitute different data sets collected at different time points, the fact that voxels within a given searchlight carry information about pairs should not bias us towards finding a linear (or any other) relationship between their responses to object pairs and responses to single objects if none really exists. In contrast, if such a relationship did exist, then we would expect that it should be most evident among voxel clusters with the highest signal-to-noise ratios (see "Voxel Selection" in Supplemental Discussion).

Nevertheless, we undertook several additional analyses to ensure that the positive correlations we observed between pair classification performance and both R^2 and slope in LOC were not artifacts of voxel selection. First, we recomputed both searchlight classification accuracies and regression analyses using data from separate scans. Classification accuracy was measured for data drawn from four of the six total scans, and regression analysis was performed on the remaining two. This process was repeated for each possible draw of 4 scans for each subject (e.g., scans 1,2,3,4 versus 5 and 6, then scans 1, 3, 4, 5 versus 2, 6, etc.). Data were accumulated across all 15

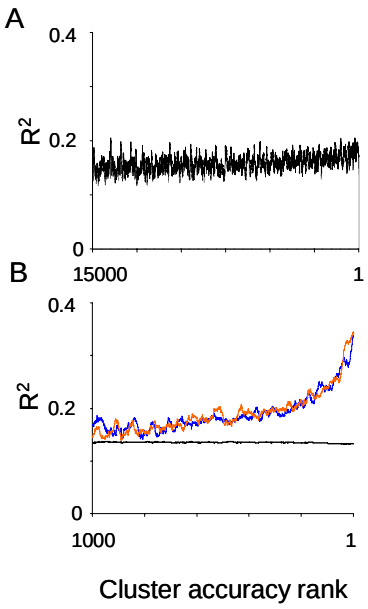

Figure S3. Validation of regression analysis. A) R^2 as a function of searchlight cluster pair classification rank when R^2 and classification accuracy are computed from separate data sets, averaged across subjects. Best-performing clusters are those with low rank numbers, at the right end of the x-axis. As in Figure 3 of the main text, R^2 showed a significant positive correlation to cluster accuracy. Ranking extends to higher numbers than in Figure 2 in the main text due to the 15 possible ways in which non-overlapping data sets could be drawn from 6 scans (see Supplemental Results). B) Comparison of relationship between R^2 and pair classification accuracy rank when R^2 is computed from original data (red) and after permuting single object condition labels (black). Permuting single-object labels completely abolishes the upward trend in R^2 as classification accuracy improves, indicating that this trend is not a trivial outcome of high classification accuracy. Also shown is R^2 as a function of single-object classification rank (blue), which does not differ meaningfully from R^2 based on pair rank.

possible draws for each subject, but searchlight classification accuracies from each draw were associated exclusively with the results of regression analysis from the corresponding two remaining scans.

Regression R^2 values for all searchlight positions across all draws were placed in a single pool for each subject and then ranked according to their associated classification accuracy, and the resulting sorted values were averaged across subjects (Supplemental Figure 3A). If the positive correlation we observed between R^2 and searchlight accuracy shown in Figure 2 were a result of peeking, we should see no such correlation when classification scores and regression analysis were performed on non-overlapping data sets. Instead, we observed a small but highly significant ($p < 10^{-6}$) positive correlation, qualitatively similar to the data shown in Figure 2. Further, 10 of 12 subjects still had correlation coefficients significantly greater than 0 ($p < 0.05$).

Though this result indicates that the positive correlation between searchlight accuracy and R^2 is not *solely* attributable to voxel selection, it does not eliminate the possibility that voxel selection made *some* contribution to the R^2 trend in Figure 2. To assess what bias, if any, was contributed by voxel selection, we recomputed our original regression analysis (which included data from all 6 scans for each subject) after randomly permuting the condition labels for single-object responses independently for each LOC voxel. It is critical to note that this permutation step in no way altered each searchlight's classification score, which was based solely on data for *pairs*. If the trend observed in Figure 2 were even in part a result of voxel selection, we should see a positive correlation between searchlight rank and R^2 even after label permutations. Instead, we find that R^2 from permuted regressions remains constant as a function of classification accuracy, showing no bias towards higher values as classification

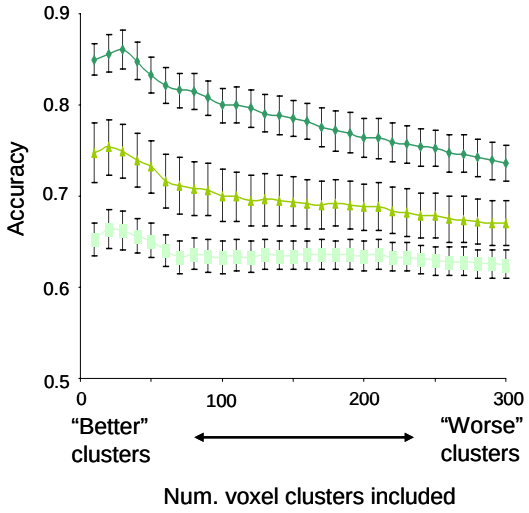

Figure S4. LOC classification accuracy for pairs and synthetic pairs as a function of pattern voxel count, averaged across subjects. Data are accuracy in classifying pair patterns based actual pair patterns (dark green), patterns derived from the means of single object patterns (light green), and patterns derived from a MAX-function combination of single object patterns (cyan). X-axis denotes the size of the pattern, expressed in terms of the number of searchlight positions included, binned in 10-searchlight increments. Searchlights were added in descending order of classification accuracy. Error bars are s.e.m.

improves. This is shown by the black line in Supplemental Figure 3B, which represents subject-averaged data after 1000 permutation cycles for each subject. Also plotted in Supplemental Figure 3B is R^2 as a function of pair classification accuracy rank (repeated from Figure 2B), as well as R^2 as a function of single object classification accuracy rank. These two functions are almost identical, further demonstrating that voxel selection did not bias our assessment of a linear relationship between single objects and object pairs.

We also wished to confirm that our regression analysis truly reflected pair responses that were the simple averages of the responses to their constituent objects, rather than weighted averages. To do so, we began by performing simulations that allowed us to examine the expected outcome of the linear regression analysis for a set of artificial voxels. In the first simulation, the responses of each voxel to pairs were the weighted averages (rather than simple average) of its single object responses, with the response to the more-preferred object in a pair weighted more strongly (i.e. weights for each object were proportional to their single-object responses). Linear regressions between these pair responses and the sum of their constituent object responses produced a unique pattern of predictive accuracy. Specifically, we found that residual error between pair response predicted by linear fits and actual pair responses was distributed unevenly across pairs, with the largest residuals attaching to pairs composed of objects with the largest differences in their single-object responses. This contrasted with the results of a a second simulation, in which pair responses were constructed from the simple averages of component responses (with noise added). In this case, regressions between pairs and the sum of their constituent objects had residuals that were uniformly distributed across pairs.

With this in mind, we next examined the distribution of residuals from linear regressions performed on real fMRI data from voxels within the 30 highest accuracy searchlights within each subject. For each voxel, we computed the squared error between the actual response to each object pair and the response predicted for that pair by linear regression. We then ranked these error terms by the magnitudes of the response differences between each pair's constituent objects. Based on the results of our simulations, we expected to see the largest residual error among pairs that ranked

high by this measure. However, a one-way repeated measures ANOVA showed no significant effect of this ranking on the magnitude of residual error for paired objects across subjects ($F(5,11) = 1.17$, $p = 0.34$). Thus we find no evidence that responses to pairs were weighted, as opposed to simple, averages of their constituent object responses. However, it is unclear whether this finding would generalize to a situation in which one of the two objects of a pair were "preferred" in terms of cortical specialization (e.g. face or house). In this case, we might expect the mean rule to break down, perhaps reflecting an absence of multistimulus normalization across separate domain-specific systems.

Supplemental Discussion

Voxel Selection

One challenge in assessing the ability of any model to characterize the relationship between single and pair responses is that of voxel selection. For at least two reasons it would be unreasonable to expect the responses of every voxel to vary reliably enough across the stimuli to allow us to discern a relationship between responses to different stimuli. The first reason is the extreme sparseness of any experimental stimulus set. Given the vast number of objects which the visual system is forced to encode, the odds are high that the four object categories we chose lie outside the tuning envelope of many neurons. Response variability among voxels corresponding to populations of such neurons will necessarily be dominated by noise, and will therefore defy any model. (It is worth noting that this challenge is not unique to fMRI. In single-unit studies of macaque IT, neurons are often excluded from further study if they do not respond to at least one element in an experimental stimulus set [S6, S7].) A second reason is partial-volume effects. Each voxel represents neural activity among a large and potentially heterogeneous population of neurons. Even though individual elements of these populations may differentiate reliably between stimuli, their aggregate activity may show considerably less selectivity. As such, response variability among these voxels will also tend to be dominated by noise.

In light of these factors, it is not surprising that, on average, 54% of LOC searchlight clusters and 69% of PPA clusters had pair classification accuracies below the 95th percentile of searchlight accuracies in the non-brain ROI. In other words, these LOC and PPA clusters had accuracies indistinguishable from chance, indicating that the responses of voxels within them were not strongly related to pair identity. Therefore, their pair responses could not be expected have any relationship, linear or otherwise, to responses evoked by single objects, and consideration of these voxels will tend to artificially reduce the predictive ability of any model. Given these factors, we used searchlight classification performance as an independent estimate of the signal-to-noise ratio of voxels within each searchlight, as explained in the Results.

Classification performance in ROIs outside LOC

We found that voxel patterns from regions outside the LOC often showed poor classification performance for single objects and object pairs. While patterns in PPA

differentiated between single objects, they did not differentiate between pair patterns at a rate above chance, and FFA patterns did not differentiate among either single objects or pairs.

These results might at first appear to contradict previous work, most notably by Haxby et al. [S8], showing that information about object identity could be gleaned from patterns in the FFA and PPA. Similar results were reported by Reddy and Kanwisher [S2]. However, both of these studies reported average classification accuracy across a stimulus set that included stimuli preferred by FFA and PPA as well; thus, above-chance accuracy for non-preferred objects may have been driven to a large extent by correct classifications between preferred and non-preferred objects (e.g., house versus brush), rather than among non-preferred objects (e.g., brush versus chair). Although Reddy and Kanwisher confirmed that patterns in both FFA and PPA did discriminate between shoes and cars, these were the only non-preferred objects that were used, making the generality of the result unclear. In contrast, Spiridon and Kanwisher [S9], using a stimulus set similar to ours (including objects in four categories not preferred by either FFA or PPA), found that patterns in the FFA and PPA did not reliably discriminate among four categories of single objects.

Given the fact that PPA patterns did reliably discriminate among single objects, it might seem somewhat surprising that they did not discriminate among object pairs (at least when whole-PPA patterns were examined; as noted above, subsets of voxels within the PPA were more discriminative). After all, object pairs might be taken to form a minimal scene, a stimulus class which robustly activates the PPA. It is worth noting, however, that there are at least three important differences between the real-world scenes that robustly activate the PPA and the object pairs shown here. First, real-world scenes contain fixed background elements such as walls and ground planes, which are believed to be critical for eliciting a strong PPA response [S10]. Second, the objects in the current study were not presented in their typical spatial relationships. Finally, the stimuli in the current experiment were equated for angular subtense even though they possessed very different real-world sizes. The close apposition of a brush and car of similar angular subtense, in an unnatural spatial relationship to each other, and in the absence of surrounding background information is unlikely to convey any meaningful sense of a scene.

Supplemental References

S1. Epstein, R.A., and Higgins, J.S. (2006). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. Cereb. Cortex.

S2. Reddy, L., and Kanwisher, N. (2007). Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr Biol *17*, 2067-2072.

S3. Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. Proc Natl Acad Sci U S A *103*, 3863-3868.

S4. Malach, R., Levy, I., and Hasson, U. (2002). The topography of high-order human object areas. Trends in Cognitive Sciences *6*, 176-184.

S5. Schwarzlose, R.F., Swisher, J.D., Dang, S., and Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. Proc Natl Acad Sci U S A *105*, 4447-4452.

S6. Zoccolan, D., Cox, D.D., and DiCarlo, J.J. (2005). Multiple object response normalization in monkey inferotemporal cortex. Journal of Neuroscience *25*, 8150-8164.

S7. Sheinberg, D.L., and Logothetis, N.K. (2001). Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. Journal of Neuroscience *21*, 1340-1350.

S8. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425-2430.

S9. Spiridon, M., and Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. Neuron *35*, 1157-1165.

S10. Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature *392*, 598-601.