

Object Ensemble Processing in Human Anterior-Medial Ventral Visual Cortex

Jonathan S. Cant and Yaoda Xu

Vision Sciences Laboratory, Department of Psychology, Harvard University, Cambridge, Massachusetts 02138

Our visual system can extract summary statistics from large collections of similar objects without forming detailed representations of the individual objects in the ensemble. Such object ensemble representation is adaptive and allows us to overcome the capacity limitation associated with representing specific objects. Surprisingly, little is known about the neural mechanisms supporting such object ensemble representation. Here we showed human observers identical photographs of the same object ensemble, different photographs depicting the same ensemble, or different photographs depicting different ensembles. We observed fMRI adaptation in anterior-medial ventral visual cortex whenever object ensemble statistics repeated, even when local image features differed across photographs. Interestingly, such object ensemble processing is closely related to texture and scene processing in the brain. In contrast, the lateral occipital area, a region involved in object–shape processing, showed adaptation only when identical photographs were repeated. These results provide the first step toward understanding the neural underpinnings of real-world object ensemble representation.

Introduction

Many everyday visual tasks require the encoding of single objects. Such object-specific processing has been the core of past neuroscientific research. For example, it has been shown that the lateral occipital area (LO), together with the posterior fusiform gyrus (pFs), processes the shapes of single objects (Malach et al., 1995; Grill-Spector et al., 1998; Kourtzi and Kanwisher, 2001) and that the parietal cortex is involved in individuating and selecting multiple objects for detailed processing (Xu and Chun, 2006, 2009). There are also many occasions when our visual system extracts summary statistics from a collection (ensemble) of objects without representing any specific object in great detail. Such object ensemble representation can aid rapid scene segmentation and guide object-specific processing. Yet despite their common occurrence, the neural mechanisms underlying ensemble processing remain poorly understood.

Here we investigated the neural representation for a simple yet omnipresent form of real-world object ensemble, namely, ensembles containing homogeneous and repeating objects such as leaves on a tree. Behavioral research has shown that observers can quickly extract average features of a homogeneous ensemble, such as its mean size, direction of motion, speed, orientation, and center location without encoding details of the individual objects composing the ensemble (Williams and Sekuler, 1984; Watamaniuk

and Duchon, 1992; Ariely, 2001; Parkes et al., 2001; Chong and Treisman, 2003; Alvarez and Oliva, 2008). Despite these advances, the neural underpinnings of object ensemble representation remain largely unknown. Moreover, behavioral studies have used displays containing simple geometric drawings (e.g., black dots) and have not directly investigated how real-world object ensembles are represented. Here we used photographs of real-world object ensembles and investigated where in the brain they may be represented.

Neuropsychological studies have documented a dissociation between shape and texture processing, such that, after bilateral LO damage, impaired shape processing but largely intact texture perception were observed (Humphrey et al., 1994; Goodale and Milner, 2004). Consistent with this finding, fMRI studies have shown that attention to an object's shape activates LO, whereas attention to texture activates the collateral sulcus (Peuskens et al., 2004; Cant and Goodale, 2007). Interestingly, this texture-sensitive collateral sulcus region overlaps to a large extent with the parahippocampal place area (PPA), a region involved in scene perception (Epstein and Kanwisher, 1998).

Although object ensembles contain individuated objects with closed contours and surface textures may not, ensembles and textures nonetheless both contain repeating structures with slight variations in features such as size, orientation, and color (Portilla and Simoncelli, 2000). Thus, although little is known about the neural representation of object ensembles, the similarity between ensembles and textures and the observed PPA activation during texture perception prompted us to choose PPA as the candidate region of interest (ROI) in which to investigate object ensemble processing in the brain. In four experiments, we used fMRI adaptation (Grill-Spector et al., 2006) to examine PPA's response to real-world homogeneous object ensembles. We also examined LO's response to object ensembles and included surface textures as stimuli in

Received June 30, 2011; revised March 5, 2012; accepted March 25, 2012.

Author contributions: J.S.C. and Y.X. designed research; J.S.C. performed research; J.S.C. analyzed data; J.S.C. and Y.X. wrote the paper.

This work was supported by National Science Foundation Grants 0719975 and 0855112 (Y.X.) and by a Natural Sciences and Engineering Research Council of Canada postdoctoral fellowship (J.S.C.). We thank Tao Gao, Sonia Poltoraski, Aaron Glick, and Iris Lee for their assistance in this study.

Correspondence should be addressed to Jonathan S. Cant, Vision Sciences Laboratory, Department of Psychology, Harvard University, William James Hall, Room 744, 33 Kirkland Street, Cambridge, MA 02138. E-mail: jcant@wjh.harvard.edu.

DOI:10.1523/JNEUROSCI.3325-11.2012

Copyright © 2012 the authors 0270-6474/12/327685-16\$15.00/0

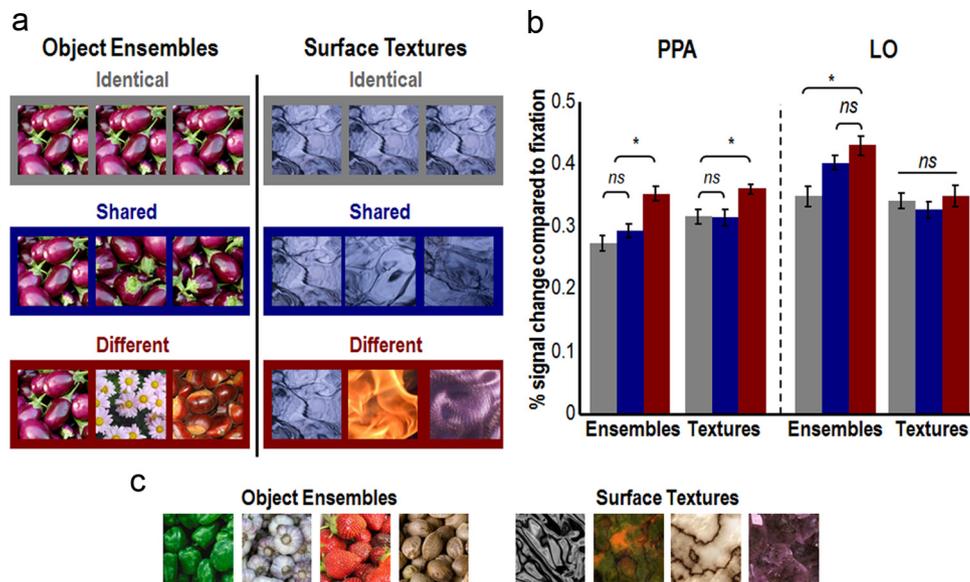


Figure 1. Example stimuli and results ($N = 12$) from experiment 1. **a**, Example stimuli used in the experiment. Based on common occurrence and ease of availability, 20 different living object ensemble images and 20 different nonliving texture images were used. In each trial, observers saw a sequential presentation of three images that were all identical (gray boxes), all different (red boxes), or shared object ensemble or surface texture statistics (blue boxes). To ensure attention to the images, observers were required to press a button on the disappearance of the third image in the sequence. **b**, Results from experiment 1. fMRI responses were extracted from independently localized object- (LO) and scene-sensitive (PPA) areas of cortex. PPA exhibited similar response patterns to both object ensembles and surface textures and showed equivalent levels of adaptation (i.e., a reduction in activation compared with the different condition) in the identical and the shared conditions when either object ensemble or surface texture statistics were repeated. In contrast, LO response patterns differed between object ensembles and surface textures, exhibiting an equivalent release from adaptation in the shared and different object ensemble conditions, where changes to local shape information are evident, but exhibited insensitivity to changes in surface textures due to a lack of closed contours in those images. Error bars represent within-subject SEs. ns, Not significant. **c**, Additional examples of stimuli used in experiment 1. * $p < 0.05$.

the first three experiments. Finally, we performed whole-brain analyses to validate our results.

Materials and Methods

Observers

Twelve paid observers (six women, six men; mean age, 25.75 years; range, 19–32 years) took part in experiment 1, 14 paid observers (two of whom took part in experiment 1; eight women, six men; mean age, 25.43 years; range, 20–33 years) took part in both experiments 2 and 3 (which were run during the same session), and 11 paid observers (1 of whom participated in experiment 1 and 4 of whom participated in experiments 2 and 3; four women, seven men; mean age, 28.45 years; range, 20–34 years) took part in experiment 4. Observers were recruited from the Harvard University community, and all were right handed, reported normal color vision and normal or corrected-to-normal visual acuity, had no history of neurological disorder, and gave their informed consent to participate in the study in accordance with the Declaration of Helsinki. The experiments were approved by the Committee on the Use of Human Subjects at Harvard University.

An additional female observer was tested in experiment 1, but the study was terminated halfway due to observer discomfort. An additional male observer was tested in experiments 2 and 3, but his data were excluded due to excessive head motion (>8 mm of translation). Finally, two additional female observers were tested but excluded from experiment 4 due to below chance level behavioral performance in the scanner (i.e., they performed below the chance level of 50% on multiple runs of a task where the group average was 95%).

Stimuli

Adaptation experiments. In experiment 1, based on common occurrence and ease of availability, we collected through the web full-color photographs of 20 different living object ensemble images and 20 different nonliving texture images (Fig. 1). All images subtended $12.5^\circ \times 12.5^\circ$ of visual angle (this applies to all images used in all subsequent experiments except where noted). In experiment 2, we removed the living and the nonliving distinction and collected through the web 20 new grayscale object ensemble images and 20 new grayscale texture images, each con-

taining 10 living and 10 nonliving images (see Fig. 3). Experiment 3 used the same stimuli as were used in experiment 2, but the images were presented in full color and the second image in each trial was approximately two-thirds the size of the first image (subtending $10.0^\circ \times 10.0^\circ$ of visual angle; see Fig. 4). Experiment 4 used object ensemble images only, created by photographing (using a D3000 digital SLR camera; Nikon) 10 different collections of black wooden beads (ordered on-line from Fire Mountain Gems, www.firemountaingems.com). The beads in each collection had the same shape, but the shape differed between collections. All beads were made of black painted wood, resulting in different object ensembles containing objects sharing the same texture/material (see Fig. 6). We ensured that the background of each bead image was uniformly white using Photoshop CS3 software (Adobe).

Object/scene localizer. Stimuli used to localize object- and scene-sensitive areas of cortex consisted of photographs of common objects (e.g., cars, chairs, food, and tools), various indoor and outdoor scenes (e.g., furnished rooms, buildings, city landscapes, and natural landscapes), male and female faces, and phase-scrambled versions of the common objects.

Ensemble/texture localizer. In addition to using the object/scene localizer, in experiments 2 and 3 we constructed a second localizer to directly identify ventral and lateral visual regions sensitive to the viewing of object ensembles and surface textures. We included both intact and phase-scrambled images. The intact images were the same as those used in the adaptation runs and consisted of full-color object ensembles and surface textures, with each category containing equal numbers of living and nonliving stimuli.

Apparatus

Stimulus presentation and the collection of behavioral responses (through a response pad placed in the observer's right hand) were controlled with a MacBook Pro (Apple) running MatLab (MathWorks) with Psychtoolbox extensions (Brainard, 1997; Pelli, 1997). Each image was rear projected with an LCD projector (Notevision XG-C465X, Sharp; resolution, 1024×768 pixels) onto a screen mounted behind the observer as he or she lay in the scanner bore. The observer viewed the images through a mirror mounted to the head coil directly above the eyes.

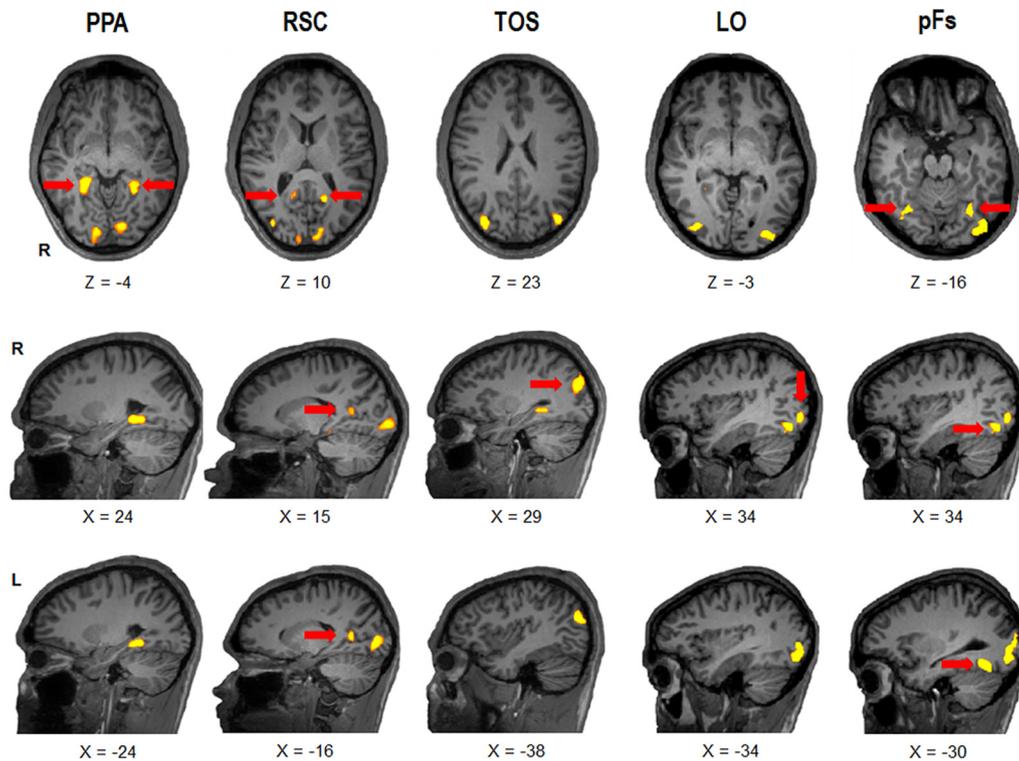


Figure 2. Examples of ROIs in individual observers. The scene-selective PPA (Talairach coordinates for the specific ROI examples shown x, y, z for right/left; $+24/-24, -40/-41, -3/-6$), RSC ($+15/-16, -53/-56, +14/+9$), and TOS ($+31/-38, -79/-77, +23/+22$) were defined by contrasting the activation for scenes against the activation for both faces and objects. The object-selective LO ($+31/-32, -77/-84, -3/-7$) and pFs ($+32/-29, -63/-61, -13/-16$) were defined by contrasting the activation for objects against the activation for scrambled objects. R, Right; L, left.

Imaging procedures

Adaptation experiments. In all experiments, we used a fast, event-related fMRI adaptation paradigm. In experiment 1, each trial lasted 6 s and contained a 500 ms fixation, three sequentially presented images (each consisting of a 200 ms image presentation and an 800 ms blank fixation), and a 2500 ms blank screen. Observers were asked to press a button as soon as the third image in the trial disappeared. The three images presented in each trial could be identical, could share object ensemble or texture features (i.e., different images of the same ensemble or texture) (Fig. 1), or could be completely different. Thus, there were six different experimental conditions (“identical,” “shared,” and “different” conditions for both the object ensemble and texture images). There were also 6 s blank fixation trials in which no images were presented. Trial order was pseudorandom and balanced for trial history (e.g., trials from all conditions, including fixation, were preceded and followed equally as often by trials from all other conditions, for one trial back and forward) (Kourtzi and Kanwisher, 2001; Xu and Chun, 2006). To further balance trial history, trial order was rotated among the conditions in different runs and among different observers. Each observer took part in five adaptation runs, each lasting 5 min 12 s with seven trials for each condition.

In experiments 2 and 3, each trial lasted 6 s and contained a 500 ms fixation, two sequentially presented images (each consisting of a 200 ms image presentation and an 800 ms blank fixation), and a 3500 ms blank screen. Observers were asked to categorize each trial as identical, shared, or different by pressing the appropriate button on the response pad. All other aspects of these experiments were identical to those of experiment 1. Experiment 4 was identical to experiments 2 and 3, except that only object ensemble images were used (see Fig. 6). Each observer took part in two adaptation runs in experiment 4, each lasting 6 min 42 s, with 16 trials for each condition.

Object/scene localizer. This localizer was used to identify the main ROIs in each observer, namely, LO and PPA, as well as three additional ROIs, the retrosplenial complex (RSC), the transverse occipital sulcus (TOS),

and pFs (Fig. 2 shows examples of these ROIs in single observers). As described by Kanwisher et al. (1997) and Epstein and Kanwisher (1998), a single run consisted of presenting four blocks each of scenes, faces, intact objects, and phase-scrambled objects with periods of fixation presented at the beginning, middle, and end of a run. Each stimulus block was 16 s long and contained 20 different images, each lasting 750 ms, and was followed by a 50 ms blank period. Each fixation block was 8 s long. There were two unique run orders, and no images were repeated in a given run. To ensure attention to the displays, observers fixated at the center and detected a slight spatial jitter, occurring randomly in 1 out of every 10 images. Each run lasted 4 min 40 s.

In experiment 1, all observers took part in four runs of this localizer, and in experiments 2 to 4, observers took part in three runs. Because the fMRI data analysis software we used allowed us to align functional data acquired across different sessions, when observers took part in two or more of the experiments, we were able to use the object/scene localizer from their earlier session for all subsequent sessions.

Ensemble/texture localizer. This localizer was used in experiments 2 and 3 to identify, in each observer, areas in the visual cortex that are selectively activated by object ensemble and surface texture images. The stimuli used in experiments 2 and 3 were used here. A single run consisted of presenting four blocks each of intact object ensembles, intact surface textures, and their phase-scrambled counterparts. All other aspects of this localizer were identical to those of the object/scene localizer, with the exception that each image in a block was presented for 500 ms and was followed by a 300 ms blank period. No images were repeated within any single block, but there were repetitions of images across blocks in a given run. All observers took part in three runs of this localizer.

Imaging parameters

This study was conducted on a 3.0 tesla Siemens MAGNETOM Tim Trio whole-body imaging MRI system at the Center for Brain Science, Harvard University (Cambridge, MA). A Siemens radiofrequency 32-channel head coil was used to collect BOLD weighted images (Ogawa et

al., 1992). For high-resolution anatomical images, T1-weighted 3D magnetization prepared rapid acquisition gradient echo sagittal slices covering the whole brain were collected [inversion time 1100 ms, echo time (TE) 1.54 ms, repetition time (TR) 2200 ms, flip angle 7°, 256 × 256 matrix size, 144 slices, 1.0 × 1.0 × 1.0 mm voxel size]. For the functional runs, a T2*-weighted echo-planar gradient echo pulse sequence (72 × 72 matrix size, field of view 21.6 cm) with TR of 1.5 s was used in experiments 1 to 4 (TE 29 ms, flip angle 90°, 208 volumes for experiments 1–3, 268 volumes for experiment 4). Another pulse sequence with TR of 2 s was used for the localizer runs (TE 30 ms, flip angle 85°, 140 volumes). Twenty-four 5 mm-thick (3 × 3 mm in-plane, 0 mm skip) slices parallel to the anterior and posterior commissure line were collected in all the functional runs.

Data analysis

fMRI data analysis. fMRI data were analyzed with Brain Voyager QX (Brain Innovation). Data preprocessing included slice acquisition time correction, 3D motion correction, linear trend removal, and Talairach space transformation (Talairach and Tournoux, 1988).

Data from both the object/scene and ensemble/texture localizers were analyzed using a general linear model (GLM), accounting for hemodynamic response lag (Friston et al., 1994). In the object/scene localizer, in accordance with Epstein and Kanwisher (1998), the PPA ROI was defined as a region in the collateral sulcus and parahippocampal gyrus whose activation was higher for scenes than for faces and objects (false discovery rate, $q < 0.05$; this threshold applies to all functional regions localized in individual observers). In addition, in accordance with Epstein and Higgins (2007), the RSC and TOS ROIs were defined as regions in retrosplenial cortex–posterior cingulate–medial parietal cortex and transverse occipital cortex, respectively, whose activations were higher for scenes than for faces and objects (Fig. 2). In accordance with Grill-Spector et al. (2000), the LO and pFs ROIs were defined as regions in the lateral occipital cortex near the posterior inferotemporal sulcus and the posterior fusiform gyrus–occipitotemporal sulcus, respectively, whose activations were higher for objects than for phase-scrambled objects. In the ensemble/texture localizer, areas sensitive to processing object ensembles and surface textures were identified as regions in collateral sulcus and parahippocampal gyrus as well as lateral occipital cortex, whose activations were higher for ensembles and textures than for phase-scrambled versions of these images.

Following the standard ROI-based analysis approach (Saxe et al., 2006), we overlaid the ROIs onto the data from our main adaptation experiments and extracted time courses from each observer. The activation levels for all conditions were then converted to the percentage of BOLD signal change from baseline by subtracting the corresponding activation from the fixation trials and then dividing by this value. Peak responses for each condition were obtained by collapsing the time courses for all of the conditions and then identifying the time point of greatest signal amplitude in the average response (Xu and Chun, 2006; Xu, 2010). This was done separately for each observer in each ROI, and the resultant peak responses were then averaged across all observers. Finally, the average levels of activation for each condition were subjected to a repeated-measures ANOVA, performed separately on each ROI in each experiment (SPSS).

Behavioral data analysis. Behavioral performance measures of reaction time and accuracy were recorded by MatLab (MathWorks) (running the Psychtoolbox) and were analyzed with SPSS. Repeated-measures ANOVAs were conducted to assess differences across the conditions in the adaptation and the localizer runs in each experiment.

Results

PPA adaptation to real world object ensemble and surface texture repetition

As our first step to investigate how real-world object ensembles are represented in the brain, in experiment 1, we used photographs of real-world object ensembles and evaluated the role of PPA and LO (the posterior-dorsal aspect of the lateral occipital complex that is distinct from pFs) in object ensemble represen-

tation. We presented observers with a sequence of either three object ensemble or three surface texture images. The images were all identical, all different, or shared object ensemble or surface texture statistics (e.g., different photographs of the same object ensemble or surface texture) (Fig. 1A). Each image was presented for 200 ms with an 800 ms interstimulus interval. Based on common occurrence and availability, for object ensembles we used images of living objects (e.g., fruits and vegetables), and for textures we used images of nonliving items (e.g., marble surfaces). Observers were asked to press a key when the third image disappeared from view. Each trial lasted 6 s. We used a counterbalanced trial history design and calculated percentage signal change compared with fixation directly from the raw MRI signal (Kourtzi and Kanwisher, 2001; Xu and Chun, 2006; Xu, 2010; Dilks et al., 2011; Todd et al., 2011).

We examined the effect of image type (ensemble vs texture) and condition (identical vs shared vs different) and their interactions in independently localized LO and PPA ROIs. Left and right hemisphere ROIs were combined in this and all subsequent experiments because no differences in activation were observed between the hemispheres. On the one hand, if PPA represents ensemble and texture statistics, then as long as the same ensemble or texture is perceived, regardless of whether identical or different images are shown, PPA will exhibit fMRI adaptation to both the identical and the shared conditions and will exhibit a release from adaptation in the different condition (when different ensembles or textures are presented). On the other hand, because LO has been shown to encode specific shape contours (for review, see Grill-Spector, 2009), for object ensembles we predict LO will show a release from fMRI adaptation whenever local shape contours change, which occurs in both the shared and different conditions, compared to the identical condition. Due to a lack of closed shape contours in the surface texture images, LO may not show sensitivity to differences in our texture images.

In PPA, the main effect of image type was not significant ($F_{(1,11)} = 2.21, p = 0.165$), but the main effect of condition did reach significance ($F_{(2,22)} = 11.19, p = 0.001$), with no significant image-by-condition interaction ($F_{(2,22)} = 0.79, p = 0.467$), indicating that PPA exhibited similar response patterns to both object ensembles and surface textures. Based on the specific predictions for object ensembles and surface textures we have outlined, we conducted planned pairwise comparisons to investigate this interaction in greater detail. For object ensembles, the identical and the shared conditions did not differ from each other ($t_{(11)} = 1.00, p = 0.49$, one tailed, and Bonferroni corrected for multiple comparisons; this applies to all subsequent planned comparisons except where noted), and each had a lower response than the different condition (different vs identical: $t_{(11)} = 3.95, p = 0.004$; different vs shared: $t_{(11)} = 3.10, p = 0.02$; Fig. 1B). The same response pattern was observed for surface textures; there was no difference between shared and identical ($t_{(11)} = 0.04, p = 0.50$), and both exhibited a lower response than different (different vs identical: $t_{(11)} = 2.93, p = 0.024$; different vs shared: $t_{(11)} = 2.25$, which approached significance at $p = 0.065$). Taken together, PPA showed equivalent levels of adaptation when either ensemble or texture statistics repeated.

A different response pattern was obtained in LO, such that both main effects and the interaction either were significant or approached significance ($F_{(1,11)} = 11.57, p = 0.045$ for image type; $F_{(2,22)} = 2.82, p = 0.081$ for condition; $F_{(2,22)} = 3.57, p = 0.045$ for the interaction between the two). This indicates that LO responded differently to object ensembles and textures. Planned comparisons revealed that, for object ensembles, the shared and

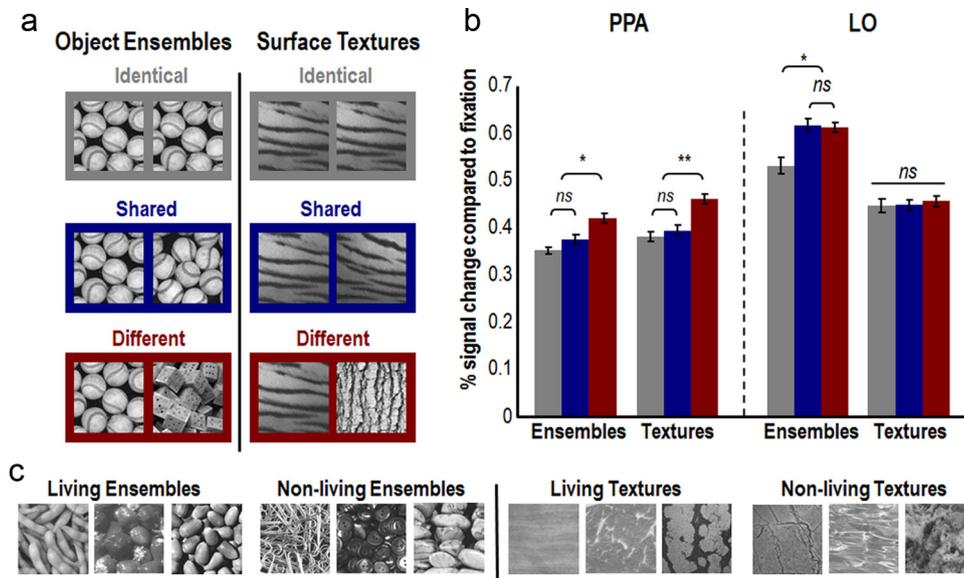


Figure 3. Example stimuli and results ($N = 14$) from experiment 2. **a**, Example stimuli used in the experiment. A new set of images, different from that used in experiment 1, was used in this experiment. This new set contained 20 object ensemble and 20 texture images, each containing 10 living and 10 nonliving examples. All stimuli were presented in grayscale to remove the repetition of colors when ensemble statistics were repeated. In each trial, observers saw a sequence of two images that were identical (gray boxes), different (red boxes), or shared object ensemble or surface texture statistics (blue boxes). Observers were asked to categorize each trial as identical, shared, or different. **b**, Results from experiment 2. Replicating results from experiment 1, PPA again showed equivalent levels of adaptation when object ensemble or texture features were repeated, and LO showed adaptation only when the local shape/contours were identical in the object ensemble images but showed no sensitivity to surface texture manipulations. Error bars represent within-subject SEs. ns, Not significant. **c**, Additional examples of stimuli used in experiment 2. * $p < 0.05$; ** $p < 0.01$.

different conditions did not differ from each other ($t_{(11)} = 1.27$, $p = 0.348$), but both were higher than the identical condition (shared vs identical: $t_{(11)} = 2.35$, which approached significance at $p = 0.063$; different vs identical: $t_{(11)} = 2.79$, $p = 0.025$). For textures, none of the three conditions differed from each other ($t < 0.85$ for all). Thus, LO exhibited an equivalent release from adaptation in the shared and different object ensemble conditions but exhibited insensitivity to changes in surface textures. Although object ensemble and texture images differ in a number of features (e.g., spatial frequency, contrast, color, semantic content), based on the well established functional properties of LO (Atlmann et al., 2003; Kourtzi and Kanwisher, 2001) (for a recent review, see Grill-Spector, 2009), we believe these adaptation findings can be explained by the role of LO in extracting shape information from closed contours. That is, because object ensembles contained well defined shape contours, changes to local shape contour were evident in both the shared and the different conditions, resulting in a release from fMRI adaptation in both conditions compared with the identical condition; meanwhile, because surface textures in general lacked closed contours, changes in texture images were ineffective in driving LO responses. Finally, although our results are consistent with these regions being involved in different types of visual processing, the differences in the patterns of adaptation between PPA and LO for processing object ensembles and surface textures did not reach significance (region-by-condition interaction for ensembles: $F_{(2,22)} = 2.54$, $p = 0.102$; region-by-condition interaction for textures: $F_{(2,22)} = 1.18$, $p = 0.327$).

To assess the reliability of these results, we also fitted a standard GLM to the data to derive β weights for each condition. Adaptation responses obtained from these β weight measures were identical to those obtained from the percent-signal change analysis in both PPA and LO. This confirms the validity of the counterbalanced trial history design and the use of the percent-

signal change analysis. As such, in all subsequent experiments, we report results using this analysis.

Although results from experiment 1 by and large followed our predictions, a few of the relevant comparisons fell short of reaching statistical significance. This was likely due to the passive viewing procedure used in experiment 1, which might not have fully engaged ensemble and texture processing in the relevant brain regions. In addition, because colored images were used in experiment 1, it is possible that PPA adaptation results were driven entirely by the repetition of color in the same and the shared conditions. To address these issues, we conducted a second experiment in which we made the task more engaging by presenting only two images sequentially, and we asked observers to process ensembles and textures more directly by categorizing whether the two images were identical, shared (ensemble or texture statistics), or different. In addition, we removed color information and used grayscale images (Fig. 3A). We also included both living and nonliving object ensemble images (e.g., paper clips and screws) as well as living surface textures (e.g., fruit and animal skins) in addition to the nonliving ones.

Despite these changes, we obtained similar response patterns in PPA and LO, as in experiment 1. Specifically, in PPA, the main effect of image type approached significance ($F_{(1,13)} = 3.68$, $p = 0.077$), and the main effect of condition was significant ($F_{(2,26)} = 21.05$, $p = 0.001$); there was no significant image-by-condition interaction ($F_{(2,26)} = 0.41$, $p = 0.667$). Planned comparisons revealed that PPA showed equivalent levels of adaptation when either object ensembles or surface textures were repeated (for object ensembles: shared vs identical: $t_{(13)} = 1.64$, $p = 0.186$; different vs identical: $t_{(13)} = 4.86$, $p = 0.001$; different vs shared: $t_{(13)} = 2.56$, $p = 0.041$; for surface textures: shared vs identical: $t_{(13)} = 0.68$, $p = 0.50$; different vs identical: $t_{(13)} = 4.71$, $p = 0.001$; different vs shared: $t_{(13)} = 3.72$, $p = 0.004$; Fig. 3B). In LO, the main effects of image type ($F_{(1,13)} = 53.83$, $p = 0.0001$) and

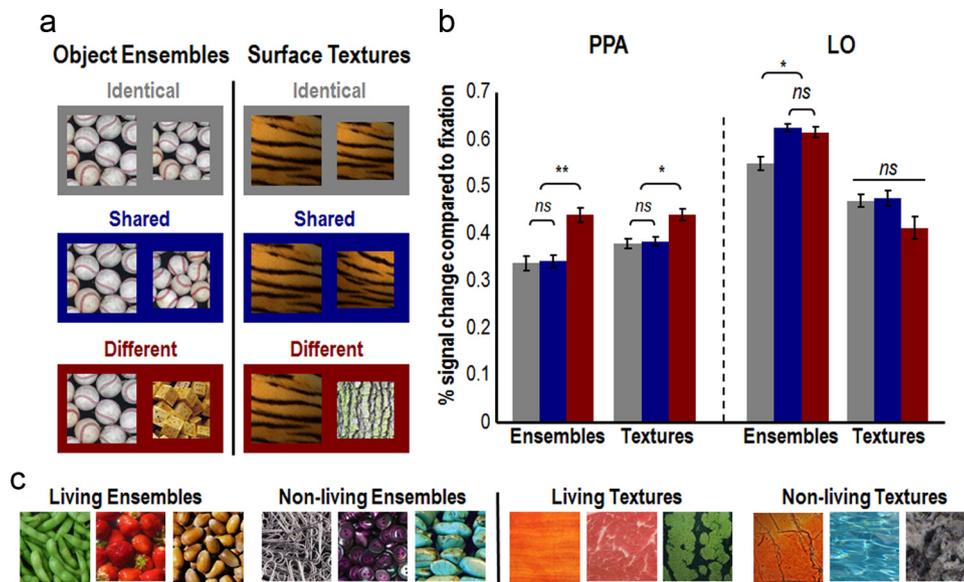


Figure 4. Stimuli and results ($N = 14$) from experiment 3. *a*, Example stimuli used in the experiment. The same stimuli, conditions, and tasks from experiment 2 were used here except the images were shown in full color, and, to investigate how image size changes affect brain responses, the second image in each trial was approximately two-thirds the size of the first image. *b*, Results from experiment 3. Replicating results from experiments 1 and 2, PPA again showed equivalent levels of adaptation when object ensemble or surface texture statistics were repeated, and LO showed adaptation only when the local shape/contours were identical in the object ensemble images but showed no sensitivity to surface texture manipulations. Error bars represent within-subject SEs. ns, Not significant. *c*, Additional examples of stimuli used in experiment 2. * $p < 0.05$; ** $p < 0.01$.

condition ($F_{(2,26)} = 6.71, p = 0.004$) were both significant, and the image-by-condition interaction approached significance ($F_{(2,26)} = 3.07, p = 0.063$). As in experiment 1, LO again showed a release from adaptation for object ensembles when local shape or contours changed in the shared and the different conditions, but it was insensitive to any changes in surface textures (for object ensembles, shared vs different: $t_{(13)} = 0.29, p = 0.50$; shared vs identical: $t_{(13)} = 2.93, p = 0.017$; different vs identical: $t_{(13)} = 3.16, p = 0.011$; for surface textures, none of the three conditions differed from each other: $t < 0.50$ for all). Importantly, differences in adaptation between PPA and LO reached significance for processing both object ensembles and surface textures, showing that these two brain areas do indeed extract different types of information from the same visual input (region-by-condition interaction for ensembles: $F_{(2,26)} = 4.28, p = 0.025$; and region-by-condition interaction for textures: $F_{(2,26)} = 4.94, p = 0.015$). Thus, it appears that focusing participants' attention directly on object ensemble and surface texture processing replicated and strengthened the results we observed in experiment 1. These results also indicate that the repetition of color information alone cannot entirely account for the PPA ensemble and texture adaptation effects obtained in experiment 1 because similar results were obtained when we removed color from the images in experiment 2. Further study is needed to understand whether color is part of the object ensemble representation in anterior-medial ventral visual cortex. It may well be. Importantly, however, what this experiment shows is that visual features, aside from color, also contribute to ensemble and texture representation in anterior-medial ventral visual cortex.

Can the PPA object ensemble adaptation effect obtained thus far be driven by the repetition of lower level image statistics present in the images, such as spatial frequency? Because PPA is situated at a relatively later stage of visual processing, its representation for object ensembles and surface textures is likely to be higher level and invariant to lower level image changes such as size (i.e., spatial frequency) because an image size change usually

corresponds to a change in viewing distance rather than to a qualitative change to an object ensemble or surface texture.

To address this issue, in experiment 3, as in experiment 2, we presented pairs of object ensemble and surface texture images but made the size of the second image approximately two-thirds the size of the first image (Fig. 4A). Because color did not impact the adaptation effect in experiment 2, all images were presented in color to make the task more engaging. We again replicated our basic findings. Namely, in PPA, the main effect of image type was not significant ($F_{(1,13)} = 2.41, p = 0.145$), but the main effect of condition was ($F_{(2,26)} = 16.48, p = 0.001$), with no significant image-by-condition interaction ($F_{(2,26)} = 1.25, p = 0.302$). Replicating results from experiments 1 and 2, PPA exhibited equivalent levels of adaptation when either object ensemble or surface texture statistics were repeated (for object ensembles: shared vs identical, $t_{(13)} = 0.17, p = 0.50$; different vs identical, $t_{(13)} = 3.68, p = 0.004$; different vs shared, $t_{(13)} = 3.96, p = 0.002$; for surface textures: shared vs identical, $t_{(13)} = 0.27, p = 0.50$; different vs identical, $t_{(13)} = 2.90, p = 0.017$; different vs shared, $t_{(13)} = 3.00, p = 0.017$) (Fig. 4B). In LO, significant results were found for the main effects of image type ($F_{(1,13)} = 61.60, p = 0.001$), condition ($F_{(2,26)} = 3.69, p = 0.039$), and the image-by-condition interaction ($F_{(2,26)} = 4.50, p = 0.021$). LO again showed a release from adaptation for object ensembles when local shape information changed in the shared and the different conditions (relative to the identical condition) but was insensitive to any changes in surface textures (for object ensembles: shared vs identical, $t_{(13)} = 3.50, p = 0.006$; different vs identical, $t_{(13)} = 2.68, p = 0.028$; different vs shared, $t_{(13)} = 0.83, p = 0.50$; for surface textures, none of the three conditions differed from each other, $t < 1.71$ for all). Finally, differences in adaptation between PPA and LO reached significance for processing both object ensembles and surface textures, again showing that different types of visual information from object ensembles and surface textures are extracted by these two brain areas (region-by-condition interaction for ensembles: $F_{(2,26)} = 12.27, p = 0.0001$; region-by-condition inter-

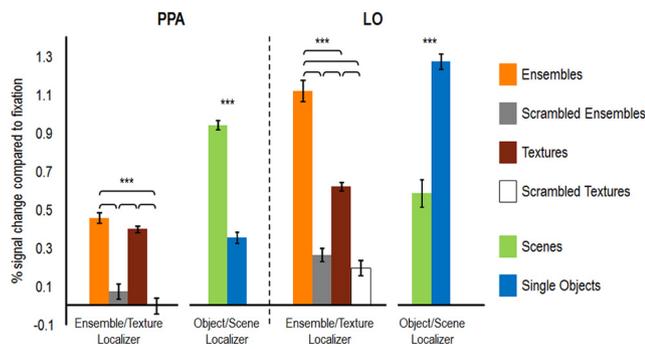


Figure 5. PPA and LO results for the four conditions in the ensemble/texture localizer. In PPA, the two intact image conditions did not differ from each other, nor did the two scrambled image conditions differ from each other. Although spatial frequency and other low-level image information (such as contrast and luminance) were equated between intact and scrambled images, intact images elicited significantly higher responses than scrambled images. For comparison, activations for scenes and single objects from the object/scene localizer are also plotted (computed in individual observers by defining PPA using the first run of the object/scene localizer and then extracting independent data from this region using the last run of the object/scene localizer). Scenes elicit the highest activation in PPA compared with objects, ensembles, and textures. Thus, although we show in this study that PPA is the key brain region mediating the representation of ensemble and texture statistics, scenes still seem to be the most effective stimuli in driving PPA response. For completeness, responses in LO, computed using the same method described for PPA, were included as well. *** $p < 0.001$.

action for textures: $F_{(2,26)} = 9.05$, $p = 0.001$). Taken together, these results indicate that the adaptation effects obtained thus far are relatively higher level and are not affected by a change in image size. This is similar to the size invariance effect previously reported in PPA for scene processing and the lateral occipital complex, of which LO is a subregion, for object processing (Grill-Spector et al., 1999; Andrews and Ewbank, 2004; Ewbank et al., 2005; Lee et al., 2006).

To directly localize regions involved in ensemble and texture processing, we included an ensemble/texture localizer that contained object ensembles, textures, and their phase-scrambled counterparts (matched in overall spatial frequency, luminance, and contrast). Although the main results obtained from this localizer are described in detail in the next section, here we used the stimulus conditions included in this localizer to further examine whether PPA adaptation for object ensembles and surface textures was driven by the repetition of low-level image features such as spatial frequency, luminance, and contrast. To do so, we extracted averaged responses for the four conditions in this ensemble/texture localizer from individual observers' PPA (defined using the object/scene localizer). We obtained a significant main effect of image conditions ($F_{(3,39)} = 27.77$, $p = 0.001$) (Fig. 5) such that the intact images elicited higher responses than the scrambled images (ensembles vs scrambled ensembles: $t_{(13)} = 6.56$, $p = 0.001$; ensembles vs scrambled textures: $t_{(13)} = 6.25$, $p = 0.001$; textures vs scrambled textures: $t_{(13)} = 7.64$, $p = 0.001$; textures vs scrambled ensembles: $t_{(13)} = 5.07$, $p = 0.001$; all two tailed and Bonferroni corrected). The two intact image conditions did not differ from each other ($t_{(13)} = 1.88$, $p = 0.472$), and the two phase-scrambled image conditions also did not differ from each other ($t_{(13)} = 1.11$, $p = 1.00$). Thus, although the intact and the scrambled images shared the same overall spatial frequency, luminance, and contrast, they elicited different amounts of PPA activation. This further suggests that PPA adaptation to the repetition of object ensemble and surface texture statistics cannot be solely attributed to the repetition of low-level visual information (such as spatial frequency, luminance and contrast) across images.

Although we have examined responses from PPA as a whole, it is possible that PPA may contain heterogeneous regions, with more posterior parts of PPA (i.e., within the collateral sulcus) more involved in object ensemble and surface texture processing than anterior parts (i.e., closer to the anterior aspect of the parahippocampal gyrus). To examine this possibility, we divided every observer's PPA ROI into anterior and posterior parts. This was done by constructing a line through the center of activation in each observer's PPA and then assigning all voxels extending in front of this line to the anterior part and all voxels extending behind this line to the posterior part. In all three experiments, we found virtually identical adaptation results in both the anterior and the posterior parts of PPA for object ensembles and surface textures (region-by-image-by-condition interaction: for experiment 1, $F_{(2,22)} = 0.57$, $p = 0.58$; for experiment 2, $F_{(2,22)} = 1.68$, $p = 0.21$; for experiment 3, $F_{(2,22)} = 0.42$, $p = 0.66$), suggesting that the anterior and the posterior parts of PPA do not differ in how they process object ensembles and surface textures.

Taken together, results from experiments 1 to 3 demonstrate that the processing of object ensembles in PPA is remarkably similar to the processing of surface textures. In light of this, one might question whether or not object ensemble processing in PPA simply reflects the processing of the surface texture of the individual objects in the ensemble rather than ensemble features per se. To address this question, in experiment 4 we used black wooden beads of different shapes as stimuli (Fig. 6A). We showed observers two images that were either identical, shared object ensemble features (i.e., different photographs of the same beads), or different (i.e., photographs of beads with different shapes). Here, in all three conditions, the surface texture of the individual objects in the ensembles was identical (i.e., painted black wood). Despite this surface texture repetition, PPA again exhibited different patterns of adaptation across the three conditions (main effect of condition: $F_{(2,20)} = 5.37$, $p = 0.014$), with planned pairwise comparisons revealing that the identical and the shared conditions did not differ from each other ($t_{(10)} = 1.05$, $p = 0.486$), but both showed a lower response compared to the different condition (different vs identical: $t_{(10)} = 2.56$, $p = 0.038$; different vs shared: $t_{(10)} = 2.77$, $p = 0.027$) (Fig. 6B).

Interestingly, in experiment 4, there was no response difference in LO (main effect of condition: $F_{(2,20)} = 0.07$, $p = 0.930$; $t < 0.39$ for all planned pairwise comparisons). This null result might be attributed to the shape of the beads used in this experiment. Because half the beads used were circular or very close to circular (Fig. 6C), in a lot of the trials there were likely minimum contour changes between different photographs depicting the same ensemble. This may explain why we failed to observe a release from adaptation in LO between the shared and the identical conditions. This might also have resulted in minimum contour changes in photographs depicting different ensembles, resulting in our failure to observe a release from adaptation in LO between the identical and the different conditions. Despite this null result, however, differences in response patterns between PPA and LO reached significance (region-by-condition interaction: $F_{(2,20)} = 5.27$, $p = 0.015$), replicating the results from experiments 2 and 3 and again demonstrating that these regions process object ensembles in significantly different ways. Overall, results of experiment 4 indicate that the processing of object ensembles in PPA does not simply reflect the processing of the surface texture of the individual objects composing an ensemble. Texture processing can certainly play a role in ensemble representation, but these results show that the shapes of the individual objects making up

an ensemble are also an important part of ensemble representation in anterior-medial ventral visual cortex.

Directly localizing ventral brain areas involved in real-world object ensemble and surface texture processing

To localize visual areas that would be naturally activated when object ensembles and surface textures are processed, we contrasted brain responses obtained when observers viewed blocks of object ensemble and surface texture images used in our adaptation experiments with responses obtained when they viewed blocks of phase-scrambled versions of these same images. Using a random-effects group analysis ($p < 0.001$, uncorrected), we identified two main regions of activation in the occipitotemporal cortex, with one located laterally in the vicinity of LO and the other located ventrally along the collateral sulcus/parahippocampal gyrus in the vicinity of PPA and extending posteriorly along the collateral sulcus. (Other regions of the brain also became active using this contrast, such as regions along the intraparietal sulcus in parietal cortex [one bilateral activation in the inferior intraparietal sulcus (Talairach x, y, z coordinates for right/left are $+26/-28, -80/-85, +19/+16$) and another bilateral activation in the superior intraparietal sulcus (Talairach x, y, z coordinates for right/left are $+25/-24, -61/-64, +39/+40$)], and one region in the right frontal cortex [$x = 48, y = 22, z = 28$]. These activations were likely driven by differences in attention as the intact ensemble and texture images were more attentionally engaging than the phase-scrambled versions of these same images.) In fact, PPA overlapped 61% with the ventral ensemble/texture area (calculated as the number of overlapping voxels divided by the total number of voxels in PPA; all other overlap values reported below were similarly calculated except where noted), and LO overlapped 55% with the lateral ensemble/texture area (Fig. 7A). When we relaxed the statistical threshold, the amount of overlap increased. At $p < 0.01$, uncorrected, PPA overlap increased to 80% and LO overlap increased to 70%. Similar results were obtained when ensemble and texture images were analyzed separately with their phase-scrambled counterparts, with the exception that LO overlap was greater for object ensembles than for surface textures, likely due to the lack of closed contours in the latter images (at $p < 0.001$ and $p < 0.01$, both uncorrected, object ensemble activation overlapped with PPA 45% and 63%,

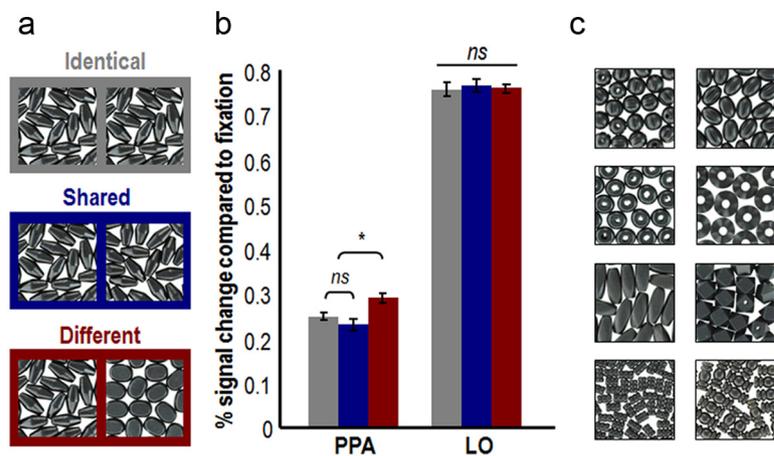


Figure 6. Stimuli and results ($N = 11$) from experiment 4. **a**, Example stimuli used in the experiment. Only object ensembles made of black wooden beads were used here. In each trial, observers saw two images that were identical (gray boxes), shared object ensemble features (i.e., different photographs of the same type of beads; blue boxes), or different (i.e., photographs of beads that differed in the shape of the individual ensemble elements; red boxes). The surface texture and the material properties of the individual objects in the ensembles were thus identical in all conditions. The same image categorization task used in experiments 2 and 3 was used here. **b**, Results from experiment 4. Despite texture/material repetition of the ensembles across the three conditions, PPA again showed adaptation when ensemble statistics were repeated but a release from adaptation when the shape of the beads changed between ensembles. Unlike the previous three experiments, LO did not show any sensitivity to our manipulations, possibly because half of the beads used in the experiment were approximately circular, resulting in minimal contour changes between different images depicting either the same or different ensembles. Error bars represent within-subject SEs. ns, Not significant. **c**, Additional stimuli used in experiment 4. * $p < 0.05$.

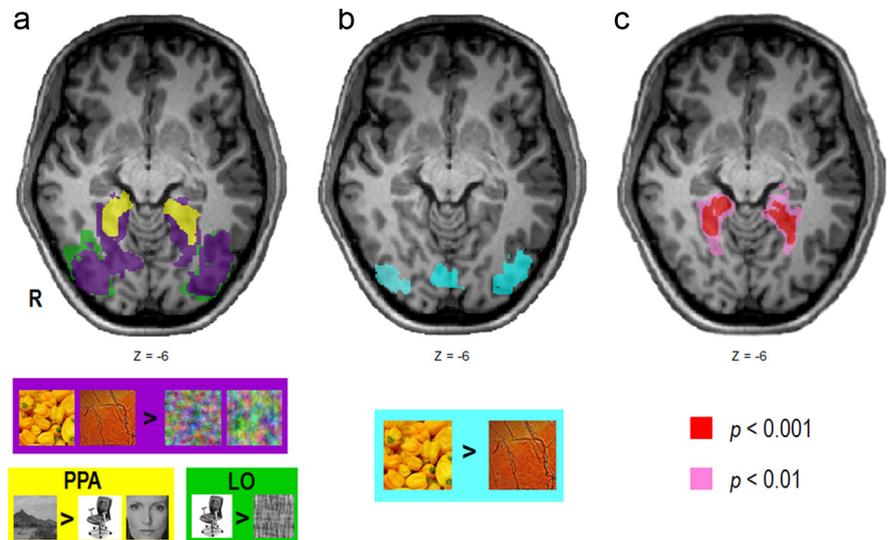


Figure 7. **a**, Group overlap of the object ensemble/surface texture regions and PPA and LO. To localize visual areas naturally activated when object ensemble and surface textures are processed, brain responses for viewing object ensembles and surface textures were contrasted with those for viewing phase-scrambled versions of these same images. Two main regions of activation were located (shown in purple), with one located laterally and the other located ventrally originating from the parahippocampal gyrus and extending posteriorly along the collateral sulcus. Although the anterior-medial ventral region (Talairach coordinates, x, y, z for right/left: $+21/-23, -50/-52, -6/-9$) overlapped greatly with PPA (yellow; $+21/-21, -39/-40, -7/-7$, defined by contrasting scenes with faces and everyday objects), the lateral region ($+33/-37, -77/-74, -1/-6$) overlapped greatly with LO (green; $36/-36, -75/-75, -3/-4$, defined by contrasting everyday objects with phase-scrambled versions of these same images). The large overlap between the different brain regions justifies our selection of PPA and LO as the main ROIs in investigating the neural underpinnings of object ensemble processing. All regions are displayed at $p < 0.001$, uncorrected. **b**, Regions differentially activated for object ensembles or surface textures. The only regions that were more active (group data, displayed at $p < 0.001$, uncorrected) in the visual occipitotemporal cortex for object ensembles than for surface textures were located in LO ($+33/-35, -80/-78, -1/-4$) and early visual cortex ($8, -86, 0$). No regions were more active for textures than for ensembles. **c**, Common region of overlap between PPA (defined using the scenes vs faces and objects contrast) and the ventral ensemble/texture region (defined using the intact vs scrambled ensembles and textures contrast), both at the group level and displayed at $p < 0.001$ ($+21/-22, -41/-44, -7/-6$) and $p < 0.01$ ($+20/-21, -44/-41, -6/-6$), both uncorrected. As the statistical threshold is relaxed, the common region of activation extends more posteriorly. R, Right.

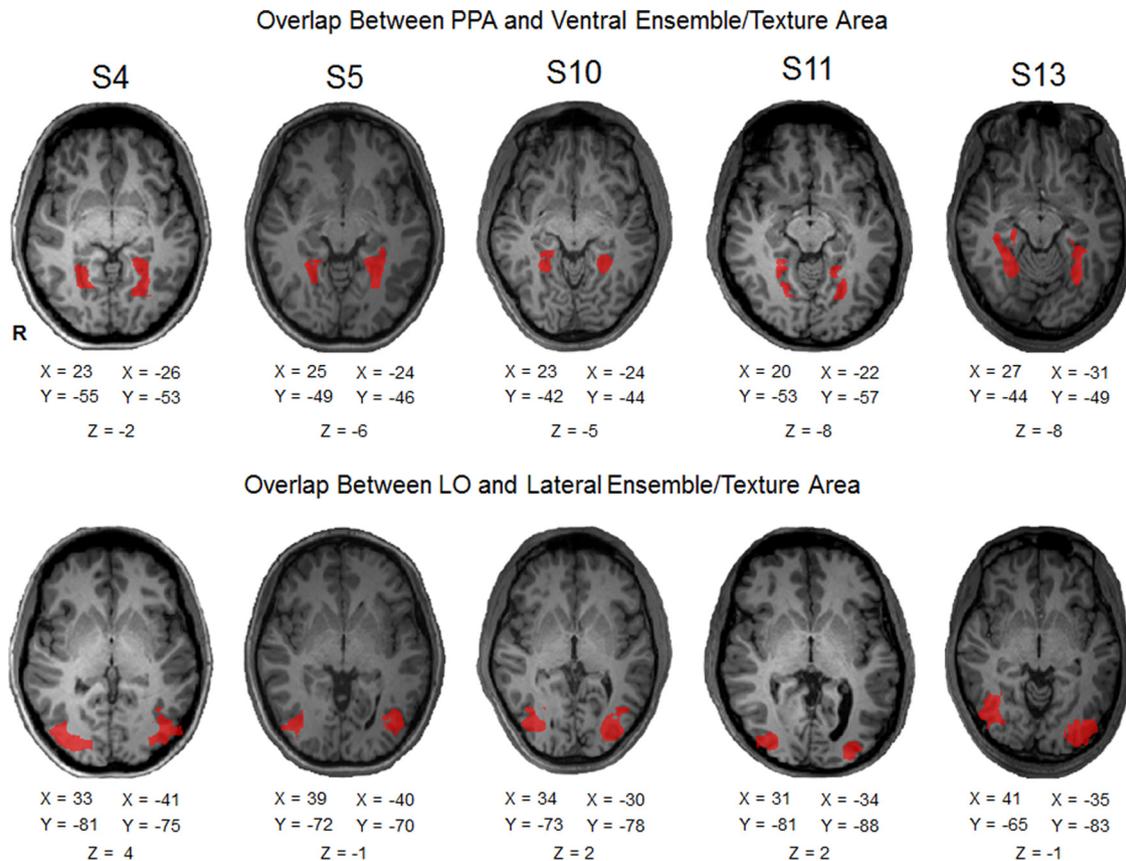


Figure 8. Top, Common regions of overlap for PPA (defined using the object/scene localizer) and the ventral ensemble/texture region (defined using the ensemble/texture localizer), shown in five representative observers. Bottom, Common regions of overlap for LO (defined using the object/scene localizer) and the lateral ensemble/texture region (defined using the ensemble/texture localizer), shown in the same five observers. All regions are displayed at $p < 0.001$, uncorrected. Talairach coordinates are given under each brain. R, Right; S, subject.

respectively, and with LO 58% and 72%, respectively; surface texture activation overlapped with PPA 50% and 76%, respectively, and with LO 22% and 41%, respectively).

We also assessed the amount of overlap between ventral regions activated by object ensembles and by surface textures (by contrasting intact vs scrambled images to first localize these regions). At the group level, the overlap between the two was 69% and 82% at $p < 0.001$ and at $p < 0.01$, both uncorrected, respectively.

To better interpret the meaning of the overlap in these analyses, we calculated the amount of overlap between PPA regions defined by the first and the last run of the same object/scene localizer. Because the same stimuli and the same observers were used, an ideal result would be near 100% overlap between the two PPA regions at the group level. In reality, however, the overlap was only 56% at $p < 0.01$, uncorrected (overlap at $p < 0.001$ could not be calculated due to insufficient power because fewer runs were included in this analysis). The failure to observe a near 100% overlap was likely due to random factors such as head motion, breathing rate, attention, and scanner noise (among others). This demonstrates that a very high degree of spatial correspondence between regions is difficult to achieve with fMRI at the group level, even when defining the same region using the same stimuli and observers. Importantly, these overlap values are similar to those reported above when the overlap between PPA, LO, and the ensemble/texture regions were compared, indicating that the values reported above reflect a high degree of overlap between the different brain regions compared.

Calculating overlap values between regions on group data can potentially overestimate the degree of overlap because averaging

individual activations together would result in spatial smoothing and would blur the boundaries between functionally distinctive regions. To address this concern, we recalculated the overlap between our functional regions in individual observers at our most conservative threshold ($p < 0.001$) and then averaged these values from individual observers to derive a group average overlap value. With this procedure, PPA overlapped 46% with the ventral ensemble/texture area, and LO overlapped 61% with the lateral ensemble/texture area (Fig. 8; illustrations of the overlap regions in individual observers). Compared with the overlap values obtained directly from the group data (61% and 55%, respectively, for the two overlaps; see above), it seems that calculating overlap based on group data does not necessarily overestimate the amount of functional overlap between regions. To better evaluate the degree of overlap between brain regions, we also recalculated the amount of overlap between PPA regions defined using the first and last runs of the same object/scene localizer using data from individual observers. At $p < 0.001$ (uncorrected), the overlap between these areas was 60%, a value comparable to the values reported above using the group data (which was 56%). Together, the similarity between the individual and group overlap analyses justifies our initial use of group statistics to report overlap between functional regions. To further validate how we calculated overlap, we compared our method, which uses a single ROI in the denominator, with that proposed by Kung et al. (2007), which uses the average of the two ROIs in comparison in the denominator. We did not find any significant differences between these two methods: overlap between PPA and ventral ensemble/texture area was 46% (our method) and 41% (Kung et al.,

2007), $t_{(13)} = 1.11$, $p = 0.276$; overlap between LO and lateral ensemble/texture area was 61% (our method) and 66% (Kung et al., 2007), $t_{(13)} = 1.07$, $p = 0.296$; and overlap between first and last run defined PPA was 60% (our method) and 62% (Kung et al., 2007), $t_{(13)} = 0.14$, $p = 0.887$. All comparisons were two tailed.

In addition to examining the overlap between regions, we also localized the ventral and the lateral ensemble/texture regions directly in each observer and used them as ROIs to reanalyze the results from our grayscale (experiment 2) and size-change (experiment 3) adaptation experiments. Results from these two ensemble/texture regions were very similar to those obtained from the PPA and LO ROIs. Specifically, the ventral region (which was in the vicinity of PPA) exhibited similar response patterns to both object ensembles and surface textures (image-by-condition interaction: in experiment 2, $F_{(2,26)} = 0.01$, $p = 0.99$; in experiment 3, $F_{(2,26)} = 1.01$, $p = 0.378$) and showed equivalent levels of adaptation when either object ensembles or surface textures were repeated (planned pairwise comparisons in experiment 2 for object ensembles: shared vs identical, $t_{(13)} = 2.16$, $p = 0.084$; different vs identical, $t_{(13)} = 3.87$, $p = 0.003$; different vs shared, $t_{(13)} = 2.72$, $p = 0.027$; planned pairwise comparisons in experiment 2 for surface textures: shared vs identical, $t_{(13)} = 1.79$, $p = 0.137$; different vs identical, $t_{(13)} = 5.36$, $p = 0.001$; different vs shared, $t_{(13)} = 3.00$, $p = 0.017$; planned pairwise comparisons for experiment 3 for object ensembles: shared vs identical, $t_{(13)} = 1.66$, $p = 0.178$; different vs identical, $t_{(13)} = 3.36$, $p = 0.008$; different vs shared, $t_{(13)} = 4.40$, $p = 0.001$; planned pairwise comparisons for experiment 3 for surface textures: shared vs identical, $t_{(13)} = 1.11$, $p = 0.425$; different vs identical, $t_{(13)} = 2.93$, $p = 0.016$; different vs shared, $t_{(13)} = 2.48$, $p = 0.046$). The lateral region, which was in the vicinity of LO, showed a release from adaptation for object ensembles when local shape or contours changed in the shared and the different conditions but was insensitive to any changes in surface textures (overall difference between object ensembles and surface textures, from the image-by-condition interaction: in experiment 2, $F_{(2,26)} = 0.87$, $p = 0.43$; in experiment 3, $F_{(2,26)} = 1.81$, $p = 0.183$; planned pairwise comparisons in experiment 2 for object ensembles: shared vs different, $t_{(13)} = 0.94$, $p = 0.50$; shared vs identical, $t_{(13)} = 3.26$, $p = 0.01$; different vs identical, $t_{(13)} = 3.43$, $p = 0.007$; planned pairwise comparisons in experiment 2 for surface textures: none of the three conditions differed from each other, all $t < 2.15$, NS; planned pairwise comparisons in experiment 3 for object ensembles, shared vs different: $t_{(13)} = 0.13$, $p = 0.50$; shared vs identical, $t_{(13)} = 2.93$, $p = 0.018$; different vs identical, $t_{(13)} = 2.26$, which approached significance at $p = 0.066$; for surface textures: none of the three conditions differed from each other, all $t < 0.38$, NS). Finally, differences in adaptation between the ventral and lateral ensemble/texture regions reached significance for processing both object ensembles (region-by-condition interaction for ensembles: in experiment 2, $F_{(2,26)} = 5.94$, $p = 0.008$; in experiment 3, $F_{(2,26)} = 4.53$, $p = 0.021$) and surface textures (region-by-condition interaction for textures: in experiment 2, $F_{(2,26)} = 4.53$, $p = 0.02$; in experiment 3, $F_{(2,26)} = 15.50$, $p = 0.001$), showing that these two brain areas extract different types of information from the same visual input.

Finally, to directly compare the processing of object ensembles and surface textures, we used the group ensemble/texture localizer data and contrasted the activations for ensembles with surface textures (Fig. 7B). At $p < 0.001$ (uncorrected), the only regions more active for object ensembles than surface textures were in the vicinity of LO and early visual cortex. The preference of LO for processing object ensembles over surface textures is consistent with our adaptation data from individual observers

and with the blocked localizer results shown in Figure 5, as well as with previous reports showing preferential processing of shapes with closed contours in LO (Kourtzi and Kanwisher, 2001; Altman et al., 2003). Greater activation in early visual cortex for ensembles than for textures may reflect the presence of more high spatial frequency information in the ensemble than in the texture images. Meanwhile, no region showed higher activation for surface textures than for object ensembles. These findings lend further support to the notion that object ensembles and surface textures share similar neural and computational mechanisms in anterior-medial ventral visual cortex.

Taken together, these results independently show that visual processing of object ensembles naturally activates two distinctive regions in ventral and lateral visual cortex, corresponding well to the location of PPA and LO, respectively. Moreover, they also show that the processing of object ensembles and surface textures activates a common region in anterior-medial ventral visual cortex. These results are consistent with our adaptation results and justify our selection of PPA as the candidate ROI to target in investigating the neural underpinnings of object ensemble processing.

Whole-brain analysis of real-world object ensemble and texture adaptation

To further assess whether additional visual areas are involved in object ensemble and surface texture processing, we performed a whole-brain group random-effects analysis on the adaptation data separately for each experiment. Specifically, we looked for regions that showed a higher response for the different compared with the identical or the shared conditions for both object ensembles and surface textures. Despite weaker effects typically associated with event-related adaptation paradigms and variability in observers' responses, at $p < 0.01$, uncorrected, we observed bilateral activation in anterior-medial ventral visual cortex in all four experiments (Fig. 9), corresponding well to the location of our PPA ROI and the ventral region activated in our ensemble/texture localizer (Fig. 7A). We want to emphasize that the medial location of this activation, although independently obtained from the four adaptation experiments involving different stimuli, tasks, and observers, was remarkably consistent and replicable across experiments. Moreover, these were the only regions activated in the ventral posterior part of the brain. (Additional regions were activated along the anterior inferior temporal sulcus, superior and inferior frontal gyrus, and medial and lateral partial cortex, likely reflecting task- and attention-related processing differences among the conditions; see Table 1.) Similar activation in anterior-medial ventral visual cortex was observed when object ensemble and surface texture data were analyzed separately and when the identical condition was excluded from the analysis. These results provide converging evidence that object ensembles and surface textures share a common neural substrate in the ventral visual cortex and confirm the involvement of PPA (which is located in anterior-medial ventral visual cortex) in object ensemble processing.

Interestingly, the continuous ventral activation that we saw in our ensemble/texture localizer (Fig. 7A) broke into two separate patches of activations in the right hemisphere in this whole-brain analysis (Fig. 9). On the one hand, this suggests that the ventral ensemble/texture processing region may be further divided into two separate regions. Indeed, a number of recent studies have reported activation of the right posterior collateral sulcus region in low-level visual texture and material property processing (Cavina-Pratesi et al., 2010a,b; Cant and Goodale, 2011). It is also

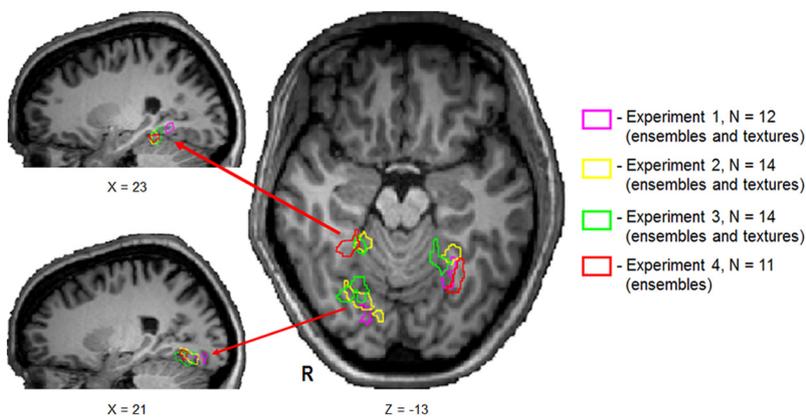


Figure 9. Regions that exhibit adaptation for repetitions of object ensemble and surface texture statistics (i.e., lower response for the identical and the shared conditions than for the different condition), plotted separately for each of the four adaptation experiments and shown as outlines to illustrate the overlap across experiments. The locations of the regions are consistent across the four experiments and reside in the anterior-medial part of ventral visual cortex, extending along the collateral sulcus and the parahippocampal gyrus (Talairach coordinates for experiment 1, anterior collateral sulcus/parahippocampal gyrus [aCoS/PG], x, y, z for right/left: +24/−25, −49/−56, −5/−11; right posterior collateral sulcus [pCoS]: 21, −76, −12; for experiment 2, aCoS/PG: +21/−26, −37/−46, −15/−8; right pCoS: +21, −69, −12; for experiment 3, aCoS/PG: +22/−20, −41/−49, −14/−12; right pCoS: 24, −66, −13; for experiment 4, aCoS/PG: +28/−30, −40/−53, −16/−11; right pCoS: 22, −63, −9). These results provide additional support for our choice of PPA as the main ROI for examining object ensemble and surface texture processing in the visual cortex. R, Right.

Table 1. Additional regions uncovered in the group random effects whole-brain analyses conducted in experiments 1 through 4

	Experiment 1	Experiment 2	Experiment 3	Experiment 4
R MPL	2, −62, 36	6, −53, 26	7, −53, 29	8, −52, 27
L MPL	−11, −63, 36	−6, −54, 29	−7, −54, 30	−7, −54, 20
R LPL	47, −68, 21	48, −63, 25	40, −67, 35	40, −67, 35
L LPL	−53, −57, 23	−49, −60, 22	−43, −61, 26	−38, −70, 32
R anterior ITS	52, −9, −6	51, 11, −17	54, −3, −6	42, −7, −5
L anterior ITS		−51, −2, −14	−60, −9, −6	−52, −3, −12
R SFG	20, 25, 39	23, 23, 49	23, 17, 41	21, 18, 49
L SFG	−21, 24, 37	−21, 24, 47	−18, 25, 43	−23, 21, 44
R IFG	37, 32, 4	19, 29, 9	17, 34, 22	14, 31, 10
L IFG	−39, 31, 5	−35, 26, −11	−42, 27, 8	−37, 33, −7

In each experiment, we looked for regions that showed a higher response for the different compared with the identical and the shared conditions for both object ensembles and surface textures (only object ensembles were used in experiment 4). Talairach coordinates (x, y, z) are given for each region, and all regions were uncovered at $p < 0.01$, uncorrected. IFG, inferior frontal gyrus; ITS, inferior temporal sulcus; L, left; LPL, lateral parietal lobe; MPL, medial parietal lobe; R, right; SFG, superior frontal gyrus.

possible that the posterior collateral sulcus, along with regions in the fusiform gyrus, represent a transition zone (Cant et al., 2009) whose function evolves from the prominence of visual shape processing more laterally to the prominence of ensemble statistical and textural processing more medially and anteriorly. On the other hand, however, these two potentially separate regions in the right ventral visual cortex could also be part of one larger functional region involved in object ensemble and texture processing. In support of this idea, when we plotted the overlap between PPA defined by the object/scene localizer and the ventral region defined by the ensemble/texture localizer, we observed that the overlap extended more posteriorly, partially encompassing both right hemisphere regions identified in our adaptation experiments, as the statistical threshold was relaxed from $p < 0.001$ to $p < 0.01$ (Fig. 7C). Further research is needed to fully understand whether the anterior and posterior regions of the right collateral sulcus represent common or distinct regions for processing object ensemble and texture features.

Response to ensembles and textures in the broader scene-processing and object-processing networks

To examine whether or not all scene-processing regions in the brain participate in object ensemble and visual texture processing in a similar manner, we examined adaptation responses to ensembles and textures in two additional scene-selective regions, namely RSC (for review, see Epstein, 2008) and TOS (Epstein et al., 2005). Both regions were defined in individual observers using the same statistical contrast and threshold as PPA. The main results within each region are reported in Table 2 and Figure 10. Although PPA and RSC showed similar response patterns in some experiments (region-by-condition interaction for ensembles: in experiment 1, $F_{(2,22)} = 0.45$, $p = 0.641$; in experiment 3, $F_{(2,24)} = 1.28$, $p = 0.297$; in experiment 4, $F_{(2,18)} = 1.18$, $p = 0.331$; region-by-condition interaction for textures: in experiment 1, $F_{(2,22)} = 1.67$, $p = 0.212$; in experiment 2, $F_{(2,22)} = 1.16$, $p = 0.331$), they differed in other experiments (region-

by-condition interaction for ensembles in experiment 2: $F_{(2,24)} = 4.77$, $p = 0.018$; region-by-condition interaction for textures in experiment 3: $F_{(2,24)} = 10.65$, $p = 0.001$). This lack of consistency in RSC response patterns, together with the overall low RSC responses in all four experiments (compared with those obtained from PPA and TOS), suggests that RSC is unlikely to play as important a role in ensemble and texture processing as PPA does. With the exception of experiment 1, response patterns differed consistently between PPA and TOS in experiments 2 through 4 (region-by-condition interaction for ensembles in experiment 1: $F_{(2,22)} = 0.38$, $p = 0.686$; in experiment 2: $F_{(2,26)} = 14.57$, $p = 0.001$; in experiment 3: $F_{(2,26)} = 37.55$, $p = 0.001$; in experiment 4: $F_{(2,20)} = 9.95$, $p = 0.001$; region-by-condition interaction for textures in experiment 1: $F_{(2,24)} = 1.40$, $p = 0.268$; in experiment 2: $F_{(2,26)} = 4.71$, $p = 0.018$; in experiment 3: $F_{(2,26)} = 12.84$, $p = 0.001$). This indicates that PPA and TOS differ in how they process object ensembles and surface textures (see also the main results within each region presented in Table 2 and Fig. 10).

To compare response amplitude for ensembles, textures, and scenes in PPA, we defined PPA in each individual using the first run of their object/scene localizer and then extracted the activations for scenes and objects using the last run of their object/scene localizer. We also extracted PPA responses for ensembles and textures in the ensemble/texture localizer. Although scenes, objects, ensembles, and textures were not included in the same run and could not be directly compared with statistical tests, it is evident that the PPA response to scenes was much greater than the response to ensembles, textures, or objects (Fig. 5; for completeness, responses in LO, computed using the same method described for PPA above, were also included). Thus, although we show in this study that PPA is the key brain region mediating the representation of ensemble and texture statistics, scenes still seem to be the most effective stimuli in driving the response in PPA.

Taken together, our results indicate that there appears to be a functional dissociation in the human scene-processing network. Specifically, while PPA is involved in both the spatial (e.g., spatial expanse) (Kravitz et al., 2011) and the nonspatial aspects of visual

Table 2. Effects of object ensemble and surface texture adaptation in RSC, TOS, and pFs for all four experiments

Main effect		Planned comparisons					
		Object ensembles			Textures		
Image	Condition	Identical vs shared	Identical vs different	Shared vs different	Identical vs shared	Identical vs different	Shared vs different
RSC							
Experiment 1	$F_{(2,22)} = 3.91 p = 0.035$		$t_{(11)} = 3.83 p = 0.005$	$t_{(11)} = 2.71 p = 0.03$			
Experiment 2	$F_{(2,24)} = 3.86 p = 0.035$					$t_{(12)} = 2.76 p = 0.025$	
Experiment 3	$F_{(2,24)} = 6.03 p = 0.008$						
Experiment 4	N/A	$F_{(2,18)} = 5.11 p = 0.017$		$t_{(9)} = 3.53 p = 0.011$	N/A	N/A	N/A
TOS							
Experiment 1			$t_{(11)} = 2.95 p = 0.02$				
Experiment 2	$F_{(2,26)} = 5.85 p = 0.008$	$t_{(13)} = 3.05 p = 0.013$					
Experiment 3							
Experiment 4	N/A				N/A	N/A	N/A
pFs							
Experiment 1	$F_{(2,22)} = 20.90 p = 0.001$		$t_{(11)} = 3.88 p = 0.004$	$t_{(11)} = 2.55 p = 0.042$		$t_{(11)} = 3.95 p = 0.004$	
Experiment 2	$F_{(1,13)} = 66.77 p = 0.001$	$F_{(2,26)} = 12.17 p = 0.001$	$t_{(13)} = 2.76 p = 0.027$	$t_{(13)} = 3.05 p = 0.013$			
Experiment 3	$F_{(1,13)} = 30.88 p = 0.001$						
Experiment 4	N/A				N/A	N/A	N/A

Image indicates ensemble and texture. Condition indicates identical, shared, and different. Image-by-condition interaction was not significant in any of the experiments. Main effects, interactions, and planned pairwise comparisons were assessed (one tailed, Bonferroni corrected). Left and right hemispheres were combined because no difference in activation was observed between the hemispheres. Only significant results are shown. N/A, Not applicable.

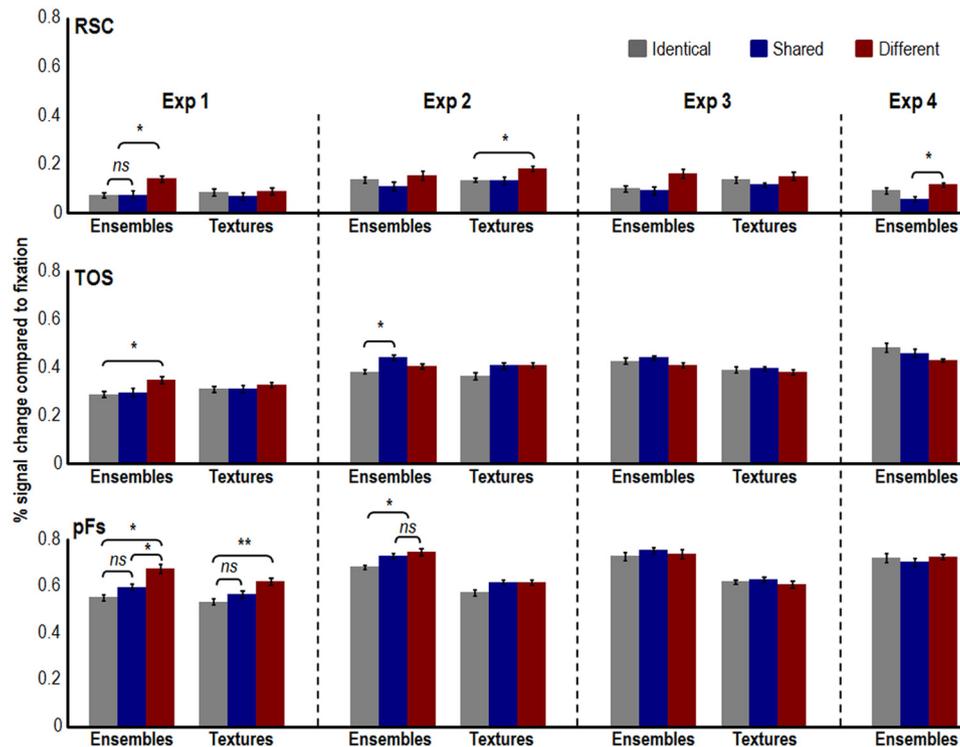


Figure 10. Adaptation results in RSC, TOS, and pFs, shown for all four adaptation experiments. Responses to object ensembles and textures in RSC and TOS were not as consistent as those observed in PPA, suggesting that RSC and TOS are unlikely to play significant roles in ensemble and texture processing. This suggests that PPA is involved in both spatial and nonspatial aspects of visual processing, but RSC and TOS may only participate in spatial aspects of visual processing. Depending on the experiment and the stimulus condition, pFs responses were either similar to LO (ensembles in experiments 1 and 2) or PPA (ensembles in experiment 1, and basic adaptation effect for textures in experiment 1). This suggests that pFs, which is anterior to LO but posterior to PPA, may be a “transition zone” whose function is transitioning from processing shapes to processing the statistical information contained in ensembles and textures. Exp, Experiment; ns, not significant. Error bars represent within-subject SEs. * $p < 0.05$; ** $p < 0.01$.

processing (e.g., object ensembles and textures), RSC and TOS may participate in only the spatial aspects of visual processing.

The broader object-processing network in the human brain consists of two connected regions, LO and pFs, which together constitute the lateral occipital complex (Grill-Spector et al., 1999; for review, see Grill-Spector, 2009). Both pFs and LO are known to respond to high-level shape information (Vinberg and Grill-Spector, 2008); however, compared with LO, pFs is more resistant to various image transformations such as size and position

(Grill-Spector et al., 1999; Kourtzi and Huberle, 2005; for review, see Grill-Spector, 2009). To investigate the role of pFs in object ensemble and surface texture processing, we extracted adaptation responses from pFs (defined in individual observers using the same statistical contrast and threshold as LO). The detailed results are reported in Table 2 and Figure 10. Overall, LO and pFs share some functional similarities in the processing of ensembles and textures (nonsignificant region-by-condition interaction for ensembles: in experiment 1, $F_{(2,22)} = 3.24, p = 0.058$; in experi-

Table 3. Accuracy (percent correct) for the localizer runs

	Objects	Scenes	Faces	Scrambled objects	Object ensembles	Scrambled ensembles	Textures	Scrambled textures
Object/scene localizer								
Experiment 1	92.19 ± 2.39	93.49 ± 1.82	92.71 ± 2.12	91.67 ± 2.96				
Experiments 2 and 3	94.94 ± 1.25	97.32 ± 0.83	96.73 ± 0.99	97.02 ± 1.11				
Experiment 4	94.01 ± 2.41	93.75 ± 3.05	94.27 ± 1.56	94.27 ± 2.72				
Ensemble/texture localizer								
Experiments 2 and 3					92.62 ± 1.90	93.16 ± 1.42	90.18 ± 1.55	93.16 ± 1.55

All reported values represent mean ± SE for percent correct. No difference across conditions reached significance in any of the localizer runs.

Table 4. Accuracy (percent correct) and response latency (milliseconds) of correct trials for experiments 2 through 4

	Object ensembles			Textures		
	Identical	Shared	Different	Identical	Shared	Different
Accuracy (% correct)						
Experiment 2	95.96 ± 1.57	93.70 ± 2.62	97.39 ± 0.63	90.37 ± 3.17	88.20 ± 2.32	95.97 ± 1.30
Experiment 3	96.77 ± 0.65	91.46 ± 2.18	97.77 ± 1.24	89.60 ± 3.10	86.79 ± 2.42	97.39 ± 0.63
Experiment 4	95.46 ± 1.06	93.18 ± 2.10	97.73 ± 0.85			
Response latency (ms)						
Experiment 2	786 ± 45	891 ± 49	836 ± 50	808 ± 39	938 ± 63	858 ± 59
Experiment 3	812 ± 45	930 ± 58	821 ± 46	858 ± 49	954 ± 65	827 ± 55
Experiment 4	915 ± 38	1021 ± 43	885 ± 28			

All values represent mean ± SE. Significant pairwise comparisons (all two tailed and Bonferroni corrected) for accuracy are as follows: textures in experiment 2, shared vs different, $t_{(13)} = 3.83$, $p = 0.006$; textures in experiment 3, shared vs different, $t_{(13)} = 4.53$, $p = 0.002$. Significant pairwise comparisons for response latency are as follows: object ensembles in experiment 2, shared vs identical, $t_{(13)} = 5.05$, $p = 0.001$; textures in experiment 2, shared vs identical, $t_{(13)} = 4.06$, $p = 0.004$; shared vs different, $t_{(13)} = 2.86$, $p = 0.036$; object ensembles in experiment 3, shared vs identical, $t_{(13)} = 5.36$, $p = 0.001$; shared vs different, $t_{(13)} = 3.52$, $p = 0.011$; textures in experiment 3, shared vs identical, $t_{(13)} = 3.01$, $p = 0.03$; shared vs different, $t_{(13)} = 3.34$, $p = 0.017$; object ensembles in experiment 4, shared vs identical, $t_{(10)} = 3.89$, $p = 0.01$; shared vs different, $t_{(10)} = 3.49$, $p = 0.017$.

ment 2, $F_{(2,26)} = 1.53$, $p = 0.236$; in experiment 3, $F_{(2,26)} = 2.36$, $p = 0.114$; in experiment 4, $F_{(2,20)} = 1.25$, $p = 0.308$; nonsignificant region-by-condition interaction for textures: in experiment 2, $F_{(2,26)} = 2.41$, $p = 0.110$; in experiment 3, $F_{(2,26)} = 1.29$, $p = 0.293$), but pFs also demonstrates some sensitivity to processing texture [significant region-by-condition interaction for textures in experiment 1: $F_{(2,22)} = 4.51$, $p = 0.023$; and a basic adaptation effect for textures was observed in experiment 1 (i.e., greater activation in the different compared with the identical condition; see Table 2 and Fig. 10)], a property found in PPA but not in LO. Given that pFs is located between LO and PPA, its role in visual processing may be in the transition from shape-specific processing to ensemble and textural statistic-specific processing. This echoes our discussion (at the end of the Results section titled “Whole-brain analysis of real world object ensemble and texture adaptation”) and is reminiscent of the proposal from a recent study arguing that the region in posterior to mid fusiform gyrus can be thought of as a transition zone between the processing of shapes in LO to the processing of surface properties (i.e., texture and color) in collateral sulcus and parahippocampal gyrus (Cant et al., 2009).

Behavioral results

In both the object/scene and the ensemble/texture localizer runs, observers were asked to detect an occasional spatial jitter of the images. Behavioral results and statistical comparisons among the conditions for the localizer runs are reported in Table 3. In experiment 1, observers were asked to press a response key when the third image in each trial disappeared from view. The overall response accuracy was 97.76% (range, 91.91–100%). Response accuracies were not recorded separately for each stimulus condition, nor were response latencies recorded. In experiments 2 through 4, observers were asked to categorize the pair of images in each trial as identical, shared (ensemble or texture statistics), or different. Behavioral results and statistical comparisons for these experiments are reported in Table 4. Some of the statistical com-

parisons reached significance in the main adaptation experiments; notably, the shared condition in experiments 2 to 4 was often harder than the identical or the different condition. However, behavioral response patterns in the main adaptation experiments did not match the fMRI response patterns, making it unlikely that behavioral responses directly contributed to the observed fMRI results. This is consistent with the findings by Xu et al. (2007), who also showed that fMRI adaptation responses in PPA are dissociable from behavioral responses. Further support of this conclusion is provided by the finding that although different tasks were used in experiment 1 and in experiments 2 to 4, the same fMRI response patterns in PPA and LO were obtained in all of the experiments.

Discussion

Object ensemble perception is an important and adaptive aspect of visual perception that can guide and complement the individuation and encoding of specific objects in a complex visual scene. Yet presently it has largely been unexplored by the neuroimaging community. Here we investigated the neural underpinnings of real-world object ensemble perception. We found that regions in anterior-medial ventral visual cortex, including the collateral sulcus and parahippocampal gyrus, which overlap to a large extent with the scene-sensitive PPA, show fMRI adaptation when ensemble statistics are repeated. This adaptation effect still holds with the removal of color information from the image and with changes in the size of the image. We also found similar adaptation effects for surface textures in this brain region, consistent with previous neuropsychological and fMRI studies on texture processing (Humphrey et al., 1994; Goodale and Milner, 2004; Cant and Goodale, 2007; Cant et al., 2009). Importantly, the object ensemble adaptation observed in this brain region was not driven entirely by the repetition of the surface texture of the individual objects in an ensemble because this region was still sensitive to object ensemble changes when objects in two ensembles had the same texture but differed in shape.

In contrast, regions in the lateral occipital cortex, which overlap with area LO (an object shape-selective area), show a release from fMRI adaptation when local contour changes regardless of whether object ensemble statistics repeat or not and does not show sensitivity to changes in surface texture. This is consistent with previous studies showing this brain region's involvement in shape/contour processing (Malach et al., 1995; Grill-Spector et al., 1998; Kourtzi and Kanwisher, 2001). Although previous studies have largely focused on single objects, here we show that this brain region is also sensitive to changes in the local shape/contour of multiple objects in an ensemble.

Our results are replicated across four different experiments, with both independently localized PPA and LO ROIs and ROIs localized by contrasting ensemble and texture images directly with their phase-scrambled images. Our results are further confirmed by *post hoc* whole-brain analyses.

Contribution of lower-level image statistics and semantic information

Our object ensemble adaptation in anterior-medial ventral visual cortex is unlikely driven by lower-level image statistics such as spatial frequency. First, in experiment 3, we used images that varied in size and obtained virtually the same adaptation results (Fig. 4). Second, in a *post hoc* analysis, using PPA as our ROI, we examined responses to intact object ensembles, intact textures, and their phase-scrambled counterparts from the ensemble/texture localizer. Although spatial frequency and other lower-level image statistics (i.e., luminance and contrast) were preserved between intact and phase-scrambled images, intact images showed significantly greater fMRI responses in PPA than their phase-scrambled counterparts (Fig. 5). Thus, preserving the lower-level image statistic profiles by itself is not sufficient to drive responses in this brain region. Third, although ensembles likely have higher spatial frequency content than textures, they exhibited identical adaptation patterns in PPA and contrasting the two types of images did not activate any region in anterior-medial ventral visual cortex (Fig. 7B). These results further suggest that ensemble/texture representation in anterior-medial ventral visual cortex is not particularly sensitive to differences in spatial frequency. Finally, Hiramatsu et al. (2011) recently demonstrated that the processing of material properties in anterior-medial ventral visual cortex (which is similar to the processing of surface texture; Cant and Goodale, 2011) is based on high-level perceptual information rather than low-level, imaged-based information. Taken together, the processing of ensembles and textures in our study is likely based more on high-level (e.g., object shape and texture statistics) than low-level visual information.

When object ensemble statistics repeat, the semantic label (i.e., its name and semantic category) also repeats. Nonetheless, it is unlikely that the repetition of such semantic labels can account for our adaptation results. First, although object ensembles have more salient semantic information (i.e., nameable identities) than surface textures, identical adaptation results were obtained for both image types. Second, in our final experiment, in which we used black wooden beads that varied only in shape and thus had similar semantic labels, we still obtained robust sensitivity to object ensemble changes in PPA. Third, Epstein et al. (2003) have demonstrated that PPA is sensitive to viewpoint changes of a scene, even though the semantic content of the scene remains the same.

Object ensemble, surface texture, and scene processing in PPA

As we discussed earlier, real-world homogeneous object ensembles and surface textures both contain multiple repeating elements (Portilla and Simoncelli, 2000), and the processing of both requires the extraction of summary statistics without encoding each repeating element in great detail. This may explain why both types of images share a common neural processing mechanism in anterior-medial ventral visual cortex, as we observed in our study.

PPA is well known to represent scenes by processing their 3D spatial layouts (Epstein and Kanwisher, 1998; Kravitz et al., 2011). Given that surface textures and object ensembles contain minimal 3D spatial layouts and in general do not invoke scene imagery, why would these stimuli share a common processing mechanism with visual scenes? In addition to encoding 3D space, scene perception often involves processing the gist of a scene by extracting abstract information without representing in great detail the individual objects composing the scene (Oliva and Schyns, 2000; Oliva and Torralba, 2001). In this regard, scene perception also involves the extraction of ensemble statistical information. This may explain why the processing of object ensembles, surface textures, and scenes all activate this common region. Thus, studying the neural underpinnings of object ensemble representation holds the possibility of bridging distinct lines of research and allows better understanding of the role of anterior-medial ventral visual cortex in visual perception. This approach is especially promising given that object ensembles are relatively easier to create and manipulate than either surface textures or visual scenes.

Although it is possible that the processing of the 3D scene structure and the processing of visual statistics may be two distinctive forms of visual processing that both involve PPA, their colocalization within PPA may not be accidental. The 3D structure of a scene may just be another form of high-order ensemble statistic (e.g., one may be able to extract the global 3D structure of a city scene containing many buildings without processing the details of each specific building in the scene) such that anterior-medial ventral visual cortex (where PPA is located) may play a more general role of extracting higher-order statistical information from any visual display. Further research is needed to fully understand the connection between the processing of a 3D scene structure and that of visual statistics in anterior-medial ventral visual cortex.

What may be represented during object ensemble processing?

A growing body of behavioral research has shown that observers can quickly extract useful ensemble statistics, such as mean size, speed, and orientation, from a display without encoding specific details of the objects composing the display (Williams and Sekuler, 1984; Watamaniuk and Duchon, 1992; Ariely, 2001; Parkes et al., 2001; Chong and Treisman, 2003, 2005a,b; Alvarez and Oliva, 2008, 2009; Alvarez, 2011). It is very possible that anterior-medial ventral visual cortex participates in such computations, although further experiments are needed to verify this.

It is also possible that anterior-medial ventral visual cortex is sensitive to second-order image statistics, such as changes in the luminance or geometric properties of the elements composing the object ensemble or surface texture. Other features such as element density, the arrangement of the elements, and the homogeneity of the elements may also be diagnostic and essential to the representation of object ensembles and surface textures in this brain region. Our findings thus provide exciting opportunities

for future research to systematically explore the various factors that contribute to the neural computations performed in this brain region.

Two independent and complementary pathways for visual object processing

Together with previous neuropsychological and neuroimaging studies, our results indicate that there are two independent and complementary visual object processing mechanisms in the brain. One such mechanism involves anterior-medial ventral visual cortex (encompassing the collateral sulcus and parahippocampal gyrus) and is specialized in extracting summary statistics from both object ensembles and surface textures without encoding the detailed features of each element composing the ensemble or texture. The other mechanism involves the lateral occipital cortex, which, together with regions in the parietal cortex (Xu and Chun, 2009), enables us to attend, individuate, and encode the detailed shape features of the individual objects in an ensemble. Together, these two object processing mechanisms allow an observer to perceive both the “individual trees” and the “entire forest” from a visual scene.

To conclude, our study showed that anterior-medial ventral visual cortex is involved in real-world object ensemble perception, an important but largely unexplored aspect of visual perception that complements object-specific representation. Moreover, object ensemble representation is closely related to surface texture and visual scene representation. Understanding object ensemble representation in the brain is thus both important and necessary if we want to fully comprehend how object perception from natural scenes is accomplished.

References

- Altmann CF, Bühlhoff HH, Kourtzi Z (2003) Perceptual organization of local elements into global shapes in the human visual cortex. *Curr Biol* 13:342–349.
- Alvarez GA (2011) Representing multiple objects as an ensemble enhances visual cognition. *Trends Cogn Sci* 15:122–131.
- Alvarez GA, Oliva A (2008) The representation of simple ensemble visual features outside the focus of attention. *Psychol Sci* 19:392–398.
- Alvarez GA, Oliva A (2009) Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proc Natl Acad Sci U S A* 106:7345–7350.
- Andrews TJ, Ewbank MP (2004) Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage* 23:905–913.
- Ariely D (2001) Seeing sets: representation by statistical properties. *Psychol Sci* 12:157–162.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Cant JS, Goodale MA (2007) Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cereb Cortex* 17:713–731.
- Cant JS, Goodale MA (2011) Scratching beneath the surface: new insights into the functional properties of the lateral occipital area and parahippocampal place area. *J Neurosci* 31:8248–8258.
- Cant JS, Arnott SR, Goodale MA (2009) fMR-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream. *Exp Brain Res* 192:391–405.
- Cavina-Pratesi C, Kientz RW, Heywood CA, Milner AD (2010a) Separate processing of texture and form in the ventral stream: evidence from fMRI and visual agnosia. *Cereb Cortex* 20:433–446.
- Cavina-Pratesi C, Kientz RW, Heywood CA, Milner AD (2010b) Separate channels for processing form, texture, and color: evidence from fMRI adaptation and visual object agnosia. *Cereb Cortex* 20:2319–2332.
- Chong SC, Treisman A (2003) Representation of statistical properties. *Vision Res* 43:393–404.
- Chong SC, Treisman A (2005a) Attentional spread in the statistical processing of visual displays. *Percept Psychophys* 67:1–13.
- Chong SC, Treisman A (2005b) Statistical processing: computing the average size in perceptual groups. *Vision Res* 45:891–900.
- Dilks DD, Julian JB, Kubilius J, Spelke ES, Kanwisher N (2011) Mirror-image sensitivity and invariance in object and scene processing pathways. *J Neurosci* 31:11305–11312.
- Epstein RA (2008) Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci* 12:388–396.
- Epstein RA, Higgins JS (2007) Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb Cortex* 17:1680–1693.
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601.
- Epstein RA, Higgins JS, Thompson-Schill SL (2005) Learning places from views: variation in scene processing as a function of experience and navigational ability. *J Cogn Neurosci* 17:73–83.
- Epstein R, Graham KS, Downing PE (2003) Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron* 37:865–876.
- Ewbank MP, Schluppeck D, Andrews TJ (2005) fMR-adaptation reveals a distributed representation of inanimate objects and places in human visual cortex. *Neuroimage* 28:268–279.
- Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ (1994) Statistical parametric maps in functional imaging: a general linear model approach. *Hum Brain Mapp* 2:189–210.
- Goodale MA, Milner AD (2004) *Sight unseen: an exploration of conscious and unconscious vision*. Oxford: Oxford UP.
- Grill-Spector K (2009) What has fMRI taught us about object recognition? In: *Object categorization: computer and human vision perspectives* (Dickinson S, Leonardis A, Schiele B, Tarr M, eds), pp 102–128. Cambridge: Cambridge UP.
- Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzhak Y, Malach R (1998) A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp* 6:316–328.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203.
- Grill-Spector K, Kushnir T, Hendler T, Malach R (2000) The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 3:837–843.
- Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10:14–23.
- Hiramatsu C, Goda N, Komatsu H (2011) Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *Neuroimage* 57:482–494.
- Humphrey GK, Goodale MA, Jakobson LS, Servos P (1994) The role of surface information in object recognition: studies of a visual form agnostic and normal subjects. *Perception* 23:1457–1481.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Kourtzi Z, Huberle E (2005) Spatiotemporal characteristics of form analysis in the human visual cortex revealed by rapid event-related fMRI adaptation. *Neuroimage* 28:440–452.
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* 293:1506–1509.
- Kravitz DJ, Peng CS, Baker CI (2011) Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *J Neurosci* 31:7322–7333.
- Kung CC, Peissig JJ, Tarr MJ (2007) Is region-of-interest overlap comparison a reliable measure of category specificity? *J Cogn Neurosci* 19:2019–2034.
- Lee Y, Matsumiya K, Wilson HR (2006) Size-invariant but viewpoint-dependent representation of faces. *Vision Res* 46:1901–1910.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135–8139.
- Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A* 89:5951–5955.
- Oliva A, Schyns PG (2000) Diagnostic colors mediate scene recognition. *Cogn Psychol* 41:176–210.

- Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int J Comput Vis* 42:145–175.
- Parkes L, Lund J, Angelucci A, Solomon JA, Morgan M (2001) Compulsory averaging of crowded orientation signals in human vision. *Nat Neurosci* 4:739–744.
- Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10:437–442.
- Peuskens H, Claeys KG, Todd JT, Norman JF, Van Hecke P, Orban GA (2004) Attention to 3-D shape, 3-D motion, and texture in 3-D structure from motion displays. *J Cogn Neurosci* 16:665–682.
- Portilla J, Simoncelli EP (2000) A parametric texture model based on joint statistics of complex wavelet coefficients. *Int J Comput Vis* 40:49–71.
- Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: a defense of functional localizers. *Neuroimage* 30:1088–1096.
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. New York: Thieme Medical Publishers.
- Todd JJ, Han SW, Harrison S, Marois R (2011) The neural correlates of visual working memory encoding: a time-resolved fMRI study. *Neuropsychologia* 49:1527–1536.
- Vinberg J, Grill-Spector K (2008) Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. *J Neurophysiol* 99:1380–1393.
- Watamaniuk SN, Duchon A (1992) The human visual system averages speed information. *Vision Res* 32:931–941.
- Williams DW, Sekuler R (1984) Coherent global motion percepts from stochastic local motions. *Vision Res* 24:55–62.
- Xu Y (2010) The neural fate of task-irrelevant features in object-based processing. *J Neurosci* 30:14020–14028.
- Xu Y, Chun MM (2006) Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature* 440:91–95.
- Xu Y, Chun MM (2009) Selecting and perceiving multiple visual objects. *Trends Cogn Sci* 13:167–174.
- Xu Y, Turk-Browne NB, Chun MM (2007) Dissociating task performance from fMRI repetition attenuation in ventral visual cortex. *J Neurosci* 27:5981–5985.