

Neural Representations of Contextual Guidance in Visual Search of Real-World Scenes

Tim J. Preston,^{1,2} Fei Guo,^{1,2} Koel Das,^{1,2,3} Barry Giesbrecht,^{1,2} and Miguel P. Eckstein^{1,2}

¹Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, California 39106-9660, ²Institute for Collaborative Biotechnologies, University of California, Santa Barbara, Santa Barbara, California 93106-5100 and ³Indian Institute of Science Education and Research, Kolkata

Exploiting scene context and object–object co-occurrence is critical in guiding eye movements and facilitating visual search, yet the mediating neural mechanisms are unknown. We used functional magnetic resonance imaging while observers searched for target objects in scenes and used multivariate pattern analyses (MVPA) to show that the lateral occipital complex (LOC) can predict the coarse spatial location of observers' expectations about the likely location of 213 different targets absent from the scenes. In addition, we found weaker but significant representations of context location in an area related to the orienting of attention (intraparietal sulcus, IPS) as well as a region related to scene processing (retrosplenial cortex, RSC). Importantly, the degree of agreement among 100 independent raters about the likely location to contain a target object in a scene correlated with LOC's ability to predict the contextual location while weaker but significant effects were found in IPS, RSC, the human motion area, and early visual areas (V1, V3v). When contextual information was made irrelevant to observers' behavioral task, the MVPA analysis of LOC and the other areas' activity ceased to predict the location of context. Thus, our findings suggest that the likely locations of targets in scenes are represented in various visual areas with LOC playing a key role in contextual guidance during visual search of objects in real scenes.

Introduction

Successful visual search is paramount to the survival of animals and is important in the daily life of humans. The human brain has implemented a variety of strategies to optimize visual search (Najemnik and Geisler, 2005; Navalpakkam et al., 2010; Eckstein, 2011; Ma et al., 2011). Uncertainty about the locations of targets and the presence of distractors slows search and often leads to search errors (Swenson and Judy, 1981; Wolfe, 1998; Palmer et al., 2000). Utilization of global scene properties as well as highly visible objects predictive of the target are critical for humans in guiding and facilitating search (Eckstein et al., 2006; Neider and Zelinsky, 2006; Torralba et al., 2006; Ehinger et al., 2009; Castelano and Heaven, 2010; Wolfe et al., 2011).

Low-level features that determine the spatial layout of a scene and are predictive of the target location will guide eye movements. For example, when searching for pedestrians, humans bias their eye movements toward areas of scenes containing features corresponding to sidewalks and pavement (global scene properties; Torralba et al., 2006). In addition, objects that frequently co-occur with target objects will also guide search

(Castelano and Heaven, 2011; Mack and Eckstein, 2011). If searching for a chimney, humans will move their eyes toward the roof of a house even if the chimney is absent from the scene.

The neural basis of contextual guidance during visual search of real-world scenes is unknown. Numerous studies have shown that there are various brain areas that encode scene information, objects, and contextual relationships across objects. Both parahippocampal cortex (PHC) and retrosplenial cortex (RSC) (Epstein and Kanwisher, 1998; Yi and Chun, 2005; Epstein and Higgins, 2007; Henderson et al., 2008; Greicius et al., 2009) preferentially respond to scenes relative to faces or single objects. The lateral occipital area (LOC) encodes object information that is invariant to size, position, and viewpoint (Grill-Spector et al., 2001; Kourtzi and Kanwisher, 2001; Konen and Kastner, 2008). Importantly, PHC has been implicated in encoding contextual associations across objects when these are viewed simultaneously (Bar and Aminoff, 2003; Bar, 2004; Bar et al., 2008). Critically, the use of contextual information to guide saccades and modulate covert attention requires that the brain represent the spatial location within the scene, which is predictive of a searched target that might not even be present in the scene. Areas in the frontoparietal attention network, which include the frontal and supplementary eye fields (FEF, SEF), and an area in the intraparietal sulcus (IPS) correlate with deployments of covert attention and target-related activity during visual search with synthetic displays (Corbetta and Shulman, 2002; Ptak, 2012), but their ability to code the contextual location of objects and features in real scenes has not been demonstrated. Thus, no study to our knowledge has identified an area that can flexibly encode the locations likely to contain a searched object in real scenes. Here, we use functional

Received Dec. 20, 2012; revised Feb. 27, 2013; accepted March 11, 2013.

Author contributions: T.J.P., F.G., B.G., and M.P.E. designed research; T.J.P. and F.G. performed research; K.D. contributed unpublished reagents/analytic tools; T.J.P. and F.G. analyzed data; T.J.P., B.G., and M.P.E. wrote the paper.

This work was supported by National Institutes of Health grant R21EY023097 and Army grant W911NF-09-D-0001.

The authors declare no competing financial interests.

Correspondence should be addressed to Miguel P. Eckstein, Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106-9660. E-mail: miguel.eckstein@psych.ucsb.edu.

DOI:10.1523/JNEUROSCI.5840-12.2013

Copyright © 2013 the authors 0270-6474/13/337846-10\$15.00/0

magnetic resonance imaging (fMRI) in conjunction with multivariate pattern analyses (MVPA) while observers searched for targets in real scenes to identify brain regions representing the likely location of the target.

Materials and Methods

Subjects

Twelve observers (four male and eight female; mean age 24; range 19–31) from the University of California, Santa Barbara (UCSB), participated in Experiment 1. Six additional observers (four male and two female; mean age 22; range 21–25) participated in a saliency control experiment (Experiment 2). All observers were naive to the purpose of the experiments. All observers had normal or corrected to normal vision (based on self-report). Before participation, participants provided written informed consent that had been approved by the UCSB Human Subjects Committee. The observers were reimbursed for their participation with course credit or money.

Stimuli

A set of 640 images of natural scenes (both indoor and outdoor) were used as stimuli. Each image subtended a viewing angle of 17.5° (727×727 pixels). An LCD video projector (Hitachi CPX505, 1024×768 resolution) inside a Faraday cage in the scanner room was used to display images on a rear projection screen behind the head coil inside the magnet bore (peak luminance: 455 cd/m^2 , mean luminance 224 cd/m^2 minimum luminance: 5.1 cd/m^2). Observers viewed the screen via a mirror angled at 45° attached to the head coil above their heads. The viewing distance was 108 cm providing a 24.6° field of view.

Experimental Design

Experiment 1

Observers ($n = 12$) were presented with 640 images of indoor and outdoor scenes. Before the presentation of each image a cue word was presented for 400 ms specifying the target that observers had to search (e.g., cup, boat, bicycle, etc.; for a total of 384 different targets across all images; see Fig. 1A). Following 300 ms of fixation, an image scene was presented (250 ms) that contained the target object in 50% of the trials (see Fig. 1B, C for sample images). Subsequent to a period of fixation lasting 250–700 ms, observers were required to indicate whether they believed the target was present in the scene stimulus using an 8-point confidence scale.

Each cue word corresponded to an object present in the image or an object absent from the image with the two conditions split equally. All targets (for both the target-present and target-absent conditions) were semantically consistent with the scene. The spatial location and appearance of the target objects in the scenes were highly variable. The observers had no prior information of the size or position of the target object in a single image, and, as a result, the presence/absence of a target object was not predictable. Among the target-present images, $\sim 50\%$ of the images contained the target on the left side and 50% on the right side. The experimental design consisted of a 2×2 factorial design with one factor of target with target absent or present and the other of cued location left or right. In addition, another condition consisted of a period of central fixation for the duration of the trial (these fixation trials were included to add jitter to the onsets of the experimental trials).

Each run of the fMRI experiment lasted 6 min and 46.62 s (251 TRs) and consisted of 4 TRs of initial and final fixation, 32 trials of target-present images, 32 trials of target-absent images, and 16 trials of fixation. The order of the trials was pseudorandomly constrained so that each trial type had a matched trial history one trial back (over all trials in the run) (Buracas and Boynton, 2002). Thus, an additional trial was included for one of the conditions (depending on the trial design) at the start of each run to give the first real trial a matched history (i.e., on each experimental run one trial type had 17 repetitions with the first trial being discarded).

Observers indicated their confidence that the cued target object was present or absent in the scene using an 8-point confidence scale by pressing one of the buttons of the response boxes (Lumina fMRI Response Pads; Cedrus). The 8-point rating scale consisted of the four buttons on

the response pad for each hand. We controlled the effect of motor activity associated with subjects' behavioral response; the response mapping between present or absent and each hand was reversed depending on which response cue was displayed. The response cues consisted of either a red square or a green triangle, and were randomized over trials with and equal probability for either cue type. The association between the response hand and the response cue was counterbalanced across observers.

Experiment 2

Saliency control experiment. We controlled for the low-level physical properties between the left-target-absent and right-target-absent image sets with a control experiment that used the same image set but with the cue word replaced by a fixation that persisted until the scene stimulus was displayed. Rather than searching for a target object, observers were instructed to search for salient regions within the image and reported which side of the scene stimulus (left or right visual field) they believed contained the most salient feature. Responses were recorded with a single button press either with the left or the right hand. As with the main experiment the hand the observers used to respond was determined by a response cue presented after the scene stimulus. The saliency search task was conducted to control for the possibility of a correlation between saliency and the location of the target object in the main study.

Definition of retinotopic regions. We identified regions of interest (ROIs) for each subject individually using standard localizer scans in conjunction with each subject's anatomical scan, which were both acquired in a single scanning session lasting ~ 1 h 45 min. These ROIs included areas in early visual cortex, higher level visual cortex, and the frontoparietal attention network. The ROIs in retinotopic visual cortex were defined using a rotating wedge and expanding concentric ring checkerboard stimulus in two separate scans (Sereno et al., 1995; DeYoe et al., 1996). Each rotating wedge and concentric ring scan lasted 512 s consisting of 16 s of initial and final fixation with eight full rotations of the wedge or eight full expansions of the ring stimulus, each lasting 64 s. Wedge stimuli in the polar mapping scan had a radius of 14° and subtended 75° consisted of alternating black and white squares in a checkerboard pattern. The colors of the checkerboard squares flickered rapidly between black and white at 4 Hz to provide constant visual stimulation. To ensure that subjects maintained central fixation, a demanding fixation task in which a gray dot at fixation darkened for 200 ms at pseudorandom intervals (with subjects indicating with a button press when this occurred) was used. The eccentricity mapping procedure was very similar except that a ring expanding from fixation (ring width 4°) was used instead of a rotating wedge. By correlating the blood oxygenation level-dependent (BOLD) response resulting from the activation caused by the wedge and ring stimuli we determined which voxels responded most strongly to particular regions in visual space and produced both polar and eccentricity maps on the surface of visual cortex that were mapped onto meshes of the individual subject's anatomy.

The center of V1 overlays the calcarine sulcus and represents the whole contralateral visual field with its edges defined by changes in polar map field signs designating the start of V2d and V2v. Areas V2d, V2v, V3d, and V3v all contain quarter-field representations with V3d adjacent to V2d and V3v adjacent to V2v. Area hV4 contains a full-field representation and shares a border with V3v (Tootell and Hadjikhani, 2001; Tyler et al., 2005). The anterior borders of these regions were defined using the eccentricity maps. Area V3A and V3B are dorsal and anterior to area V3d, with which they share a border, and contain a full hemifield representation of visual space. To separate them it is necessary to refer to the eccentricity map, which shows a second foveal confluence at the border between V3A and V3B (Tyler et al., 2005).

Definition of functionally defined visual areas. We used functional localizers in a separate scanning session to identify several visual areas including: human motion area (hMT+)/V5, LOC, the fusiform face area (FFA), and the parahippocampal place area (PPA). The PPA is referred to in the rest of the text as parahippocampal cortex (PHC) reflecting recent studies showing more generalized responses to contextual associations rather than scenes alone (Bar, 2004; Bar et al., 2008). We also defined areas implicated in the processing of eye movements and spatial attention: the FEF and SEF and an area in IPS. Area hMT+ /V5 was defined as

the set of voxels in lateral temporal cortex that responded significantly higher ($p < 10^{-4}$) to a coherently moving array of dots than to a static array of dots (Zeki et al., 1991). The scan lasted for 376 s consisting of 8 s of initial and final fixation with 18 repeats of 20 s blocks. There were three block types consisting of black dots on a mid gray background viewed through a circular aperture. Dots were randomly distributed within the circular aperture, which had a radius of 13° with a dot density of 20 dots/deg². During the moving condition all dots moved in the same direction with a speed of $3^\circ/s$ for 1 s before reversing direction. In the edge condition strips of dots (width 2°) moved with opposite motion to each other creating kinetically defined borders. To define area hMT+/V5 a general linear model (GLM) analysis was performed and the activation resulting from the contrast moving > stationary dots was used to define the region constrained by individual anatomy to an area within the inferior temporal sulcus.

The LOC, FFA, and PPA localizers were combined into a single scan to maximize available scanner time. In the combined localizer scan, referred to as the LFP localizer, the scan duration was 396 s consisting of three, 12 s periods of fixation at the beginning, end, and the middle of the run, and five repeats of the four experimental conditions each lasting 18 s. These conditions were as follows: intact objects, phase scrambled objects, intact faces, and intact scenes. During the 18 s presentation period, each stimulus was presented for 300 ms followed by 700 ms of fixation before the next stimulus presentation. To maximize statistical power, the LFP scan was run twice for each individual using different trial sequences. To define areas of functional activity a GLM analysis was performed on the two localizer scans. Area LOC was defined as the activation revealed by the intact objects > scrambled objects contrast (Kourtzi and Kanwisher, 2001). The FFA region was isolated by the face stimuli > intact object stimuli contrast (Kanwisher et al., 1997). The PPA regions by the scenes > faces + objects contrast (Epstein and Kanwisher, 1998). All localizer regions were guided by the known anatomical features of these areas reported by previous groups.

We localized brain areas in the frontoparietal attention network by adapting an eye-movement task developed by Connolly et al. (2002). The task consisted of eight repeats of two blocks each lasting 20 s: one where a fixation dot was presented centrally and a second where every 500 ms the fixation dot was moved to the opposite side of the screen. During the moving dot condition, the dot could be positioned anywhere along a horizontal line perpendicular to the vertical meridian and between 4 and 15° away from the screen center. Subjects were required to make saccades to keep the moving dot fixated. Contrasting the two conditions in a GLM analysis revealed activation in the FEF, SEF, and a region in the dorsal IPS thought to be a candidate for a putative human lateral intraparietal area (LIP; Connolly et al., 2002).

We defined a control ROI centered on each subject's hand motor area from BOLD activity resulting from hand use identified using a GLM contrasting all task trials versus fixation trials. During fixation trials subjects were not required to respond and so there would be no hand use. The hand area ROIs were then defined about the activity resulting from this contrast in the central sulcus (primary motor cortex) guided by reference to studies directly investigating motor activity resulting from hand or finger use (Lotze et al., 2000; Alkadhi et al., 2002).

fMRI data acquisition

Data were collected at the UCSB Brain Imaging Center using a 3 T TIM Trio Siemens Magnetom with a 12-channel phased-array head coil. An echo-planar sequence was used to measure BOLD contrast (TR = 1620 ms; TE = 30 ms; flip angle = 65° ; FOV = 192 mm; slice thickness = 3.5 mm, matrix = 78×78 , 29 axial slices) for experimental runs. For one subject we used a sequence with a 210 mm FOV (with all other parameters remaining constant) to accommodate the subject's larger brain. Localizer scans used a higher resolution sequence (TR = 2000 ms; TE = 35 ms; flip angle = 70° ; FOV = 192 mm; slice thickness = 2.5 mm, matrix = 78×78 , 30 coronal slices). A high-resolution T1-weighted MPRAGE scan (1 mm^3) was also acquired for each participant (TR = 2300 ms; TE = 2.98 ms; flip angle = 9° ; FOV = 256 mm; slice thickness = 1.1 mm, matrix = 256×256).

fMRI data analysis

We used FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>) to process each observer's anatomical scans to determine the gray-white matter and gray matter–pial boundaries, which were then used to reconstruct inflated and flattened 3D surfaces. We preprocessed functional data using FSL 4.1 (<http://www.fmrib.ox.ac.uk/fsl/>) to perform 3D motion correction, alignment to individual anatomical scans, high-pass filtering (3 cycles per run), and linear trend removal. No spatial smoothing was performed on the functional data used for the multivariate analysis to avoid washing out variability between voxels. We used SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>) to perform GLM analyses of the localizer and experimental scans. ROIs were defined on the inflated mesh and projected back into the space of the functional data.

Multivoxel pattern analysis

MVPA has been successfully applied to fMRI data to evaluate the information context of multivoxel activation patterns in targeted brain regions (Sapuro and Serences, 2010; Weil and Rees, 2010). Here we used regularized linear discriminant analysis (LDA; Duda et al., 2000) to classify the patterns of fMRI data within each ROI. The regions we investigated included V1, V2d, V2v, V3A, V3B, V3d, V3v, hV4, LOC, hMT+/V5, RSC, FFA, PHC, FEF, SEF, and IPS. We used the union of corresponding ROIs in the left and right hemispheres to construct a single bilateral area for each ROI. We normalized (*z*-score) each voxel time course separately for each experimental run to minimize baseline differences between runs and different ROIs. The initial data vectors for the multivariate analysis were then generated by shifting the fMRI time series by 3 TRs to account for the hemodynamic response lag. All classification analysis was performed on the mean of the three data points collected for each trial with the first data point taken from the start of the trial when the cue word is presented. For each single trial, the output of the multivariate pattern classifier was a single scalar value generated based on the weighted sum of the input values across all voxels in one specific ROI. We used a leave-one-out cross-validation scheme across runs. The classifier learned a function that mapped between voxel activity patterns and experimental conditions from 9 of the 10 runs. Given a new pattern of activity from a single trial in the left-out testing run, the trained classifier determined whether this trial belonged to context left versus context right target-absent condition.

To improve classification accuracy we also investigated the effect of taking the mean of eight trials per run and training and classifying on the resulting mean data. Averaging trials before classification can boost discrimination performance at the cost of increased variance. In this case we took the mean of eight trials per runs resulting in two training/test patterns per condition per run (for the data this resulted in 40 patterns over all conditions and runs). As the number of ways that the eight trials, which were included in the averaging, has a large number of permutations, a permutation analysis was conducted in which a random sampling of eight trials (without replacement) was used to generate the mean patterns. To obtain the final classification result the mean of 1000 permutation analyses was taken.

Eye-tracking data collection and analysis

Nine of the 12 observers had their eye position recorded during the fMRI experimental scans using an Eyelink 1000 eye tracker (Eyelink). This system uses fiber optics to illuminate the eye with infrared light and tracks the eye orientation using the pupil position relative to the corneal reflection. Observers completed a 9-point calibration before the first experimental run, and repeated the calibration before subsequent runs if the calibration began to degrade due to head motion. All visual stimuli presented on the screen were within the limits of the calibration region. An eye movement was recorded as a saccade if both velocity and acceleration exceeded a threshold (velocity > $30^\circ/s$; acceleration > $8000^\circ/s^2$). A saccade outside an area extending 1° from the fixation was considered an eye movement away from fixation. Data collected during the whole trial of 4.86 s were analyzed, starting from the frame corresponding to the onset of the stimuli display of one trial until the frame of the end of a trial (before the display onset of the following trial). An interval of 4.86 s was

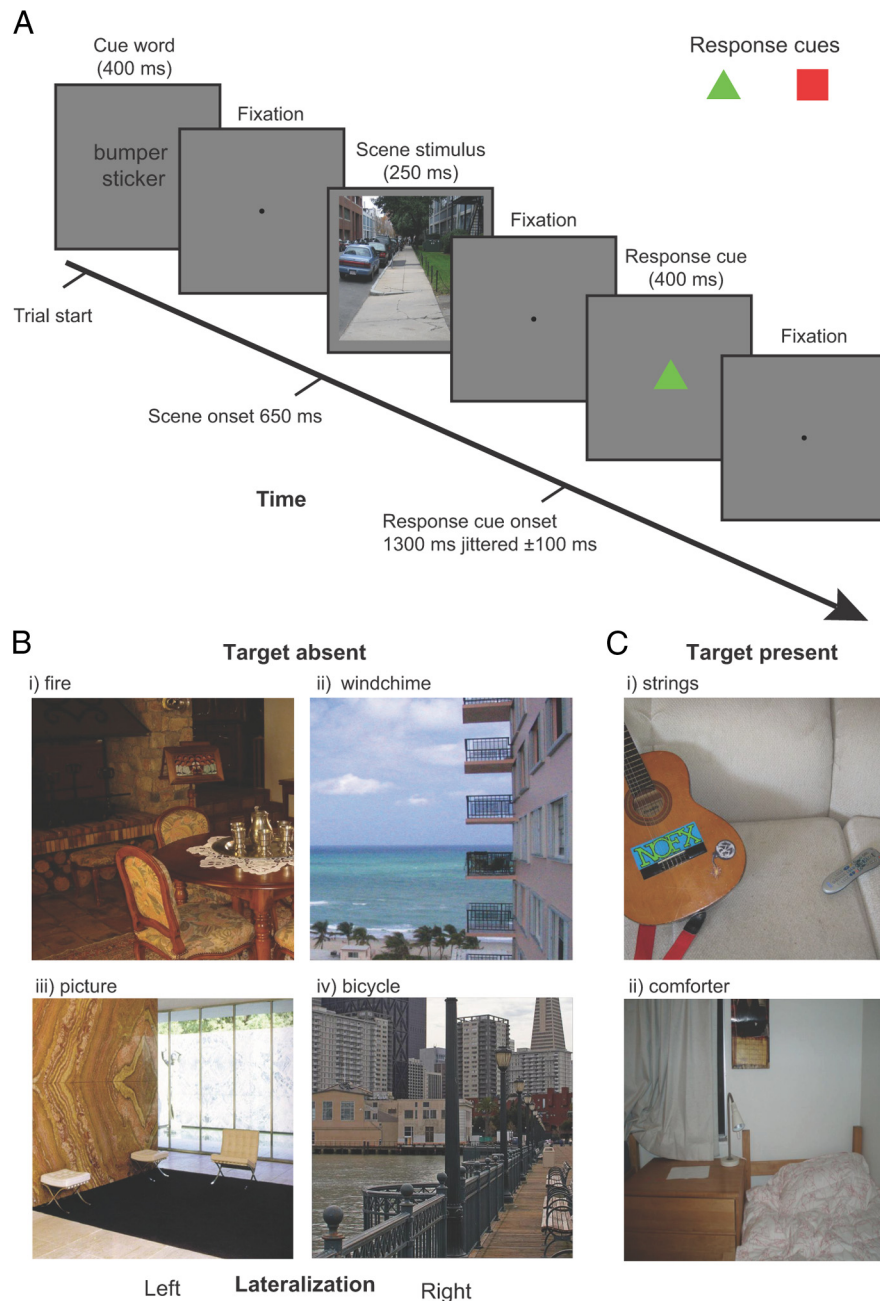


Figure 1. Experimental design. **A**, Example time line of a single trial. Observers responded with their decision confidence whether the cued target object was present or absent using an 8-point confidence scale. To dissociate decision from motor responses we presented a color cue (red square or green triangle), randomized on a trial-to-trial basis, which indicated the mapping between the observer’s hands and the confidence rating. **B**, Target-absent images were deliberately selected to contain a contextual location that was lateralized either toward the left (*i, iii*) or right (*ii, iv*) of the image. **C**, Examples of target present stimuli.

examined for any eye movements occurring within the trial, which could potentially affect the BOLD signal, starting with the onset of the cue word and including the display and response periods. The first analysis examined three measures across left-target-absent and right-target-absent conditions: (1) the mean total number of saccades, (2) the mean of saccade amplitude (in degrees of visual angle), and (3) the SD of distance of eye position from central fixation on a trial. The second analysis examined the measure of average distance from eye position to central fixation, which is an absolute value of relevant eye positions. We used the average distance of horizontal and vertical coordinates of eye positions to the horizontal and vertical coordinates of the center fixation as a 2D

input to the pattern classifier to discriminate the presence or absence of the target objects in the natural scenes.

Results

To identify brain areas coding the spatial location of contextually relevant objects/scene regions, 12 observers searched for arbitrary targets in real-world scenes while we recorded neural activity using fMRI in an event-related design. We show that using MVPA (Tong and Pratte, 2012) we can predict from BOLD activity in LOC whether the likely location of 213 different targets, which themselves are absent from the scene, were to the right or left from the vertical centerline of the images. LOC activity could be also related to the degree of contextual guidance inherent to an image as judged by independent raters. The ability to predict the relevant contextual location from LOC activity correlated with whether a larger percentage of independent human raters agreed on the expected location for the searched targets. Areas IPS and RSC also showed some degree of encoding of scene context during search.

Experiment 1

We presented observers ($n = 12$) with 640 images of indoor and outdoor scenes. Before the presentation of each image a cue word was presented specifying the target that observers had to search for (Fig. 1A). Following 300 ms of fixation, an image scene was presented (250 ms) that contained the target object in 50% of the trials (Fig. 1B,C). Observers were then required to indicate whether they believed the target was present in the scene stimulus using an 8-point confidence scale. Mean behavioral performance for target detection was 80.0%; significantly greater than chance (50%) performance ($t_{(11)} = 26.5$; $p \ll 0.05$). Importantly, unknown to participants, target-absent images were deliberately selected to contain a contextual location that was lateralized either toward the left or right half of the image (Fig. 1B). An independent measure of human expectations about contextual locations was obtained by having 100 separate observers select the most likely target location for each target-absent image. Figure 4A shows examples of selections (yellow dots) for two images. We assigned a context label (left or right) based on the image hemifield containing the majority of selections from the 100 observers. Figure 1C shows examples of target present images.

Decoding left versus right contextual locations in scenes

We applied an MVPA (regularized LDA, RLDA; Duda et al., 2000) to the fMRI voxels from ROIs defined by retinotopic and functional localizer scans (Fig. 2) to predict the labels of the hemifield (left or

right) contextually relevant to the cued targets that were absent in the images. Figure 3A shows the MVPA performance for a single trial analysis (area under the receiver operating characteristic curve, AUC) discriminating between the left and right contextual locations for single trials. Of all the ROIs, LOC showed the highest discrimination performance followed by RSC and IPS. Only LOC reached statistical significance ($p < 0.01$, false discovery rate corrected; FDR). Figure 3B also shows MVPA performance after averaging the fMRI response of eight trials. Results are consistent with the single trial analysis, but with trial averaging performance in RSC and IPS also reached statistical significance ($p < 0.01$ FDR).

Correlation between inherent contextual guidance in an image and observer ensemble neural activity

We also investigated whether there was a relationship between the amount of contextual guidance inherent in individual images and the neural activity that each of these images elicit across observers. We quantified the inherent contextual guidance in an image by using the context selections of the 100 independent observers. If the activity in LOC for a given image reflects inherent amounts of contextual guidance in that image, then we might expect that certain images for which the majority of the 100 independent raters agreed on the likely target location might elicit a consistent neural response in LOC across a majority of observers' brains. We took the proportion of context selections from the observers falling into one image side (right vs left from the vertical center line) as a quantitative measure of the inherent contextual guidance of the image (e.g., 0.6 of the selections falling in the right hemifield of the image would be low inherent guidance content). We correlated the proportion of context selections with the proportion of MVPAs from all observers predicting a contextual location on the right side of the image (see Fig. 4 for LOC). Figure 4B shows that there is a reliable relationship between the ensemble brain activity (as quantified by the MVPA) and the amount of contextual guidance as defined by the proportion of lateralized selections. Even though the correlations reached statistical significance for a number of areas (V1, V3v, LOC, hMT+, RSC, IPS) the areas showing the strongest correlations (LOC, RSC, and IPS) are the same as those which give significant above chance prediction of contextually relevant image hemifield (Fig. 3).

Scene context for open space versus closed space scenes

Previous studies have shown that some areas can discriminate scene type and also whether the scene contains an open space versus a closed space (Kravitz et al., 2011; Park et al., 2011) In this analysis we evaluated whether there was interaction between the areas coding context and the scene type (open vs closed). We used 10 independent observers to rate our scenes as open or closed. Of the 320 target absent images, 162 were rated as closed, 158 were considered open scenes. We verified using MVPA that, as in pre-

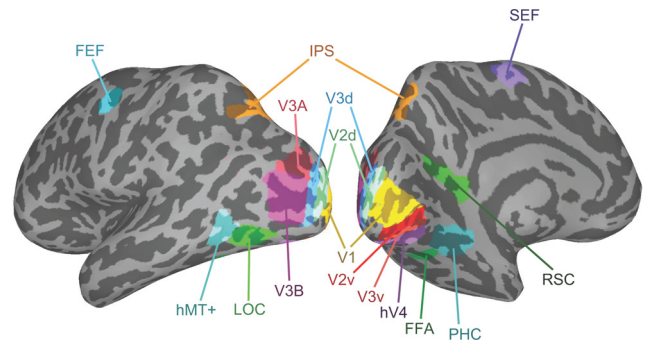


Figure 2. ROIs defined by retinotopic and functional localizer scans.

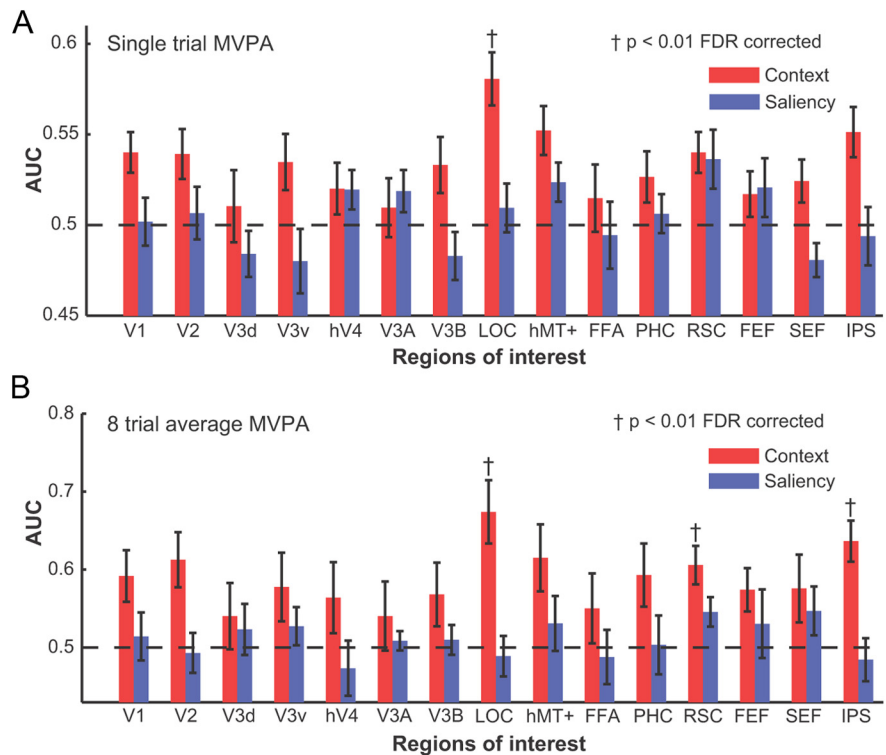


Figure 3. MVPA discrimination performance between the left and right contextual locations for the main study (red columns) and the saliency control study (blue columns). MVPA results are shown for both single trial (A) and eight average analyses (B). Error bars indicate SEM. Area LOC is significantly above chance ($p < 0.01$; FDR corrected) for the single trial analysis (A) while areas LOC, RSC, and IPS are significantly above chance ($p < 0.01$; FDR corrected) for the eight-trial average analysis. † indicates significance at $p < 0.01$; FDR corrected.

vious studies (Kravitz et al., 2011), PHC was able to discriminate open versus closed scenes (AUC = 0.598, $p < 0.01$ FDR). We then evaluated whether MVPA performance in LOC predicting context varied for open versus closed scenes (Fig. 5A). MVPA performance predicting the contextual location (Fig. 5A) was not significantly different for open space scenes versus close space scenes (Kravitz et al., 2011). Examples of the subcategories are shown in Figure 5B.

Global scene properties versus object–object co-occurrence

To assess whether context defined by global scene properties (Fig. 5A) or object–object co-occurrence (Mack and Eckstein, 2011) (Fig. 5A, object–object) are mediated by similar or different neural mechanisms, we separated the images into these two types of statistical regularities using the majority opinion of 12 independent raters. Although the definitions of these two categories are

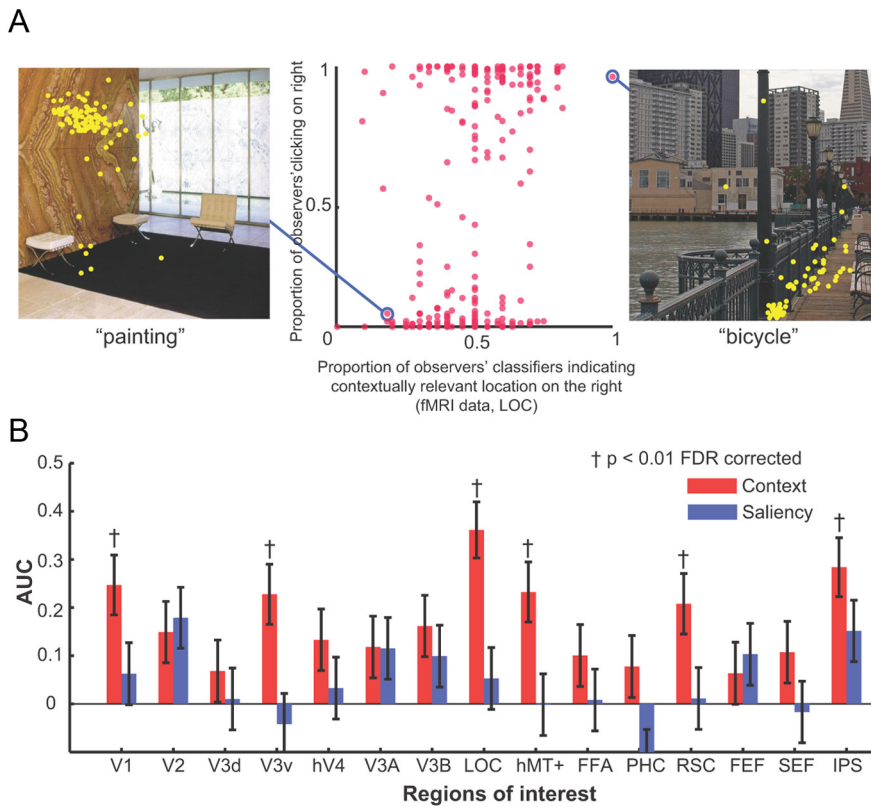


Figure 4. *A*, Scatter plot showing the relationship between the proportion of observers' contextually relevant selections ($n = 100$) that fell into the right image hemifield and proportion of MVPA classifiers' AUC (area LOC) across 12 brains predicting that the contextual location was on the right side of the image. *B*, Correlations for all ROIs for the same analysis (*A*) for all ROIs including results for the saliency control experiment. Areas V1, V3v, LOC, hMT+, RSC, and IPS show significantly above-chance correlations ($p < 0.01$; FDR corrected). Error bars indicate SEM. † indicates significance at $p < 0.01$; FDR corrected.

not unequivocal, we described and illustrated to the raters the two types of images by showing an example in which context was defined by the presence of another spatially localized object and global scene properties in terms of the general structure of the scene and not a specific object per se (Fig. 5*B*). Of the 320 target absent images, using a majority opinion, 147 images were classified as object–object co-occurrence and 154 as global scene properties. There were 19 images in which there was a tie across raters; these images were excluded from the analysis. Of the 320 target absent images, 181 images were classified with a 90% agreement across raters. We then assessed the MVPA performance discriminating context (left vs right) in these two sets of images. Figure 5*A* shows similar ($p = 0.46$) MVPA performance across image sets suggesting that LOC might mediate both types of contextual information. A reanalysis using only images for which there was a consensus across raters about the category of the image yielded similar results.

Eye movements and response times

One possible concern is that the results might be caused by involuntary eye movements toward the contextual location of the images, even though observers were instructed to maintain central fixation during the trial. Eye position was monitored for 10 of the 12 observers and the data show that observers were successful at maintaining central fixation for a majority of the trials. Specifically, during the 800 ms period from the onset of the cue word to the end of the presentation of the scene, fixations $> 2^\circ$ from central fixation occurred in 5.7% of the trials (averaged across observers). If the whole trial was considered (~ 5 s), observers

moved their eyes $> 2^\circ$ from central fixation on 13.6% of the trials. To evaluate whether differential eye movement across context left versus right images could account for our results, we quantified two additional eye movement behaviors. First, the mean number of eye movements (averaged across the 10 observers) was 3.7 ± 0.67 for trials for which images with the contextual location on the left and it did not differ significantly from the 3.96 ± 0.67 mean movements for those with the contextual location on the right ($p = 0.850$; paired t test). Second, the mean saccade amplitude was 2.70 ± 0.34 degrees for images with the context on the left and it did not differ significantly from the 2.68 ± 0.33 mean amplitude for images with context on the right ($p = 0.323$; paired t test). Finally, an MVPA applied to the eye position data for each observer could not reliably predict the contextual hemifield (average AUC = 0.513), suggesting that the fMRI results cannot be explained by eye movements toward the contextual locations.

We also investigated the possibility that differential response times (RT) for context-left versus context-right images might mediate the differences in patterns of BOLD activity. However, the RTs were not statistically significant (target-absent context right mean RT across observers = 1.18 s; target-absent

context left mean RT across observers = 1.20 s; $p = 0.147$; $t_{(11)} = 1.56$).

Experiment 2

To verify that the measured neural activity is related to the spatial location of objects/features that are predictive of the location of the searched target and is not due to physical low-level differences across images, we ran a separate study that used the same set of images but made context behaviorally irrelevant. Six new observers viewed the same images but were instructed to indicate the side of the image that contained the most salient region or features. If our measured patterns of neural activity in the main target-search experiment were driven by contextual information then the MVPA accuracy should be close to chance predicting the contextually relevant hemifield when observers are engaged in a different task for which context is irrelevant. Figure 3*A* shows that MVPA accuracy is near chance (for all areas $p > 0.3$; e.g., LOC, $p = 0.744$) when observers searched for the most salient region in the image. Furthermore, averaging voxel activity across eight trials in the MVPA test phase does not improve MVPA accuracy and does not result in any statistically significant results (for all areas $p > 0.5$; e.g., LOC, $p = 0.848$). Figure 4*B* also shows that when context is behaviorally irrelevant the correlation between the proportion of selections in one image hemifield and the MVPA scalar responses were not significantly different from zero for all areas (for all areas $p > 0.5$; e.g., LOC, $p = 0.602$). To demonstrate that the lack of a significant result in Experiment 2 is not related to a lower statistical power given the fewer observers participating in

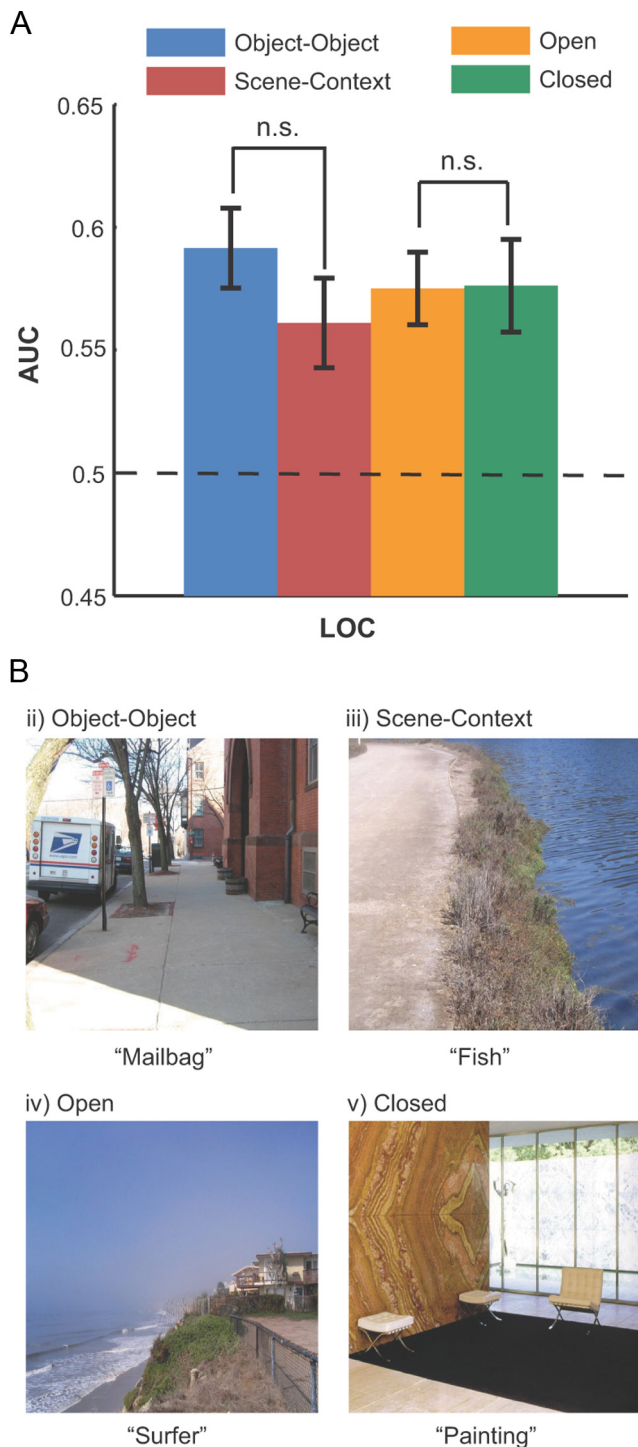


Figure 5. *A*, MVPA performance (AUC) predicting the contextual location of subsets of images containing contextual information based on scene context or object–object co-occurrence, or images with open or closed scenes (*B*). Images were classified based on the majority opinion of 10 raters. There is no significant difference between the subcategories' AUC ($p > 0.5$; FDR corrected). Error bars indicate SEM.

Experiment 2 versus Experiment 1 ($n = 6$ vs $n = 12$) we conducted a bootstrap resampling analysis of the data from Experiment 1. We created 5000 bootstrap samples by selecting 6 subjects from the pool of 12 (with replacement) and calculated the t statistic for the MVPA scene context discrimination performance of the group of 6 selected subjects against chance. Then by calculating the proportion of bootstrap samples (from the distri-

bution of 5000) whose t statistic was less than the observed t statistic for the 12 subjects (a Monte Carlo test; Efron and Tibshirani, 1993) we determined whether the discrimination accuracy for 6 subjects was significantly less than for 12. We found that in no areas was the discrimination accuracy significantly less for 6 subjects compared with 12 (for all areas $p > 0.2$; e.g., LOC, $p = 0.200$).

Discussion

The neural representation of scene context

Synthetic cues predictive of a target location modulate activity throughout visual cortex (Corbetta et al., 1990; Brefczynski and DeYoe, 1999; Gandhi et al., 1999; Maunsell and Cook, 2002; Carrasco, 2006, 2011) but areas in the frontoparietal attention network, which include the frontal and supplementary eye fields (FEF, SEF) and an area in the IPS (Kastner et al., 1998; Corbetta et al., 2000; Hopfinger et al., 2000; Corbetta and Shulman, 2002; Yantis et al., 2002; Giesbrecht et al., 2003; Kelley et al., 2008; Ptak, 2012) are believed to be the source of the cue-related modulation in activity in the visual areas (Silver et al., 2005). However, no study has demonstrated its involvement in contextual guidance in real scenes. Our results suggest that IPS encodes the spatial location of areas within the scene that are contextually relevant to the target being searched for. However, much stronger effects were found in the LOC. Furthermore, when focusing on individual images the MVPA decision variable in IPS, LOC, and RSC (and to a lesser degree other areas including V1, V3v, and hMT+) were positively correlated with the amount of contextual guidance in an image independently quantified by the explicit context judgments of 100 independent observers. Our eye position control analyses confirm that the MVPA BOLD results cannot be accounted for by eye movements (Konen and Kastner, 2008). In addition, to disambiguate possible contributions of differences in varying low-level features between the contextually relevant and irrelevant halves of the images, we conducted a control experiment that used the same images but for which new observers had to decide which side of the image contained the most salient region. In the control condition, MVPA analysis did not predict contextual locations from activity in all areas including LOC and IPS, suggesting that our main results cannot be attributed to low-level physical differences between right and left halves of the images. There were also no significant differences in response times between trials for which the contextual location was lateralized on the right versus the left side of the image.

The ability to decode coarse locations of contextually relevant regions of the image from activity in LOC and IPS is consistent with previous studies showing that these areas have a retinotopic organization (Wandell et al., 2007; Silver and Kastner, 2009). Utilization of context to guide search in real scenes requires processing of objects and places and subsequent orienting of attentional processes to the likely target location. Thus, our results suggest that activity across object (LOC), scene (RSC), and attention areas (IPS) may mediate contextual guidance of search, perhaps comprising a scene context network. One possible scenario is that neural activity in LOC and RSC in response to objects and scene regions that are likely to co-occur with the searched target is modulated by higher level areas including the medial prefrontal cortex (Greicius et al., 2009). The location of the predictive objects and scene areas might then be relayed from LOC to the frontoparietal network, including IPS, which then directs attention to the contextually relevant location. Based on the location information, IPS, which has been shown to represent target detection in real scenes (Guo et al., 2012), might increase target-

related activity at the contextual locations likely to predict the presence of the target, a computation that can approximate optimal Bayesian (Torralla et al., 2006; Gold and Shadlen, 2007; Eckstein et al., 2009).

An alternative interpretation is that IPS modulates activity in LOC. Although the interactions between IPS and LOC are not fully understood, a recent study looking at the increased LOC activity for interacting objects relative to noninteracting objects also considered the possible effects of attentional feedback from IPS to LOC (Kim and Biederman, 2011; Kim et al., 2011). They showed that transcranial magnetic stimulation (TMS) of LOC, but not IPS abolished a behavioral facilitation (in the absence of TMS) in identifying interacting objects compared with noninteracting depictions (Kim et al., 2011). This result indicates that the neural activity in LOC has a functional role in identifying object relationships (Kim et al., 2011) and suggests that context-related activity in LOC in the present study is integrated by IPS to direct attention and is not solely the result of feedback from the IPS.

Roles of the parahippocampal and retrosplenial cortices in processing scene context

A finding of interest is that PHC did not predict the location of scene context. PHC preferentially responds to scenes (Epstein and Kanwisher, 1998; Greicius et al., 2009), visuospatial structure of individual scenes (Epstein and Higgins, 2007), contextual associations across objects (Bar, 2004; Bar et al., 2008), and spatial congruency across two objects (Gronau et al., 2008). However, contextual associations across objects would not suffice to provide contextual guidance during search. In complex scenes, many objects appear that are contextually related with varying degrees of strength and it is unclear how the current understanding of PHC responses to associated objects would guide search in real scenes. Furthermore, PHC contextual association activity is triggered by the simultaneous presentation of two objects that are related while contextual guidance during search (and the differential LOC activity in our experiment) occurs even if there is only one object present in the scene that is spatially associated to a searched target object that is absent. In addition, contextual guidance requires explicit encoding of the spatial location of the contextually relevant regions or object(s) that has not been observed in PHC in the current or previous studies.

An additional analysis used MVPA to show that PHC could reliably discriminate between open versus closed scenes replicating previous results (Kravitz et al., 2011). This result is important because it suggests that the inability to predict the likely location of searched objects from PHC in our study cannot be attributed to inappropriate localization or segmentation of PHC.

Our results also show an ability to decode context location from RSC (Epstein and Kanwisher, 1998; Yi and Chun, 2005; Epstein and Higgins, 2007; Henderson et al., 2008), although these effects were much weaker than those obtained from LOC. A representation of scene context in RSC is consistent with an interpretation that RSC supports processes that allow scenes to be localized within a larger extended environment (Epstein and Higgins, 2007) and a viewpoint-invariant scene representation used to create a more integrative cognitive map of the environment (Park and Chun, 2009).

Scene context effects in lower visual areas

In addition to the effects in LOC, IPS, and RSC, small but significant effects were found for the correlations between the discrimination performance of the areas V1, hMT+, and V3v and the

degree of contextual guidance in the scenes. Unlike our interpretation of the activity in LOC, we argue that these results reflect the effects of feedback from IPS (Bisley, 2011) based on numerous previous studies. Attentional enhancement of neural activity in monkey middle temporal area has been shown to follow modulations of responses in LIP (Saalmann et al., 2007). In addition, Granger casualty analysis of human BOLD responses (Bressler et al., 2008) and the application of TMS (Ruff et al., 2008) also suggest top-down effects exerted by IPS on the BOLD response of lower visual areas (V1–hV4).

Object–object co-occurrence versus global scene properties

Recently, scene context has been categorized in terms of scenarios in which an object is predictive of the location of the target (object–object co-occurrence) (Castelhano and Heaven, 2011; Mack and Eckstein, 2011) or those in which the general features or the scene gist guides search (e.g., search for an airplane can trigger an eye movement toward the top of the image if the image is recognized as a downtown city scene). We evaluated whether the ability of LOC to predict contextual location varied for these two categories of scene context. We might have expected to obtain a higher performance predicting context for object–object co-occurrence given LOC's specialization for processing objects (Grill-Spector et al., 2000; Kourtzi and Kanwisher, 2001). Instead, our results show that LOC activity was equally predictive of contextual locations for both object–object co-occurrence and general scene properties. The result is consistent with recent findings showing LOC's activity to scenes can be related to a composite of the individual activity to objects composing the scene (MacEvoy and Epstein, 2011).

Multifunctionality of LOC during search

Our findings support the current view that LOC does not simply encode object categories but also represents the functional importance of objects for an impending search task including modulation of activity of a searched target (Peelen et al., 2009), preparatory patterns of activity representing the search object (Peelen and Kastner, 2011), and suppression of activity to distractor objects (Seidl et al., 2012). Our current results demonstrate that LOC activity in response to objects or scene regions is altered if these are predictive of the search target suggesting that LOC might be critical in contextual guidance during search which is arguably one of the most important strategies used by humans to allow for efficient visual search.

References

- Alkadi H, Crelier GR, Boendermaker SH, Golay X, Hepp-Reymond MC, Kollias SS (2002) Reproducibility of primary motor cortex somatotopy under controlled conditions. *AJNR Am J Neuroradiol* 23:1524–1532. [Medline](#)
- Bar M (2004) Visual objects in context. *Nat Rev Neurosci* 5:617–629. [CrossRef Medline](#)
- Bar M, Aminoff E (2003) Cortical analysis of visual context. *Neuron* 38:347–358. [CrossRef Medline](#)
- Bar M, Aminoff E, Schacter DL (2008) Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *J Neurosci* 28:8539–8544. [CrossRef Medline](#)
- Bisley JW (2011) The neural basis of visual attention. *J Physiol* 589:49–57. [CrossRef Medline](#)
- Brefczynski JA, DeYoe EA (1999) A physiological correlate of the “spotlight” of visual attention. *Nat Neurosci* 2:370–374. [CrossRef Medline](#)
- Bressler SL, Tang W, Sylvester CM, Shulman GL, Corbetta M (2008) Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J Neurosci* 28:10056–10061. [CrossRef Medline](#)
- Buracas GT, Boynton GM (2002) Efficient design of event-related fMRI ex-

- periments using M-sequences. *Neuroimage* 16:801–813. [CrossRef Medline](#)
- Carrasco M (2006) Covert attention increases contrast sensitivity: psychophysical, neurophysiological and neuroimaging studies. *Prog Brain Res* 154:33–70. [CrossRef Medline](#)
- Carrasco M (2011) Visual attention: the past 25 years. *Vision Res* 51:1484–1525. [CrossRef Medline](#)
- Castelhano MS, Heaven C (2010) The relative contribution of scene context and target features to visual search in scenes. *Atten Percept Psychophys* 72:1283–1297. [CrossRef Medline](#)
- Castelhano MS, Heaven C (2011) Scene context influences without scene gist: eye movements guided by spatial associations in visual search. *Psychon Bull Rev* 18:890–896. [CrossRef Medline](#)
- Connolly JD, Goodale MA, Menon RS, Munoz DP (2002) Human fMRI evidence for the neural correlates of preparatory set. *Nat Neurosci* 5:1345–1352. [CrossRef Medline](#)
- Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215. [Medline](#)
- Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1990) Attentional modulation of neural processing of shape, color, and velocity in humans. *Science* 248:1556–1559. [CrossRef Medline](#)
- Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat Neurosci* 3:292–297. [CrossRef Medline](#)
- DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J (1996) Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc Natl Acad Sci U S A* 93:2382–2386. [CrossRef Medline](#)
- Duda RO, Hart PE, and Stork DG (2000) *Pattern classification*. Ed. 2. New York: Wiley-Interscience.
- Eckstein MP (2011) Visual search: a retrospective. *J Vis* 11(5). pii:14. [CrossRef Medline](#)
- Eckstein MP, Drescher BA, Shimozaki SS (2006) Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychol Sci* 17:973–980. [CrossRef Medline](#)
- Eckstein MP, Peterson MF, Pham BT, Droll JA (2009) Statistical decision theory to relate neurons to behavior in the study of covert visual attention. *Vision Res* 49:1097–1128. [CrossRef Medline](#)
- Efron B, and Tibshirani R. (1993) *An introduction to the bootstrap*. New York: Chapman & Hall/CRC.
- Ehinger KA, Hidalgo-Sotelo B, Torralba A, Oliva A (2009) Modeling visual search in a thousand scenes: the roles of saliency, target features, and scene context. *Vis Cogn* 17:945–978. [CrossRef Medline](#)
- Epstein RA, Higgins JS (2007) Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb Cortex* 17:1680–1693. [Medline](#)
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601. [CrossRef Medline](#)
- Gandhi SP, Heeger DJ, Boynton GM (1999) Spatial attention affects brain activity in human primary visual cortex. *Proc Natl Acad Sci U S A* 96:3314–3319. [CrossRef Medline](#)
- Giesbrecht B, Woldorff MG, Song AW, Mangun GR (2003) Neural mechanisms of top-down control during spatial and feature attention. *Neuroimage* 19:496–512. [CrossRef Medline](#)
- Gold JJ, Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci* 30:535–574. [CrossRef Medline](#)
- Greicius MD, Supekar K, Menon V, Dougherty RF (2009) Resting-State functional connectivity reflects structural connectivity in the default mode network. *Cereb Cortex* 19:72–78. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. *Vision Res* 41:1409–1422. [CrossRef Medline](#)
- Gronau N, Neta M, Bar M (2008) Integrated contextual representation for objects' identities and their locations. *J Cogn Neurosci* 20:371–388. [CrossRef Medline](#)
- Guo F, Preston TJ, Das K, Giesbrecht B, Eckstein MP (2012) Feature-independent neural coding of target detection during search of natural scenes. *J Neurosci* 32:9499–9510. [CrossRef Medline](#)
- Henderson JM, Larson CL, Zhu DC (2008) Full scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: an fMRI study. *Brain Cogn* 66:40–49. [CrossRef Medline](#)
- Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. *Nat Neurosci* 3:284–291. [CrossRef Medline](#)
- 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. [Medline](#)
- Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* 282:108–111. [CrossRef Medline](#)
- Kelley TA, Serences JT, Giesbrecht B, Yantis S (2008) Cortical mechanisms for shifting and holding visuospatial attention. *Cereb Cortex* 18:114–125. [Medline](#)
- Kim JG, Biederman I (2011) Where do objects become scenes? *Cereb Cortex* 21:1738–1746. [CrossRef Medline](#)
- Kim JG, Biederman I, Juan, CH (2011) The benefit of object interactions arises in the lateral occipital cortex independent of attentional modulation from the intraparietal sulcus: a transcranial magnetic stimulation study. *J Neurosci* 31:8320–8324. [CrossRef Medline](#)
- Konen CS, Kastner S (2008) Two hierarchically organized neural systems for object information in human visual cortex. *Nat Neurosci* 11:224–231. [CrossRef Medline](#)
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* 293:1506–1509. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Lotze M, Erb M, Flor H, Huelsmann E, Godde B, Grodd W (2000) fMRI evaluation of somatotopic representation in human primary motor cortex. *Neuroimage* 11:473–481. [CrossRef Medline](#)
- MacEvoy SP, Epstein RA (2011) Constructing scenes from objects in human occipitotemporal cortex. *Nat Neurosci* 14:1323–1329. [CrossRef Medline](#)
- Mack SC, Eckstein MP (2011) Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *J Vis* 11(9):1–16. [CrossRef Medline](#)
- Maunsell JH, Cook EP (2002) The role of attention in visual processing. *Philos Trans R Soc Lond B Biol Sci* 357:1063–1072. [CrossRef Medline](#)
- Ma WJ, Navalpakkam V, Beck JM, Berg Rv, Pouget A (2011) Behavior and neural basis of near-optimal visual search. *Nat Neurosci* 14 783–790. [CrossRef Medline](#)
- Najemnik J, Geisler WS (2005) Optimal eye movement strategies in visual search. *Nature* 434:387–391. [CrossRef Medline](#)
- Navalpakkam V, Koch C, Rangel A, Perona P (2010) Optimal reward harvesting in complex perceptual environments. *Proc Natl Acad Sci U S A* 107:5232–5237. [CrossRef Medline](#)
- Neider MB, Zelinsky GJ (2006) Scene context guides eye movements during visual search. *Vision Res* 46:614–621. [CrossRef Medline](#)
- Palmer J, Verghese P, Pavel M (2000) The psychophysics of visual search. *Vision Res* 40:1227–1268. [CrossRef Medline](#)
- Park S, Chun MM (2009) Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage* 47:1747–1756. [CrossRef Medline](#)
- Park S, Brady TF, Greene MR, Oliva A (2011) Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *J Neurosci* 31:1333–1340. [CrossRef Medline](#)
- Peelen MV, Kastner S (2011) A neural basis for real-world visual search in human occipitotemporal cortex. *Proc Natl Acad Sci U S A* 108:12125–12130. [CrossRef Medline](#)
- Peelen MV, Fei-Fei L, Kastner S (2009) Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* 460:94–97. [CrossRef Medline](#)
- Ptak, R (2011) The frontoparietal attention network of the human brain: action, saliency, and a priority map of the environment. *Neuroscientist* 18:502–515. [Medline](#)
- Ruff CC, Bestmann S, Blankenburg F, Bjoertomt O, Josephs O, Weiskopf N, Deichmann R, Driver J (2008) Distinct causal influences of parietal ver-

- sus frontal areas on human visual cortex: evidence from concurrent TMS-fMRI. *Cereb Cortex* 18:817–827. [Medline](#)
- Saalmann YB, Pigarev IN, Vidyasagar TR (2007) Neural mechanisms of visual attention: how top-down feedback highlights relevant locations. *Science* 316:1612–1615. [CrossRef Medline](#)
- Saproo S, Serences JT (2010) Spatial attention improves the quality of population codes in human visual cortex. *J Neurophysiol* 104:885–895. [CrossRef Medline](#)
- Seidl KN, Peelen MV, Kastner S (2012) Neural evidence for distracter suppression during visual search in real-world scenes. *J Neurosci* 32:11812–11819. [CrossRef Medline](#)
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268:889–893. [CrossRef Medline](#)
- Silver MA, Ress D, Heeger DJ (2005) Topographic maps of visual spatial attention in human parietal cortex. *J Neurophysiol* 94:1358–1371. [CrossRef Medline](#)
- Silver MA, Kastner S (2009) Topographic maps in human frontal and parietal cortex. *Trends Cogn Sci* 13:488–495. [CrossRef Medline](#)
- Swenson RG, Judy PF (1981) Detection of noisy visual targets: models for the effects of spatial uncertainty and signal-to-noise ratio. *Percept Psychophys* 29:521–534. [CrossRef Medline](#)
- Tong F, Pratte MS (2012) Decoding patterns of human brain activity. *Annu Rev Psychol* 63:483–509. [CrossRef Medline](#)
- Tootell RB, Hadjikhani N (2001) Where is “dorsal V4” in human visual cortex? Retinotopic, topographic and functional evidence. *Cereb Cortex* 11:298–311. [CrossRef Medline](#)
- Torralla A, Oliva A, Castelhanos MS, Henderson JM (2006) Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol Rev* 113:766–786. [CrossRef Medline](#)
- Tyler CW, Likova LT, Chen C-C, Kontsevich LL, Schira MM, Wade AR (2005) Extended concepts of occipital retinotopy. *Curr Med Imag Rev* 1:319–329. [CrossRef](#)
- Wandell BA, Dumoulin SO, Brewer AA (2007) Visual field maps in human cortex. *Neuron* 56:366–383. [CrossRef Medline](#)
- Weil RS, Rees G (2010) Decoding the neural correlates of consciousness. *Curr Opin Neurol* 23:649–655. [CrossRef Medline](#)
- Wolfe JM (1998) What can 1 million trials tell us about visual search? *Psychol Sci* 9:33–39. [CrossRef](#)
- Wolfe JM, Võ ML, Evans KK, Greene MR (2011) Visual search in scenes involves selective and nonselective pathways. *Trends Cogn Sci* 15:77–84. [CrossRef Medline](#)
- Yantis S, Schwarzbach J, Serences JT, Carlson RL, Steinmetz MA, Pekar JJ, Courtney SM (2002) Transient neural activity in human parietal cortex during spatial attention shifts. *Nat Neurosci* 5:995–1002. [CrossRef Medline](#)
- Yi DJ, Chun MM (2005) Attentional modulation of learning-related repetition attenuation effects in human parahippocampal cortex. *J Neurosci* 25:3593–3600. [CrossRef Medline](#)
- Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS (1991) A direct demonstration of functional specialization in human visual cortex. *J Neurosci* 11:641–649. [Medline](#)