

# Clutter Modulates the Representation of Target Objects in the Human Occipitotemporal Cortex

Yaara Erez and Galit Yovel

## Abstract

■ Target objects required for goal-directed behavior are typically embedded within multiple irrelevant objects that may interfere with their encoding. Most neuroimaging studies of high-level visual cortex have examined the representation of isolated objects, and therefore, little is known about how surrounding objects influence the neural representation of target objects. To investigate the effect of different types of clutter on the distributed responses to target objects in high-level visual areas, we used fMRI and manipulated the type of clutter. Specifically, target objects (i.e., a face and a house) were presented either in isolation, in the presence of a homogeneous (identical objects from another category) clutter (“pop-out” display), or in the presence of a heterogeneous (different objects) clutter, while participants performed a target identification task. Using multivoxel pattern

analysis (MVPA) we found that in the posterior fusiform object area a heterogeneous but not homogeneous clutter interfered with decoding of the target objects. Furthermore, multivoxel patterns evoked by isolated objects were more similar to multivoxel patterns evoked by homogeneous compared with heterogeneous clutter in the lateral occipital and posterior fusiform object areas. Interestingly, there was no effect of clutter on the neural representation of the target objects in their category-selective areas, such as the fusiform face area and the parahippocampal place area. Our findings show that the variation among irrelevant surrounding objects influences the neural representation of target objects in the object general area, but not in object category-selective cortex, where the representation of target objects is invariant to their surroundings. ■

## INTRODUCTION

Natural scenes are typically comprising multiple objects presented simultaneously. A central goal of our visual system is to extract task-relevant objects from such cluttered displays. Nonetheless, the majority of neuroimaging studies of high-level visual cortex have examined the neural representation of isolated objects, and therefore, relatively little is known about the extent to which simultaneously presented irrelevant objects affect the neural responses to task-relevant objects.

Recently, a few neuroimaging studies have examined the responses to objects presented simultaneously with a second nontarget object (Reddy, Kanwisher, & VanRullen, 2009; Reddy & Kanwisher, 2007). These studies showed that response patterns to target objects in the general object area—the lateral occipital complex (LOC)—were interrupted by a nontarget object presented simultaneously. Another study showed increased change in BOLD response in LOC when a target object was presented with several nontarget objects (Jeong & Xu, 2013). A few other studies have focused on objects presented within natural scenes. Neural patterns in object general cortex evoked by target objects embedded within a scene were similar to patterns elicited by same-category objects presented

in isolation and could be discriminated from those evoked by objects from other categories (Seidl, Peelen, & Kastner, 2012; Peelen & Kastner, 2011; Peelen, Fei-Fei, & Kastner, 2009). Interestingly, it has been shown that the neural patterns of scenes were predicted by the patterns evoked by their constituent objects (MacEvoy & Epstein, 2011).

One important aspect that has not been addressed in these previous studies is the effect of different types of clutters on the representation of target objects and specifically the effect of the variation among the surrounding irrelevant objects. In particular, many behavioral studies have shown that detection of target stimuli is better when presented among homogeneous surrounding stimuli (i.e., pop-out displays) compared with heterogeneous stimuli (Duncan & Humphreys, 1989; Treisman & Gormican, 1988). Consistent with these behavioral studies, a recent fMRI study employed a neural competition paradigm to assess competition among low-level stimuli comprising Gabor gratings in a homogeneous clutter (pop-out display) compared with a heterogeneous clutter (Beck & Kastner, 2005). To measure neural competition, stimuli were presented either simultaneously or sequentially, whereas a lower fMRI signal to simultaneous compared with sequential presentation reflects competition among multiple stimuli in the former but not the latter display (Kastner, De Weerd, Desimone, & Ungerleider, 1998). Competition effect was evident for heterogeneous clutter but was

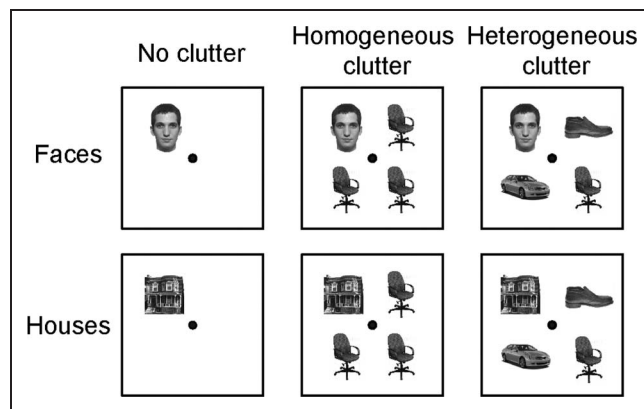
eliminated for a pop-out display in early visual cortical areas V1 and V2/VP as well as area V4. This paradigm, however, does not allow to directly examine the effect of the type of clutter on the representation of the target stimulus.

The current study directly examined the effect of the type of clutter on the representation of the target stimuli by manipulating the type of clutter and assessing how it may influence the decoding and representational pattern of target objects. We used objects rather than low-level visual stimuli to compare the effect of clutter in object general areas and object category-selective cortex, which showed different sensitivity to effects of clutter (Reddy et al., 2009; Reddy & Kanwisher, 2007). To that effect, we presented target objects either in isolation, in the presence of a homogeneous clutter comprising three identical objects from another category, or in the presence of a heterogeneous clutter comprising three objects from different categories (Figure 1). On the basis of previous behavioral (Nothdurft, 1993; Duncan & Humphreys, 1989; Treisman & Gormican, 1988) and neuroimaging (Beck & Kastner, 2005) studies, we predicted that heterogeneous clutter will interfere with decoding and representation of target objects more than homogeneous clutter. On the basis of recent findings showing that responses to preferred objects in a clutter display were preserved in category-selective areas (Reddy & Kanwisher, 2007), we expected no such effect of the type of clutter in these category-selective areas.

## METHODS

### Participants

Nineteen healthy volunteers (age = 19–32 years, 11 women) with normal or corrected-to-normal vision participated in



**Figure 1.** Experimental conditions. The experiment included two target objects (faces, houses) presented in three clutter conditions (no clutter, homogeneous clutter, and heterogeneous clutter). Nontarget objects were cars, chairs, and shoes. Images for each of these six conditions were presented in blocks. Participants were asked to perform a 1-back task on the identity of the target while maintaining their eyes on a fixation dot.

the experiment. All participants gave written informed consent to participate in the study approved by the Tel-Aviv Sourasky Medical Center. Three participants were excluded from the analysis: one because of excessive head movements in the scanner, one because of long periods of sleep during the experiment, and one because object general brain areas could not be identified during analysis.

### Experimental Procedure

High-resolution fMRI data were collected in a 3T GE MRI scanner, using an eight-channel head coil. EPI sequence was used to collect fMRI data with a repetition time of 2 sec, echo time of 35 msec, 23 slices per repetition time, slice thickness of 2.4 mm with no gap, and field of view of 20 cm. The acquisition matrix was  $96 \times 96$  (in-plane resolution  $2.08 \times 2.08$  mm), which was reconstructed into  $128 \times 128$  matrix (in-plane resolution  $1.56 \times 1.56$  mm). Brain coverage included the entire occipital and temporal lobes. Stimuli were presented using Psychtoolbox2 for Matlab (Brainard, 1997) and projected on an MRI-compatible screen inside the scanner. In the main experiment, target stimuli from two categories were presented among irrelevant surrounding objects (clutter). Target stimuli were presented under three clutter conditions (Figure 1): (a) in isolation (no-clutter condition); (b) in the presence of a homogeneous clutter comprising three identical objects from a category that is different from the target object categories, thus yielding a pop-out display; (c) a heterogeneous clutter comprising three objects from different categories all different from the target object categories. Targets were faces and houses, and nontarget objects were cars, chairs, and shoes. Two of the six exemplars from each category were used in each run, and the stimuli presented in the first three runs were then presented again in the three last runs. The nontargets in the homogeneous display varied throughout a block. Stimuli were grayscale images of  $3.9^\circ \times 4.1^\circ$  and were presented on a  $768 \times 1024$  pixel screen around a fixation dot at four locations: top left, top right, bottom right, and bottom left. The stimuli were presented as close as possible to the fixation dot and were centered  $2.4^\circ$  away from the fixation dot on both  $x$  and  $y$  axes. They were presented for 200 msec, followed by a 1300-msec intertrial interval, arranged in blocks of eight trials of the same condition (12-sec block duration). Each run of the main experiment consisted of initial 6-sec fixation dummy scans, two blocks for each condition, and three blocks of a baseline fixation dot (a total of 186 sec). Locations of target and nontargets were counterbalanced within block, and categories and order of experimental conditions were counterbalanced across blocks and runs. Participants were asked to perform 1-back task on the target by pressing a response box button while fixating on the fixation dot and were explicitly instructed to maintain fixation and not to move their gaze away from the fixation dot. Responses were

collected to assess the effect of clutter on identification of the target stimuli.

Additionally to the main experiment, a standard functional localizer was used to provide an independent data set to define ROIs. The functional localizer included four object categories presented in a blocked design: faces, scenes, objects, and scrambled objects. Each block lasted 12 sec and included 15 stimuli, each presented for 300 msec with a 500-msec interstimulus interval. Category block order was counterbalanced within and across runs. Each localizer run consisted of an initial 6-sec fixation dummy scans, four blocks for each category, and five blocks of a baseline fixation dot (a total of 258 sec). To maintain vigilance, participants were asked to perform a 1-back task on the presented stimuli by pressing a response box button. Each participant completed three localizer runs.

## Data Analysis

### *Preprocessing*

Statistical parametric mapping (SPM5) software was used for the data analysis. The first three dummy volumes in each run were discarded from the analysis. The data were then preprocessed using coregistration to the anatomical scan, slice timing correction, and realignment. Spatial smoothing with a  $5 \times 5 \times 5$  mm kernel was applied for the functional localizer only. A general linear model was estimated for each participant using a canonical hemodynamic response function with seven regressors for the main experiment and five regressors for the functional localizer to account for the experimental conditions and fixation blocks.

### *ROI Analysis*

Four ROIs were defined using the functional localizer data. Object general regions (posterior fusiform gyrus [pFs] and lateral occipital [LO]) were defined using objects > scrambled objects  $t$  contrast maps ( $p < .00001$ ), with exclusion masking of faces > objects and scenes > objects  $t$  contrasts ( $p < .05$ ). These two areas comprise the LOC, and we considered them separately as previous evidence suggests they show different effects for cluttered displays (MacEvoy & Epstein, 2011). Category-selective regions were defined using faces > objects and scenes > objects  $t$  contrast maps ( $p < .00001$ ) for the fusiform face area (FFA) and parahippocampal place area (PPA), respectively.

### *Voxelwise-based Analysis*

Voxelwise-based analysis (classification using multivariate pattern analysis [MVPA] and similarity analysis using correlations) was performed using raw intensity values, which are equivalent to  $\beta$  estimates in a blocked design (Misaki, Kim, Bandettini, & Kriegeskorte, 2010). Following standard

preprocessing using SPM5, a full-scan raw intensity values were extracted for each ROI, condition, and run. Each voxel's full-scan values in each run were normalized to  $z$  scores. Then, two volumes at the beginning of each block (4 sec) were excluded from analysis to account for the hemodynamic lag.

### *Controlled ROI Size for Voxelwise-based Analysis*

To allow the comparison between participants and ROIs, the size of ROI was controlled. For the analysis, the 47 most activated voxels in each ROI (based on the  $t$  contrast maps used for defining each ROI) were used. ROIs with a smaller number of voxels were excluded from analysis. This number of voxels was chosen to maximize both the number of participants that can be included in the analysis and the number of voxels used for each ROI, mainly based on the size of the pFs, which had the smallest number of voxels on average. However, to assure the robustness of the results, the analysis was also conducted using a large range of ROI sizes (30–80 voxels).

### *Classification Using MVPA*

Voxelwise-based classification was performed to measure decoding performance for discrimination between the target objects. For each condition, the mean intensity across all voxels was subtracted from each voxel intensity value (Axelrod & Yovel, 2012; Misaki et al., 2010; Serences, Saproo, Scolari, Ho, & Muftuler, 2009). This procedure was performed for each run separately. Then the intensity values of each block were averaged for each voxel. Each run consisted of two blocks for each condition and a total of six runs yielded 12 voxelwise patterns for each condition that were used for classification. Support vector machine classification was conducted using LibSVM library for Matlab. A leave-one-run-out (two patterns) cross-validation procedure was repeated six times, and the decoding performance was averaged. Classification was performed between the two target categories (faces, houses) for each clutter condition (no clutter, homogeneous clutter, heterogeneous clutter) separately, and decoding performance for the different clutter conditions was then averaged across participants. To verify classification chance level of 50%, we conducted a permutation test with 1000 repetitions, in which the patterns were randomly labeled. The average decoding performance of the permutation test across participants in all the ROIs and clutter conditions was in the range of 49.8–50.2%, confirming chance level of 50%.

### *Voxelwise Pattern Similarity Analysis*

Similarity between pairs of conditions was assessed using correlations between the elicited voxelwise response patterns. We computed correlations based on both the entire data set and split-halves of the data. We first computed

correlations based on the entire data set. For each condition, the average voxelwise pattern was computed across all blocks and runs. Correlation between the no-clutter condition and each of the clutter conditions (homogeneous, heterogeneous) was computed and compared within each target category (faces, houses) separately. To be able to measure the within-condition similarity, thus provide a baseline of the similarity of the no-clutter condition with itself, we also computed split-half correlations. Correlations between the no-clutter condition and itself as well as between the no-clutter condition and each of the clutter conditions (homogeneous, heterogeneous) were computed and compared within each target category (faces, houses) separately. For each pair of conditions, the voxelwise pattern was computed across all blocks in half of the data (three runs) and correlated with the average pattern that was computed based on the other half of the data. The split-half correlations across all possible split-halves of the data were then averaged to get the similarity measure between the two conditions.

#### *Average fMRI Signal Analysis*

Time courses were extracted for each of the six conditions of the main experiment using the MarsBaR ROI toolbox for SPM (Brett, Anton, Valabregue, & Poline, 2002) within each of the predefined ROIs. The average percentage signal change at repetition times of 3–6 from block onset were averaged across all blocks for each condition and ROI and was used as the dependent measure for the average activity.

Statistical analysis was conducted using SPSS 20. Two-tailed paired *t* test was used for statistically comparing two samples, unless otherwise specified.

#### **Eye Tracking Control**

Eye movements of six participants were tracked in the scanner. We used iView X MRI-LRsystem (SMI Sensor-motoric Instruments) with a sampling rate of 50 Hz. For four participants, we did not manage to acquire data because of unsuccessful calibration and misidentification of the pupil. One of these four participants was also excluded from the whole analysis because we were unable to identify object general areas in his data. Therefore, eye tracking data were analyzed for the two remaining participants. The system output files were converted into a text format and analyzed using a custom-made Matlab code for each participant separately. For each run, the median *x* and *y* eye position coordinates during fixation blocks were computed and subtracted from the *x* and *y* eye coordinates during the experiment blocks. The *x* and *y* coordinates for each condition were then pooled across the two experiment blocks of each condition in each run, and their mean and standard deviation were computed.

## **RESULTS**

### **Behavioral Effects of Clutter**

We assessed the effect of clutter on behavioral performance by computing both accuracy and RTs. The overall accuracy level was very high (mean  $\pm$  SEM: 95.2  $\pm$  0.9%). Repeated-measures ANOVA of the proportion of correct responses with clutter condition (no clutter, homogeneous clutter, and heterogeneous clutter) and category (face, house) as within-subject factors showed no effect for the clutter condition,  $F(2, 30) < 1$ , and no interaction,  $F(2, 30) < 1$ . Repeated-measures ANOVA of the RTs with clutter condition and category as within-subject factors revealed a marginally significant main effect for clutter condition,  $F(2, 30) = 3.2$ ,  $p = .05$ , and no interaction,  $F(2, 30) = 2.9$ ,  $p > .05$ . RTs across categories were shorter for the no-clutter (mean  $\pm$  SEM: 0.54  $\pm$  0.02 sec) condition compared with heterogeneous clutter (mean  $\pm$  SEM: 0.56  $\pm$  0.02 sec),  $t(15) = 2.5$ ,  $p = .02$ . There was no significant difference in RTs between the no-clutter and homogeneous clutter (mean  $\pm$  SEM: 0.55  $\pm$  0.02 sec) conditions,  $t(15) = 1$ ,  $p = .4$ , or between the two clutter displays,  $t(15) = 1.7$ ,  $p = .1$ .

### **Effect of Clutter on Target Object Representation in Object General Areas**

#### *Effect of Clutter on Decoding Performance*

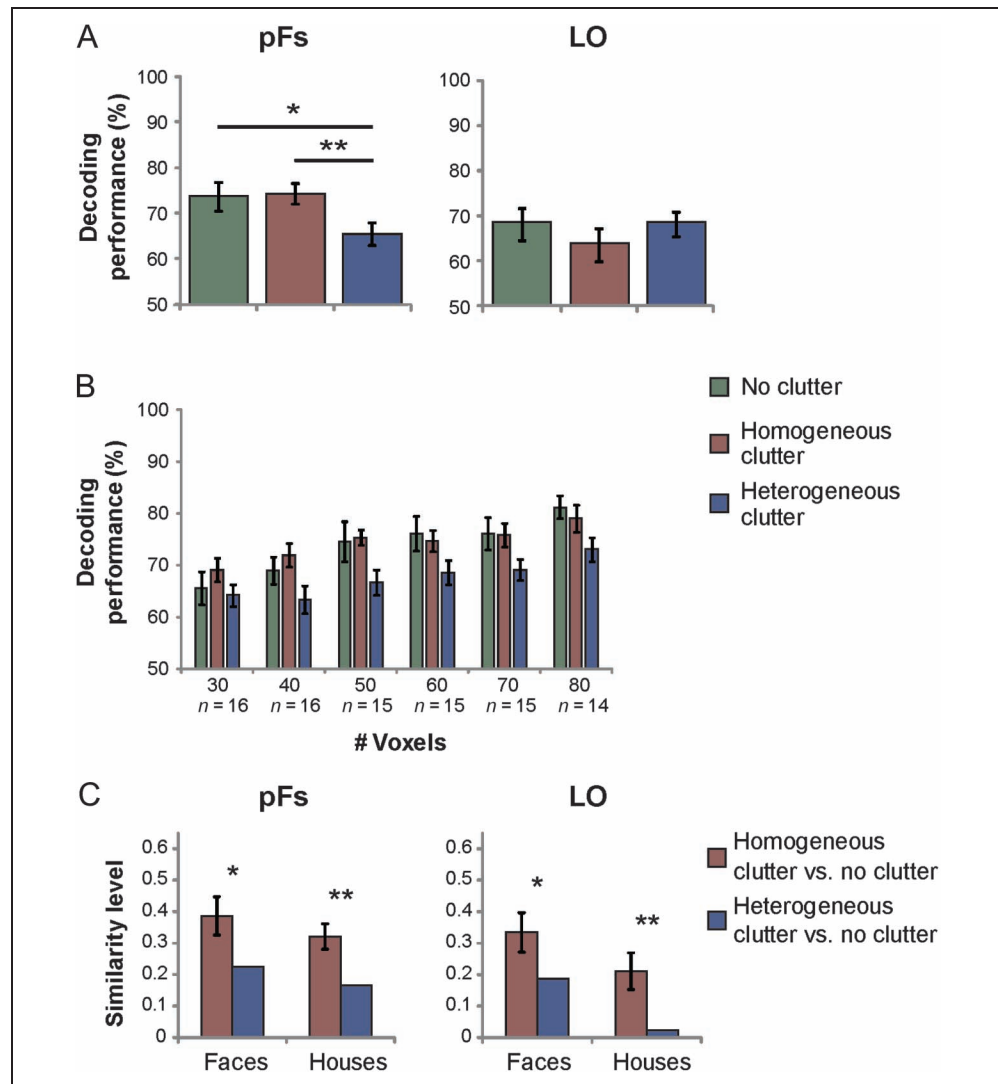
We first asked whether clutter interferes with the decoding of the target stimuli. Discrimination between faces and houses was evaluated for the three clutter conditions based on the voxelwise pattern, resulting in the accuracy level of classification (decoding performance) for each condition and ROI (see Methods for details). A two-way repeated-measures ANOVA of decoding performance with hemisphere and clutter condition as within-subject factors revealed no interaction between the factors in both pFs and LO (pFs:  $F(2, 12) < 1$ ,  $p = .6$ ; LO:  $F(2, 28) < 1$ ,  $p = .6$ ); therefore, decoding performance was averaged across hemispheres.

Classification between faces and houses was highly successful even in the presence of clutter: Decoding performance for each of the three clutter conditions was significantly above chance level in the two ROIs (one-tailed one-sample *t* test against 50% on decoding performance for all clutter conditions:  $t > 3.5$ ,  $p < .004$ ).

We then examined the effect of clutter on the level of decoding (Figure 2A) and found a significant difference between the clutter conditions in the pFs, but not in LO. Two-way repeated-measures ANOVA revealed a significant interaction between clutter condition and ROI,  $F(2, 28) = 4.65$ ,  $p = .02$  (one-way repeated-measures ANOVA for each ROI: pFs:  $F(2, 30) = 3.8$ ,  $p = .03$ ; LO:  $F(2, 28) = 1.4$ ,  $p = .27$ ). We next asked whether the presence of a clutter reduced decoding performance relative to the no-clutter condition in the pFs. Decoding did not differ significantly when a homogeneous clutter was added, compared with



**Figure 2.** Neural representations of target objects are modulated by the type of clutter in object general areas. (A) Voxelwise pattern-based discrimination between the two target categories under the three clutter conditions in object general areas. (B) Decoding performance in the pFs for the two target categories is presented for the three clutter conditions across a range of ROI sizes used for classification. (C) Voxelwise pattern-based similarity in object general areas. Similarity was computed between the no-clutter condition and the homogeneous or heterogeneous clutter conditions for each target category separately. Error bars indicate *SEM*. For A and C, \* $p < .05$ , \*\* $p < .01$ .



the no-clutter condition,  $t(15) = 0.1$ ,  $p = .9$ . However, a heterogeneous clutter significantly reduced classification performance,  $t(15) = 2.4$ ,  $p = .03$ . To further investigate the effect of the type of clutter, we asked whether decoding performance in the pFs was interrupted by the heterogeneous display to a larger extent compared with the homogeneous display and found that decoding performance was significantly better for the homogeneous clutter condition compared with the heterogeneous one,  $t(15) = 3.2$ ,  $p = .006$  (Figure 2A). The same analysis was conducted using varying sizes of ROIs (30–80 voxels) to verify that the effect is not limited to a specific ROI size. Decoding performance increased as a function of ROI size but showed similar effects of clutter on decoding performance across the range of ROI sizes (Figure 2B), indicating the robustness of this finding. Decoding performance in the pFs for the homogeneous clutter condition was larger compared with the heterogeneous clutter ( $p < .05$  for 40, 50, and 70 voxels;  $.05 < p < .08$  for 30, 60, and 80 voxels). In the LO, there were no differences between the clutter conditions across the range of ROI

sizes ( $F < 2.5$ ,  $p > .1$ ). Thus, the discrimination between the two categories is interrupted by a heterogeneous clutter, but not by a homogeneous clutter in the pFs.

#### Effects of Clutter on Pattern Similarity

To further investigate the effect of the type of clutter on the neural representation, we tested the similarity between neural patterns of responses evoked under the different clutter displays. The magnitude of similarity between each of the clutter conditions and the no-clutter presentation was assessed using correlation between their evoked voxelwise patterns across the entire data set within each category (see Methods for details). Repeated-measures ANOVA with hemisphere, clutter condition (homogeneous clutter, heterogeneous clutter), and category (face, house) as within-subject factors revealed no interaction between the factors in both pFs and LO ( $F < 1$ ); therefore, the similarity values were averaged across hemispheres. In both pFs and LO, the multivoxel pattern evoked by the no-clutter condition was more similar to

the pattern evoked by the homogeneous than the heterogeneous displays, for both categories (pFs:  $t(15) = 2.7$ ,  $p = .018$  for faces,  $t(15) = 3.9$ ,  $p = .001$  for houses; LO:  $t(14) = 2.4$ ,  $p = .03$  for faces,  $t(14) = 3.2$ ,  $p = .006$  for houses; Figure 2C). In other words, the neural representation of the homogeneous clutter resembled the no-clutter pattern more than that of the heterogeneous clutter. These findings imply that the content of the task-irrelevant clutter is represented in LOC, at least partially, and not just the number of stimuli. To ensure that the results we obtained are not limited to a specific ROI size, we conducted the same analysis for each ROI using a broad range of ROI sizes. The same pattern of findings was evident, showing greater similarity between homogeneous and no-clutter displays compared with the heterogeneous clutter and the no-clutter conditions for the two categories across all ROI sizes ( $t > 2.2$ ,  $p < .05$ ).

To provide a baseline for the between-condition similarities, we also computed correlations based on split-halves of the data, which allowed measuring the similarity between the no-clutter condition and itself. This analysis indicates whether the representation of a single stimulus is different from each of the 2 four-stimulus displays but does not assess the effect of clutter on the discriminability of the target objects. Repeated-measures ANOVA with hemisphere, clutter condition (no-clutter, homogeneous clutter, heterogeneous clutter), and category (face, house) as within-subject factors revealed no interaction with hemisphere in both pFs and LO ( $p > .1$ ); therefore, the similarity values were averaged across hemispheres. Repeated-measures ANOVA with clutter condition and category as within-subject factors revealed no interaction with category in both pFs and LO (pFs:  $F(2, 30) < 1$ ,  $p = .5$ ; LO:  $F(2, 28) = 2.4$ ,  $p = .1$ ) and a significant effect of clutter condition (pFs:  $F(2, 30) = 10.5$ ,  $p < .001$ ; LO:  $F(2, 28) = 30.8$ ,  $p < .001$ ). In both pFs and LO, the multivoxel pattern evoked by the no-clutter condition was more similar to itself than to the pattern evoked by both the homogeneous and the heterogeneous displays across categories (pFs:  $t(15) > 3.3$ ,  $p < .005$ ; LO:  $t(15) > 6$ ,  $p < .001$ ). These findings indicate that the representation of a single-stimulus display is not the same as the four-stimulus display. In contrast to the clear significant differences between the two clutter displays observed in pFs and LO when correlations were computed based on the entire data set as described above, there was no difference in similarity between each of the clutter displays to the no-clutter condition, in both pFs and LO, when the split-half correlations were used (pFs:  $t(15) = 0.9$ ,  $p = .4$ ; LO:  $t(14) = 0.9$ ,  $p = .38$ ). This is most likely because of the use of only half of the data, which may yield noisier voxelwise patterns across runs that are less sensitive to effects that can be revealed when the entire data are considered. Indeed, the level of similarity obtained by the split half analysis was much lower than the level of similarity found when all the data set was included (averaged across ROIs, clutter conditions, and categories:  $t(14) = 11.5$ ,  $p < .001$ ).

Similar results were obtained when computing similarity using varying ROI sizes.

### **Effect of Clutter on Target Object Representation in Category-selective Areas**

We next assessed the effects of clutter on responses to preferred objects in category-selective areas. On the basis of previous findings that showed robustness of response to clutter in these areas (Reddy & Kanwisher, 2007), we expected a smaller or no effect of clutter on the representation of target objects compared with the object general areas.

#### *Effect of Clutter on Decoding Performance*

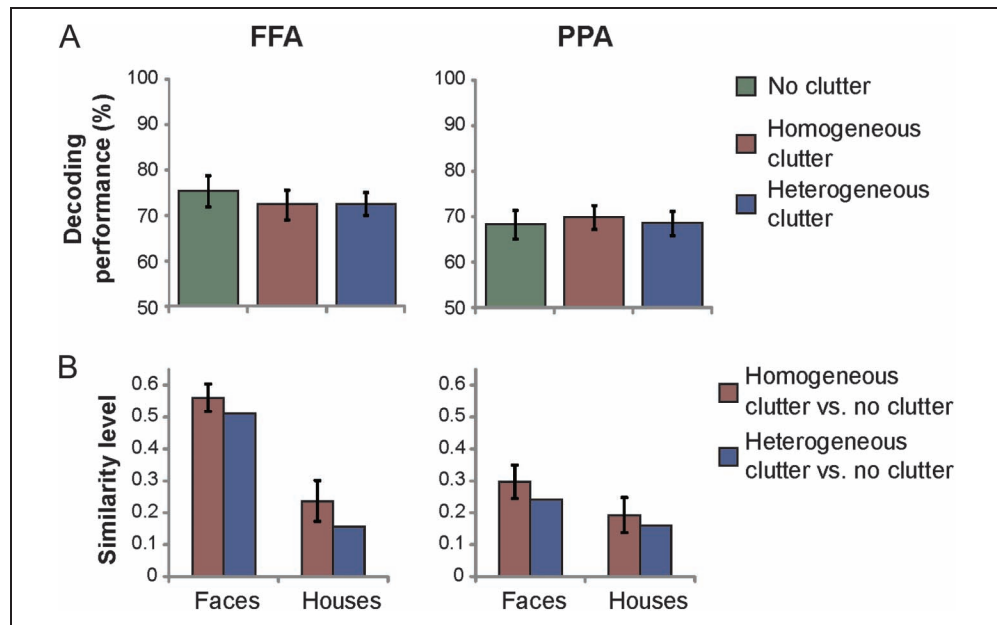
We used MVPA to assess the effect of clutter on decoding performance in category-selective areas (Figure 3A). A two-way repeated-measures ANOVA of decoding performance with hemisphere and clutter condition as within-subject factors revealed no interaction between the factors in the FFA and PPA (FFA:  $F(2, 22) = 2.2$ ,  $p = .1$ ; PPA:  $F(2, 30) = 2.4$ ,  $p = .11$ ). Therefore, decoding performance was averaged across hemispheres.

Decoding performance in the FFA and PPA for each of the three clutter conditions was significantly above chance level (one-tailed one-sample  $t$  test against 50% on decoding performance for all clutter conditions:  $t > 5.5$ ,  $p < .001$ ) and was not affected by the clutter condition (one-way repeated-measures ANOVA: FFA:  $F(2, 26) < 1$ ,  $p = .7$ ; PPA:  $F(2, 30) < 1$ ,  $p = .9$ ). These results imply that in category-selective areas target discrimination is not affected by either the presence of a clutter or the type of the clutter. The lack of effects was also evident when we conducted classification in each ROI using varying ROI sizes (one-way repeated-measures ANOVA,  $p > .29$ ).

#### *Effects of Clutter on Patterns Similarity*

The voxelwise pattern similarity between each of the two clutter conditions and the no-clutter condition in the FFA and PPA was assessed using correlation to test for the effect of the type of clutter on the response pattern (Figure 3B). Repeated-measures ANOVA with hemisphere, clutter condition, and category as within-subject factors revealed no interaction between the factors in both FFA and PPA ( $F < 1$ ); therefore, similarity level was averaged across hemispheres. In both ROIs, there was no effect of the clutter type on the neural representations: The voxelwise pattern similarity between the heterogeneous clutter and the no-clutter condition was not significantly different from the similarity between the homogeneous clutter and the no-clutter conditions for both categories (FFA:  $t(13) < 1.3$ ,  $p > .23$ ; PPA:  $t(15) < 1.1$ ,  $p > .3$ ). Similar results were obtained when computing similarity using varying ROI sizes. These findings imply that in category-selective areas

**Figure 3.** Neural representations of preferred target objects are insensitive to clutter in category-selective areas. (A) Voxelwise pattern-based discrimination between faces and houses under the three clutter conditions in category-selective areas. (B) Voxelwise pattern-based similarity in category-selective areas. Similarity was computed between the no-clutter condition and the homogeneous or heterogeneous clutter conditions for each target category separately. Error bars indicate *SEM*.



the pattern of response does not contain information about the content of the clutter.

To further explore the representation of single-stimulus display compared with four-stimulus display, we also computed correlations based on split-halves of the data, which allowed measuring the similarity between the no-clutter condition and itself to serve as a baseline for the between-condition similarities. Repeated-measures ANOVA with hemisphere, clutter condition (no clutter, homogeneous clutter, heterogeneous clutter), and category (face, house) as within-subject factors revealed no interaction with hemisphere in both FFA and PPA ( $p > .05$ ); therefore, the similarity values were averaged across hemispheres. Repeated-measures ANOVA with clutter condition and category as within-subject factors revealed no interaction with category in both FFA and PPA (FFA:  $F(2, 26) < 1, p = .8$ ; PPA:  $F(2, 30) < 1, p = .5$ ) and a significant effect of clutter condition (FFA:  $F(2, 26) = 3.4, p < .05$ ; PPA:  $F(2, 30) = 14.9, p < .001$ ). In the FFA, the multivoxel pattern evoked by the no-clutter condition was more similar to itself than to the pattern evoked by the homogeneous clutter across categories,  $t(13) = 3.2, p = .007$ , and there

was no difference between the similarity of the no-clutter condition and itself and the similarity of the no-clutter and heterogeneous clutter conditions,  $t(13) = 0.9, p = .4$ . In the PPA, the multivoxel pattern across categories evoked by the no-clutter condition was more similar to itself than to the pattern evoked by both the homogeneous and the heterogeneous displays,  $t(15) > 4.2, p < .001$ . These findings suggest that the representation of a four-stimulus display is different than a single-stimulus display in the PPA and, to some extent, also in the FFA. Similar to the results obtained when the entire data set was used for the correlations, there was no difference in similarity between each of the clutter displays to the no-clutter condition in both FFA and PPA (FFA:  $t(13) = 1.7, p = .1$ ; PPA:  $t(15) = 0.9, p = .4$ ). Similar results were obtained when computing similarity using varying ROI sizes.

### Effect of Clutter on Average fMRI Responses

We computed the average fMRI responses for the two categories under the three clutter conditions in each of the ROIs (Table 1). In all the ROIs, data were pooled

**Table 1.** The Effect of Clutter on the Average Level of fMRI Signal

	Faces			Houses		
	No clutter	Homogeneous Clutter	Heterogeneous Clutter	No clutter	Homogeneous Clutter	Heterogeneous Clutter
pFs	0.188 ± 0.1	0.45 ± 0.08	0.44 ± 0.12	0.54 ± 0.12	0.67 ± 0.12	0.77 ± 0.1
LO	0.52 ± 0.08	0.67 ± 0.1	0.66 ± 0.1	0.55 ± 0.1	0.67 ± 0.11	0.76 ± 0.09
FFA	1.1 ± 0.1	1.1 ± 0.12	1.1 ± 0.12	0.26 ± 0.07	0.22 ± 0.08	0.33 ± 0.08
PPA	-0.52 ± 0.05	-0.33 ± 0.05	-0.33 ± 0.06	0.1 ± 0.06	0.15 ± 0.06	0.27 ± 0.05

Percent signal change (%) is presented for the two categories under the three clutter conditions for each of the ROIs (mean ± *SEM*).

across hemispheres using a weighted average based on the relative volumes of the areas, as no differences were found between the two hemispheres.

In object general areas, the overall response was modulated by the presence of the clutter regardless of its type. Two-way repeated-measures ANOVA with clutter condition and category as within-subject factors revealed significant effect for clutter condition (pFs:  $F(2, 30) = 15.85, p < .001$ ; LO:  $F(2, 30) = 15.97, p < .001$ ) and no interaction,  $F(2, 30) < 1.5, p > .25$ . In both pFs and LO, the response to the no-clutter condition was significantly smaller than the homogeneous and heterogeneous clutter conditions,  $t(15) > 4.1, p < .001$ , and there was no difference between the two types of clutters,  $t(15) < 1.4, p > .19$ .

In category-selective areas, the FFA and PPA, we found different effects: clutter had no effect on the average fMRI response in the FFA ( $p > .3$ ) but did modulate the average response to the target stimuli in the PPA. In the PPA, repeated-measures ANOVA with clutter condition and category as within-subject factors revealed a significant effect of clutter condition,  $F(2, 30) = 27.2, p < .001$ , and no interaction,  $F(2, 30) = 2.85, p > .07$ . The three clutter conditions were significantly different from each other,  $t(16) > 2.4, p < .03$ , with the smallest response to the no-clutter condition and the largest response to the heterogeneous clutter condition.

We concluded that, although the average response carry some information about the clutter display in the LOC and PPA, it might comply to the number of presented stimuli, rather than the type of clutter, in particular in the LOC. This may be consistent with previous evidence for retinotopic organization in these areas (e.g., Brewer, Liu, Wade, & Wandell, 2005; Hasson, Levy, Behrmann, Hendler, & Malach, 2002; Sereno, Pitzalis, & Martinez, 2001). However, the voxelwise pattern provides additional data compared with the average signal regarding the information conveyed in the neural response, as was demonstrated by both the current study and many previous studies (e.g., Haynes & Rees, 2006; Haxby et al., 2001).

### Analysis of Eye Movements

Eye tracking data inside the scanner were collected and analyzed for two participants during scanning to ensure that the participants maintained fixation throughout the experiment. Analysis of eye position showed that participants were fixating on the fixation dot without moving their eyes toward the center of the target stimuli across all the experimental conditions. The mean  $x$  and  $y$  coordinates across conditions and runs overlapped with the fixation dot for both participants. We further wanted to ensure that there were no differences in eye positions across conditions. For each participant, the  $x$  and  $y$  coordinates were averaged for each condition and run. Then two-way repeated-measures ANOVA was computed for each coordinate, with category and clutter condition as within-subject factors, for each participant separately (Axelrod & Yovel,

2012; Schwarzlose, Swisher, Dang, & Kanwisher, 2008). For the two participants and two coordinates, there was no main effect of category or clutter condition and no interaction between the two ( $p > .05$ ). Similar analysis of the standard deviation of each coordinate in each condition and run showed no main effect of category or clutter condition and no interaction between the factors for both participants ( $p > .05$ ), implying that the variability of eye position was similar across the experimental conditions.

## DISCUSSION

The current study investigated the neural representation of target objects when presented simultaneously with irrelevant nontarget objects (clutter), simulating real-world cluttered scenes containing target objects required for goal-directed behavior. Furthermore, by manipulating the variation among irrelevant objects, we examined how different types of clutter may influence the representation of target objects in high-level visual object areas. Specifically, we showed that a heterogeneous but not homogeneous clutter (pop-out) interfered with decoding of target objects in the pFs part of the object area. Consistent with that, complementary analysis of the similarity among the response patterns of the two clutter types and the isolated display revealed that the response pattern to an isolated target object was more similar to its pattern when presented with a homogeneous clutter than when presented with a heterogeneous clutter. Interestingly, decoding of preferred target objects in category-selective areas (i.e., faces and houses in the face and place-selective areas) were not affected by the presence of the clutter or the type of the clutter.

The effect of homogenous and heterogeneous clutter on the representation of target stimuli has been shown long ago in behavioral studies (Duncan & Humphreys, 1989; Treisman & Gormican, 1988), and their neural correlates have been demonstrated for simple visual stimuli in both neurophysiological (Kastner, Nothdurft, & Pigarev, 1999; Nothdurft, Gallant, & Van Essen, 1999; Knierim & van Essen, 1992) and neuroimaging (Beck & Kastner, 2005) studies in early visual areas. More recent studies have examined the representation of multiple objects in higher-level visual areas. Decreased decoding performance was found for target objects presented simultaneously with one other object, compared with objects presented in isolation (Reddy & Kanwisher, 2007), indicating the cost of clutter and in line with a previous evidence for the processing of context information in LOC (Altmann, Deubelius, & Kourtzi, 2004). Here, we expanded these findings by using clutters comprising multiple objects as well as different types of clutters and showed that heterogeneous but not homogenous clutter interferes with the target representation. Findings from other studies, mainly focusing on target objects presented in natural scenes containing multiple objects, have demonstrated that categorical information about single objects embedded in a



scene could be well extracted from their multivoxel pattern response (Peelen et al., 2009). Furthermore, a scene could be well represented based on the information obtained by its constituent objects (MacEvoy & Epstein, 2011; Park, Brady, Greene, & Oliva, 2011). Manipulation of the type of clutter in our study allowed us to show that different types of clutter may differently influence the representation of a task-relevant target object.

In contrast to the effect of clutter that we found in LOC, our results showed that, although the addition of the nontarget objects to a preferred target object (faces and houses) was represented in their category-selective brain areas (FFA and PPA), the decoding of preferred target objects was not modulated by the presence of clutter. These findings are in line with the robustness to clutter previously demonstrated in these areas when a target was presented with one nontarget object (Reddy & Kanwisher, 2007). This dissociation between general object areas and category-selective areas may imply that, although object general areas represent information about both targets and surrounding objects, specialized category-selective areas disregard the irrelevant information of non-preferred stimuli. These findings may be consistent with recent behavioral studies showing pop-out effect for faces even when presented among heterogeneous object display (Hershler & Hochstein, 2005). Future studies are needed to assess whether the same remains when the irrelevant clutter is composed of multiple preferred objects, like displays of faces in a crowd.

Another issue to consider is the difference we observed between the pFs and LO. For both areas, we found that clutter information was represented, but it significantly interfered with decoding of the target objects in the pFs but not in the LO. The decoding and similarity analyses provide different and complementary information about the effect of clutter on object representation. The decoding analysis measures the discriminability between conditions, and we used it here to assess the level of discrimination between the target objects (face and house) within each of the three clutter types. Therefore, it reflects the nature of the representation of the target objects themselves. The similarity analysis measures the relative distance between the representations of two conditions in the representational space. We compared within-category correlations between the no-clutter condition and each of the clutter conditions, thus measuring the representation of the entire display (i.e., target and clutter), rather than the target objects per se. Therefore, the effect of clutter type found in LO using the similarity analysis, but not the decoding analysis, implies that clutter is represented in LO to some extent but does not modulate the discriminability of the target objects as it does in the pFs. Other studies also showed differences between the two areas in regard to representation of multiple-object displays (MacEvoy & Epstein, 2011): Prediction of scene representations from their constituent objects failed to exceed chance level in the pFs but was above chance level in the LO. Both these

results and the results presented in the current study suggest that target object representations are modulated by the surrounding objects in pFs, but not in LO. However, further studies are required to clarify the distinctive roles of the pFs and LO in the processing of multiple-object displays. It is noteworthy that all ROIs exhibited discrimination between the target object categories significantly above chance level for all clutter conditions: Even when information about the clutter interfered with decoding, the target objects could still be well discriminated, in accordance with the high success rate in the task across all conditions. This implies that irrelevant clutter information could be still filtered out efficiently, even in the pFs, though to a lesser extent.

Finally, our findings raise an important question regarding the neural coding scheme that may underlie the effect of irrelevant clutter objects on the response to target objects. One possibility is that information about each of the irrelevant nontarget objects is represented as randomized noise added to the pattern of the target object representation, without an actual representation for the category of each object. According to this coding scheme, the magnitude of added noise increases with the number of objects, and a large amount of noise then leads to an interference with decoding. This view predicts that the number of objects in a clutter, rather than the variation of their identity or category, will influence decoding of target stimuli. Thus, homogeneous and heterogeneous clutters are expected to interfere with decoding to a similar extent relative to an isolated display as long as they contain the same number of irrelevant objects. A second coding scheme is based on previous theoretical (Bundesen, Habekost, & Kyllingsbaek, 2005, 2011) and experimental (MacEvoy & Epstein, 2009, 2011; Reddy et al., 2009; Zoccolan, Cox, & DiCarlo, 2005; Reynolds, Chelazzi, & Desimone, 1999) studies that suggested weighted average mechanism for the representation of multiple simultaneously presented objects. According to this coding scheme, each of the task-irrelevant nontarget objects is represented with its own pattern, additionally to the pattern of the target object, and these patterns sum up, possibly via a weighted average mechanism, based on their relative behavioral significance, to generate a combined response pattern. The relative weights may vary between areas, also depending on the category of the target object and the nontarget objects, as well as other factors such as attention, thus leading to reduced decoding in some of the areas, and preserved performance in others. Importantly, this coding scheme predicts that the representations of target objects may be sensitive to the variability among nontarget objects, thus yielding different effects for homogeneous and heterogeneous clutters. For example, nontargets in a heterogeneous display may attract more attention than in a homogeneous display; thus, some of the areas have a larger weight in the final response pattern. Our findings, showing differences in the responses to the two types of clutters in the pFs and LO, together

with previous evidence demonstrating the processing of multiple objects and clutter information in object general areas (MacEvoy & Epstein, 2009, 2011; Reddy et al., 2009; Reddy & Kanwisher, 2007; Altmann et al., 2004) seem consistent with this latter putative mechanism.

Although the nontarget objects may gain some considerable representational weight in the object general area, these weights may be minimized in the more fine-grained selective processing in the PPA and FFA, thus resulting in the lack of effect for clutter type that we found in these areas. However, given there was no differential effect for the two clutter types in these regions, it may be that the nontarget objects are represented as randomized noise, and our data would not allow us to conclude which coding scheme may be at work there. In conclusion, neural representations of objects embedded in clutter displays have gained considerable interest in the past few years. This study is the first to show the effect of different types of clutters on the neural representation of target objects. Specifically, we demonstrated that heterogeneous clutter, but not homogeneous one (pop-out display), interfered with the representation of target objects in the object general area. Further questions that should be investigated in future studies are how representations of target objects are affected by other types of clutters and contextual displays, which are perceptually and/or semantically related to the target object.

## Acknowledgments

This study was supported by an Israel Science Foundation grant 446/12 and a Wolfson Foundation grant to G. Y. We thank Ido Tavor and Jonathan Oron for their help with scanning and setting up the eye tracking system.

Reprint requests should be sent to Yaara Erez, Department of Psychological Sciences, Tel-Aviv University, P.O. Box 39040, Ramat Aviv, Tel Aviv, 6997801 Israel, or via e-mail: yaara.erez@gmail.com.

## REFERENCES

- Altmann, C. F., Deubelius, A., & Kourtzi, Z. (2004). Shape saliency modulates contextual processing in the human lateral occipital complex. *Journal of Cognitive Neuroscience*, *16*, 794–804.
- Axelrod, V., & Yovel, G. (2012). Hierarchical processing of face viewpoint in human visual cortex. *Journal of Neuroscience*, *32*, 2442–2452.
- Beck, D. M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, *8*, 1110–1116.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). *Region of interest analysis using an SPM toolbox* [abstract]. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan. Available on CD-ROM in Neuroimage, Vol. 16, No. 2, Abstract 497.
- Brewer, A. A., Liu, J., Wade, A. R., & Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience*, *8*, 1102–1109.
- Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review*, *112*, 291–328.
- Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2011). A neural theory of visual attention and short-term memory (NTVA). *Neuropsychologia*, *49*, 1446–1457.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.
- Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*, *34*, 479–490.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.
- Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*, 523–534.
- Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, *45*, 1707–1724.
- Jeong, S. K., & Xu, Y. (2013). Neural representation of targets and distractors during object individuation and identification. *Journal of Cognitive Neuroscience*, *25*, 117–126.
- Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, *282*, 108–111.
- Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1999). Neuronal responses to orientation and motion contrast in cat striate cortex. *Visual Neuroscience*, *16*, 587–600.
- Knierim, J. J., & van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, *67*, 961–980.
- MacEvoy, S. P., & Epstein, R. A. (2009). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology*, *19*, 943–947.
- MacEvoy, S. P., & Epstein, R. A. (2011). Constructing scenes from objects in human occipitotemporal cortex. *Nature Neuroscience*, *14*, 1323–1329.
- Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage*, *53*, 103–118.
- Nothdurft, H. C. (1993). The role of features in preattentive vision: Comparison of orientation, motion and color cues. *Vision Research*, *33*, 1937–1958.
- Nothdurft, H. C., Gallant, J. L., & Van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: Correlates of “popout” under anesthesia. *Visual Neuroscience*, *16*, 15–34.
- Park, S., Brady, T. F., Greene, M. R., & 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. *Journal of Neuroscience*, *31*, 1333–1340.
- Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*, 94–97.
- Peelen, M. V., & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, 12125–12130.
- Reddy, L., & Kanwisher, N. (2007). Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Current Biology*, *17*, 2067–2072.
- Reddy, L., Kanwisher, N. G., & VanRullen, R. (2009). Attention and biased competition in multi-voxel object representations.

- Proceedings of the National Academy of Sciences, U.S.A.*, 106, 21447–21452.
- Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, 19, 1736–1753.
- Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, 105, 4447–4452.
- Seidl, K. N., Peelen, M. V., & Kastner, S. (2012). Neural evidence for distracter suppression during visual search in real-world scenes. *Journal of Neuroscience*, 32, 11812–11819.
- Serences, J. T., Saproo, S., Scolari, M., Ho, T., & Muftuler, L. T. (2009). Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *Neuroimage*, 44, 223–231.
- Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science*, 294, 1350–1354.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15–48.
- Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *Journal of Neuroscience*, 25, 8150–8164.