

Published in final edited form as:

J Neurosci. 2012 August 22; 32(34): 11812–11819. doi:10.1523/JNEUROSCI.1693-12.2012.

Neural evidence for distracter suppression during visual search in real-world scenes

Katharina N. Seidl^{a,b,*}, Marius V. Peelen^{a,b,c,1,*}, and Sabine Kastner^{a,b}

^aDepartment of Psychology, Princeton University, Princeton, NJ 08544

^bPrinceton Neuroscience Institute, Princeton University, Princeton, NJ 08544

^cCenter for Mind/Brain Sciences, University of Trento, 38068 Rovereto, Italy

Abstract

Selecting visual information from cluttered real-world scenes involves the matching of visual input to the observer's attentional set -- an internal representation of objects that are relevant for current behavioral goals. When goals change, a new attentional set needs to be instantiated, requiring the suppression of the previous set to prevent distraction by objects that are no longer relevant. In the present fMRI study, we investigated how such suppression is implemented at the neural level. We measured human brain activity in response to natural scene photographs that could contain objects from (i) a currently relevant (target) category, and/or (ii) a previously but not presently relevant (distracter) category, and/or (iii) a never relevant (neutral) category. Across conditions, multi-voxel response patterns in object-selective cortex (OSC) carried information about objects present in the scenes. However, this information strongly depended on the task relevance of the objects. As expected, information about the target category was significantly increased relative to the neutral category, indicating top-down enhancement of task-relevant information. Importantly, information about the distracter category was significantly reduced relative to the neutral category, indicating that the processing of previously relevant objects was suppressed. Such active suppression at the level of high-order visual cortex may serve to prevent the erroneous selection of, or interference from, objects that are no longer relevant to ongoing behavior. We conclude that the enhancement of relevant information and the suppression of distracting information both contribute to the efficient selection of visual information from cluttered real-world scenes.

Introduction

At any point in time, only a small portion of our cluttered visual environment contains information that is relevant for ongoing behavior, requiring us to extract the subset of information that is currently important to us. During real-world visual search the selection of relevant information may be achieved by matching incoming visual input to a top-down attentional set – an internal representation or ‘template’ of objects that are currently relevant (Duncan and Humphreys, 1989). At the neural level, top-down attentional sets modulate responses in inferior temporal cortex, biasing responses in favor of behaviorally relevant objects (Chelazzi et al., 1993; Chelazzi et al., 1998; Peelen et al., 2009; Peelen and Kastner, 2011).

¹To whom correspondence should be addressed. Center for Mind/Brain Sciences, University of Trento Corso Bettini 31 38068 Rovereto (TN), Italy phone +39 0464 808718 marius.peelen@unitn.it.

*K.N.S. and M.V.P. contributed equally to this work

In real-world settings goals change frequently, requiring that the current attentional set be exchanged for a new one. Previous studies have shown that visual search is more difficult when the distracters in the display are items that were targets on earlier trials (Shiffrin and Schneider, 1977; Dombrowe et al., 2011; Evans et al., 2011), an effect that can be long-lasting when the previous attentional set was well established (Shiffrin and Schneider, 1977). This suggests that the instantiation of a new attentional set may require the sustained suppression of previous attentional sets in order to prevent selection of information that is no longer relevant (Shiffrin and Schneider, 1977).

The current study was aimed at providing neural evidence for the hypothesis that adopting a new attentional set involves not only the enhancement of the new task-relevant attentional set, but also the active suppression of the previously but no longer relevant attentional set. Specifically, using fMRI we compared the extent to which objects embedded in scenes (i.e. within-scene objects) were processed as a function of task relevance during a real-world visual search task. By requiring participants to detect different object categories over the course of the experiment, we could measure brain activity in response to scene photographs that contained objects from a currently relevant category (target category) or objects from a category that was previously but no longer relevant (distracter category). Crucially, in order to obtain a baseline against which to compare responses to the target and distracter categories, we also included scenes containing objects from a third category that was never task-relevant (neutral category). If changing attentional sets solely requires the enhancement of task-relevant information, the distracter and neutral categories should be processed to the same extent, while processing of the target category should be enhanced. Alternatively, changing attentional sets may require both suppression of the previous attentional set and enhancement of the current attentional set, resulting in increased processing of the target category but decreased processing of the distracter category, both relative to the neutral category (Fig. 1).

Materials and Methods

Participants

Twenty-six healthy adult volunteers (14 females, age: 19 – 32) participated in the study. All participants (normal or corrected-to normal vision; no history of neurological or psychiatric disease) gave informed written consent for participation in the study, which was approved by the Institutional Review Panel of Princeton University. The data from two participants were discarded due to excessive head movement. All participants were naïve as to the goals of the study.

Visual stimuli & experimental design

Participants were tested in three different tasks that were run within the same scanning session and will be referred to as (i) category detection task, (ii) category pattern localizer and (iii) object-selective cortex localizer.

Category detection task—The category detection task included 192 photographs depicting landscapes and cityscapes that were selected from an online database (Russell et al., 2008) according to the presence of three object categories: people, cars and trees. Overall there were eight different scene types that differed with regard to the number and combination of depicted object categories (Fig. 4-6). Scenes could contain either (i) a single category (people only, cars only, trees only; Fig. 4), (ii) two of the three categories (i.e. trees and people, trees and cars, cars and people; Fig. 5), (iii) all three categories (people, cars, trees; Fig. 6), or (iv) none of the three categories (no category). 24 photographs of each scene type were included in the experiment. Across scene photographs, the category

exemplars embedded in the scenes varied in appearance, spatial locations, size and viewpoint. All scenes were presented in grey-scale. The scenes subtended $10^{\circ} \times 13^{\circ}$ visual angle and were presented against a white background. Natural scene photographs were directly followed by perceptual masks of equal size. Masks were generated by combining white noise at different spatial frequencies and superimposing naturalistic structure on the noise (Walther et al., 2009). A projector outside the scanner room displayed the stimuli onto a translucent screen located at the end of the scanner bore. Participants viewed the screen through a mirror attached to the head coil.

During the category detection task, the three object categories (people, cars, trees) were assigned to three different levels of behavioral relevance: (i) target, (ii) distracter and (iii) neutral. The target category was the category that participants were instructed to detect in a given run. The distracter category was the category that had been the target category in the preceding run. Thus, for any given participant two object categories alternated in serving as the target and distracter categories across the four runs of the experiment. Finally, the neutral category was never task-relevant and remained unmentioned to the participant for the entire duration of the experiment.

In order to ensure that any differences between the target, neutral and distracter category were not attributable to differences between the three specific object categories, the assignment of the three object categories to their levels of behavioral relevance was counterbalanced across participants, creating three different participant groups (Fig. 2A). To allow for statistical analyses across the three participant groups, all scene types were relabeled according to the behavioral relevance of the embedded object categories: target (T), neutral (N), distracter (D), target and neutral (T-N), target and distracter (T-D), neutral and distracter (N-D), all categories (T-N-D).

Each run began and ended with a fixation period of 14 s. At the beginning of each trial, the central fixation cross changed from black to grey for 500 ms to indicate the impending onset of the natural scene picture (Fig. 2B). Subsequently, the scene was presented centrally for 70 ms and followed by a randomly chosen perceptual mask that remained on the screen for 330 ms. Participants were required to give a speeded two-alternative forced choice response upon presentation of the scene, indicating the absence or presence of the target category. Each trial ended with a fixation period that varied in length from 950 ms to 1650 ms. Each run consisted of twelve instances of each of the eight scene types and 24 fixation-only trials, on which no natural scene or mask was presented, yielding a total of 120 trials per run. The target was present on 50% of the trials. Participants performed four runs. The average trial length was 2.2 s. In order to familiarize participants with the task and the task-relevant categories a practice session comprising four runs of the category detection task took place prior to the scanning session. A separate set of 192 stimuli from the same online database was used in the practice session.

Category pattern localizer—Following the category detection task, participants completed four runs of a category localizer in order to obtain category-selective response patterns in visual cortex. The localizer was run after the main experiment to ensure that participants remained unaware of the neutral category during the category detection task. Stimuli for the category localizer consisted of isolated exemplars from six object categories (trees, cars, headless bodies, faces, whole people, outdoor scenes) depicted on a white background. Of the six included object categories, only whole people, trees and cars were relevant to the present experiment. The stimuli were presented monochromatically and subtended $12^{\circ} \times 12^{\circ}$ visual angle.

One run consisted of 29 blocks of 14 s each. Blocks 1, 8, 15, 22 and 29 were fixation-only blocks, in which a central fixation cross was presented on a white background. One block of each condition was presented between two fixation epochs. For the first half of the experiment the order of blocks was determined randomly. The block order for the second half of the run mirrored the first half of the run in order to equate the mean serial position of each condition within the run. Each block contained ten sequential presentations of exemplars from the respective categories. Stimuli were presented in a jittered manner at random positions within 1° of the central fixation cross. They were shown for 400 ms and followed by a white fixation-only screen of equal duration. Participants performed a one-back repetition detection task. Stimulus repetitions occurred twice during each block.

Object-selective cortex localizer—At the end of the scanning session, participants performed two runs of a standard object-selective cortex localizer identical to that used in previous studies (Peelen et al., 2009; Peelen and Kastner, 2011). Pictures of 20 daily-life objects (e.g. telephone, cheese, slide, alarm clock) were presented in one condition and scrambled versions of the same objects in a second condition. The stimuli were presented monochromatically and subtended 12°×12° visual angle.

Data acquisition and analysis

Imaging data were obtained with a 3 Tesla MRI scanner (Siemens Allegra; Siemens, Erlangen, Germany) equipped with a standard head coil (Nova Medical, Wilmington, MA, USA). For all functional runs, thirty-four axially oriented slices (gap 1mm, 3×3×3mm³ voxels, field of view (FoV): 192, 64×64 matrix) covering the whole brain were acquired with a T2*-weighted echoplanar sequence (time of repetition (TR): 2000 ms, echo time (TE): 30 ms, flip angle (FA): 90°). A high spatial resolution T1-weighted anatomical scan (magnetization-prepared rapid acquisition gradient echo [MPRAGE], TR = 2500 ms, TE = 4.3 ms, FA = 8°, 256×256 matrix, voxel size: 1×1×1mm³) was obtained at the end of each session.

Data were analyzed using the AFNI software package (Cox, 1996) and MATLAB (Mathworks, Natick, MA). Images from functional runs were slice-time corrected and motion-corrected with reference to an image in the run closest to the anatomical scan. A temporal high-pass filter (cutoff of 0.006 Hz) was applied in order to remove low-frequency drifts from the data. Data from the object-selective cortex localizer were smoothed with a Gaussian kernel with a full-width half maximum of 4 mm. No spatial smoothing was applied to the data from the category detection task or the category localizer.

For each participant statistical analyses were performed using multiple regression within the framework of the general linear model (GLM). GLMs that served as a basis for MVPA were computed separately for the odd and even runs of the category detection task (due to the varying target and distracter categories across runs) as well as for the two localizer experiments. Time series for the category detection task were modeled by including regressors for each of the eight natural scene types. Regression models for the two localizer experiments included predictors for every condition. In order to compare beta weights across the different scene types, a separate GLM was computed on all four runs of the category detection task. The model included regressors for each of the eight scene types, relabeled for the level of behavioral relevance. All regressors of interest were convolved with a standard model of the hemodynamic response. Regressors of non-interest were included in order to account for head motion, linear drifts and quadratic drifts.

Our analysis focused on object-selective cortex (OSC) (Fig. 3A), where multivoxel response patterns have previously been shown to carry object category information (Haxby et al., 2001; Peelen et al., 2009). OSC was functionally defined in each individual participant.

Voxels in the temporo-occipital regions that were activated more strongly for intact compared to scrambled objects ($p < 0.05$, uncorrected for multiple comparisons) were included in the OSC ROI. Across participants, the average OSC size after combining both hemispheres was 349 voxels ($SD = 142$).

In order to compare results between OSC and early visual cortex, we defined Brodmann areas (BA) 17 and 18 using the Talairach atlas implemented in AFNI ('TT_Daemon'). Coordinates for BA17 and BA18 voxels were then projected back into each participant's native space and voxels that showed significant visually-evoked activation (versus fixation baseline; $p < 0.05$, uncorrected for multiple comparisons) during the category localizer were included in the two ROIs. ROIs of the left and right hemisphere were combined. Across participants, the average sizes of BA17 and BA18 were 86 voxels ($SD = 19$) and 295 voxels ($SD = 66$), respectively.

For each ROI, we obtained separate activation patterns for each scene type in the category detection task and object category in the category localizer by extracting the corresponding t-values from each voxel in the ROI. The activation patterns from the eight scene types of the detection task were then correlated with the patterns from the three categories in the category localizer, yielding an 8×3 correlation matrix for each participant. All correlations were Fisher-transformed and weighted such that the mean correlation was equal for the three category localizer conditions. These correlations were then used to estimate how much information the scene-induced activation pattern carried for the object categories that were present in the scene (Fig. 3B).

Results

Task and behavioral performance

Twenty-four participants performed a category detection task during which they detected exemplars of a target category embedded in natural scene photographs (Fig. 2B). In order to dissociate effects of behavioral relevance from effects of object category, the assignment of the object categories (cars, people and trees) to the levels of behavioral relevance (target, neutral, distracter) varied across three different participant groups (*Materials and Methods*; Fig. 2A). Averaged across participants, accuracy (percent correct) was high and similar for the three object categories (people: 85.3%; cars: 81.4%; trees: 85.2%). Significant differences in accuracy were observed between cars and people [group 1; $t_{(7)} = -5.08$; $p < 0.005$], and between cars and trees [group 2; $t_{(7)} = -3.25$; $p < 0.05$], but not between trees and people [group 3; $t_{(7)} = -0.39$; $p = 0.71$]. It is important to note that, due to the counterbalanced assignment of object category to behavioral relevance, all three categories occurred equally often as the target, neutral, and distracter category. Differences in accuracy could thus not systematically bias effects related to task relevance. No differences in response times across target object categories were observed ($p > 0.05$, for all groups).

Category information in single-category scenes

The degree of categorical processing of object categories in natural scenes was estimated by correlating multivoxel response patterns evoked by scenes in the category detection task with response patterns evoked by category exemplars that were presented in independent localizer scans (*Materials and Methods*; Fig. 3B). For scenes containing a single category (people, cars, or trees; Fig. 4), information values were calculated by subtracting the average correlation between the scene response pattern and the two non-present category response patterns from the correlation between the scene pattern and the pattern for the category that was embedded in the scene (Fig. 3B). Mean information values were entered into a one-way repeated-measures analysis of variance (ANOVA) with the factor 'task relevance' (target,

neutral, distracter). A significant main effect [$F(2,46) = 8.42; p < 0.001; \eta^2_p = 0.27$] indicated that information values differed as a function of the task relevance of the within-scene objects (Fig. 4). Scene-evoked response patterns carried significantly more information for target objects than for neutral objects ($p < 0.05$). Importantly, information for distracter objects was significantly lower than information for neutral objects ($p < 0.05$). Further, information values were significantly greater than zero for both the target category [$t_{(23)} = 4.35; p < 0.001; \text{Cohen's } d = 0.89$] and the neutral category [$t_{(23)} = 3.65; p < 0.005; \text{Cohen's } d = 0.74$], indicating that both target and neutral objects were processed up to the categorical level. In contrast, information for the distracter category did not differ significantly from zero [$t_{(23)} = -0.23; p = 0.82; \text{Cohen's } d = -0.04$].

To test whether effects of enhancement and suppression are also evident in the overall BOLD response amplitude in OSC we compared the estimated beta weights for the three single-category scenes. A one-way ANOVA of response amplitude with the factor 'scene type' (target-only, neutral-only, distracter-only) revealed a significant main effect [$F(2,46) = 10.28; p < 0.001; \eta^2_p = 0.31$]. Target-only scenes elicited the highest response, and planned pairwise comparisons showed that responses elicited by target - only scenes significantly exceeded those elicited by neutral-only ($p < 0.001$) and distracter-only ($p < 0.001$) scenes. However, response amplitudes did not differ between neutral-only and distracter-only scenes ($p = 0.37$). Thus, MVPA investigating category information was a more sensitive measure than overall response amplitudes for revealing effects of distracter suppression in OSC.

Category information in two-category scenes

For scenes depicting two categories (Fig. 5), we calculated information values for each of the two categories that were embedded in the scene. We first computed the correlation between the scene-evoked pattern and the pattern evoked by the category of interest. From this, we subtracted the correlation between the scene-evoked pattern and the pattern evoked by the category that was not depicted in the scene.

For scenes containing the target and the neutral categories, information for each of the two categories was calculated by subtracting the scene-distracter correlation from the scene-target and scene-neutral correlations. Category information values were then entered into a one-way repeated measures ANOVA with the factor 'task relevance' (target, neutral). The ANOVA revealed a significant main effect of 'task relevance' [$F(1,23) = 12.4; p < 0.005; \eta^2_p = 0.35$] indicating that target information was enhanced relative to information for the neutral category (Fig. 5A). Replicating the effects obtained for single category scenes, activation patterns in OSC evoked by scenes displaying the target and the neutral category carried significant information for both categories [target: $t_{(23)} = 5.68; p < 0.001; \text{Cohen's } d = 1.16$; neutral: $t_{(23)} = 2.78; p < 0.05; \text{Cohen's } d = 0.57$].

For scenes containing the target and the distracter categories, information values were obtained by subtracting scene-neutral correlations from the scene-target and scene-distracter correlations. The one-way repeated measures ANOVA on category information with the factor 'task relevance' (target, distracter) revealed a significant main effect [$F(1,23) = 16.06; p < 0.001, \eta^2_p = 0.41$], reflecting higher target information than distracter information (Fig. 5B), again confirming the results for the single-category scenes. Moreover, while the target category was processed to the category level [$t_{(23)} = 5.68; p < 0.001; \text{Cohen's } d = 0.86$], information for the distracter category did not differ significantly from zero [$t_{(23)} = -0.07; p = 0.95; \text{Cohen's } d = -0.01$].

For scenes containing the neutral and distracter categories information values were obtained by subtracting correlations between the scene and target response patterns from scene-neutral and scene-distracter correlations. No significant difference was found between

information for the two categories [$F(1,23) = 0.51$; $p = 0.48$; $\eta^2_p = 0.02$] and the mean information value did not differ significantly from zero [test of the intercept: $F(1,23) = 1.25$; $p = 0.28$; $\eta^2_p = 0.05$]

Category information in three-category scenes

Scenes containing all three categories (Fig. 6) particularly lend themselves to investigating the presence of enhancement and suppression processes, as they allow for the direct comparison of target and distracter information against the neutral baseline. The scene-evoked response was correlated with the three object category patterns. Subsequently, the correlation involving the neutral category was subtracted from the correlations involving the target and distracter categories. Consequently, positive information values are indicative of target enhancement while negative information values are indicative of distracter suppression relative to the neutral category baseline. A one-way repeated measures ANOVA revealed that information for the target category was significantly larger than information for the distracter category [$F(1,23) = 20.56$; $p < 0.001$, $\eta^2_p = 0.47$]. Moreover, while information for the target category was significantly above the neutral baseline [$t_{(23)} = 2.79$; $p < 0.05$; *Cohen's d* = 0.57], information for the distracter category was significantly below the neutral baseline [$t_{(23)} = -4.15$; $p < 0.001$; *Cohen's d* = -0.85] (Fig. 6).

Category information in no category scenes and early visual cortex

Finally, we determined information values for each of the three categories (target, distracter, neutral) in response patterns evoked by scenes in which none of the three categories were actually depicted. We first computed the correlation between the scene-evoked pattern and the pattern evoked by the category of interest and then subtracted the average of the correlations between the scene-evoked pattern and the patterns evoked by the other two categories. There was a marginally significant main effect of task relevance [$F(2,46) = 3.25$; $p = 0.048$; $\eta^2_p = 0.12$], that is accounted for by lower distracter compared to neutral information ($p < 0.05$). Critically, however, there was no difference in category information for the target category compared to either the distracter or the neutral category ($p > 0.1$).

In contrast to OSC, information values in early visual cortex (BA17, BA18; *Materials and Methods*) did not differ as a function of task relevance for any of the scene types ($p > 0.05$ for all scene types, for both ROIs), and the mean information value did not significantly differ from zero ($p > 0.05$ for all scene types, for both ROIs).

Behavioral interference

The behavioral data did not reveal evidence for interference from the distracter category on task performance. The presence of a distracter object in a scene did not lead to performance decrements compared to the presence of a neutral object in either RT [N vs. D: $t_{(23)} = 0.11$; $p = 0.91$; T-N vs. T-D: $t_{(23)} = 1.15$; $p = 0.26$] or in accuracy [N vs. D: $t_{(23)} = -0.19$; $p = 0.85$; T-N vs. T-D: $t_{(23)} = 0.22$; $p = 0.83$]. Thus, it appears that participants were successful in counteracting interference from distracters, possibly as a consequence of the reduced processing of distracter items in OSC. To provide evidence for this interpretation, we tested for a possible relationship between neural suppression and behavioral measures of interference in an across-subjects correlation analysis. Interestingly, a significant negative correlation was obtained between behavioral interference, measured as the RT difference between distracter-only and neutral-only scenes, and neural suppression, measured as the category information difference between neutral and distracter categories in neutral-only and distracter-only scenes ($r = -0.45$; $p < 0.05$). In other words, participants who showed less neural distracter suppression showed more behavioral interference. However, no significant correlations were found when using accuracy as a measure of interference or when comparing behavioral interference with neural suppression in two-category scenes. Thus,

future work will be necessary to firmly establish a link between neural measures of suppression and behavioral interference effects.

Discussion

The current study revealed that detecting object categories in cluttered natural scenes entails not only the enhancement of the attentional set that matches the currently task-relevant category, but also the suppression of a previous attentional set that has become obsolete after a change in the search target had occurred. Specifically, we found that response patterns in OSC evoked by scene images depicting objects with varying degrees of behavioral relevance contained more information for a target category and less information for a distracter category, relative to a neutral baseline category. This pattern of results was robust across different scene types that varied with regard to the number and combination of object categories they depicted. Critically, the differences between the processing of target, neutral and distracter information cannot be explained by incidental differences in the processing of the three object categories that we used (i.e. trees, cars, and people), as the assignment of object category to level of behavioral relevance was counterbalanced across participants (Fig. 2A). Thus, the current study provides evidence that real-world visual search involves mechanisms for both target-related enhancement and distracter-related suppression.

Both facilitative and suppressive processes have been implicated in theories of top-down attentional control. The notion of target enhancement has found support from the observation that attended stimuli evoke stronger neuronal responses even if there are no surrounding distracter objects competing for limited processing resources (Carrasco et al., 2000). Evidence for the idea that the suppression of distracter-induced interference plays an integral role in attentional selection derives from studies demonstrating that the benefits of directing attention to a stimulus are larger in cases where there is competition from nearby distracter objects (Awh and Pashler, 2000; Kastner and Ungerleider, 2001). In the past, the lack of an appropriate baseline against which to compare responses to target and distracters has made it difficult to disentangle the relative contributions of enhancement and suppression in top-down attentional modulation. In particular, previous studies have typically compared responses evoked by attended stimuli to responses evoked by ignored stimuli (O'Craven et al., 1999; Peelen et al., 2009). The evidence provided by these studies leaves open whether the observed differences in neural responses evoked by attended and ignored conditions are caused by enhanced processing of relevant information, reduced processing of distracting information, or a combination of both processes. In the realm of working memory, Gazzaley and colleagues (2005) found evidence for enhanced and suppressed processing of to-be-remembered and to-be-ignored stimuli, respectively, compared to the degree to which the same stimuli were processed under passive viewing conditions. However, a passive viewing baseline is not optimal, as the degree to which participants attend to the displayed stimuli is uncontrolled. The failure to truly passively view stimuli in the baseline condition may lead to an overestimation of suppressive effects and a concurrent underestimation of facilitative processes. In the present study, the baseline condition was matched with the distracter condition in terms of task demands, behavioral response, and perceptual input. The sole difference between these conditions was their behavioral relevance, with the distracter, but not the neutral object category having been previously relevant. Thus, the current design allows for making a strong case for both facilitative and suppressive mechanisms in real-world visual search.

Top-down distracter-related suppression as a means to improve the processing of relevant information has been suggested by findings from a number of behavioral paradigms. If the distracters in a visual search array are presented in a time-staggered manner, the subset of distracters shown first does not influence the efficiency of target detection (Watson and

Humphreys, 1997). According to the visual marking hypothesis, this effect is due to the deprioritization of old distracters by a top-down inhibitory template that is selective for familiar objects in the display. On a similar note, visual search can be guided to novel objects in a search array (Mruczek and Sheinberg, 2005). This effect too has been speculated to rely on the development of search templates that guide search away from old distracters and concomitantly to novel objects in the display (Yang et al., 2009). Finally, the processing of objects that were previously ignored is impaired relative to the processing of new (neutral) objects, an effect termed ‘negative priming’ (Tipper, 1985). Negative priming is thought to reflect lingering effects from the suppression the target had received in its past role as a distracter item.

Which changes at the neuronal level can account for our results? Previous studies on feature-based attention suggest that the stronger representations of task-relevant objects result from the pre-activation of neurons that are selective for the target object (Desimone and Duncan, 1995). Once a target object enters the visual field, the visual representation is biased towards the task-relevant object through the enhanced firing rates of these target-selective neurons (Chelazzi et al., 1993). The recent finding of target-selective responses during preparation for real-world visual search (Peelen and Kastner, 2011) suggests that these mechanisms involved in feature-based attention also play a role in the selection of object categories from naturalistic stimuli. It is conceivable that the suppression of the discarded attentional set, observed here, can be explained by similar mechanisms. Rather than being pre-activated, neurons coding for the distracter category may be actively suppressed prior to the onset of the stimulus, resulting in decreased processing of the distracter category. In the present design it is likely that the suppressive attentional set is constant for the duration of an entire block, rather than being newly instantiated on each trial. Future work will be necessary to determine the upper and lower bounds of the time scale at which distracter suppression is effective.

Alternatively, suppression of the distracter category could occur at a later processing stage. That is, information for the distracter category may be present at earlier time points that escape the time window probed by fMRI and then be filtered out by a top-down biasing mechanism in subsequent processing stages. Future work, probing the presence of distracter information during search preparation may help shed light upon the nature of the top-down signals involved in distracter suppression.

Selective attention has also been shown to influence the representation of task-relevant locations and features by modulating the synchrony of neural oscillations both within local circuits and across distant neuronal populations (Womelsdorf and Fries, 2007). Specifically, in macaque monkeys, increases in spike-field coherence within the gamma-frequency band have been observed when the stimulus located in the receptive fields of V4 neurons shared features with a target object during visual search (Bichot et al., 2005). Conversely, a down-regulation of synchrony has been reported for neurons processing distracting information (Womelsdorf et al., 2006). Importantly, attentional modulation of neuronal synchronization appears to predict behavioral task performance at least partially (Womelsdorf et al., 2006). Thus, it will be an exciting venue for future research to explore the idea that task-dependent increases and decreases in neuronal synchronization among category-selective neurons in OSC may account for the effects of enhanced target and suppressed distracter processing observed in the current study.

An important question that remains to be resolved concerns the source region of the attentional set. A distributed network of higher-order areas including the Superior Parietal Lobule (SPL), the Frontal Eye Fields (FEF), Supplementary Eye Fields (SEF) and Intraparietal Sulcus (IPS) has consistently been implicated in the exertion of top-down

attentional control signals. Although the majority of results on the frontoparietal attention network stem from research on the control of spatial attention (Kastner et al., 1999; Hopfinger et al., 2000) the frontoparietal attention network has also been implicated in the deployment of feature-based attention (Giesbrecht et al., 2003; Liu et al., 2003; Egner et al., 2008). Top-down control signals from the frontoparietal attention network may be involved in both the enhancement of the active attentional set and the suppression of the outdated attentional set. Indication for a common source of enhancing and suppressive control signals has been provided in a study by Moore and Armstrong (2003), which showed that microstimulation of the FEF simultaneously increases firing rates at the attended location and reduces activity at distracter locations in V4. It is unclear, however, whether and how these results apply to non-spatial attention biases. Finally, a region in medial prefrontal cortex (mPFC) has been shown to differentially represent target and distracter categories during preparation for real-world visual search in a manner similar to OSC, suggesting that it plays a role in the implementation of attentional sets (Peelen and Kastner, 2011). Future studies, possibly using perturbation methods, will be needed to identify where and when the attentional sets that are enhanced and suppressed during real-world visual search are instantiated in the brain (Zanto et al., 2011; Feredoes et al., 2011).

To conclude, our results demonstrate that previously but no longer relevant attentional sets are suppressed, which may help to switch to new goals and serve to minimize interference from otherwise distracting information. More generally, our results show that the visual representation of complex real-world scenes is shaped not only by current task goals but also by recent past experience.

Acknowledgments

This work was supported by National Institutes of Health Grants R01-EY017699 and R01-MH064043 and National Science Foundation Grant BCS-1025149.

References

- Awh E, Pashler H. Evidence for split attentional foci. *J Exp Psychol Hum Percept Perform.* 2000; 26:834–846. [PubMed: 10811179]
- Bichot NP, Rossi AF, Desimone R. Parallel and serial neural mechanisms for visual search in macaque area V4. *Science.* 2005; 308:529–534. [PubMed: 15845848]
- Carrasco M, Penpeci-Talgar C, Eckstein M. Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision Res.* 2000; 40:1203–1215. [PubMed: 10788636]
- Chelazzi L, Duncan J, Miller EK, Desimone R. Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol.* 1998; 80:2918–2940. [PubMed: 9862896]
- Chelazzi L, Miller EK, Duncan J, Desimone R. A neural basis for visual search in inferior temporal cortex. *Nature.* 1993; 363:345–347. [PubMed: 8497317]
- Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res.* 1996; 29:162–173. [PubMed: 8812068]
- Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Annu Rev Neurosci.* 1995; 18:193–222. [PubMed: 7605061]
- Dombrowe I, Donk M, Olivers CN. The costs of switching attentional sets. *Atten Percept Psychophys.* 2011; 73:2481–2488. [PubMed: 21879418]
- Duncan J, Humphreys GW. Visual search and stimulus similarity. *Psychol Rev.* 1989; 96:433–458. [PubMed: 2756067]
- Egner T, Monti JM, Trittschuh EH, Wieneke CA, Hirsch J, Mesulam MM. Neural integration of top-down spatial and feature-based information in visual search. *J Neurosci.* 2008; 28:6141–6151. [PubMed: 18550756]

- Evans KK, Horowitz TS, Wolfe JM. When categories collide: accumulation of information about multiple categories in rapid scene perception. *Psychol Sci.* 2011; 22:739–746. [PubMed: 2155522]
- Feredoes E, Heinen K, Weiskopf N, Ruff C, Driver J. Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proc Natl Acad Sci U S A.* 2011; 108:17510–17515. [PubMed: 21987824]
- Gazzaley A, Cooney JW, McEvoy K, Knight RT, D’Esposito M. Top-down enhancement and suppression of the magnitude and speed of neural activity. *J Cogn Neurosci.* 2005; 17:507–517. [PubMed: 15814009]
- Giesbrecht B, Woldorff MG, Song AW, Mangun GR. Neural mechanisms of top-down control during spatial and feature attention. *Neuroimage.* 2003; 19:496–512. [PubMed: 12880783]
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science.* 2001; 293:2425–2430. [PubMed: 11577229]
- Hopfinger JB, Buonocore MH, Mangun GR. The neural mechanisms of top-down attentional control. *Nat Neurosci.* 2000; 3:284–291. [PubMed: 10700262]
- Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron.* 1999; 22:751–761. [PubMed: 10230795]
- Kastner S, Ungerleider LG. The neural basis of biased competition in human visual cortex. *Neuropsychologia.* 2001; 39:1263–1276. [PubMed: 11566310]
- Liu T, Slotnick SD, Serences JT, Yantis S. Cortical mechanisms of feature-based attentional control. *Cereb Cortex.* 2003; 13:1334–1343. [PubMed: 14615298]
- Moore T, Armstrong KM. Selective gating of visual signals by microstimulation of frontal cortex. *Nature.* 2003; 421:370–373. [PubMed: 12540901]
- Mruczek RE, Sheinberg DL. Distractor familiarity leads to more efficient visual search for complex stimuli. *Percept Psychophys.* 2005; 67:1016–1031. [PubMed: 16396010]
- O’Craven KM, Downing PE, Kanwisher N. fMRI evidence for objects as the units of attentional selection. *Nature.* 1999; 401:584–587. [PubMed: 10524624]
- Peelen MV, Fei-Fei L, Kastner S. Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature.* 2009; 460:94–97. [PubMed: 19506558]
- Peelen MV, Kastner S. A neural basis for real-world visual search in human occipitotemporal cortex. *Proc Natl Acad Sci U S A.* 2011; 108:12125–12130. [PubMed: 21730192]
- Russell BC, Torralba A, Murphy KP, Freeman WT. LabelMe: a database and web-based tool for image annotation. *International journal of computer vision.* 2008; 77:157–173.
- Shiffrin RM, Schneider W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol Rev.* 1977; 84:127.
- Tipper SP. The negative priming effect: inhibitory priming by ignored objects. *Q J Exp Psychol A.* 1985; 37:571–590. [PubMed: 4081101]
- Walther DB, Caddigan E, Fei-Fei L, Beck DM. Natural scene categories revealed in distributed patterns of activity in the human brain. *J Neurosci.* 2009; 29:10573–10581. [PubMed: 19710310]
- Watson DG, Humphreys GW. Visual marking: prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychol Rev.* 1997; 104:90–122. [PubMed: 9009881]
- Womelsdorf T, Fries P. The role of neuronal synchronization in selective attention. *Curr Opin Neurobiol.* 2007; 17:154–160. [PubMed: 17306527]
- Womelsdorf T, Fries P, Mitra PP, Desimone R. Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature.* 2006; 439:733–736. [PubMed: 16372022]
- Yang H, Chen X, Zelinsky GJ. A new look at novelty effects: guiding search away from old distractors. *Atten Percept Psychophys.* 2009; 71:554–564. [PubMed: 19304646]
- Zanto TP, Rubens MT, Thangavel A, Gazzaley A. Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nat Neurosci.* 2011; 14:656–661. [PubMed: 21441920]

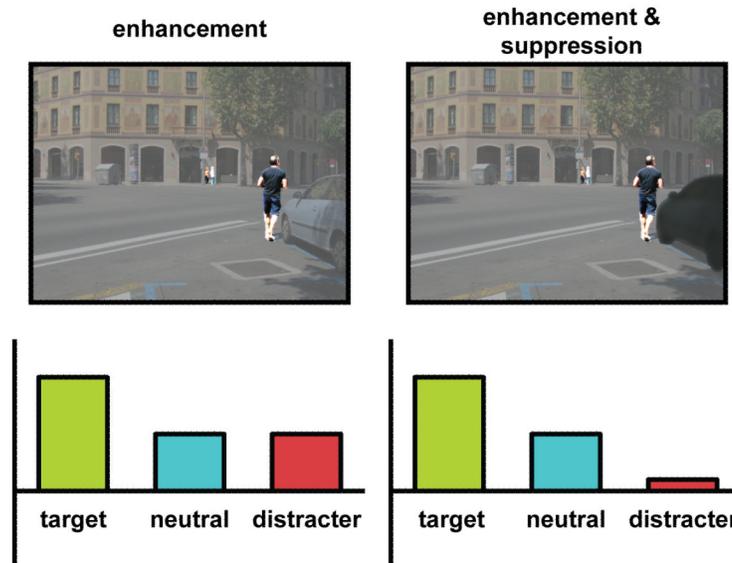


Figure 1. Hypotheses

Different neural processes and their interactions may mediate the selection of behaviorally relevant information from natural scenes. In the example, people are the object category that is relevant to ongoing behavior (i.e. ‘target’ category) and cars are the object category that was previously, but is not presently relevant (i.e. ‘distracter’ category), while all other object categories present in the scene are never task-relevant (i.e. ‘neutral’ category, e.g. trees). If visual search in natural scenes is accomplished solely through the enhancement of relevant information, processing of the target category should be enhanced relative to both neutral and distracter categories (left panel). In contrast, if category detection additionally requires the suppression of a previous attentional set, distracter processing should be reduced compared to processing of the neutral categories (right panel).

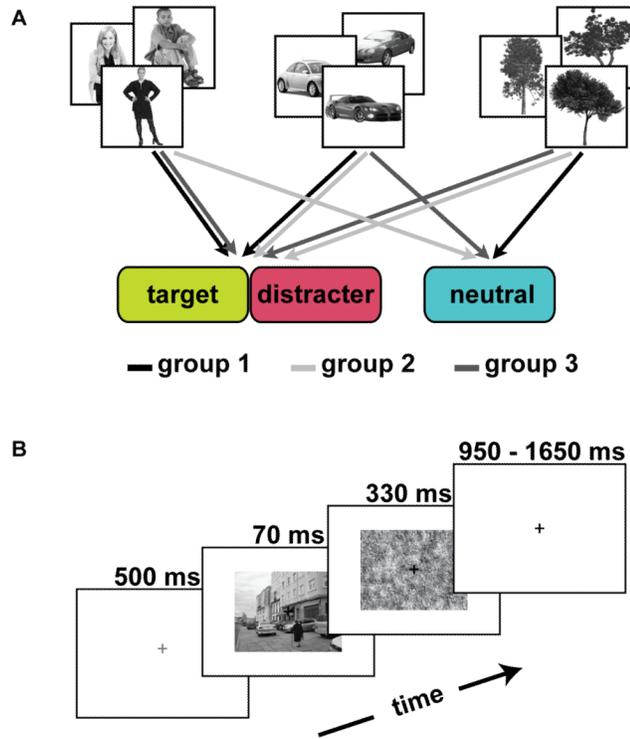


Figure 2. Experimental design

(a) Counterbalancing of object category to task relevance across participants. The assignment of the three object categories to their levels of behavioral relevance was counterbalanced across participants, creating three different participant groups. For group 1 (black), people and cars were the target and distracter categories and trees were the neutral category. For group 2 (light grey) cars and trees were assigned to be the target and distracter categories while people constituted the neutral category. For group 3 (dark grey), people and trees were the target and distracter categories while cars constituted the neutral category. (b) *Trial structure*. Participants indicated on each trial with a button press whether or not the target category was present in the photograph.

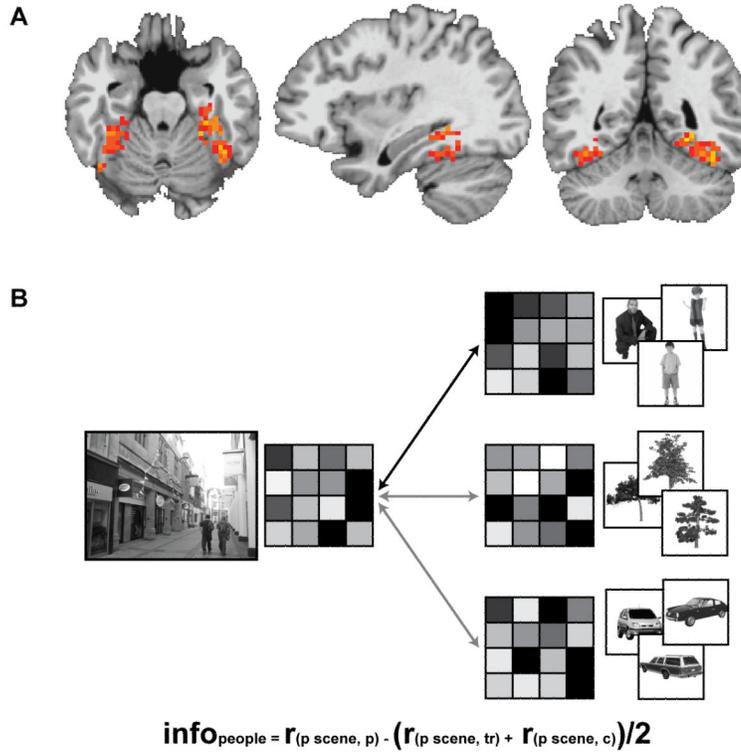


Figure 3. Analytical approach

(a) Definition of a region-of-interest in object-selective cortex (OSC). OSC was defined separately in each individual participant by contrasting responses to intact and scrambled objects that were presented in an independent localizer experiment. The figure depicts the ventral cluster of the OSC ROI in a group-level analysis (at $p < 0.001$). (b) *Schematic overview of analysis approach.* The general approach was to correlate multivoxel patterns evoked by scene images in the category detection task with response patterns from the independent category localizer. The example illustrates how information was calculated for the category people in a response pattern evoked by a scene containing people. The scene response pattern was first correlated with the people response pattern from the category localizer (black arrow; $r_{(\text{p scene}, \text{p})}$, p=people). Next, the correlations between the scene-elicited activation pattern and the tree and car response patterns were computed (grey arrows; $r_{(\text{p scene}, \text{tr})}$, $r_{(\text{p scene}, \text{c})}$; tr=tree category, c=car category). Categorical information related to people was then defined as the difference between the scene-people correlation and the average of the scene-car and scene-tree correlations.

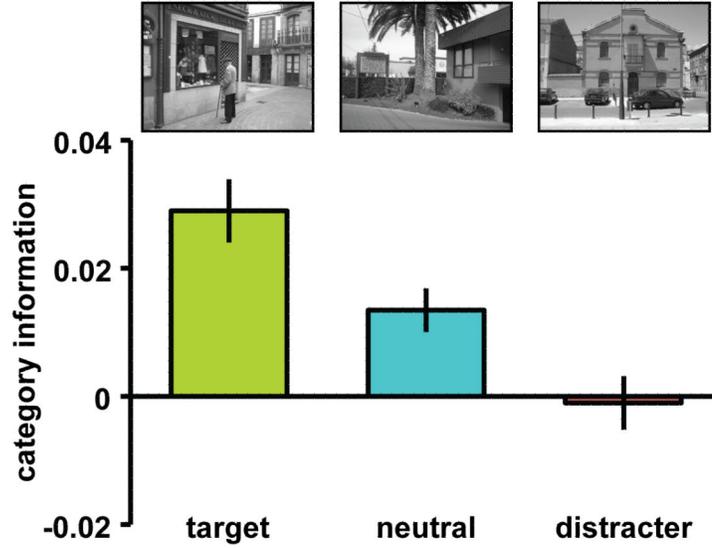


Figure 4. Category information in single-category scenes

Examples for single-category scenes are shown for a run in which people served as the target category while trees constituted the neutral category and cars the distracter category. Category information was calculated by first computing the correlation between the scene response and the response elicited by the category present in the scene. Then the average of the correlations between the scene response and the responses evoked by the two non-present categories was subtracted. Response patterns in OSC carried significant information for the target and neutral categories but not for the distracter category. Error bars denote \pm standard errors of the mean (s.e.m.) corrected for within-subject comparisons.

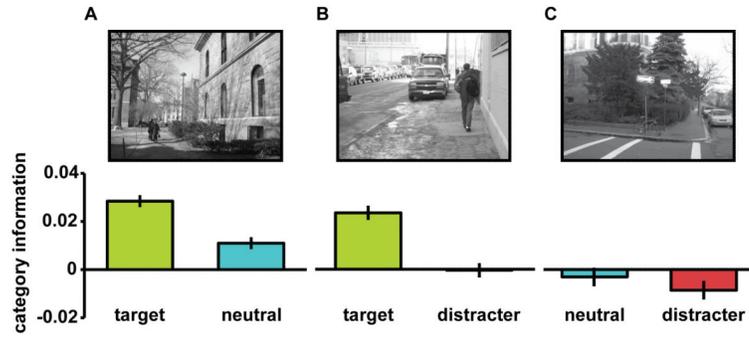


Figure 5. Category information in scenes depicting two categories

Examples for two-category scenes are shown for a run in which people served as the target category while trees constituted the neutral category and cars the distracter category. Category information was calculated separately for the two categories embedded in the scene. The correlation between the scene and the category not present in the scene was subtracted from the correlation between the scene and one of the two categories present in the scene. (a) In T-N scenes both the target and the neutral category were processed to the category level. (b) In T-D scenes the target but not the distracter category was processed to the category level. (c) Information values did not differ for the neutral and distracter category in N-D scenes. Error bars denote \pm s.e.m. corrected for within-subject comparisons.

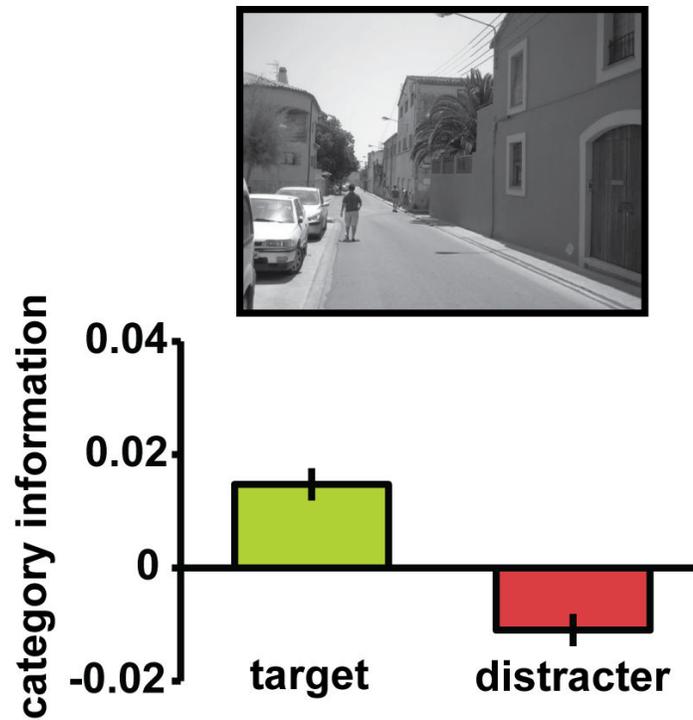


Figure 6. Category information in scenes depicting all three object categories

The correlation between the scene-elicited response pattern and the response pattern evoked by the neutral category was subtracted from the correlations involving the target and distracter categories. Target information was enhanced relative to the neutral category while distracter information was reduced. Error bars denote \pm s.e.m. corrected for within-subject comparisons.