Integrated Contextual Representation for Objects' Identities and Their Locations

Nurit Gronau, Maital Neta, and Moshe Bar

Abstract

■ Visual context plays a prominent role in everyday perception. Contextual information can facilitate recognition of objects within scenes by providing predictions about objects that are most likely to appear in a specific setting, along with the locations that are most likely to contain objects in the scene. Is such identity-related (''semantic'') and location-related (''spatial'') contextual knowledge represented separately or jointly as a bound representation? We conducted a functional magnetic resonance imaging (fMRI) priming experiment whereby semantic and spatial contextual relations between prime and target object pictures were independently manipulated. This method allowed us to determine whether the two contextual factors affect object recognition with or without interacting, supporting a unified versus independent representations, respectively. Results revealed a Semantic × Spatial interaction in reaction times for target object recognition. Namely, significant semantic priming was obtained when targets were positioned in expected (congruent), but not in unexpected (incongruent), locations. fMRI results showed corresponding interactive effects in brain regions associated with semantic processing (inferior prefrontal cortex), visual contextual processing (parahippocampal cortex), and object-related processing (lateral occipital complex). In addition, activation in fronto-parietal areas suggests that attention and memory-related processes might also contribute to the contextual effects observed. These findings indicate that object recognition benefits from associative representations that integrate information about objects' identities and their locations, and directly modulate activation in object-processing cortical regions. Such *context frames* are useful in maintaining a coherent and meaningful representation of the visual world, and in providing a platform from which predictions can be generated to facilitate perception and action. ■

INTRODUCTION

Objects are typically seen embedded in specific contextual settings, such as a desk in an office or a car on a street. These environmental regularities can be useful in facilitating object identification by increasing predictability in the sensory input, and thus, streamlining the process of recognition. Indeed, studies using behavioral (Davenport & Potter, 2004; Chun & Jiang, 1998; Bar & Ullman, 1996; Biederman, Mezzanotte, & Rabinowitz, 1982; Friedman, 1979; Mandler & Johnson, 1976; Palmer, 1975; Biederman, 1972), as well as functional magnetic resonance imaging (fMRI) (Cox, Meyers, & Sinha, 2004; Bar & Aminoff, 2003) approaches, have documented the importance of contextual information in the facilitation and enhancement of visual processing and perceptual memory (for a review, see Bar, 2004).

The influence of contextual processes on object recognition may stem from multiple sources of information, including knowledge about the expected identity, size, position, and relative depth of an object within a scene (e.g., Biederman et al., 1982; Mandler & Johnson, 1976). Among these factors, two important sources of contextual knowledge are the focus of the present study: information about which object *identities* are most likely to appear within a specific visual setting (e.g., an office typically contains a desk, a desk lamp, and a computer), and information about which *locations* within a visual setting are most likely to contain objects (e.g., phones are typically placed above, and not below desks). Although both identity-based and location-based associative knowledge have been shown to enhance object detection and recognition (Chun & Jiang, 1998, 1999; Sanocki & Epstein, 1997; Bar & Ullman, 1996; Biederman et al., 1982), the exact nature of the relationship between these contextual factors remains largely unclear. Do identity- and location-related associations have distinct effects on visual perception, suggesting independent underlying representations for the two sources of knowledge? Or do these contextual factors interact in the course of object recognition, supporting a joint representation of identities and locations in visual associative processing?

Recently, we have proposed that contextual knowledge is represented within memory structures, or *context frames*, that contain knowledge about specific objects that are typical for a given visual setting, as well as about the

Martinos Center for Biomedical Imaging at MGH, Harvard Medical School

probable spatial relations between these objects (Bar, 2004; Bar & Ullman, 1996). The notion of context frames is reminiscent of earlier concepts such as *schemata* (Biederman et al., 1982; Hock, Romanski, Galie, & Williams, 1978; Mandler & Johnson, 1976), *scripts* (Schank, 1975), and *frames* (Minsky, 1975), which all imply a unified, global representation for identity- and location-based associative information. Such a joint representation is appealing as a theoretical construct because it suggests that objects are organized in memory within structures that depict typical scenes. However, not all empirical findings are necessarily in agreement with this concept.

For instance, studies examining associative processing during visual object recognition have shown that a prime object (desk) can facilitate recognition of a successive, semantically related object identity (lamp), even when both pictures appear at the same physical location (e.g., at the screen center) where there is often a violation of the real-world spatial relations between the two items (i.e., a lamp typically appearing *above* a desk) (McPherson & Holcomb, 1999; Carr, McCauley, Sperber, & Parmelee, 1982; Sperber, McCauley, Ragain, & Weil, 1979). Similar priming effects are obtained for target picture objects when using semantically related words as primes (e.g., the word ''DESK''), implying that the meaning of both verbal and pictorial stimuli is encoded in a common, amodal representation (e.g., Carr et al., 1982; Sperber et al., 1979). Such an amodal representation for visual objects is presumably abstracted of specific sensory details and of metric coordinates, suggesting that associative links between object identities are not necessarily encoded in a point-to-point scene-like representation (Potter & Faulconer, 1975; Pylyshyn, 1973; Sperber et al., 1973). Furthermore, studies directly investigating processing of objects within scenes have shown that immediate memory for object identities and their visual details is unaffected by the spatial relations between the objects within the scene, that is, whether items are organized in a coherent (scene-like) or incoherent (jumbled) spatial organization (e.g., Mandler & Ritchey, 1977; Mandler & Johnson, 1976). In accordance with these findings, the representation of object identity appears to be largely independent of spatial layout and object location in on-line visual scene processing, as well as in visual short-term memory (e.g., Rensink, 2000; Simons, 1996). Taken together, these findings suggest that contextual associations between visual objects can be independent of spatial location representations. Furthermore, visual contextual knowledge might be conceptual or abstract in nature, rather than encoded in a strictly picture-like representation.

Additional evidence that might imply the independence of identity- and location-related contextual knowledge comes from studies showing that the spatial properties of a scene or an object can be processed independently of knowledge of an associated object identity. For instance, rapid extraction of the spatial layout of a scene can constrain the probable location(s) of a target object within the scene, regardless of the object's identity (Chun & Jiang, 1998; Sanocki & Epstein, 1997). This rapid extraction of spatial properties may rely on basic visual features of the scene, such as global structure, perceived depth, and surface configuration (e.g., Chun & Jiang, 1998; Sanocki & Epstein, 1997; Schyns & Oliva, 1994). Alternatively, location-related associative knowledge may rely on semantic comprehension of a visual stimulus, such as when a static picture of an object in action (e.g., a tool pointing down) produces a sense of motion (Kourtzi & Kanwisher, 2000; Reed & Vinson, 1996; Freyd, 1983) and guides visual attention to the direction of the implied action (Gronau & Bar, unpublished data; Riddoch, Humphreys, Edwards, Baker, & Willson, 2003). Likewise, a picture of a table can guide attention to an implied, upper location, based on one's knowledge that tables generally serve as surfaces *above* which objects are placed. Note that in the last two examples, knowledge of the coarse semantic and spatial properties of a context does not necessarily dictate which particular object identities are associated with it. For example, although all desks and tables are used for placing objects on, location-based contextual knowledge does not differentiate between specific objects that are more likely to appear on an office desk (e.g., desk lamp, computer), on a picnic table (e.g., picnic basket, paper plates), or on a coffee table (e.g., vase, decorative bowl). Thus, spatial guidance based on an object's direction of action, or on its basic function, does not necessarily provide the detailed knowledge required to activate an associated object identity. Location-based knowledge, then, may generate expectations about a position of an upcoming object, in the absence of specific information regarding the specific object's identity.

Although the evidence above might suggest that contextual knowledge of object identities and locations can be represented separately, there is, nevertheless, a firm basis to postulate that the two types of information are integrated at some level of representation. In everyday life, objects' locations often correlate with their identities, such that certain objects are more expected to be positioned at certain locations in a scene (e.g., sofas and carpets are more expected to be seen at floor level of a room, whereas computer screens and curtains are more expected to be seen at hand or eye level of the room). Given the potential usefulness of such environmental regularities to visual recognition and to action planning, it is reasonable to expect that the underlying contextual representations will reflect these regularities, and that the brain will exploit them. Consequently, knowledge of an object's identity should constrain the possible locations in which the object is to be searched, and vice versa. Thus, although one might maintain independently some form of an abstract, conceptual association between objects (e.g., a desk lamp is related to a desk), along with separate knowledge about the spatial

relations between objects in the world, we propose that a unified *context frame* encompasses both these two subtypes of contextual associations (e.g., a desk lamp is necessarily placed on a desk, by virtue of its function). Such a global contextual representation can provide a highly coherent and meaningful representation of the visual environment because it depicts with greater accuracy the relations between objects within a certain context.

Behavioral studies investigating the effects of visual contextual information on object recognition have not allowed a systematic investigation of the relations between identity-based and location-based associative factors because these factors were typically examined in isolation (Davenport & Potter, 2004; Bar & Ullman, 1996; De Graef, Christiaens, & d'Ydewalle, 1990; Biederman et al., 1982; Palmer, 1975; Biederman, 1972). For instance, the effects of the relatedness between an object identity and a scene were examined only within expected, but not unexpected, locations in the scene (Davenport & Potter, 2004; De Graef et al., 1990; Biederman et al., 1982). Similarly, effects of spatial consistency between objects were examined only among conceptually related, and not unrelated, identities (e.g., Hollingworth, 2006; Bar & Ullman, 1996; De Graef et al., 1990; Biederman et al., 1982; Biederman, 1972). In addition, visual search studies using real-world objects and scenes have typically manipulated targets' location within conceptually related scenes only (Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006). Thus, in the absence of a simultaneous manipulation of both identity and location contextual factors, the dependency between the two types of associative knowledge could not be explicitly assessed. Similar limitations underlie fMRI studies (Aminoff, Gronau, & Bar, in press; Cox et al., 2004; Goh et al., 2004; Bar & Aminoff, 2003) and event-related potential studies (Ganis & Kutas, 2003; McPherson & Holcomb, 1999) investigating the neural correlates of contextual effects on visual recognition.

The goal of the present study, therefore, was to determine whether identity- and location-related associative knowledge is represented separately or jointly, examining both the behavioral and the neural levels. To this end, we used an event-related fMRI priming task in which object primes served as contextual cues for recognition of target objects. Most importantly, identity and location relations between prime and target were independently manipulated, allowing a direct examination of the dependency of the two associative factors in visual object recognition. If the two factors are represented independently, we predicted that additive effects would be found in behavioral measures (e.g., reaction time [RT]) as well as in corresponding fMRI measures (blood oxygenation level dependent [BOLD]) for target object recognition. If, however, identity- and location-related associative knowledge is combined at some level of processing to form a unified contextual representation, we should obtain *interactive* effects of the two factors (possibly in addition to each of the factors' separate effects). Thus, identity-based contextual effects should be larger when targets appear in spatially congruent (i.e., plausible, or expected) locations than in incongruent (i.e., implausible, or unexpected) locations, suggesting that an object's position in space is tightly related to its perceived identity within a particular visual setting.

Note that although both location- and identity-associative factors may rely on semantic analysis of a contextual prime, they nevertheless have different associated outputs under an assumption of their independence: Location-based associations guide attention to a specific spatial location (regardless of the target's expected identity), whereas identity-based associations prime a target's identity (regardless of its specific location in space). We use here the term ''semantic knowledge'' to denote the representation of object relations that are abstracted of specific visual details and spatial coordinates. This term is adopted from the semantic priming literature, in which both verbal and pictorial priming effects are typically accounted for by conceptual (e.g., categorical) relatedness. Likewise, we use the term ''spatial knowledge'' to denote information about an expected location of an upcoming target object (independent of its specific identity). In effect, we considered an object to be spatially congruent with its contextual environment if it appeared in a plausible, rather than an implausible, location, as predicted from the context (see Methods). Importantly, the use of ''semantic'' and ''spatial'' associative terms is somewhat arbitrary and should be considered merely as working definitions for identity- and location-based contextual factors, under the null hypothesis of independent representations. Both definitions are based on statistical co-occurrences of specific object properties in the natural environment.

As for the neural manifestation of visual contextual representation, we were particularly interested in the interactive effects of identity- and location-associative factors within three levels of cortical representation. First, we aimed to investigate whether brain regions typically associated with semantic processing would be modulated by spatial contextual information. Therefore, we examined activity in the inferior prefrontal cortex (IPC), a region previously shown to be associated with a wide variety of semantic and conceptual tasks (both verbal and nonverbal), and specifically associated with the phenomenon of semantic priming (for a recent review, see Van Petten & Luka, 2006). To the extent that semantic (or conceptual) priming effects are influenced by spatial factors, the IPC should be sensitive to global visual properties of a contextual setting, exhibiting increased differences between semantically related and unrelated targets when these appear in an expected (congruent) location. Second, we examined activation in medial-temporal lobe regions (hippocampus, parahippocampal cortex [PHC])

that have been strongly implicated in contextual associative processing (see Bar, 2004; Eichenbaum, 2004, for reviews). The PHC, in particular, was found to play a key role in a visual cortical network for analyzing contextual associations (Aminoff et al., 2007; Bar, 2004; Bar & Aminoff, 2003), and thus, it is a natural candidate for exploring the interactive effects of identity- and location-based contextual factors on object processing. Third, we investigated the effects of associative knowledge on perceptual regions directly involved in visual object perception. Specifically, we examined brain activity in object-selective regions within the ventrolateral occipital–temporal lobes, typically known as the lateral occipital complex (LOC; see, e.g., Grill-Spector, Kourtzi, & Kanwisher, 2001; Kanwisher, Chun, McDermott, & Ledden, 1996; Malach et al., 1995). We sought to explore whether evidence for an integrated contextual representation would be manifested already in the LOC, indicating that top–down contextual influences may directly affect visual object processing.

Finally, it should be noted that by proposing a unified *context frame* for identity- and location-related associative information, we are not suggesting that such a representation necessarily resides within a specific brain region. Rather, we employ a more global approach according to which rich information about object identities and their spatial relations can reside within and/or exert influence on a network of brain regions. Such a network potentially encompasses areas subserving semantic processing, visual–contextual analysis, and possibly other high order functions (Bar, 2007). The concept of a context frame, then, should be viewed as a general framework for investigation of high-level associative processing, rather than a construct represented within a highly localized cortical region.

METHODS

Participants

Twenty volunteers (10 women; mean age = 25 years, range = 23–28 years) participated in the experiment. Nineteen of the subjects were right-handed. All participants gave written, informed consent before participation in the study (all procedures were approved by Massachusetts General Hospital Human Studies Protocol number 2001P-001754).

Materials and Task Procedures

Subjects underwent fMRI scanning while performing a fast event-related priming task. Prime stimuli consisted of real-world objects and target stimuli consisted of either real objects (50%) or nonsense objects (50%). Subjects' task was to classify rapidly whether the target in each trial was a real or a nonsense object by using a two-alternative forced-choice keypress. The real/nonsense object task was adopted from the lexical decision task in the verbal semantic priming domain, in which subjects typically judge whether a target letter string is a real or a nonsense word. Real-object stimuli (including both primes and targets) consisted of everyday objects, such as furniture, tools, and clothing. Nonsense-object targets consisted of artificial, colorful shapes that were manually designed with Adobe Photoshop. Primes always appeared at the center of the screen, whereas targets either appeared at an upper (50%) or lower (50%) screen location, regardless of their identity. Critically, primes served as spatial cues, as they implied that a target would appear, under real-world conditions, in one of the two task-relevant locations (e.g., a picture of a dresser implied that the target would appear in an upper location because an object would typically be placed above, and not below the dresser; whereas a picture of a drill pointing down suggested that the target would appear in a lower location because the drill was directed toward an object potentially located below it; see Figure 1A). Accordingly, a target was defined as spatially congruent if it appeared in a *plausible* (expected) location, and it was defined as spatially incongruent if it appeared in an *implausible* (unexpected) location, with respect to the prime. In effect, half of the targets appeared in a spatially congruent location and half appeared in a spatially incongruent location, with an equal proportion of congruent and incongruent targets in upper and lower screen positions. In addition to manipulation of target location, within the real-object target trials, half of prime and target pairs were semantically related (e.g., dresser and mirror), and half were semantically unrelated (dresser and pot), forming an orthogonal two-by-two factorial design with four equally probable conditions (see Figure 1B).

Notice that the plausibility, or congruency, of the target's position relative to the prime was independent of whether the prime and the target were semantically related or unrelated. In the example shown in Figure 1B, for instance, a prime dresser would typically rest on the floor (as is the case with tables, beds, sofas, etc), and thus, any target object would potentially be placed above, and not below, the dresser. Prime and target object pictures appeared on a 8.3 × 8.3-in. gray background square, corresponding to a visual angle of approximately 14.6° × 14.6°. The gray square appeared at the center of a black screen and was divided into three imaginative horizontal sections in which a prime or a target object could be presented (the former occupying the central section, whereas the latter occupying either the upper or the lower sections; prime and target stimuli were centered within each horizontal plane). Object pictures spanned approximately 4.9° × 4.9°, with a 4.9° center-to-center distance between prime and target stimuli.

There were, altogether, 144 real-object target pairs (36 in each condition) and 144 nonsense-object target

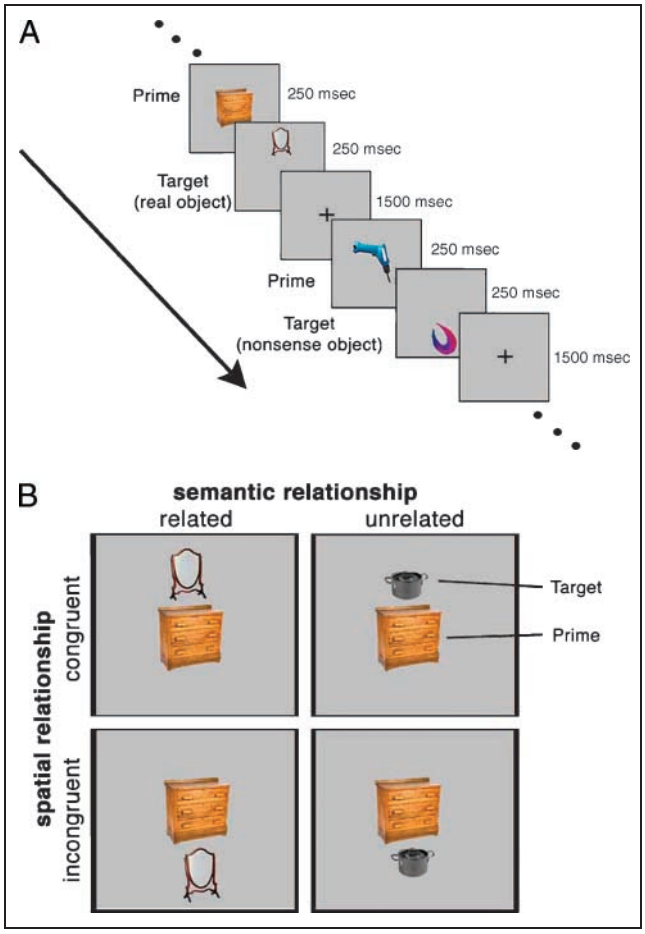

Figure 1. (A) Examples of stimuli in the priming paradigm. Each trial lasted 2 sec and consisted of a prime, a target (each presented for 250 msec), and a fixation cross, presented until onset of the next trial. Subjects' task was to classify whether the target was a real or a nonsense object (the prime was always a real object). Primes appeared at the center of the screen, whereas targets appeared either at an upper or a lower screen location, and were either spatially congruent or incongruent with respect to the prime. (B) The two-by-two factorial design within the real-object target trials. Each subject was exposed to a particular pairing of prime and target in only one of the four conditions (counterbalanced across subjects).

pairs (for which a separate set of primes than that of the real-object target pairs was used). In addition, 156 fixation trials were randomly interleaved between the prime–target trials to allow optimal deconvolution of the BOLD signal (stimuli randomization was determined by the Optseq program within the FS-FAST software tools; see http://surfer.nmr.mgh.harvard.edu/optseq). Upon presentation of all 444 trials (288 real and nonsense target pairs, plus 156 fixation trials), stimuli were repeated in a new randomized order, such that each prime–target pair was presented twice in the course of the experiment. Importantly, each subject was exposed to a prime stimulus in only one of the four experimental conditions (counterbalanced across subjects), and the same prime–target pair seen at initial presentation was also seen upon repetition. The temporal difference between the first and second stimulus presentation was approximately 15 min, and it was equal across all four conditions.

Trials in which subjects made errors (5%) or exhibited extreme RTs (above or below three standard deviations, 3%) were excluded from further statistical analyses.

Prior to participation in the study, subjects performed a practice session of 52 trials that differed from the experimental trials.

Localizer Task for Defining Object-processing Brain Regions

In addition to the main priming task, a functional localizer task was administered to define object-selective regions within the LOC. Subjects performed a 1-back task in which they viewed alternating blocks of faces, houses, outdoor scenes, and everyday objects (excluding objects presented in the main priming task). Within each block, subjects were required to indicate via keypress whether a picture was presented twice successively during the block (10% of trials). There were 12 presentation blocks for the object picture category and 6 presentation blocks for each of the other picture categories. Pictures within a block were presented for 700 msec with a 300-msec blank interstimulus interval, for a total of 20 pictures presented at a duration of 20 sec. During rest periods between blocks (20 sec), subjects passively fixated on the screen.

fMRI Data Acquisition

Stimuli were presented with a Matlab software and were back-projected onto a translucent screen for viewing in the MRI scanner via a mirror attached to a custom-built head coil. All images were acquired in a 3-T Siemens–Allegra scanner at the MGH-Martinos Center in Charlestown, Massachusetts. For each subject, two high-resolution structural images were acquired for spatial normalization and cortical surface reconstruction using a 3-D MPRAGE (T1 weighted) sequence (128 sagittal slices, TR = 2.53 sec, TE = 3.25 msec, flip angle = 7°, FOV = 256, in-plane resolution 1 × 1 mm, slice thickness = 1.33 mm). Subsequently, a series of functional images was collected using a T2*-weighted EPI sequence of 33 interleaved slices oriented along the AC–PC line (TR = 2 sec, TE = 25 msec, flip angle = 90°, FOV = 64 × 64 matrix, in-plane resolution 3.125 × 3.125 mm, slice thickness = 3 mm + 1 mm skip). Functional imaging parameters were identical for both the priming and the localizer tasks.

Statistical Image Analysis

Functional data were analyzed using the FS-FAST analysis tools (see detailed description in Bar & Aminoff, 2003). Functional images were motion corrected (using

the AFNI package; Cox, 1996) and spatially smoothed with an 8-mm Gaussian full-width half-maximum filter. The intensities for all runs were then normalized to correct for signal intensity changes and temporal drift, with global rescaling for each run to a mean intensity of 1000. In order to obtain whole-brain group activation maps in the event-related priming task, a Finite Impulse Response (FIR) model was used. This model does not make any a priori assumptions about the shape of the hemodynamic response (HDR). Linear models are used to estimate the response amplitude at each time point of the HDR (see Burock & Dale, 2000, for details of this technique). Because the BOLD response typically begins 2 sec after stimulus onset and peaks between 4 and 7 sec (e.g., Dale & Buckner, 1997), and in order to obtain a reliable estimation of the HDR that is based on an optimal time window of the BOLD activation, the signal intensity in each condition was averaged across 2 to 8 sec from trial onset (see, e.g., Manoach, Greve, Lindgren, & Dale, 2003). Motion parameters derived from realignment correction were also entered to the model as covariates of no interest. The data were then tested for statistical significance and activation maps were constructed for specific contrasts of interest. Average group results were obtained using a random-effect statistical model, and were projected onto an ''inflated'' brain (Fischl, Sereno, & Dale, 1999) with an average curvature of 80 different brains. Cortical activation maps were corrected for multiple comparisons by employing a Monte Carlo simulation and obtaining significance levels with a clusterwise threshold. Specifically, we ran 10,000 Monte Carlo simulations of the smoothing, resampling, averaging, and thresholding procedures using synthesized white Gaussian noise data. This allowed us to derive a p value for each of the clusters in our analysis which indicated how likely it was that a cluster of a certain size would be found by chance given an initial voxelwise threshold of $p = .005$. Only clusters significant at a clusterwise p value of .01 were presented in the final whole-brain analysis.

Region-of-interest Analysis

ROIs for the IPC, PHC, hippocampus, and LOC were constrained both structurally and functionally. The structural constraint was based on a hand labeling of these brain structures for each subject. In the IPC, the anatomical label was restricted to the posterior IPC (i.e., frontal operculum and inferior precentral sulcus), as only little activation was obtained for the functional mask (see below) within the anterior IPC. In the PHC, the anatomical label consisted of the collateral sulcus and the parahippocampal gyrus, and in the LOC it consisted of the fusiform gyrus, the occipital–temporal sulcus, and the more dorsal lateral occipital (LO) region. The additional functional constraint for the ROI analysis in each subject was based on an unbiased mask selecting only the subset of the voxels within each anatomical label that were activated in a positive direction by any component of the priming task (i.e., all conditions > fixation-baseline). For the LOC, the functional mask was defined by comparing activation for objects versus faces, houses, and scenes within the localizer task (see e.g., Hasson, Harel, Levy, & Malach, 2003). All functional masks were obtained by using an estimated HDR that was defined by a gamma function of 2.25 sec hemodynamic delay and 1.25 sec dispersion. For each ROI, a minimum of 20 voxels, active at a threshold of $p < .01$ (uncorrected), was required in order to conduct the ROI analysis. For subjects who did not comply with these criteria (ranging from 0 to 4 subjects, depending on the specific brain region), a less stringent threshold of $p < .05$ was used for the functional mask. Subjects who still showed no activation under these conditions were discarded altogether from the ROI analysis (1 subject in the left IPC, 2 subjects in the left PHC, 3 subjects in the left and right hippocampus, and 2 subjects in the right LOC; altogether 6 subjects were excluded from one or more analyses, with a maximal exclusion of 3 analyses for one of the subjects). Voxels from the ROIs were then averaged for each time point in each condition, within each anatomical structure. Percent signal changes from baseline were computed and averaged across 2 to 8 sec from stimulus onset, allowing to conduct statistical tests for specific contrasts of interest across subjects.

RESULTS

Behavioral Results

The primary goal of the study was to assess whether identity-based and location-based contextual factors affect target object recognition with or without interacting, implying unified versus independent representations, respectively. Each prime–target pair was presented twice during the course of the experiment, and thus, a secondary goal was to investigate whether effects of repetition (e.g., repetition priming [RP]) were modulated by the different contextual factors. Because semantic relations could be manipulated only between prime objects and real target objects, and to simplify results presentation, we discarded all nonsense target object trials from the statistical analyses. The results, therefore, focus on the four contextual conditions within the real target object trials (Figure 1B).

Overall accuracy level in the target classification task was high (95%), with no trend for speed–accuracy tradeoff. A repeated measures analysis of variance (ANOVA) of RT for the real target objects, with semantic (related, unrelated) and spatial (congruent, incongruent) relations as within-subject factors, revealed shorter latencies for the semantically related than the unrelated condition [main effect of the semantic factor, $F(1, 19) = 15.35$, $p < .001$]. This finding replicates typical semantic priming effects

obtained with both verbal and pictorial stimuli (e.g., Van Petten & Luka, 2006; Ganis & Kutas, 2003; McPherson & Holcomb, 1999). There was no main effect for the spatial factor [$F(1, 19) = 0.32$], however, a (marginally) significant interaction emerged between the semantic and spatial factors [$F(1, 19) = 4.17, p < .055$], suggesting that the former was modulated by prime–target spatial relations. Namely, shorter RTs were obtained for semantically related than for unrelated targets in the spatially congruent [$t(19) = 4.42, p < .001$, two-tailed], but not in the incongruent condition [$t(19) = 0.09$; see Figure 2A]. All other simple main effects were not significant ($p > .05$).

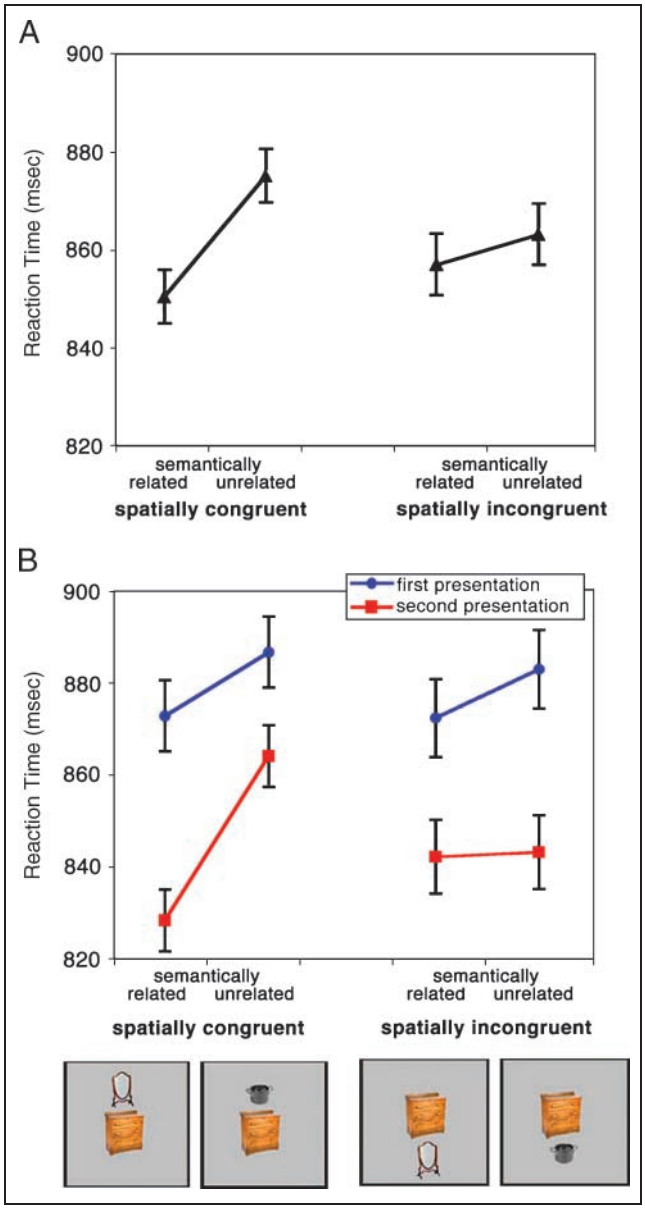

Figure 2. (A) RT data collapsed across the two experimental presentations for the four contextual conditions. (B) RT data for the four contextual conditions within each experimental presentation. Error bars indicate the standard error of the difference between responses to the semantically related and the unrelated conditions.

A closer inspection of the results revealed that the interaction above was further qualified by differences between the first and second presentations of the stimuli (Figure 2B). Specifically, whereas a semantic main effect was obtained at first presentation [$F(1, 19) = 4.52, p < .05$], semantic and spatial factors interacted only upon stimulus repetition [i.e., in the second presentation; $F(1, 19) = 9.48, p < .006$]. A three-way ANOVA confirmed that the Semantic (related, unrelated) × Spatial (congruent, incongruent) × Presentation (first, second) interaction was statistically significant [$F(1, 19) = 5.77, p < .03$]. This interaction could also be described as a Semantic × Spatial interaction in RP effects (i.e., the difference in RT between the first and second presentations; see Figure 2B), where a greater RP effect was found for semantically related than for unrelated targets within the spatially congruent condition [$t(19) = 2.36, p < .03$], but not in the incongruent condition [$t(19) = -0.94$]. RP effects are often considered to reflect a form of learning that improves perceptual processing and/or response selection for a repeated target (see Schacter, Dobbins, & Schnyer, 2004, for a review). The Semantic × Spatial interaction in RP indicates that semantic relatedness affected efficiency processing of a repeated target, however, this effect was dependent on the target's location (whether congruent or incongruent with respect to the prime).

Note that we describe the interaction effects obtained as a modulation of semantic effects by spatial condition. However, in some cases (such as the Semantic × Spatial interaction in the second presentation), spatial contextual effects were also modulated by semantic factors, resulting in significantly faster RTs for spatially congruent than incongruent targets in the semantically related condition, along with a reversed pattern of results in the semantically unrelated condition. These findings indicate that spatial consistency can facilitate recognition of a semantically related object, yet it may also hinder recognition of a semantically unrelated object (possibly due to violation of expectations regarding the most probable object identity to appear in a spatially congruent location). Interestingly, and in contrast to expectations, nonsense object targets (not shown in Figure 2) displayed a similar pattern of results to that of the semantically unrelated objects, that is, slower RTs for the spatially congruent than the incongruent locations, suggesting that these nonsense patterns were actually perceived as unrelated (rather than neutral) objects within the specific task context (we discuss possible implications of this outcome in the Discussion section). Despite the spatial effects mentioned above, we refer throughout the manuscript mainly to the modulation of semantic contextual effects by spatial location because it was most consistent across analyses, both in the RT and in the fMRI data below.

Our behavioral findings suggest, therefore, that semantic and spatial associations can affect object recognition

in an interactive manner. Although there was no Semantic × Spatial interaction in the first stimulus presentation, subjects were, nevertheless, sensitive to the global configuration of prime and target, as reflected by the interactive results in the second experimental exposure. Furthermore, a semantic main effect of target identity was obtained during first presentation, reflected by shorter latencies to semantically related than unrelated objects, *regardless* of target location. These findings suggest that subjects may have used different types of schematic knowledge during initial and repeated presentations. Specifically, whereas responses to first presentation were likely mediated by a ''pure'' semantic (or conceptual) contextual schema that is abstracted of specific spatial relations, responses during the second presentation were dominated by a combined semantic–spatial contextual representation, as reflected by the Semantic × Spatial interaction in RTs. We propose that activation of such a rich visual contextual representation was triggered by the exposure to the prime–target pairs and the encoding of their spatial relations during the initial presentation. Whereas subjects were equally exposed to stimuli in all contextual conditions, however, our results imply that encoding of prime–target relations was most efficient for stimuli pairs that formed a coherent percept (i.e., objects that were consistent with each other in both semantic and spatial dimensions). We will further elaborate on the differences in results between the initial and repeated experimental presentations in the Discussion section.

fMRI Results

Semantic Processing in the Inferior Prefrontal Cortex

To account for the differences between the first and second presentations obtained in the behavioral (RT) results, we analyzed the BOLD responses from the two presentations separately. Results for the first presentation within the posterior IPC showed a clear interaction of Semantic × Spatial conditions in both hemispheres [LH: $F(1, 18) = 12.48, p < .002$; RH: $F(1, 19) = 6.27, p < .03$], reflecting a larger percent signal change for the semantically related than for the unrelated targets within the spatially congruent [LH: $t(18) = 3.82, p < .001$; RH: $t(19) = 2.12, p < .05$], but not the incongruent ($p > .05$ in both hemispheres) condition. Interestingly, this pattern of results was reversed in the second presentation, where semantically related targets showed reduced percent signal change compared with unrelated targets. Once again, this difference was significant only in the spatially congruent condition [LH: $t(18) = -3.31, p < .004$; RH: $t(19) = -2.42, p < .03$; see Figure 3A].

Note that the interaction obtained in the BOLD responses for the first presentation was in contrast to the lack of interaction found in the behavioral results, whereas both measures (BOLD signal and RT) showed a similar pattern of responses in the second presentation (i.e., reduced/faster responses for semantically related than for unrelated targets, in the spatially congruent, but not the incongruent, condition). The fMRI findings for the repeated items (in the second presentation) mirror typical semantic priming effects, in which response reductions are usually obtained for semantically related, relative to unrelated items (for a review, see Van Petten & Luka, 2006). In contrast, the response enhancement found for the semantically related targets in the first presentation (within the spatially congruent condition) suggests that rather than being primed, and thus, activated to a lesser degree, these items may have benefited from deeper encoding than the unrelated targets, resulting in significantly enhanced activation. This account is in accordance with our prior suggestion that subjects encoded more efficiently prime–target pairs that constructed a coherent percept, than an incoherent one, during the first experimental presentation. The fact that prime and target stimuli were presented within very short temporal intervals, and at adjacent spatial locations, further supports a postrecognition associative account for the increased activation. Namely, subjects ''actively'' associated, or linked, prime and target into a global visual percept subsequent to the presentation and recognition of prime–target pairs (as is the case with postlexical matching in the verbal semantic priming domain; e.g., Neely, 1991). Indeed, studies investigating associative encoding of words and pictures (e.g., using word pair associates, or face–name associates) have typically shown increased activation in the prefrontal cortex during successful encoding processing (Sperling et al., 2001; Montaldi et al., 1998; Dolan & Fletcher, 1997). Most importantly, the unique encoding-related activity found for meaningful contextual configurations suggests that the latter benefited from rich preexisting long-term memory representations.

As mentioned previously, the differences between the fMRI data in the first and second presentations can also be portrayed as a Semantic × Spatial interaction in RP effects [LH: $F(1,18) = 16.77, p < .001$; RH: $F(1, 19) = 11.11, p < .003$], reflecting greater response-reduction for the semantically related than for the unrelated targets in the spatially congruent [LH: $t(18) = 4.48, p < .001$; RH: $t(19) = 2.84, p < .01$], but not the incongruent condition. These results replicate our behavioral findings in emphasizing that the efficiency at which targets were processed during their repeated presentation was strongly affected by a joint influence of semantic and spatial factors. Interestingly, the fMRI pattern for RP also mirrored the fMRI results for the first presentation, revealing maximal response-reduction for targets that initially showed maximal encoding-related activation (in the semantically related, spatially congruent, condition), and relatively little, if any, response-reduction for targets that initially showed low levels of activation (in all other conditions). Note that the little response-reduction

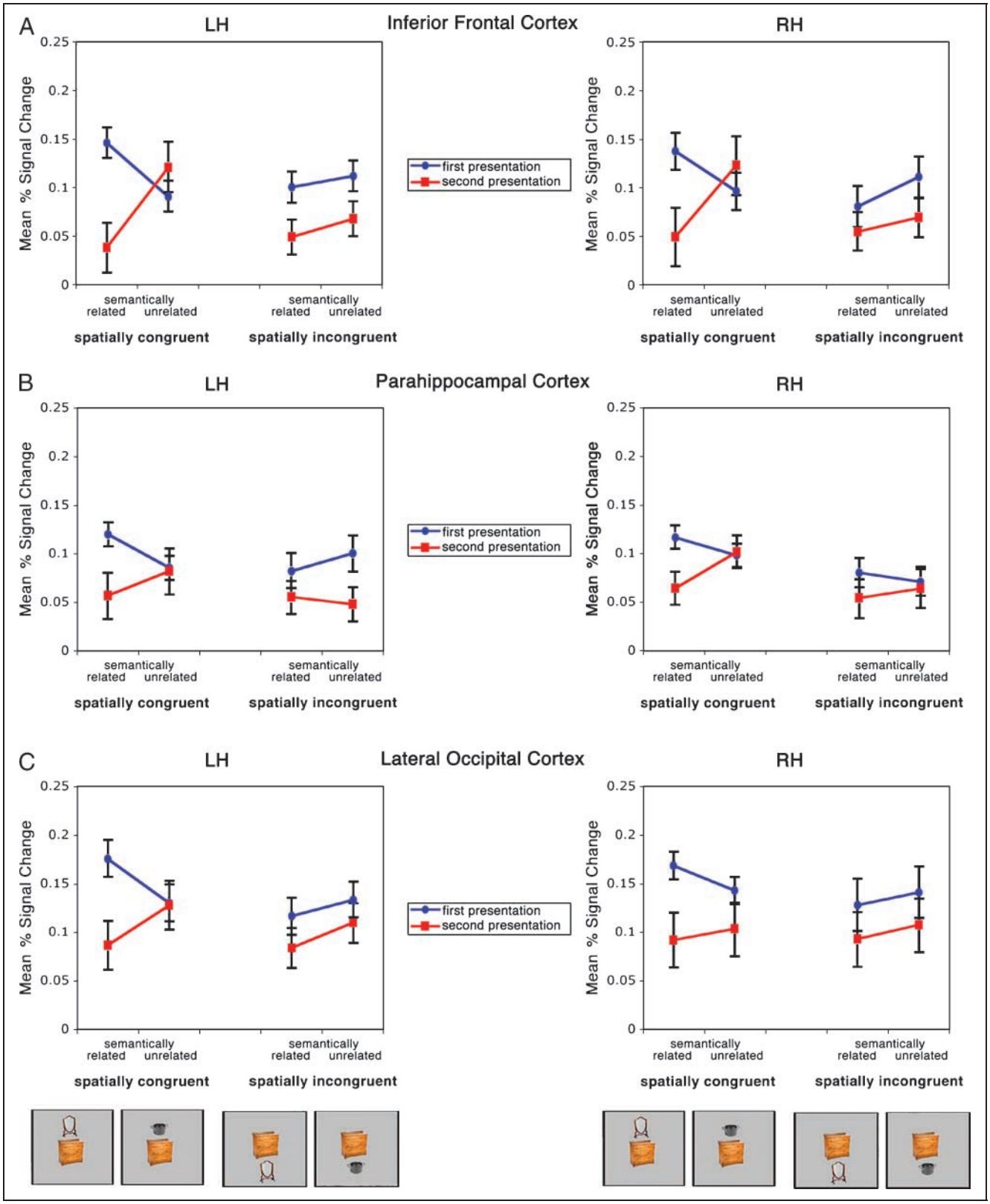

Figure 3. ROI analyses for the four contextual conditions within each experimental presentation. Mean percent signal changes from baseline were collapsed across 2 to 8 sec from trial onset. LH = left hemisphere; RH = right hemisphere. Error bars indicate the standard error of the difference between responses to the semantically related and the unrelated conditions.

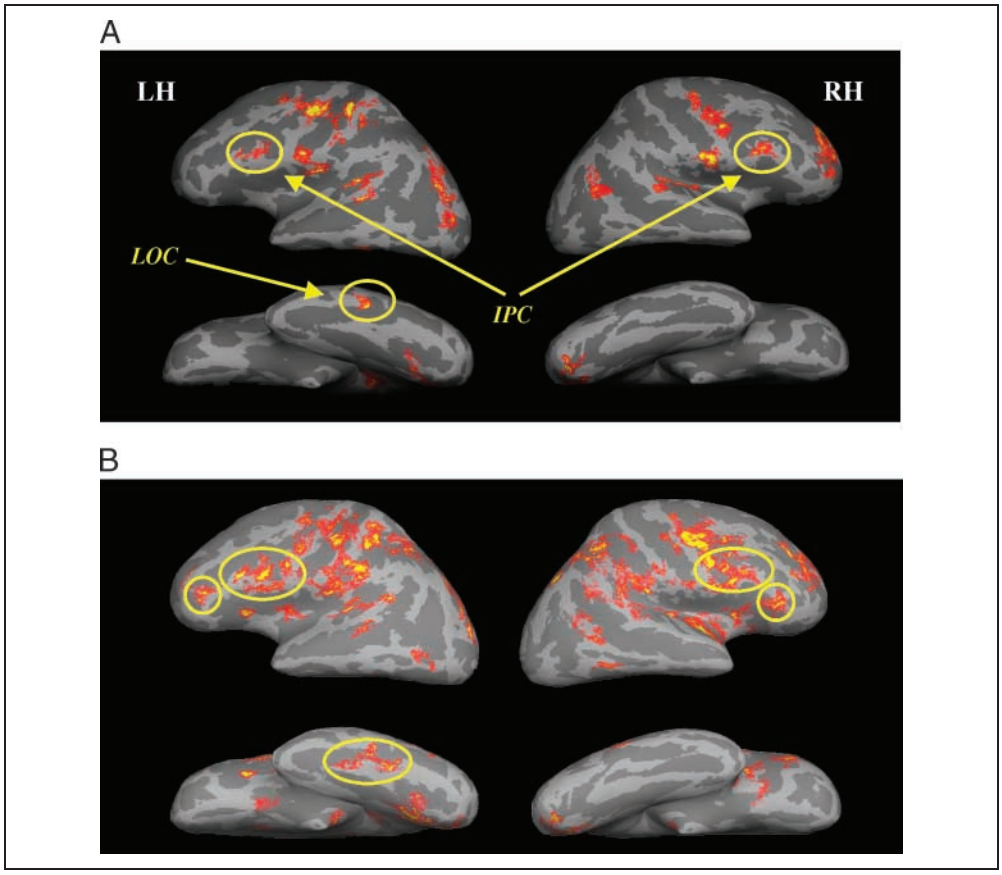

Figure 4. Statistical activation maps for the Semantic × Spatial interaction effect (A) in the first experimental presentation, (B) for the RP effect. Bilateral inferior prefrontal cortex (IPC) and left lateral occipital complex (LOC) regions are denoted in both figures. Activity in the first presentation was restricted in these areas to the posterior IPC and to the lateral occipito-temporal sulcus. Upon stimulus repetition activation extended to the fusiform gyrus and to the anterior IPC. The activations shown were obtained by conducting a random effect group analysis (n = 20) and are presented on the lateral (upper) and ventral (lower) cortical surfaces of an ''inflated'' brain. Sulci are shown in dark gray and correspond to the averaged curvature of 80 different brains. Activations are significant at $p < 0.01$, corrected for cluster size, using a Monte Carlo simulation (see Methods).

mentioned cannot be attributed to a floor effect, as activation levels in the second presentation for these conditions exceeded that of the semantically related, spatially congruent, condition (in fact, in the semantically *unrelated*, spatially congruent, condition there was a response-enhancement upon stimulus repetition, suggesting that greater neural resources were required to process an unexpected object identity appearing in an expected location). The selectivity of responses to prime–target pairs that constituted a coherent percept is in agreement with ''classical'' RP findings obtained with single-cell recordings, in which greatest neural reduction was found for repeated stimuli among neurons that were most active at first presentation (Li, Miller, & Desimone, 1993). Thus, a differential pattern of enhanced activation, followed by a significant drop in the BOLD signal, was obtained for targets consistent with primes in both semantic and spatial dimensions, further supporting our hypothesis that these two associative factors are linked within a unified contextual representation.

Visual Contextual Processing in the Parahippocampal Cortex

The overall pattern of results in the PHC resembled that of the IPC (see Figure 3B), however, only the left hemisphere showed statistically significant Semantic × Spatial interaction effects. Namely, a significant interaction was found for the first presentation in the left PHC [$F(1, 17) = 7.74$, $p < .02$], indicating a larger activation for the semantically related than for the unrelated targets in the spatially congruent condition [$t(17) = 2.76$, $p < .02$], but not in the incongruent condition. Similar to findings in the IPC, a reversed pattern of activation was seen in the second presentation (i.e., stronger activation for the semantically unrelated than for the related targets, only in the spatially congruent condition), resulting in a three-way Semantic × Spatial × Presentation interaction [$F(1, 17) = 9.15$, $p < .008$]. These findings replicate our previous behavioral and neural results of greater response-reduction for the semantically related than for the unrelated items appearing in a congruent [$t(17) = 2.79$, $p < .02$], but not in an incongruent, location. Paralleling findings were obtained within the left hippocampus, reflecting a significant Semantic × Spatial interaction in the first presentation [$F(1, 16) = 7.76$, $p < .02$], as well as in the RP effects for the second presentation [$F(1, 16) = 5.66$, $p < .03$]. Results from the right hemisphere of both the PHC and the hippocampus showed a similar trend to those of the left hemisphere, although none of the interactions reached significance ($p > .05$).

Object-related Processing in the Lateral Occipital Complex

Analysis of activation from the LOC replicated our findings from previous ROI analyses, however, similar to the PHC and the hippocampus, the Semantic × Spatial interactions obtained were restricted to the left hemisphere (see Figure 3C). Specifically, a semantic effect, depicting increased activation for semantically related items in the first presentation, was found for the spatially congruent condition [$t(19) = 2.47, p < .03$], but not the incongruent condition [$F(1, 19) = 5.21, p < .04$, for the Semantic × Spatial interaction]. In addition, a reversed trend in the second presentation resulted in greater RP effects for the semantically related than for the unrelated targets. This difference was evident only in the spatially congruent condition [$t(19) = 3.06, p < .006$; $F(1, 19) = 5.43, p < .04$, for the three-way interaction]. A similar pattern of activation was found in the right hemisphere, however, all interaction effects failed to reach significance ($p > .05$).

Whole-brain Analysis of Semantic × Spatial Interaction Effects

In addition to the ROI analyses, we conducted a whole-brain group analysis examining contextual-dependent activation for specific contrasts of interest. The contrasts of semantic and spatial main effects (e.g., semantically related vs. unrelated, across spatial congruency; and spatially congruent vs. incongruent, across semantic relatedness) revealed very little, if any, brain activation, both within and across experimental presentations. We therefore focus on the statistical activation maps evoked by the two main interaction effects obtained in the study

Table 1. Semantic × Spatial Interaction Effect in the First Experimental Presentation

			Talairach Coordinates			
Surface	*Brain Areas*	*Hemisphere*	*x*	*y*	*z*	*p*
Lateral	Middle frontal sulcus/anterior frontal cortex	R	23	57	11	10^{-6}
	Posterior inferior prefrontal cortex	L	−48	13	23	10^{-4}
		R	59	13	13	10^{-4}
	Superior precentral sulcus/gyrus	L	−43	−5	48	10^{-6}
		R	47	−5	35	10^{-4}
	Central sulcus	L	−41	−17	50	10^{-6}
	Subcentral gyrus	L	−52	−4	16	10^{-5}
		R	53	−10	10	10^{-6}
	Superior postcentral gyrus/sulcus/superior IPS	L	−48	−22	54	10^{-4}
	Posterior lateral fissure/superior temporal gyrus	L	−66	−40	12	10^{-6}
		R	38	−31	11	10^{-5}
	Inferior angular parietal gyrus/middle temporal gyrus	R	40	−65	29	10^{-4}
	Inferior angular parietal gyrus/middle occipital gyrus	L	−47	−64	32	10^{-5}
Ventral	Lateral occipito-temporal sulcus	L	−58	−51	−16	10^{-5}
Medial	Superior frontal gyrus	R	12	24	53	10^{-4}
	Cingulate gyrus/sulcus	L	−10	−8	48	10^{-5}
		R	5	−22	33	10^{-4}
	Parieto-occipito sulcus/precuneus gyrus	L	−22	−72	31	10^{-6}
		R	4	−68	29	10^{-5}
	Posterior cingulate gyrus/retrosplenial	L	−10	−43	9	10^{-5}
	Cuneus gyrus/calcarine sulcus/lingual gyrus	L	−15	−74	15	10^{-7}
		R	2	−72	25	10^{-8}

Peak activation of clusters significant at $p < 0.01$, corrected for cluster size (Sup. = superior, Inf. = inferior, IPS = interparietal sulcus). *p* Values (right most column) correspond to the voxelwise (uncorrected) significance level of the peak voxel within each cluster (log values).

(see Figure 4A–B): (1) a Semantic × Spatial interaction during the first presentation (defined as [semantically related − semantically unrelated, within the spatially congruent location] − [semantically related − semantically unrelated, within the spatially incongruent location]); and (2) the three-way Semantic × Spatial × Presentation interaction, corresponding to the Semantic × Spatial interaction for the RP effect (defined as [RP for semantically related − RP for semantically unrelated, within the spatially congruent location] − [RP for semantically related − RP for semantically unrelated, within the spatially incongruent location]). Recall that these two interaction contrasts showed consistently robust effects in previous ROI analyses, indicating a differential semantic effect within the spatially congruent condition, during both initial stimuli presentation and upon prime–target repetition.

The statistical activation map for the Semantic × Spatial interaction effect in the first stimulus presentation (Figure 4A) revealed significant activations in regions previously subjected to an ROI analysis, namely, the bilateral posterior IPC (∼BA 44/45) and the left lateral occipito-temporal sulcus (within the LOC). Although demonstrating a significant effect in the ROI analysis, activation in the left PHC and the left hippocampus did not reach significance in the whole-brain analysis.

Table 2. Semantic × Spatial Interaction for the RP Effect

			Talairach Coordinates			
Surface	*Brain Areas*	*Hemisphere*	*x*	*y*	*z*	*p*
Lateral	Middle frontal sulcus/anterior frontal cortex	R	16	56	13	10^{-6}
	Anterior inf. prefrontal cortex	L	−44	45	1	10^{-5}
		R	54	23	−9	10^{-5}
	Posterior inf. prefrontal cortex	L	−47	14	20	10^{-6}
		R	49	3	24	10^{-6}
	Sup. precentral sulcus/gyrus	L	−55	1	35	10^{-5}
		R	51	−1	37	10^{-7}
	Central sulcus	L	−34	−18	36	10^{-6}
	Insula	L	−31	22	2	10^{-5}
		R	49	−11	−9	10^{-5}
	Inf. supramarginal parietal gyrus	L	−56	−16	36	10^{-6}
		R	52	−32	17	10^{-5}
	Sup. postcentral gyrus/sulcus/sup. IPS	L	−23	−40	44	10^{-5}
		R	26	−63	41	10^{-6}
	Posterior lateral fissure/sup. temporal gyrus	L	−58	−49	29	10^{-5}
	Inf. angular parietal gyrus/middle temporal gyrus	R	57	−59	5	10^{-4}
	Inf. IPS/middle occipital gyrus	L	−26	−66	40	10^{-5}
Ventral	Lateral occipito-temporal sulcus/fusiform	L	−47	−51	−7	10^{-5}
Medial	Sup. frontal gyrus	L	−2	−10	69	10^{-6}
		R	12	19	52	10^{-5}
	Cingulate gyrus/sulcus	L	−3	−18	32	10^{-6}
		R	5	−26	33	10^{-7}
	Parieto-occipito sulcus/precuneus gyrus	L	−19	−61	29	10^{-6}
		R	12	−68	47	10^{-6}
	Cuneus gyrus/calcarine sulcus/lingual gyrus	L	−17	−63	4	10^{-6}
		R	0	−84	21	10^{-5}

Peak activation of clusters significant at $p < 0.01$, corrected for cluster size (Sup. = superior, Inf. = inferior, IPS = intraparietal sulcus). p Values (right most column) correspond to the voxelwise (uncorrected) significance level of the peak voxel within each cluster (log values).

Other significant clusters of activation for the interaction contrast were found in the bilateral premotor cortex (precentral sulcus and gyrus), subcentral gyrus, superior temporal lobe, cingulate cortex, and precuneus and cuneus gyrus (extending to the calcarine sulcus and the lingual gyrus). In addition, activation was found in the right anterior frontal cortex (~BA 10) (see Table 1).

Notably, a very similar map of activation was obtained for the interaction contrast of the RP effect (or the Semantic × Spatial × Presentation interaction), albeit this RP contrast produced overall stronger activity spanning over larger brain regions (see Figure 4B). The relative similarity between the two activation maps indicates once again that maximal response reduction for repeated stimuli was obtained in regions that were initially most reactive to the joint effect of semantic and spatial associative factors. Among those regions were the IPC (spanning both posterior and anterior inferior frontal cortex) and the left LOC (including lateral occipital–temporal sulcus and fusiform gyrus), as well as other medial and lateral regions originally activated by the Semantic × Spatial interaction in the first presentation (see Table 2).

In addition, lateral parietal activation was observed bilaterally in both dorsal (e.g., intraparietal sulcus, straddling the postcentral sulcus) and ventral (e.g., inferior supramarginal parietal gyrus) foci, suggesting a potential role for attention in the priming effect for the repeated items (see e.g., Corbetta & Shulman, 2002, for the role of the parietal cortex in attentional allocation). One possible explanation for such attentional involvement is that perceptual processes, or processes associated with response-selection, are less attentionally demanding for (repeated) semantically related than unrelated targets. This difference in processing efficiency, however, accounts only for targets appearing in spatially congruent, but not incongruent, locations. Alternatively, the parietal activations (in conjunction with the robust frontal activations) may reflect memory retrieval processes, occurring either implicitly or explicitly upon stimuli repetition (see reviews in Wagner, Shannon, Kahn, & Buckner, 2005; Buckner & Wheeler, 2001). A third possible account for the frontal–parietal activation seen is that it reflects action-related representations associated with processing of ''active'' object pairs, such as when a prime tool is presented with its corresponding target object (e.g., a hammer with a nail). Previous studies have shown that viewing and naming tools activates action- and motor-processing regions within temporal, parietal, and frontal ''mirror-neuron'' networks (e.g., Hauk & Pulvermuller, 2004; Chao & Martin, 2000; Grafton, Fadiga, Arbib, & Rizzolatti, 1997). Most importantly, whether attention, memory, action perception, or a combination of these mechanisms underlies the frontal–parietal activation obtained, these results clearly indicate that processing of repeatedly presented stimuli is particularly beneficial for prime–target pairs that are both associated semantically and are positioned in correct spatial locations, thus forming a coherent, meaningful, contextual representation.

DISCUSSION

The present study investigated the nature of contextual representations by using a priming paradigm in which semantic and spatial relations between prime and target were independently manipulated. Our results demonstrate that identity-based and location-based contextual factors interact with each other during visual object recognition, supporting the proposal that the two types of associative knowledge are linked together in a unified representation. Such a *context frame* may be particularly beneficial because it captures the natural regularities in the visual environment and it conveys important information about the interactions between objects and their function within a contextual setting.

Although both behavioral and fMRI measurements showed evidence for a combined contextual representation, some discrepancy was found between the two measures: RT data showed an interactive effect of semantic and spatial information upon prime–target repetition, whereas fMRI data showed interactive results both at initial and at repeated presentations of stimuli. As mentioned earlier, we propose that the fMRI results for the first experimental presentation reflect the deep associative encoding of object pairs that formed contextually coherent percepts. This encoding-related activity was not reflected by the RT measure, however, possibly because encoding took place at a relatively late cognitive stage, during, or after, response execution. Namely, associating the prime and the target into a global percept presumably occurred only after both stimuli were recognized, thus affecting only the fMRI, which integrates across a longer interval than the corresponding RT measure (see, e.g., Henson, 2003, for a similar account of dissociated RT and fMRI responses in verbal semantic priming). Furthermore, the RT measure showed a significant semantic main effect during initial stimuli exposure, suggesting that subjects may have activated a rapid, identity-based contextual representation as default. This latter finding corroborates findings from earlier studies showing that the meaning of pictures, as well as the associative links between object identities, can be represented on a conceptual level that is devoid of specific visual details and of metric coordinates (e.g., Carr et al., 1982; Mandler & Johnson, 1976; Potter & Faulconer, 1975; Pylyshyn, 1973; Sperber et al., 1973). The dissociation between the behavioral and BOLD indices during the first presentation further implies that ''pure'' semantic (or conceptual) knowledge and a combined semantic–spatial contextual representation can both affect responses to a target object, albeit possibly at different processing stages (e.g., early vs. late stages of recognition, respectively). Upon exposure to prime–target pairs

and their spatial relations, however, responses in both RT and fMRI measures were dominated by a combined contextual representation (as seen in the RP effects for the second presentation). These results suggest that a result of a rich unified context frame may be encoding highly specific perceptual information, and its activation is particularly relevant during episodic retrieval of this information. Thus, although identity-based (semantic) contextual knowledge may serve as a rapid but general means to predict visual events, a unified contextual representation allows the generation and retrieval of highly detailed episodic percepts, based on exposure to these visual events.

Our findings provide similar evidence for the existence of a combined contextual representation, as well as a more abstract, semantic one. As for the existence of a purely spatial contextual representation, the results of the present study are less clear. Recall that such a representation presumably depicts the general spatial layout of a visual setting, providing information about the most probable location(s) to contain an object (regardless of the object's specific details). Previous studies using arbitrary contextual settings and artificial scenes provided support for such a representation by showing strong effects of spatial context on object identification, independent of target identity (e.g., Chun & Jiang, 1998; Sanocki & Epstein, 1997). The spatial effects in the present study were modulated by semantic relatedness, in that spatial congruency enhanced target recognition in the semantically related condition, whereas it often hindered recognition in the semantically unrelated condition. Thus, in contrast to previous studies, there were no independent spatial main effects, whether at the behavioral or at the neural levels. One possible account for this apparent discrepancy is that the studies mentioned above used tasks in which there was greater uncertainty of target location (e.g., a visual search task) because of a large stimulus array size or a wide range of potentially relevant target positions. In the contextual cueing paradigm (Chun & Jiang, 1998), for instance, subjects are extensively trained to detect a target object within a large distractor array, resulting in faster identification of targets appearing in trained (familiar) than untrained (unfamiliar) visual settings. Attentional guidance to a target location (following training) is also demonstrated when real-world scenes are used as contextual settings, although target identity and its location within these scenes are typically determined arbitrarily (Brockmole & Henderson, 2006; Summerfield, Lepsien, Gitelman, Mesulam, & Nobre, 2006). Thus, when using visual search paradigms, in combination with extensive search training, subjects are presumably more sensitive to spatial regularities that can aid in predicting target location, and thus, they can benefit from spatial context to a larger extent. Furthermore, the use of a priming task in the present study, in which both prime and target consisted of real-world objects that were either semantically related or unrelated, may have biased subjects to seek semantic relatedness and to overweigh semantic contextual factors over spatial ones. The fact that nonsense target objects elicited responses similar to these of the semantically unrelated targets may suggest that the meaningless shapes were, in fact, perceived as unrelated (rather than neutral) objects, further supporting a dominant role for semantic over spatial contextual factors in the present task. To determine the exact role of ''pure'' location-based associative knowledge in real-world visual settings, one may need to use a paradigm more directly oriented to spatial localization, as well as to control for semantic factors associated with target recognition (e.g., by using arbitrary, meaningless, targets). In line with this suggestion, recent data from our lab (Gronau & Bar, submitted) reveal that prime object pictures can indeed serve as efficient spatial cues for target location, when using a simple location-detection task and controlling for semantic relatedness between prime and target. Thus, for example, a prime picture of a tool pointing down results in faster RTs for an arbitrary target appearing in a lower, than an upper, position, although the prime cue is non-informative in nature (i.e., predicts the actual location of the target on only 50% of trials). Further research may be required to fully understand the circumstances under which spatial contextual knowledge, based on one's lifetime experience with visual objects, can affect recognition processes in real-world settings.

Neural Underpinnings of a Unified Contextual Representation

Although several fMRI studies have investigated the cortical correlates of visual contextual processing (Aminoff et al., 2007; Summerfield et al., 2006; Bar & Aminoff, 2003; see Bar, 2004, for a review), the present research is a first step toward uncovering the neural underpinnings of the interactive effects of semantic and spatial associative information. Strong interactions between identity- and location-based contextual factors were found in regions implicated in semantic processing (bilateral IPC), as well as regions implied in visual contextual and associative processing (left PHC and hippocampus), indicating that analysis in these regions is mediated by a combined semantic–spatial representation. These findings therefore provide a novel insight into the actual representation and processing of contextual associations, and more specifically, offer support to the concept of a unified *context frame* (Bar, 2004; Bar & Ullman, 1996) at the neural level.

In addition, context-related activity was found in a network of frontal–parietal areas, suggesting that attention and/or memory-related processes are involved in the activation of such unified representations. Supportive evidence for this account arises from studies showing that objects that are associated with each other tend to capture attention and to be recalled to a greater extent than

unassociated objects (e.g., Riddoch et al., 2003; Moores, Laiti, & Chelazzi, 2002; Henke, Buck, Weber, & Wieser, 1997). Furthermore, capture of attention by semantically and spatially related objects may also explain the greater RP effects obtained for these objects, because attention has shown to be an important factor in modulating RP magnitude (Vuilleumier, Schwartz, Duhoux, Dolan, & Driver, 2005; Yi & Chun, 2005; Eger, Henson, Driver, & Dolan, 2004). Alternatively, the robust frontal–parietal activation seen might be related to action-related representations triggered by prime–target object pairs that were ''actively'' associated with each other (mainly tools and their target objects). As mentioned earlier, there is ample evidence to suggest that viewing tools (as well as listening to their sounds, or generating words associated with their action) activates regions in frontal motor areas (the ''mirror-neuron'' system) as well as in parietal action-related areas (e.g., Lewis, Brefczynski, Phinney, Janik, & DeYoe, 2005; Hauk & Pulvermuller, 2004; Chao & Martin, 2000; Grafton et al., 1997). That subjects were sensitive to the potential action relations between prime and target suggests that *context frames* may not only incorporate the precise visual details of a certain context (e.g., the spatial relations between objects), but also other multisensory information such as functional use and goal-directed action afforded by the context (see also Bach, Knoblich, Gunter, Friederici, & Prinz, 2005). Taken together, our findings indicate that visual contextual analysis involves regions specialized in associative processing, as well as other attentional, memory, and multisensory representation areas.

Evidence for a unified contextual representation was also found in lower perceptual regions within the LOC, specifically in the occipito-temporal sulcus and the fusiform gyrus. The LOC has been implicated in visual object recognition (e.g., Grill-Spector et al., 2001; Malach et al., 1995), and activity in this region is known to be strongly affected by global perceptual factors such as object completion (e.g., Lerner, Hendler, & Malach, 2002; Kourtzi & Kanwisher, 2001). Here we show that the LOC is not only sensitive to the properties of individual objects but also to their contextual surroundings (see Cox et al., 2004, for similar findings in face-processing regions). Such global processing may be especially ecological, as objects in natural scenes (as opposed to laboratory settings) are rarely seen in isolation. Thus, although typically associated with processing of individual shapes and objects, the LOC might in fact be sensitive to the perceived unity of a set of associated objects.

Activation seen in the LOC might also be an outcome of top–down projections originating in higher order regions, such as the frontal lobes or the PHC. This account is in accordance with traditional views of contextual facilitation, suggesting that high-level associative processes directly modulate early perceptual stages of object recognition (e.g., Bar & Ullman, 1996; Boyce, Pollatsek, & Rayner, 1989; Sperber et al., 1979; Palmer, 1975; Biederman, 1972; but see Hollingworth & Henderson, 1998). Indeed, recent evidence from imaging studies has provided support for the notion that top–down mechanisms play a significant role in visual perception (e.g., Bar et al., 2006; Mechelli, Price, Friston, & Ishai, 2004; Chawla, Rees, & Friston, 1999; Kosslyn et al., 1993). It seems possible, therefore, that the activations seen in the LOC in the present study originate in feedback projections from higher semantic associative processing regions. Because of the low temporal resolution of the BOLD signal, however, one cannot specify the exact stage at which such projections may take place. Higher temporal resolution methods, such as magnetoencephalography, could provide a more definitive answer to this question (Bar et al., 2006).

Finally, an important question concerns the lateralization of the contextual effects to the left hemisphere, in medial-temporal lobe regions (PHC and hippocampus) and in the LOC. Although, traditionally, left and right hemispheres are implicated in ''local'' (or parts-based) and ''global'' (or holistic) visual processing, respectively (e.g., Fink et al., 1996; Marsolek, Schacter, & Nicholas, 1996; Corballis, 1989; Robertson, Lamb, & Knight, 1988; Bradshaw & Nettleton, 1981), our results reveal greater sensitivity to the global properties of an associative setting in the left hemisphere. On the face of it, one would predict the opposite result, with activation for a unified contextual representation mainly lateralized to the right. However, this apparent contradiction can be resolved under the assumption that the global nature of such a representation predominantly derives from its coherence, or its overall meaning. Alternatively, this coherence is a result of the compatibility between individual objects (i.e., local information) in the scene. Namely, only visual settings consisting of objects that are in accord with each other, both semantically and spatially, can be interpreted as meaningful or comprehensible. Thus, it may be the case that high-level semantic processes, which have typically more lateralized to the left hemisphere (e.g., Van Petten & Luka, 2006), mediate the lateralization effects seen in the present study.

In summary, our study provides novel evidence for the representation of visual context within a unified context frame that comprises information about the identities most likely to appear in a certain visual setting, as well as their typical spatial relations. Interactive effects of identity- and location-based contextual factors were obtained in semantic and visual contextual brain regions (e.g., IPC and PHC), as well as in lower object-processing regions (e.g., LOC). In addition, activation in frontoparietal areas suggested that attention, memory, and/or action-related processes might partially mediate the contextual effects seen. Taken together, these results portray a network of high-level brain regions that work in orchestration to construct a meaningful percept of the visual environment. Most relevant for the present

study, the coherence of such a percept is determined by the combined effects of semantic and spatial contextual factors.

Acknowledgments

We thank Elissa Aminoff, Jasmine Boshyan and Mark Fenske for their assistance in data analysis. In addition, we thank two anonymous reviewers for their helpful comments. This research was supported by NINDS R01-NS044319 and NS050615, the McDonnell Foundation 21002039 (all to M. B.), and the MIND Institute.

Reprint requests should be sent to Moshe Bar, Martinos Center for Biomedical Imaging at MGH, Harvard Medical School, Rm. 2301, 149 Thirteenth Street, Charlestown, MA 02129, or via e-mail: bar@nmr.mgh.harvard.edu.

REFERENCES

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and non-spatial associations. *Cerebral Cortex, 27,* 1493–1503.

Bach, P., Knoblich, G., Gunter, T. C., Friederici, A. D., & Prinz, W. (2005). Action comprehension: Deriving spatial and functional relations. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 465–479.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5,* 617–629.

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11,* 280–289.

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron, 38,* 347–358.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 449–454.

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception, 25,* 343–352.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177,* 77–80.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 556–566.

Bradshaw, J. L., & Nettleton, N. C. (1981). The nature of hemispheric specialization in man. *Behavioral and Brain Sciences, 4,* 51–91.

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition, 13,* 99–108.

Buckner, R. L., & Wheeler, M. E. (2001). The cognitive neuroscience of remembering. *Nature Reviews Neuroscience, 2,* 624–634.

Burock, M. A., & Dale, A. M. (2000). Estimation and detection of event-related fMRI signals with temporally correlated noise: A statistically efficient and unbiased approach. *Human Brain Mapping, 11,* 249–260.

Carr, T. H., McCauley, C., Sperber, R. D., & Parmelee, C. M. (1982). Words, pictures, and priming: On semantic activation, conscious identification, and the automaticity of information processing. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 757–777.

Chao, L. L., & Martin, A. (2000). Representation of manipulable manmade objects in the dorsal stream. *Neuroimage, 12,* 478–484.

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience, 2,* 671–676.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71.

Corballis, M. C. (1989). Laterality and human evolution. *Psychological Review, 96,* 492–505.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-driven and stimulus driven attention in the brain. *Nature Reviews Neuroscience, 3,* 201–215.

Cox, D., Meyers, E., & Sinha, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science, 303,* 115–117.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29,* 162–173.

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping, 5,* 329–340.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15,* 559–564.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52,* 317–329.

Dolan, R. J., & Fletcher, P. C. (1997). Dissociating prefrontal and hippocampal function in episodic memory encoding. *Nature, 388,* 582–585.

Eger, E., Henson, R. N. A., Driver, J., & Dolan, R. J. (2004). BOLD repetition decreases in object-responsive ventral visual areas depend on spatial attention. *Journal of Neurophysiology, 92,* 1241–1247.

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron, 44,* 109–120.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1996). Where in the brain does visual attention select the forest and the trees? *Nature, 382,* 626–628.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. *Neuroimage, 9,* 195–207.

Freyd, J. (1983). The mental representation of movement when static stimuli are viewed. *Perception & Psychophysics, 33,* 575–581.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108,* 316–355.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Brain Research, Cognitive Brain Research, 16,* 123–144.

Goh, J. O. S., Siong, S. C., Park, D., Gutchess, A., Hebrank, A., & Chee, M. W. L. (2004). Cortical areas involved in object, background, and object-background processing revealed with functional magnetic resonance adaptation. *Journal of Neuroscience, 24,* 10223–10228.

Grafton, S. T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage, 6,* 231–236.
Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Visual Research, 41,* 1409–1422.
Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron, 37,* 1027–1041.
Hauk, O., & Pulvermuller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Human Brain Mapping, 21,* 191–201.
Henke, K., Buck, A., Weber, B., & Wieser, H. G. (1997). Human hippocampus establishes associations in memory. *Hippocampus, 7,* 249–256.
Henson, R. N. A. (2003). Neuroimaging studies of priming. *Progress in Neurobiology, 70,* 53–81.
Hock, H. S., Romanski, L., Galie, A., & Williams, C. S. (1978). Real-world schemata and scene recognition in adults and children. *Memory & Cognition, 6,* 423–431.
Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 58–69.
Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127,* 398–415.
Kanwisher, N., Chun, M., McDermott, J., & Ledden, P. (1996). Functional imaging of human visual recognition. *Cognitive Brain Research, 5,* 55–67.
Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Chabris, C. F., Rauch, S. L., & Anderson, A. K. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of Cognitive Neuroscience, 5,* 263–287.
Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience, 12,* 48–55.
Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293,* 1506–1509.
Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J., & DeYoe, E. A. (2005). Distinct cortical pathways for processing tool versus animal sounds. *Journal of Neuroscience, 25,* 5148–5158.
Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology, 69,* 1918–1929.
Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, U.S.A., 92,* 8135–8139.
Mandler, J. M., & Johnson, N. S. (1976). Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 529–540.
Mandler, J. M., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory,* 386–396.
Manoach, D. S., Greve, D. N., Lindgren, K. A., & Dale, A. M. (2003). Identifying regional activity associated with temporally separated components of working memory using event-related functional MRI. *Neuroimage, 20,* 1670–1684.
Marsolek, C. J., Schacter, D. L., & Nicholas, C. D. (1996). Form-specific visual priming for new associations in the right cerebral hemisphere. *Memory & Cognition, 24,* 539–556.
McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36,* 53–65.
Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down: Neuronal interactions during perception and imagery. *Cerebral Cortex, 14,* 1256–1265.
Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). McGraw-Hill.
Montaldi, D., Mayes, A. R., Barnes, A., Pirie, H., Hadley, D. M., Patterson, J., et al. (1998). Associative encoding of pictures activated the medial temporal lobes. *Human Brain Mapping, 6,* 85–104.
Moores, E., Laiti, L., & Chelazzi, L. (2002). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6,* 182–189.
Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 264–336). Hillsdale, NJ: Erlbaum.
Neider, N. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46,* 614–621.
Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3,* 519–526.
Potter, M. C., & Faulconer, B. (1975). Time to understand pictures and words. *Nature, 253,* 437–438.
Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80,* 1–24.
Reed, C., & Vinson, N. (1996). Conceptual effects on representational momentum. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 839–850.
Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7,* 17–42.
Riddoch, M. J., Humphreys, G. W., Edwards, S., Baker, T., & Willson, K. (2003). Seeing the action: Neuropsychological evidence for action-based effects on object selection. *Nature Neuroscience, 6,* 82–89.
Robertson, L. C., Lamb, M. R., & Knight, R. T. (1988). Effects of lesions of temporal–parietal junction on perceptual and attentional processing in humans. *Journal of Neuroscience, 8,* 3757–3769.
Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science, 8,* 374–378.
Schacter, D. L., Dobbins, I. G., & Schnyer, D. (2004). Specificity of priming: A cognitive neuroscience perspective. *Nature Reviews Neuroscience, 5,* 853–862.
Schank, R. C. (1975). *Conceptual information processing*. New York: Elsevier.
Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5,* 195–200.
Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7,* 301–305.
Sperber, R. D., McCauley, C., Ragain, R. D., & Weil, C. M. (1979). Semantic priming effects on picture and word processing. *Memory & Cognition, 7,* 339–345.
Sperling, R., Bates, J., Cocchiarella, A., Schacter, D. L., Rosen, B. R., & Albert, M. (2001). Encoding novel face–name

associations: A functional MRI study. *Human Brain Mapping, 14,* 129–139.
Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron, 49,* 905–916.
Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113,* 766–786.
Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain & Language, 97,* 279–293.
Vuilleumier, P., Schwartz, S., Duhoux, S., Dolan, R. J., & Driver, J. (2005). Selective attention modulates neural substrates of repetition priming and ''implicit'' visual memory: Suppressions and enhancements revealed by fMRI. *Journal of Cognitive Neuroscience, 17,* 1245–1260.
Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences, 9,* 445–453.
Yi, D. J., & Chun, M. M. (2005). Attentional modulation of learning related repetition attenuation effects in human parahippocampal cortex. *Journal of Neuroscience, 25,* 3593–3600.