

Interaction between Scene and Object Processing Revealed by Human fMRI and MEG Decoding

Talia Brandman and  Marius V. Peelen

Center for Mind/Brain Sciences, University of Trento, 38068 Rovereto (TN), Italy

Scenes strongly facilitate object recognition, such as when we make out the shape of a distant boat on the water. Yet, although known to interact in perception, neuroimaging research has primarily provided evidence for separate scene- and object-selective cortical pathways. This raises the question of how these pathways interact to support context-based perception. Here we used a novel approach in human fMRI and MEG studies to reveal supra-additive scene-object interactions. Participants (men and women) viewed degraded objects that were hard to recognize when presented in isolation but easy to recognize within their original scene context, in which no other associated objects were present. fMRI decoding showed that the multivariate representation of the objects' category (animate/inanimate) in object-selective cortex was strongly enhanced by the presence of scene context, even though the scenes alone did not evoke category-selective response patterns. This effect in object-selective cortex was correlated with concurrent activity in scene-selective regions. MEG decoding results revealed that scene-based facilitation of object processing peaked at 320 ms after stimulus onset, 100 ms later than peak decoding of intact objects. Together, results suggest that expectations derived from scene information, processed in scene-selective cortex, feed back to shape object representations in visual cortex. These findings characterize, in space and time, functional interactions between scene- and object-processing pathways.

Key words: contextual processing; fMRI; MEG; object perception; scene perception; visual cortex

Significance Statement

Although scenes and objects are known to contextually interact in visual perception, the study of high-level vision has mostly focused on the dissociation between their selective neural pathways. The current findings are the first to reveal direct facilitation of object recognition and neural representation by scene background, even in the absence of contextually associated objects. Using a multivariate approach to both fMRI and MEG, we characterize the functional neuroanatomy and neural dynamics of such scene-based object facilitation. Finally, the correlation of this effect with scene-selective activity suggests that, although functionally distinct, scene and object processing pathways do interact at a perceptual level to fill in for insufficient visual detail.

Introduction

Scene context powerfully shapes our perception of objects in everyday life, facilitating recognition when objects are small, distant, shaded, or partially occluded (Oliva and Torralba, 2007). For example, Figure 1A gives the clear impression of a sailing

ship, but not when isolated from its background. The neural basis of such scene-based object perception remains largely unexplored. Previous work has focused on distinct neural mechanisms of object and scene processing, providing evidence for separate object- and scene-selective areas (Malach et al., 1995; Epstein and Kanwisher, 1998) that each has a causal role in the perception of their preferred category (Mullin and Steeves, 2011; Dilks et al., 2013). In the present study, we used fMRI and MEG to ask how these two types of information interact in the brain, probing the functional neuroanatomy and neural dynamics of contextual facilitation of objects in real-world scenes.

Previous behavioral and neuroimaging studies have examined contextual processing using scene-object congruity paradigms, measuring the effects of contextual violations on object perception and neural responses. These studies found that semantic violations and spatial violations impair object detection and recognition (Biederman et al., 1982; Boyce and Pollatsek, 1992; Bar and Ullman, 1996; Davenport and Potter, 2004). Neuroimaging

Received March 1, 2017; revised June 5, 2017; accepted June 28, 2017.

Author contributions: T.B. and M.V.P. designed research; T.B. performed research; T.B. analyzed data; T.B. and M.V.P. wrote the paper.

This work was supported by the Autonomous Province of Trento, Call "Grandi Progetti 2012," project "Characterizing and improving brain mechanisms of attention-ATTEND," and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement 659778. This manuscript reflects only the authors' view and the Agency is not responsible for any use that may be made of the information it contains. We thank Nikolaas Oosterhof and Daniel Kaiser for useful advice on data analysis; and Clayton Hickey, David Melcher, and our reviewers for valuable comments on the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Talia Brandman, Corso Bettini, 31, 38068 Rovereto (TN), Italy. E-mail: talli.brandman@gmail.com.

DOI:10.1523/JNEUROSCI.0582-17.2017

Copyright © 2017 the authors 0270-6474/17/377700-11\$15.00/0

studies show increased EEG and fMRI responses to incongruent scene-object displays, depending on the type of violation and task (Ganis and Kutas, 2003; Demiral et al., 2012; Võ and Wolfe, 2013; Mudrik et al., 2014; Rémy et al., 2014). These increased responses to incongruent conditions, potentially reflecting prediction error signals (Noppeney, 2012), suggest that contextual cues present in the scene influence the processing of objects, at some level of processing. However, extant results do not identify the mechanisms by which scene information acts to shape the perceptual interpretation of objects.

There are two competing accounts of how the brain could use scene information to shape object processing. One possibility is that scene information facilitates the processing of objects at a perceptual stage. For example, according to the contextual-facilitation model (Bar, 2004), rapidly formed predictions derived from contextual associations project back to the inferior-temporal cortex, activating candidate object representations. Such an account is broadly similar to recent views on the role of expectation in perception (Summerfield and de Lange, 2014), with scene context creating expectations about object identities.

An alternative is that visual scene and object processing proceed in parallel with information being integrated only at post-perceptual stages (Hollingworth and Henderson, 1999), in line with studies showing the functional independence of these visual pathways (Mullin and Steeves, 2011; Dilks et al., 2013). Thereby, object and scene processing regions in the visual cortex would independently represent bottom-up sensory information, with integration occurring at postperceptual regions.

To distinguish between these possibilities, we developed a new approach that boosts the facilitating effect of scenes on the perceptual interpretation of objects, by reducing internal object information. Specifically, we presented degraded objects either in their intact scenes or in isolation, as well as the scenes alone (see Fig. 1B). These three conditions allowed us to test whether perceptual representations of objects and scenes are independent or interactive by comparing responses evoked by degraded objects in scenes with the sum of responses evoked by degraded objects alone and scenes alone. This supra-additivity test of object and scene processing was applied to fMRI, MEG, and behavioral measures. We used multivariate pattern analysis (MVPA) (Haxby et al., 2001) to track how objects were represented in the brain. To this end, we trained a machine-learning algorithm to classify the neural responses elicited by intact objects, and then tested its classification of neural responses evoked by degraded objects with and without scenes, and scenes alone (see Fig. 1C). This allowed us to determine how scene context impacted the resemblance of response patterns evoked by degraded objects to those evoked by intact objects. We thus characterize, in space and time, the facilitating effects of scene context on the neural representation of objects.

Materials and Methods

In Study 1 and Study 2, we measured the multivariate representations of object animacy in fMRI and MEG signals while participants performed an oddball detection task (see General procedure). In a separate animacy pattern localizer used for classifier training, participants viewed intact animate and inanimate objects, fully visible and in high resolution, presented in isolation. In addition, for the fMRI experiment, a region-of-interest (ROI) localizer served to localize object- and scene-selective ROIs in visual cortex. Finally, Study 3 measured object perception behaviorally. All procedures were approved by the ethics committee of the University of Trento.

Experimental stimuli

The stimulus set consisted of degraded objects that were perceived as ambiguous on their own but that were easily categorized in context. We ensured that the background scenes did not contain objects that were specifically associated with one of the object categories, such as other animals or traffic signs. The main experiment included photographs of 30 animate (animals and people) and 30 inanimate (cars, boats, planes, and trains) objects in outdoor scenes. Sixty photographs of scenes from MorgueFile database (www.morguefile.com) were cropped and cleaned to include one dominant foreground object. Each object was selected and downsized in resolution (i.e., pixelated), and saved once in its original background, and once on a uniform gray background of mean luminance of the original background. The image was also edited to exclude the object and saved as background only. The final images (180 in total) included the degraded objects in scenes, the degraded objects alone, and the scenes alone (see samples in Fig. 1B). To avoid familiarity effects passing from objects in scenes to objects alone, the stimulus set was split in two, such that different objects were presented for degraded objects in scenes and for degraded objects alone within a given subject. (For example, for two stimuli Bird1 and Bird2, a given participant would either view Bird1 in isolation and Bird2 in the sky, or vice versa.) The scenes alone matched the degraded objects alone (participants who viewed Bird1 in isolation would also view its background separately, but not embedded). The two sets were counterbalanced across subjects. The animacy pattern localizer used in Studies 1 and 2 included the 60 objects from the main experiment, and an additional set of 60 new objects that were matched for category and subcategory of the main-experiment set, enlarged, in high resolution and centered on a white background. Visual angles were as follows: 6.07×4.55 behavioral, 6.15×4.61 in MRI, 5.99×4.50 in MEG (400×300 pixels on MRI and MEG displays).

Stimulus optimization and selection

The stimulus set was optimized and validated in a behavioral pilot experiment in a laboratory setting, in which participants rated the degraded objects' animacy, to compare the level of object ambiguity in isolation versus in context. Participants were asked to rate each degraded object, presented either within its original scene context or on a white background, on a scale of animacy from 1 (inanimate) to 8 (animate). Participants' ratings were normalized to a mean of 0 and SD of 1. The final stimulus set, used in the fMRI and MEG studies, consisted of objects that were perceived as more ambiguous on their own than in context ($N = 8$; animate: $t = 3.70$, $p = 0.008$; inanimate: $t = 6.42$, $p < 0.001$), with a mean difference of 0.70 between normalized scores of objects with and without context. Fifty-four additional objects, showing smaller differences in ratings (between scene background or white background), were tested in the piloting stages but were excluded from the final stimulus set.

General procedure for Studies 1 and 2

On each trial, participants viewed a single briefly presented (50 ms) stimulus. During main-experiment runs, participants performed an oddball task, in which they pressed a button each time a number was presented instead of an object or scene. They were also instructed to memorize the objects. The main experiment was followed by the animacy pattern localizer. On the first run of the pattern localizer, participants pressed a button each time they recognized a memorized object. On the following runs, they performed the oddball number task.

Study 1: fMRI

Participants. Nineteen healthy participants (12 female, mean \pm SD age, 25 ± 4.4 years) were included. All participants had normal or corrected-to-normal vision and gave informed consent. Sample size was chosen to exceed that of previous studies using similar fMRI decoding methods (e.g., Proklova et al., 2016). Two additional participants were excluded from data analysis due to excessive head motion during scanning.

Experimental design. The main experiment consisted of 5 scanner runs of 352 s duration, each composed of 4 fixation blocks and 3 blocks for each of the 6 conditions: animate/inanimate \times object/scene/object-in-scene (total 18). The animacy pattern localizer consisted of 3 scanner runs of 336 s duration, each composed of 5 fixation blocks and 4 blocks per condition: animate/inanimate \times old/new (total 16). Each block con-

sisted of 16 trials, in which a stimulus was presented for 50 ms followed by a 950 ms fixation. This resulted in 240 trials (120 2 s volumes) per condition in the main experiment, and 192 trials (96 2 s volumes) per condition in the animacy pattern localizer.

Data acquisition and preprocessing. Whole-brain scanning was performed with a 4T Bruker MedSpec MRI scanner using an 8-channel head-coil. T2*-weighted EPIs were collected (TR = 2.0 s, TE = 33 ms, 73° flip angle, 3 × 3 × 3 mm voxel size, 1 mm gap, 30 slices, 192 mm FOV). A high-resolution T1-weighted image (MP-RAGE; 1 × 1 × 1 mm voxel size) was obtained as an anatomical reference. The data were analyzed using MATLAB (The MathWorks; RRID:SCR_001622) with statistical parametric mapping (SPM; RRID:SCR_007037). Each run was preceded by 12 s fixation discarded from the analysis. Preprocessing included slice-timing correction, realignment, and spatial smoothing with a 6 mm FWHM Gaussian kernel. A GLM HRF model was estimated for each participant for the univariate analyses.

Category-selectivity localizer. A category-selectivity localizer ended the scanning session, designed to identify object-selective and scene-selective areas. The localizer included 80 grayscale images per category of objects, scenes, bodies, and scrambled objects (i.e., a random mixture of pixels of each of the object images). It consisted of 2 scanner runs of 336 s duration, each composed of 5 fixation blocks and 4 blocks per condition: object/scene/body/scrambled-object (total 16). Each block consisted of 20 trials, in which a stimulus was presented for 350 ms followed by a 450 ms fixation. Participants performed a 1-back task, in which they pressed a button every time the same image was presented twice in a row. One participant was removed from the ROI analysis due to excessive head motion during the category-selectivity localizer.

Individual ROI selection. Object-selective areas were functionally defined by stronger responses to intact objects than to scrambled objects in the category-selectivity localizer. These were defined for each participant in native space by contrasting activity evoked by intact objects against scrambled objects and against baseline activity. Lateral occipital (LO) and posterior fusiform sulcus (pFs) ROIs were generated bilaterally for each participant by identifying occipital and temporal voxels in the ventral visual stream where both contrasts garnered uncorrected p values < 0.01. Similarly, scene-selective areas were functionally defined by stronger responses to scenes than to objects and scrambled objects. These were defined for each participant in native space by contrasting activity evoked by scenes against objects, against scrambled objects, and against baseline activity. Transverse occipital sulcus (TOS) and parahippocampal place area (PPA) ROIs were generated bilaterally for each participant by identifying occipital and temporal voxels in the ventral visual stream where the three contrasts garnered uncorrected p values < 0.01. In most participants, individually selected retrosplenial complex (RSC) was too small to perform MVPA, and it was therefore excluded from the individual ROI analysis. To limit dimensionality effects on classifier performance and prevent overfitting (Hastie et al., 2001), only the most significant 100 voxels of each ROI were included in the analysis. ROIs < 20 voxels were discarded (1 participant removed from LO, 2 removed from TOS). Decoding accuracy was then averaged across hemispheres.

Multivariate analysis. The data within each voxel were detrended and normalized (mean ± SD) across the time course of each run, and shifted two volumes (4 s) to account for the hemodynamic lag. The data were then averaged across blocks within each run, resulting in one block of 8 volumes per condition per run. Multivariate analysis was performed using CoSMoMVA toolbox (Oosterhof et al., 2016) (RRID:SCR_014519). An LDA classifier discriminated between response patterns to animate versus inanimate objects. The decoding approach is illustrated in Figure 1C. First, decoding of intact object animacy was measured within the animacy pattern localizer, by training on old-object trials (i.e., objects included in the main-experiment set), and testing on new-object trials. Next, cross-decoding was achieved by training on all conditions of the animacy pattern localizer, and testing on each of the main-experiment conditions (object, scene, object-in-scene). For each participant, cross decoding was performed across all voxels of each ROI, resulting in an overall accuracy score for the ROI for each of the three conditions.

Searchlight analysis. The same cross-decoding method described above was applied in a searchlight approach (Kriegeskorte et al., 2006), across

the whole brain of each participant. The searchlight analysis was performed separately for each of the main experiment conditions, in native space, with a radius of 3 voxels, resulting in an accuracy score for each voxel for each of the three conditions. Thereafter, accuracy maps were normalized to MNI space for group significance testing. Supra-additive contextual facilitation was examined by contrasting decoding accuracy for degraded objects in scenes against the sum of accuracies for degraded objects alone and scenes alone. To ensure that the resulting clusters were driven by enhanced decoding of degraded objects in scenes rather than below-chance decoding of objects alone or scenes alone, this contrast was tested in conjunction with three additional contrasts of decoding accuracy for degraded objects in scenes against chance (50%), against degraded objects alone, and against scenes alone. For each of these four contrasts separately, voxelwise significance was tested by computing random-effect cluster statistics corrected for multiple comparisons, using threshold-free cluster enhancement (TFCE, $p < 0.05$) (Smith and Nichols, 2009; Stelzer et al., 2013). The group searchlight ROI is the full conjunction of the four TFCE maps.

Correlation analysis. We tested for a significant correlation across subjects, between the contextual facilitation revealed by the searchlight analysis, and evoked activity in five scene-selective areas: right RSC, bilateral PPA, and bilateral TOS. These ROIs were defined in a group analysis in MNI space by contrasting activity evoked by scenes against objects, against scrambled objects, and against baseline activity. RSC, PPA, and TOS were generated by identifying occipital, temporal, and retrosplenial voxels in the ventral visual stream where all three contrasts garnered uncorrected p values < 0.01 at group level (random effects). The RSC was identified only in the right hemisphere, whereas the TOS and PPA were identified bilaterally. Univariate activity in scene-selective areas was defined by the average across voxelwise T values, contrasting degraded objects in scenes with the mean of degraded objects alone and scenes alone. As a comparable measure, contextual facilitation for this analysis was calculated as the average across voxelwise decoding accuracies for degraded objects in scenes minus the mean of degraded objects alone and scenes alone, as revealed by the searchlight analysis (similar results were obtained with the supra-additive measure of contextual facilitation). Because the main cluster found in the searchlight analysis was in the right occipital-temporal cortex, this region was chosen as the seed for the correlation analysis. Thus, the seed was defined as voxels in the right hemisphere showing significant supra-additive contextual facilitation in the four-way conjunction analysis (see Searchlight analysis). In addition, we tested the same correlation using object-selective voxels in the lateral occipital cortex (LOC) as seed region, composed of LO and pFs bilaterally.

Controlling for multiple comparisons. All significant t tests and correlations reported remained significant when correcting for multiple comparisons within each section of the Results, using false discovery rate (FDR) at a significance level of 0.05.

Study 2: MEG

Participants. Twenty-five healthy participants (10 female, mean ± SD age, 25 ± 3.3 years) were included. All participants had normal or corrected-to-normal vision and gave informed consent. Sample size was chosen to exceed that of previous studies using similar MEG decoding methods (e.g., Carlson et al., 2013). Four additional participants were excluded from data analysis due to excessive signal noise, which incurred removal of many noisy trials and/or channels, leaving insufficient data for classification analysis.

Experimental design. The main experiment consisted of 8 runs of 285 s duration, each composed of 5 fixation breaks (8 s each), 6 oddball trials, and 15 trials per condition: animate/inanimate × object/scene/object-in-scene (96 trials/run). The animacy pattern localizer consisted of 3 runs of 382 s duration, each composed of 7 fixation breaks (8 s each), 8 oddball trials, and 30 trials per condition: animate/inanimate × old/new (128 trials/run). This resulted in 120 trials per condition in the main experiment and 90 trials per condition in the animacy pattern localizer. Each trial began with 500 ms fixation, followed by 50 ms stimulus presentation, and followed by a mean ITI of 2 s ± 500 ms jitter. Trials were randomly intermixed.

Data acquisition and preprocessing. Electromagnetic brain activity was recorded using an Elekta Neuromag 306 MEG system, composed of 204 planar gradiometers and 102 magnetometers. Signals were sampled continuously at 1000 Hz and bandpass filtered online between 0.1 and 330 Hz. Offline preprocessing was done using MATLAB (RRID: SCR_001622) and the FieldTrip analysis package (RRID: SCR_004849). Artifact removal was performed manually by excluding noisy trials and channels. Data were then demeaned, detrended, down-sampled to 100 Hz, and time-locked to stimulus onset. The data were averaged across trials of the same exemplar within each run, resulting in one trial per exemplar per run.

Multivariate analysis. Decoding was performed across posterior magnetometers (48 channels before noise-based exclusion) of each participant, between 0 and 500 ms. Before decoding, temporal smoothing was performed by averaging across neighboring time-points with a radius of 2 (20 ms). Decoding of intact object animacy as well as the cross-decoding analysis followed the same classification method as in Study 1. Decoding was performed for every possible combination of training and testing time-points between 0 and 500 ms, resulting in a 50×50 matrix of 10 ms time-points, for each of the three conditions, per subject. In addition, to generate a measure of same-time cross-decoding, decoding accuracy of each time-point along the diagonal of the matrix was averaged with its neighboring time-points at a radius of 2 (20 ms). Significance was tested on contrasts across the entire time \times time matrix as well as along same-time cross decoding using TFCE ($p < 0.05$).

Study 3: behavior

Object perception was measured in a series of online behavioral studies conducted via Amazon Mechanical Turk. Although the physical experimental settings (e.g., screen parameters, room lighting) in online data collection vary across participants, these settings did not systematically vary across conditions because each participant completed trials of all conditions in random order.

Study 3a: object recognition

Two experiments measured the behavioral recognition of degraded objects, degraded objects in scenes, and scenes alone. The first experiment was conducted with the original stimulus set used in Studies 1 and 2, and the second experiment tested its replicability with a cropped version of these stimuli.

Participants. Thirty-eight participants (19 female, mean \pm SD age, 35 ± 10 years) were included in the initial experiment, and 19 participants (6 female, mean \pm SD age, 37 ± 10 years), who did not participate in the initial experiment, participated in the replication. All participants had normal or corrected-to-normal vision and gave informed consent. Initial sample size was chosen to exceed that of previous studies investigating similar behavioral effects (e.g., Munneke et al., 2013). Two additional participants were excluded from data analysis due to chance-level performance throughout the task. One additional participant was excluded from the analysis of the replication due to insufficient valid responses.

Stimuli. In the replication experiment, degraded objects alone were cropped to exclude redundant pixels, including shadows and reflections of the original stimuli.

Procedure. On each trial, participants typed in the object they perceived in the briefly presented image. They were instructed to guess the appropriate object for empty scenes. The stimulus subset was presented once per subject, in random order, resulting in 90 test trials. Eight practice trials preceded the test trials. In each trial, a stimulus was presented for 50 ms, after which a blank screen with an open question field appeared until response.

Response quantification. Behavioral data were quantified by coding open answers into nine object categories: bird, fish, 4-legged mammal, human, road vehicle, train, watercraft, aircraft, and other. In addition to this 9-category coding scheme, fine-grain recognition was measured by coding open answers into 18 object categories: duck (or swan, goose), bird, whale, dog, sheep (or goat, lamb), squirrel (or chipmunk), deer (or elk, moose), dog, cat (or kitten), horse (or pony, donkey), calf (or cow, bull, buffalo), rabbit, human, boat (but not raft, barge), plane (but not

helicopter), tractor (or truck), car (but not bus, carriage), train, and other.

Study 3b: object detection

The last online experiment measured the rapid detection of animacy (yes/no) in degraded and intact objects, with and without scenes.

Participants. Forty participants (14 female, mean \pm SD age, 36 ± 10 years) were included in the experiment. Half of the participants were presented with degraded objects, and the other half were presented with intact objects. All participants had normal or corrected-to-normal vision and gave informed consent.

Stimuli. Degraded-object stimuli included the degraded objects and degraded objects in scenes used in Studies 1, 2, and 3a. Intact-object stimuli were the same objects and objects in scenes without object degradation, such that object resolution was as high as scene resolution.

Procedure. On each trial, participants pressed 1 if they detected an animate object in the image or 0 if they did not. They were instructed to respond as fast as possible without compromising accuracy. The stimulus subset was presented 3 times per subject, in random order, resulting in 180 test trials. Two practice blocks, 4 trials each, preceded the test trials. Timing feedback was given on the second practice block and on the test trials. In each trial, a stimulus was presented for 50 ms, after which a blank screen appeared until response. Following responses longer than 1 s, participants were presented with the feedback “Too slow” on the screen before proceeding to the next trial.

Results

Study 1: fMRI

Decoding intact object animacy in object-selective cortex

In a first analysis of the fMRI data, we assessed the representation of animacy in object-selective cortex for the animacy pattern localizer data. Results showed that object animacy was strongly represented in two object-selective areas: the LO area (decoding accuracy mean 81.19%; against chance: $t_{(16)} = 11.02$, $p < 0.001$, $d = 2.67$) and the pFs area (mean 79.75%; $t_{(17)} = 12.16$, $p < 0.001$, $d = 2.87$). These results replicate previous findings of animacy decoding in visual cortex (Kriegeskorte et al., 2008).

Scene-based object facilitation in object-selective cortex

Next, we examined the representation of object animacy in object-selective areas for each of the three main experiment conditions. Classifiers were trained on data from the animacy pattern localizer and tested on the conditions in the main experiment using a cross-decoding approach (Fig. 1C). Animacy of degraded objects presented within scene context could be reliably decoded in both LO and pFs (against chance; LO, $t_{(16)} = 6.43$, $p < 0.001$, $d = 1.56$; pFs, $t_{(17)} = 6.14$, $p < 0.001$, $d = 1.45$; Figure 2). Most importantly, decoding accuracy for degraded objects in scenes was higher than the sum of accuracies (minus chance: 50%) for degraded objects alone and scenes alone (LO, $t_{(16)} = 2.99$, $p = 0.008$, $d = 0.72$; pFs, $t_{(17)} = 2.62$, $p = 0.018$, $d = 0.62$), demonstrating supra-additive contextual facilitation. Moreover, the presence of the scene strongly boosted the decoding of degraded objects in scenes relative to when these same objects were shown in isolation (LO, $t_{(16)} = 3.25$, $p = 0.005$, $d = 0.79$; pFs, $t_{(17)} = 2.96$, $p = 0.009$, $d = 0.70$), even though activity patterns evoked by the scenes themselves did not carry any information about object animacy (against chance; LO, $t_{(16)} = 0.75$, $p = 0.465$, $d = 0.18$; pFs, $t_{(17)} = 0.07$, $p = 0.946$, $d = 0.02$).

In addition, as an alternative method to test for supra-additive contextual facilitation, we summed the BOLD responses for degraded objects alone and scenes alone before decoding (MacEvoy and Epstein, 2011), then tested the difference in decoding accuracy between degraded objects in scenes and the combined signal for degraded objects alone and scenes alone. All other steps were the same as in the analysis reported above. Using this approach

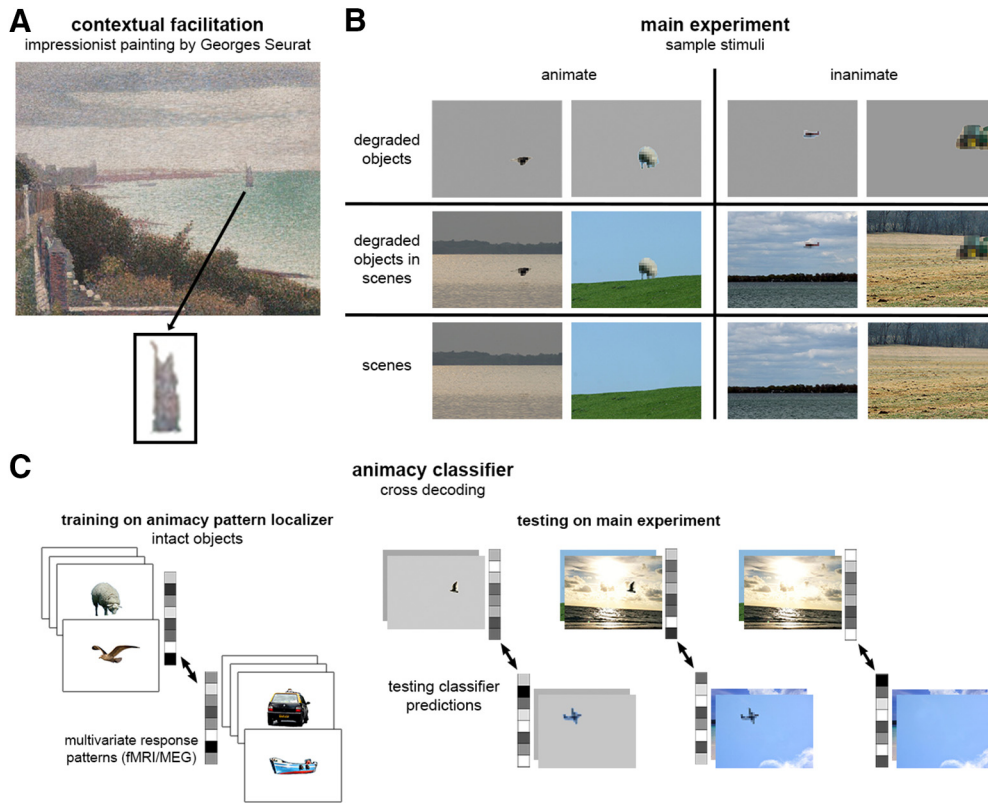


Figure 1. Measuring the representation of contextually defined objects. **A**, Context-based object perception is demonstrated in this painting by Georges Seurat, *Grandcamp, Evening* (1885), in which the boat is a coarse shape (enlarged underneath) gaining its meaning from scene context. **B**, Sample stimuli: animate and inanimate degraded objects, degraded objects in scenes, and scenes alone. **C**, Cross-decoding object animacy: classifier trained on the multivoxel (fMRI) or multisensor (MEG) response pattern to animate and inanimate intact objects, and then tested on animate and inanimate degraded objects, degraded objects in scenes and scenes alone.

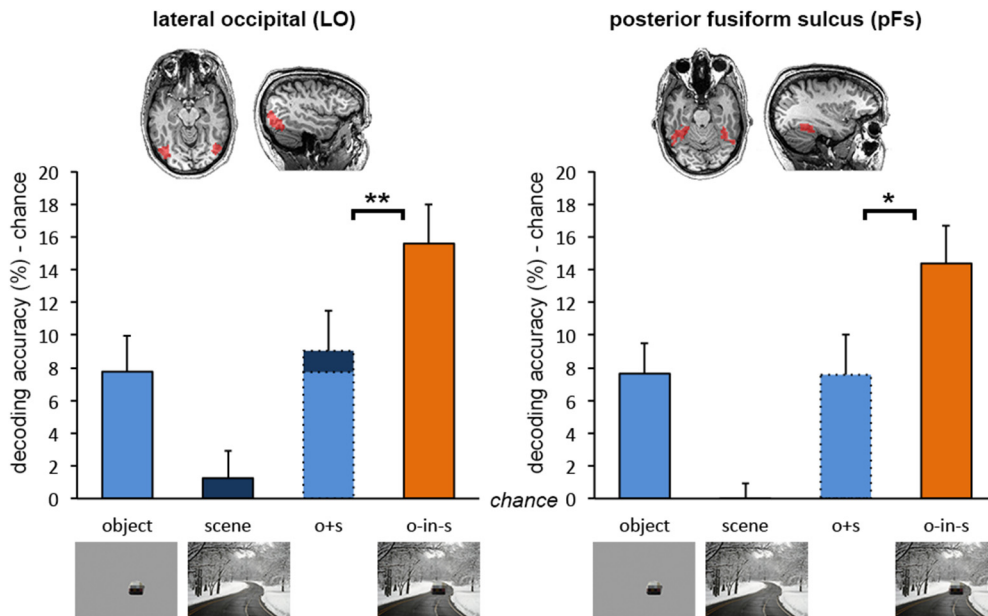


Figure 2. fMRI cross-decoding of object animacy in object-selective cortex. Analysis of object-selective areas revealed supra-additive contextual facilitation of object-animacy representation in both the LO and pFs. Data are represented as mean distance from chance (50% decoding accuracy) + SEM. * $p < 0.05$. ** $p < 0.01$.

too, supra-additive contextual facilitation was significant (LO, $t_{(16)} = 5.60, p < 0.001, d = 1.36$; pFs, $t_{(17)} = 4.09, p < 0.001, d = 0.96$).

Finally, to test whether contextual facilitation found in object-selective areas was related to differences in overall activation in

these regions, we examined animate and inanimate univariate BOLD responses within each area. Data were processed similarly as for multivariate analysis, excluding the normalization step. Results showed no differences between animate and inanimate degraded objects in scenes (LO, $t_{(16)} = 1.08, p = 0.324, d = 0.08$;

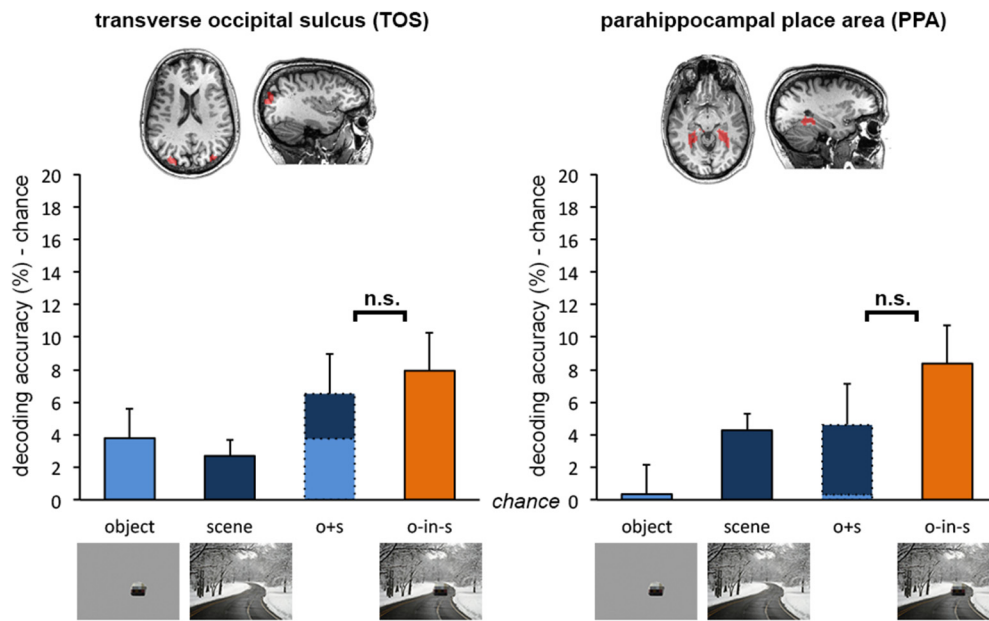


Figure 3. fMRI cross-decoding of object animacy in scene-selective cortex. Analysis of scene-selective areas revealed additive contextual facilitation of object-animacy representation in both the TOS and PPA. Data are represented as mean distance from chance (50% decoding accuracy) + SEM.

pFs, $t_{(17)} = 1.16$, $p = 0.262$, $d = 0.06$) or degraded objects alone (LO, $t_{(16)} = 1.65$, $p = 0.119$, $d = 0.03$; pFs, $t_{(17)} = 1.40$, $p = 0.179$, $d = 0.04$). Thus, multivariate representations of object animacy as measured here cannot be explained by regional response-magnitude differences between animate and inanimate objects.

Scene-based object facilitation in scene-selective cortex

Given previous reports of object decoding in scene-selective areas (e.g., Harel et al., 2013), we extended the same cross-decoding approach to TOS and PPA. Animacy of degraded objects in scenes could be reliably decoded in both TOS and PPA (against chance; TOS, $t_{(15)} = 3.80$, $p = 0.002$, $d = 0.95$; PPA, $t_{(17)} = 5.71$, $p < 0.001$, $d = 1.34$; Figure 3), in line with previous reports. However, in scene-selective areas, in contrast to object-selective areas, decoding accuracy for degraded objects in scenes was not significantly different from the sum of accuracies (minus chance) for degraded objects alone and scenes alone (TOS, $t_{(15)} = 0.49$, $p = 0.630$, $d = 0.12$; PPA, $t_{(17)} = 1.99$, $p = 0.062$, $d = 0.47$), demonstrating additive facilitation in these areas. This was confirmed by a significant interaction ($F_{(2,28)} = 11.48$, $p < 0.001$, $\eta_p^2 = 0.45$) between regions (object-selective, scene-selective) and context (degraded object in scene, degraded object, scene).

Scene-based object facilitation in the whole brain

To test for contextual facilitation outside object- and scene-selective areas, we performed the cross-decoding analysis (Fig. 1C) using a searchlight approach, in which contextual facilitation was defined by a supra-additive boost in decoding accuracy for degraded objects when presented within scene context. Voxels that reached significance (TFCE, $p < 0.05$) in a stringent conjunction test (see Materials and Methods) were found in regions of the extrastriate visual cortex (Fig. 4A). These clusters partly overlapped with object-selective cortex but not scene-selective cortex (Fig. 4B).

The role of scene-selective areas in scene-based object facilitation

In a final analysis, we tested whether the contextual facilitation observed in the extrastriate visual cortex was related to concurrent processing in scene-selective areas (Fig. 5), as previously

proposed (Bar et al., 2001; Kveraga et al., 2011; Aminoff et al., 2013). First, we tested for a correlation across subjects, between the contextual facilitation in the searchlight region and evoked activity in five scene-selective areas: right RSC, bilateral PPA, and bilateral TOS. Correlation was most significant for the right RSC ($r_{(15)} = 0.67$, $p = 0.002$), and also significant for the right and left PPA and left TOS ($r_{(15)} > 0.49$, $p < 0.038$, for all tests), but not for the right TOS ($r_{(15)} = 0.18$, $p = 0.474$). Next, we tested for a correlation between contextual facilitation in the object-selective LOC and evoked activity in the same five scene-selective areas. This similarly revealed significant correlations for the right RSC, right and left PPA, and left TOS ($r_{(15)} > 0.70$, $p < 0.002$, for all tests). These results indicate that increased activity in scene-selective areas was associated with increased contextual facilitation of objects in extrastriate visual cortex and particularly in object-selective areas, providing evidence for functional interactions between scene and object processing.

Study 2: MEG

To characterize the time course of scene-based object facilitation, we used the same experimental approach during MEG recording.

Decoding intact object animacy across time

In a first analysis, we assessed the temporal dynamics of animacy decoding of intact objects within the animacy pattern localizer. Classifiers were trained and tested on 10 ms time intervals from 0 to 500 ms relative to stimulus onset, resulting in a 50×50 time \times time matrix of decoding accuracy. Results showed significant decoding of object animacy along the diagonal of the matrix (same training and testing time-points), peaking between 180 and 200 ms after stimulus onset (Fig. 6). These results replicate recent findings of animacy decoding using MEG (Carlson et al., 2013; Cichy et al., 2014).

Scene-based object facilitation across time

Next, we examined the time course of decoding of object animacy in the main experiment, using the same cross-decoding approach as in the fMRI study (Fig. 1C). Contextual facilitation was defined

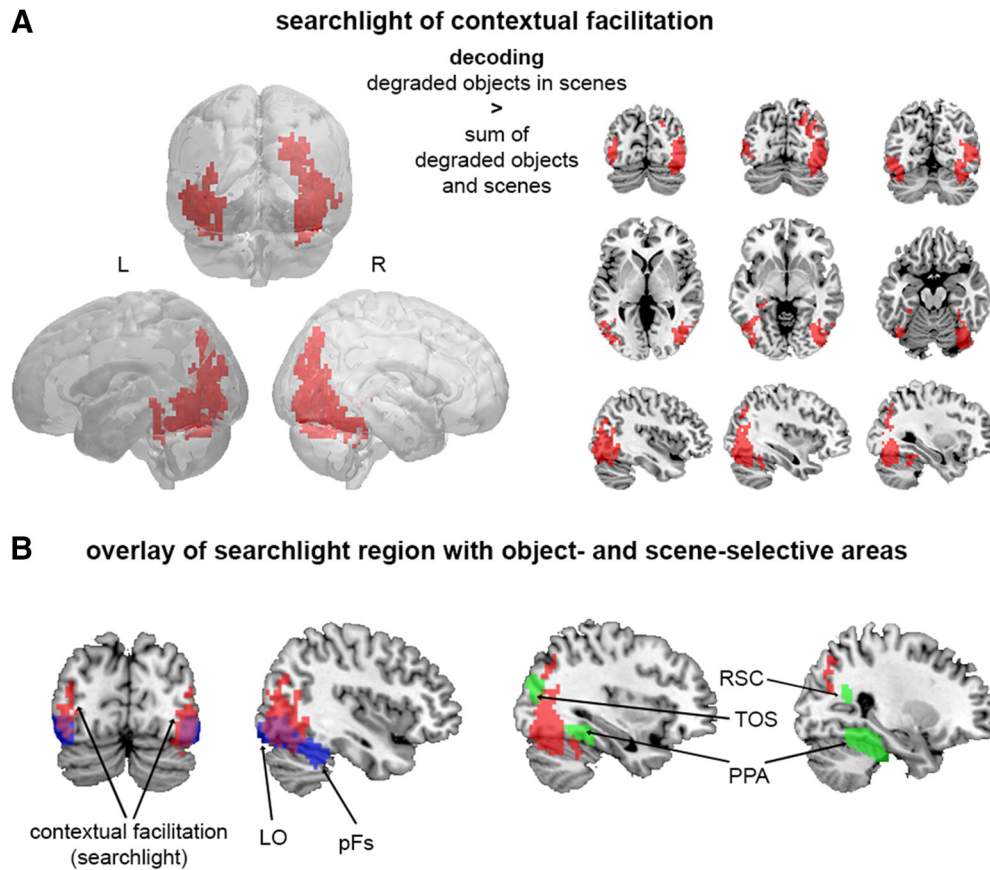


Figure 4. fMRI cross-decoding of object animacy across the whole brain. **A**, Searchlight analysis revealed supra-additive contextual facilitation of object-animacy representation in a widespread region of the extrastriate visual cortex (TFCE, $p < 0.05$). **B**, The searchlight region (red) partially overlapped object-selective areas LO and pFs (blue), but not scene-selective areas TOS, PPA, and RSC (green).

by a supra-additive boost in decoding accuracy for degraded objects when presented within scene context. Time-points that reached significance (TFCE, $p < 0.05$) in a stringent conjunction test (see Materials and Methods) were found at multiple time points between 320 and 340 ms after stimulus onset (see black outlines and asterisks in Fig. 7). Similar to the fMRI results (Fig. 2), activity patterns evoked by scenes alone did not carry any information about object animacy (TFCE, $p > 0.05$ across all time-points), yet having the scene as background to the degraded objects strongly boosted decoding relative to degraded objects alone.

Study 3: behavior

Study 3a: object recognition

The fMRI and MEG results show that neural activity evoked by the degraded objects was better categorized when the objects were presented within scenes. To test whether adding scene context also leads to better visual recognition of degraded objects, we examined the perceptual effect of these stimuli in behavioral experiments.

In an online behavioral task, we examined recognition of degraded objects with and without context, as well as expectations about objects induced by the scene alone. Results revealed strong object recognition for degraded objects in scenes, which was significantly better than the summed accuracy for degraded objects alone and scenes alone ($t_{(37)} = 3.28$, $p = 0.002$, $d = 0.53$), calculated as the distance from chance by subtracting 11.11% (100% divided by 9 object categories) (Fig. 8A). A similar effect was

found for fine-grained recognition using the alternative quantification scheme with 18 object categories ($t_{(37)} = 2.91$, $p = 0.006$, $d = 0.47$). These results show that scene context explicitly facilitated the visual recognition of degraded objects in our experimental stimuli. These results were replicated ($t_{(18)} = 2.33$, $p = 0.031$, $d = 0.52$) in a second experiment using cropped versions of the degraded objects.

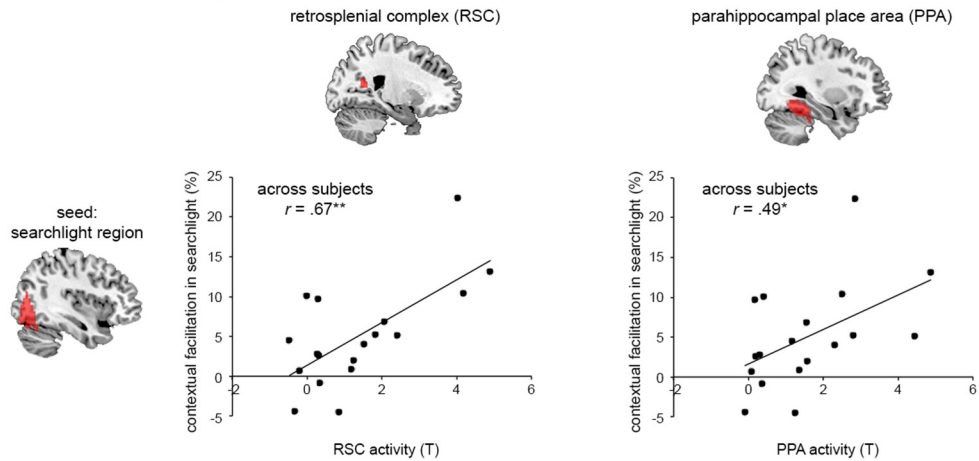
Study 3b: object detection

Results of Study 3a show that the perception of degraded objects is facilitated by scene context. To test whether the facilitating effect of scenes is unique to ambiguous objects, we compared their effect on degraded objects versus intact objects. Because intact objects were expected to be easily recognized in the paradigm used for Study 3a, we instead used a rapid detection task with degraded and intact objects, with and without scenes.

Results revealed better animacy-detection accuracy for objects in scenes than for objects alone, for both degraded objects ($t_{(19)} = 5.25$, $p < 0.001$, $d = 1.17$) and intact objects ($t_{(19)} = 3.42$, $p = 0.003$, $d = 0.76$). However, contextual facilitation, measured by the difference between objects in scenes and objects alone, was much stronger for degraded objects than for intact objects ($t_{(38)} = 3.01$, $p = 0.004$, $d = 0.96$; Fig. 8B). Similar effects were found for sensitivity (d' ; $t > 2.5$, $p < 0.02$, for all tests).

The reduced effect of facilitation for intact objects relative to degraded objects resembles accounts of inverse effectiveness observed in multisensory integration (Stein and Stanford, 2008). Thus, to better examine the relationship between input redun-

A correlating decoding in searchlight region with activity in scene-selective areas



B correlating decoding in object-selective areas with activity in scene-selective areas

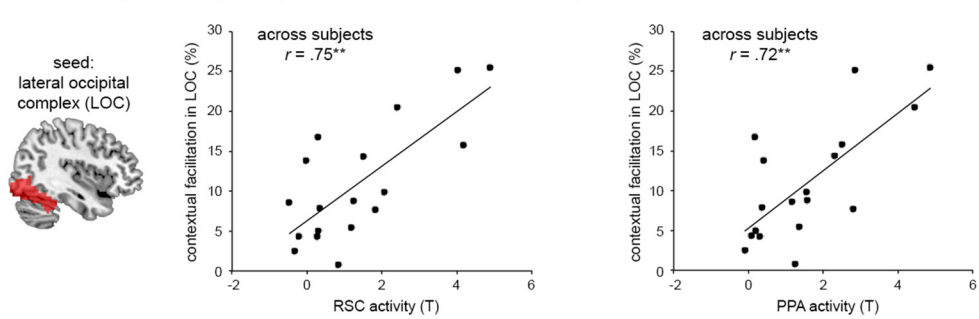


Figure 5. fMRI correlation of contextual facilitation with scene-selective activity. Scatterplots present correlations across subjects between multivariate contextual facilitation in seed region and context-dependent univariate activity in scene-selective RSC and PPA (right hemisphere). **A**, Searchlight region as seed region. **B**, Object-selective LOC as seed region. * $p < 0.05$. ** $p < 0.01$.

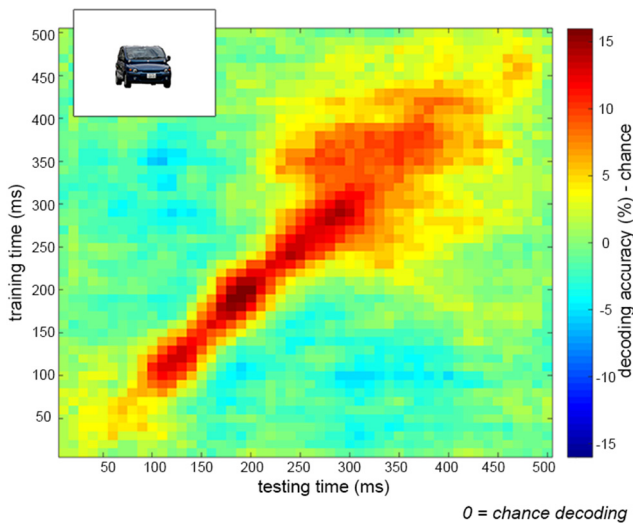


Figure 6. MEG decoding of intact object animacy. Decoding accuracy across a time \times time space from stimulus onset to 500 ms, averaged across subjects. Object animacy was successfully decoded from intact objects within the animacy pattern localizer. Data are represented as mean distance from chance (50% decoding accuracy).

dancy and integration effectiveness, we tested the correlation between object-alone ambiguity found in Study 3b and contextual facilitation found in Study 3a. Ambiguity was defined as $[1 - \text{accuracy}]$ per object categorization viewed in isolation. We then correlated these ambiguity scores with the supra-additive measure of contextual facilitation in the original behav-

ioral recognition task, across object exemplars. The result was a significant correlation between contextual facilitation and intact-object ambiguity ($r_{(28)} = 0.35, p = 0.006$) and a weak positive correlation between contextual facilitation and degraded-object ambiguity ($r_{(28)} = 0.21, p = 0.100$). In addition, the two ambiguity scores were highly correlated ($r_{(28)} = 0.58, p < 0.001$). These results suggest that the more ambiguous the object, the more effective the facilitation.

Discussion

In the current study, we investigated how scene and object processing interact to support context-based perception. We found that scene context facilitated neural representation of degraded objects in fMRI and MEG, as well as their visual recognition and detection, in a supra-additive manner; scenes facilitated the representation of degraded objects even when they did not evoke such representations on their own. Furthermore, contextual facilitation in visual cortex and particularly in object-selective areas was correlated with scene-selective activation, revealing a clear interaction between scenes and objects in their neural mechanisms of visual processing. These findings suggest that scene information facilitates the processing of objects at a perceptual stage (i.e., within the visual cortex), thereby providing direct evidence for interactive views of scene and object perception (Biederman et al., 1982; Bar and Ullman, 1996; Bar, 2004; Davenport and Potter, 2004). In the following paragraphs, we use the current findings to characterize, in space and time, the process of scene-based object facilitation.

The fMRI results showed scene-based facilitation in object-selective areas LO and pFs as well as in extended regions of the

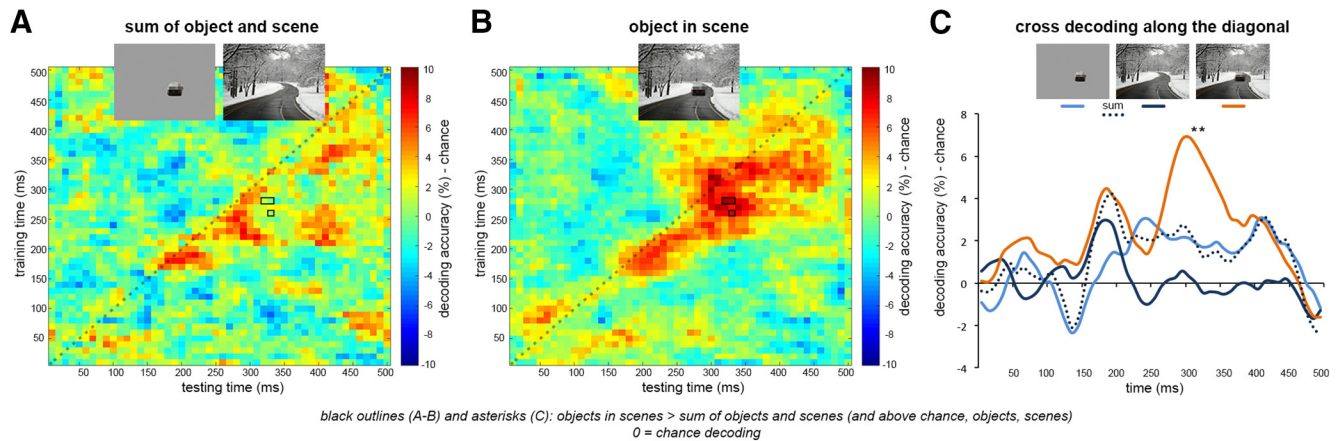


Figure 7. MEG cross-decoding of object animacy. Matrices represent the decoding accuracy across a time \times time space from stimulus onset to 500 ms, averaged across subjects. **A**, Summed accuracies for degraded objects alone and scenes alone. **B**, Cross-decoding accuracy for degraded objects in scenes. **C**, Cross-decoding accuracy along the smoothed time diagonal (matched training and testing times). Data are represented as mean distance from chance (50% decoding accuracy). *TFCE, $p < 0.05$.

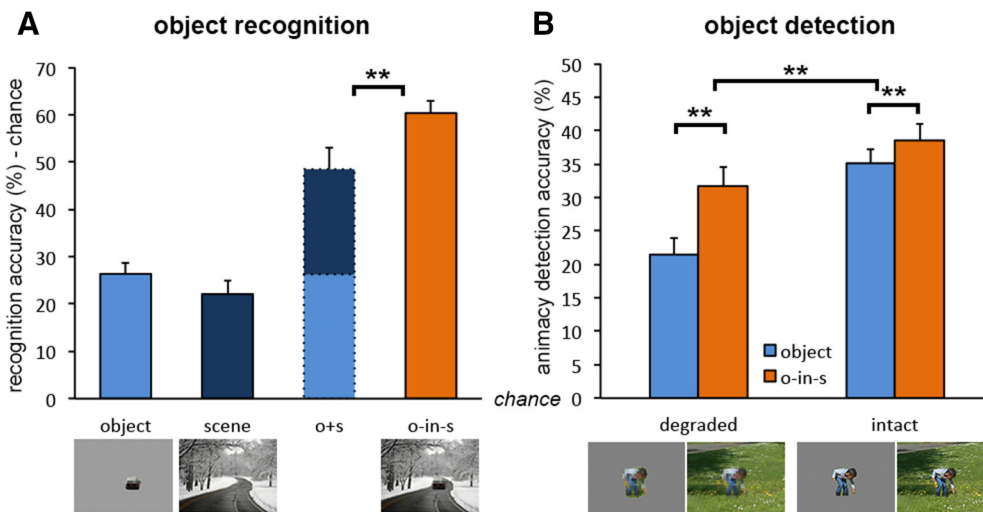


Figure 8. Visual object perception. **A**, Object recognition: Category coding of open answers revealed supra-additive contextual facilitation of object recognition. Data are represented as mean distance from chance (11.11% recognition accuracy) + SEM. **B**, Object-animacy detection: Scenes facilitated performance for degraded objects significantly more than for intact objects. Data are represented as mean distance from chance (50% animacy detection accuracy) + SEM. ** $p < 0.01$.

extrastriate visual cortex. These results indicate that object representations in visual cortex can fully rely on low-spatial-frequency object information (i.e., degraded object resolution) when facilitated by scene context, even in the absence of contextually associated objects. Importantly, the degree to which scenes facilitated object representation was correlated with concurrent activity in scene-selective areas, mostly in the RSC and PPA. This finding provides novel evidence that scene-selective areas are associated with scene-based facilitation of object representations, extending previous work on the role of RSC and PPA in object–object interactions (Bar, 2004).

Using MEG, we characterized the time course of scene-based object perception, showing supra-additive contextual facilitation starting from 320 ms after stimulus onset, >100 ms later than peak category decoding of intact objects (Carlson et al., 2013; Cichy et al., 2014) and initial category decoding of intact objects embedded in scenes (Kaiser et al., 2016). This delayed timing is unlikely to reflect feedforward effects (e.g., Johnson and Olshausen, 2003; Joubert et al., 2007). Rather, given our fMRI results, we propose it reflects a feedback loop within visual processing. Relatedly, recent findings show that high-level scene rep-

resentations emerge at ~ 250 ms after stimulus onset (Cichy et al., 2016). Together with the correlation with scene-selective areas observed in the fMRI study, this indicates a longer processing route for scene-defined objects relative to intact objects, in which scene cues are first processed along the scene pathway, then relayed back to the extrastriate visual cortex to facilitate object processing.

The present findings show that, when intrinsic object information is insufficient, extrinsic cues play an important role in shaping the representation of objects. These modulatory influences have been observed in studies investigating interactions between nearby representations of frequently co-occurring objects (Cox et al., 2004; Brandman and Yovel, 2010). However, whereas Cox et al. (2004) proposed that object representations themselves contain some embodiment of likely contexts, the current results suggest that contextual modulation of object representation can additionally come from a separate processing pathway. Particularly, the current findings of a temporal delay and a correlation between contextual facilitation and scene-selective activity point toward the interaction between separate object and scene processing pathways, rather than the direct embodiment of context within bottom-up object representations.

Altogether our results provide support for matching models of contextual processing, which postulate that contextual cues reduce the amount of perceptual evidence needed to match an object with its unique representation, thereby facilitating identification (Bar and Ullman, 1996; Bar and Aminoff, 2003; Bar, 2004; Mudrik et al., 2014). This framework is in line with predictive coding models (Srinivasan et al., 1982), by which feedback connections carry predictions of activation whereas bottom-up connections convey the residual errors in prediction (Rao and Ballard, 1999; Huang and Rao, 2011). Here, the scene may generate contextual expectations, which evoke predicted object templates that are subsequently compared with bottom-up input given by the degraded object. More generally, this may reflect Hebbian learning (Hebb, 1949): stimuli that frequently co-occur in the environment coactivate connected neural representations. Specifically, the degraded object and the scene context may both induce subthreshold activation of multiple plausible representations in the visual cortex. The intersection between scene- and object-triggered activations is then a unique neural circuit that is activated by both, resulting in an overall suprathreshold activation of that representation.

Along these lines, if contextual integration is the convergence of subthreshold feedforward and feedback activations, then it will become redundant for feedforward inputs that exceed the threshold on their own. By this logic, the supra-additive effects found in our data may depend on the insufficiency of ambiguous objects in activating bottom-up representations. In line with this notion, we found that the facilitating effect of the scene on object detection was reduced for intact objects, and that object ambiguity was correlated with the effectiveness of contextual facilitation. In other words, when feedforward object input was informative on its own, the scene contributed less to its perception. This inverse relationship, between redundancy of integrated information and the magnitude of interactive effect, corresponds to the principle of inverse effectiveness characterized in the field of multisensory integration (Stein and Stanford, 2008). We therefore propose that a similar principle may apply for scene-object integration.

Finally, we lay out a few open questions inspired by the current findings. First, we must consider that scene-object integration examined here is unidirectional, in that it tests the effects of scenes on object encoding, but not the effects of objects on scene encoding. Here, scene-selective areas did not exhibit supra-additive facilitation of object encoding, and are therefore less likely to be the locus of integration. However, this does not rule out convergence of scene and object information within scene selective regions, and should be examined also for scene encoding, when objects are the facilitator. Second, the current study targeted mid-to-high-level visual object processing, by generalizing from object-category representations evoked by large intact isolated objects to responses evoked by degraded objects in scenes. This approach precludes decoding driven by low-level visual features but does not rule out semantically driven decoding of object category. It could therefore be asked whether semantic representations of object category contribute to contextual facilitation emerging in visual cortex, and whether it would generalize to nonvisual semantic cues. Last, we point out that there may be other extra-visual factors contributing to the visual integration of scenes and objects, such as attention and task relevance.

In conclusion, the current study provides comprehensive insight into the neural basis of scene-based object perception. By demonstrating supra-additive contextual facilitation in the visual cortex, our data provide evidence for the interaction of object and scene processing at a perceptual stage, likely via feedback pro-

cesses converging at ~320–340 ms after scene onset. Together with correlated activity in scene-selective areas, these findings characterize functional interactions between scene and object neural pathways. These interactions likely play an important role in supporting efficient natural vision by facilitating object recognition in real-world scenes.

References

- Aminoff EM, Kveraga K, Bar M (2013) The role of the parahippocampal cortex in cognition. *Trends Cogn Sci* 17:379–390. [CrossRef Medline](#)
- Bar M (2004) Visual objects in context. *Nat Rev Neurosci* 5:617–629. [CrossRef Medline](#)
- Bar M, Aminoff E (2003) Cortical analysis of visual context. *Neuron* 38:347–358. [CrossRef Medline](#)
- Bar M, Ullman S (1996) Spatial context in recognition. *Perception* 25:343–352. [CrossRef Medline](#)
- Bar M, Tootell RB, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, Dale AM (2001) Cortical mechanisms specific to explicit visual object recognition. *Neuron* 29:529–535. [CrossRef Medline](#)
- Biederman I, Mezzanotte RJ, Rabinowitz JC (1982) Scene perception: detecting and judging objects undergoing relational violations. *Cogn Psychol* 14:143–177. [CrossRef Medline](#)
- Boyce SJ, Pollatsek A (1992) Identification of objects in scenes: the role of scene background in object naming. *J Exp Psychol Learn Mem Cogn* 18:531–543. [CrossRef Medline](#)
- Brandman T, Yovel G (2010) The body inversion effect is mediated by face-selective, not body-selective, mechanisms. *J Neurosci* 30:10534–10540. [CrossRef Medline](#)
- Carlson T, Tovar DA, Alink A, Kriegeskorte N (2013) Representational dynamics of object vision: the first 1000 ms. *J Vis* 13:1. [CrossRef Medline](#)
- Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. *Nat Neurosci* 17:455–462. [CrossRef Medline](#)
- Cichy RM, Khosla A, Pantazis D, Oliva A (2017) Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *Neuroimage* 153:346–358. [CrossRef Medline](#)
- Cox D, Meyers E, Sinha P (2004) Contextually evoked object-specific responses in human visual cortex. *Science* 304:115–117. [CrossRef Medline](#)
- Davenport JL, Potter MC (2004) Scene consistency in object and background perception. *Psychol Sci* 15:559–564. [CrossRef Medline](#)
- Demiral SB, Malcolm GL, Henderson JM (2012) ERP correlates of spatially incongruent object identification during scene viewing: contextual expectancy versus simultaneous processing. *Neuropsychologia* 50:1271–1285. [CrossRef Medline](#)
- Dilks DD, Julian JB, Paunov AM, Kanwisher N (2013) The occipital place area is causally and selectively involved in scene perception. *J Neurosci* 33:1331–1336a. [CrossRef Medline](#)
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601. [CrossRef Medline](#)
- Ganis G, Kutas M (2003) An electrophysiological study of scene effects on object identification. *Brain Res Cogn Brain Res* 16:123–144. [CrossRef Medline](#)
- Harel A, Kravitz DJ, Baker CI (2013) Deconstructing visual scenes in cortex: gradients of object and spatial layout information. *Cereb Cortex* 23:947–957. [CrossRef Medline](#)
- Hastie T, Tibshirani R, Friedman J (2001) *The elements of statistical learning*. New York: Springer.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430. [CrossRef Medline](#)
- Hebb DO (1949) *The organization of behavior: a neuropsychological theory*. New York: Wiley.
- Hollingworth A, Henderson JM (1999) Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. *Acta Psychol (Amst)* 102:319–343. [CrossRef Medline](#)
- Huang Y, Rao RP (2011) Predictive coding. *Wiley Interdiscip Rev Cogn Sci* 2:580–593. [CrossRef Medline](#)
- Johnson JS, Olshausen BA (2003) Timecourse of neural signatures of object recognition. *J Vis* 3:499–512. [CrossRef Medline](#)
- Joubert OR, Rousselet GA, Fize D, Fabre-Thorpe M (2007) Processing scene context: fast categorization and object interference. *Vision Res* 47:3286–3297. [CrossRef Medline](#)
- Kaiser D, Oosterhof NN, Peelen MV (2016) The neural dynamics of atten-

- tional selection in natural scenes. *J Neurosci* 36:10522–10528. [CrossRef](#) [Medline](#)
- Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103:3863–3868. [CrossRef](#) [Medline](#)
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60:1126–1141. [CrossRef](#) [Medline](#)
- Kveraga K, Ghuman AS, Kassam KS, Aminoff EA, Hämäläinen MS, Chamon M, Bar M (2011) Early onset of neural synchronization in the contextual associations network. *Proc Natl Acad Sci U S A* 108:3389–3394. [CrossRef](#) [Medline](#)
- MacEvoy SP, Epstein RA (2011) Constructing scenes from objects in human occipitotemporal cortex. *Nat Neurosci* 14:1323–1329. [CrossRef](#) [Medline](#)
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135–8139. [CrossRef](#) [Medline](#)
- Mudrik L, Shalgi S, Lamy D, Deouell LY (2014) Synchronous contextual irregularities affect early scene processing: replication and extension. *Neuropsychologia* 56:447–458. [CrossRef](#) [Medline](#)
- Mullin CR, Steeves JK (2011) TMS to the lateral occipital cortex disrupts object processing but facilitates scene processing. *J Cogn Neurosci* 23:4174–4184. [CrossRef](#) [Medline](#)
- Munneke J, Brentari V, Peelen MV (2013) The influence of scene context on object recognition is independent of attentional focus. *Front Psychol* 4:552. [CrossRef](#) [Medline](#)
- Noppeney U (2012) Characterization of multisensory integration with fMRI: experimental design, statistical analysis, and interpretation. In: *The neural bases of multisensory processes* (Murray MM, Wallace MT, eds). Boca Raton, FL: CRC Press.
- Oliva A, Torralba A (2007) The role of context in object recognition. *Trends Cogn Sci* 11:520–527. [CrossRef](#) [Medline](#)
- Oosterhof NN, Connolly AC, Haxby JV (2016) CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU octave. *Front Neuroinform* 10:27. [CrossRef](#) [Medline](#)
- Proklova D, Kaiser D, Peelen MV (2016) Disentangling representations of object shape and object category in human visual cortex: the animate-inanimate distinction. *J Cogn Neurosci* 28:680–692. [CrossRef](#) [Medline](#)
- Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87. [CrossRef](#) [Medline](#)
- Rémy F, Vayssière N, Pins D, Boucart M, Fabre-Thorpe M (2014) Incongruent object/context relationships in visual scenes: where are they processed in the brain? *Brain Cogn* 84:34–43. [CrossRef](#) [Medline](#)
- Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44:83–98. [CrossRef](#) [Medline](#)
- Srinivasan MV, Laughlin SB, Dubs A (1982) Predictive coding: a fresh view of inhibition in the retina. *Proc R Soc Lond B Biol Sci* 216:427–459. [CrossRef](#) [Medline](#)
- Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci* 9:255–266. [CrossRef](#) [Medline](#)
- Stelzer J, Chen Y, Turner R (2013) Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage* 65:69–82. [CrossRef](#) [Medline](#)
- Summerfield C, de Lange FP (2014) Expectation in perceptual decision making: neural and computational mechanisms. *Nat Rev Neurosci* 15:745–756. [CrossRef](#) [Medline](#)
- Võ ML, Wolfe JM (2013) Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychol Sci* 24:1816–1823. [CrossRef](#) [Medline](#)