Encoding-Stage Crosstalk Between Object- and Spatial Property-Based Scene Processing Pathways

Drew Linsley and Sean P. MacEvoy

Department of Psychology, Boston College, Chestnut Hill, MA 02467, USA

Address correspondence to Drew Linsley, Department of Psychology, Boston College, 300 McGuinn Hall, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA. Email: linsleyd@bc.edu

Scene categorization draws on 2 information sources: The identities of objects scenes contain and scenes’ intrinsic spatial properties. Because these resources are formally independent, it is possible for them to lead to conflicting judgments of scene category. We tested the hypothesis that the potential for such conflicts is mitigated by a system of “crosstalk” between object- and spatial layout-processing pathways, under which the encoded spatial properties of scenes are biased by scenes’ object contents. Specifically, we show that the presence of objects strongly associated with a given scene category can bias the encoded spatial properties of scenes containing them toward the average of that category, an effect which is evident both in behavioral measures of scenes’ perceived spatial properties and in scene-evoked multivoxel patterns recorded with functional magnetic resonance imaging from the parahippocampal place area (PPA), a region associated with the processing of scenes’ spatial properties. These results indicate that harmonization of object- and spatial property-based estimates of scene identity begins when spatial properties are encoded, and that the PPA plays a central role in this process.

Keywords: fMRI, multivoxel pattern analyses, parahippocampal place area, scene perception, spatial layout

Introduction

Fast and accurate scene recognition is critical to daily life, allowing us to navigate from one place to another and to interact efficiently and appropriately with the environment at each point along the way. Objects have often been cast as the fundamental building blocks of scenes, with scene recognition proposed to emerge from a cataloging of the types of objects in a scene and the spatial relationships among them (Biederman 1972, 1987; Friedman 1979; Biederman et al. 1982; De Graef et al. 1990). Consistent with this view, behavioral studies have shown that scene recognition falters when highly informative objects (e.g., refrigerators or toilets) are removed from their associated scenes (MacEvoy and Epstein 2011) or when scenes contain incongruent objects (Davenport and Potter 2004; Joubert et al. 2007). More recently, however, theoretical and behavioral studies of scene recognition have demonstrated a substantial role for scenes’ intrinsic global properties, particularly spatial properties such as depth, openness, and navigability, complementing the category cues provided by objects (Schyns and Oliva 1994; Oliva and Schyns 2000; Oliva and Torralba 2001, 2006; Renninger and Malik 2004; McCotter et al. 2005; Fei-Fei et al. 2007; Vogel and Schiele 2007; Greene and Oliva 2009).

From a physical perspective, the kinds of objects scenes contain and scenes’ spatial properties are often unrelated. To be sure, these factors place constraints on each other. Objects help define a scene’s spatial dimensions, and a scene's spatial dimensions may place limits on the types of objects it can contain. Yet, many scene categories that differ from each other in their spatial properties can nevertheless accommodate each other’s associated objects without grossly altering their properties or violating physical law. For instance, although bathrooms and kitchens may differ in their average dimensions at the category level, there is usually no physical (as opposed to semantic) impediment to replacing a kitchen’s refrigerator with a shower, or a bathroom's bathtub with a stove, and little change in either scene's spatial properties as a consequence of the switch. In general, all other things (e.g., object size) being equal, the identities of objects in a scene and the scenes’ spatial properties are, with a few exceptions, logically independent.

The reliance of scene recognition on 2 cues, that can vary fairly freely with respect to each other raises a problem. How do scene categorization decisions cope with the potential for conflicts between the categories most associated with each cue? Consider, for example, the problem of categorizing a large bathroom. While the objects present (a sink, toilet, and bathtub) might be closely associated with bathrooms, the spatial dimensions of the room might be more closely associated with a different room category, such as a kitchen. Drawing from models positing that both objects and scenes' spatial properties can activate schemata or context frames for specific scene categories (Biederman 1972; Palmer 1975; Friedman 1979; Antes et al. 1981; Biederman et al. 1982; Loftus et al. 1983; Boyce and Pollatsek 1992; Bar and Ullman 1996; Hollingworth and Henderson 1998; Bar 2004; Mudrik et al. 2010), we might imagine that this conflict is resolved via some negotiation between the schemata activated by each resource. In this view, objects and spatial properties are processed independently through the stage at which each triggers a “hypothesis” of the room’s category. An alternative is that potential conflicts between the room’s object contents and spatial properties are blunted at the stage at which those resources are encoded, that is, before schemata are activated by each. Information from whichever resource is likely to be more reliable under a given set of circumstances (likely objects in this scenario) how the other resource is encoded, with the goal of maximizing the likelihood that each resource ultimately activates the same schema.

In the present study, we used a combination of behavioral and neuroimaging techniques to assess whether any such encoding-stage bias occurs during scene viewing, specifically asking whether the encoded values of scenes’ spatial properties are influenced by scenes’ object contents. Taking advantage of the susceptibility of perceived spatial properties to...
negative aftereffects (Greene and Oliva 2010), we first asked participants in a series of behavioral experiments to judge the spatial scales of average-sized bathrooms and kitchens after prolonged exposure to either very large or very small scenes from the same category. We find that the magnitudes of aftereffects produced by both small and large adapting rooms were significantly smaller when objects which were strongly informative of scene category, such as refrigerators or toilets, were visible in adapting scenes versus when they were masked. These results indicate that the presence of informative objects in adapting scenes biased their encoded spatial properties toward values associated with the average of their category. Next, using functional magnetic resonance imaging (fMRI), we observed an essentially identical bias in scene-evoked activity patterns in the parahippocampal place area (PPA), a region which has been associated with the encoding of scenes' spatial properties (Aguirre et al. 1998; Epstein and Kanwisher 1998; Kravitz et al. 2011; Mullally and Maguire 2011, 2013; Park et al. 2011), as well as processing of scenes' contextual associations (Aminoff et al. 2007; Hassabis et al. 2007; Summerfield et al. 2010; Howard et al. 2011). We propose that these behavioral and physiological biases reflect “crosstalk” between object- and spatial property-processing pathways that serves scene recognition by harmonizing estimates of scene identity derived from each of these information resources, and that this process is mediated by the PPA.

Materials and Methods

Behavioral Experiments

Participants
A total of 119 participants (25 males) between 18 and 27 years old were recruited for the study, chiefly from among Boston College undergraduates enrolled in introductory psychology courses. All had normal or corrected-to-normal vision and provided written informed consent in accordance with the procedures of the Boston College Institutional Review Board. Participants were either paid $15 or received course credit.

Stimuli
We assessed the influence of objects on scenes’ encoded spatial properties by comparing the strength of aftereffects produced by scenes with and without informative objects visible. Visual stimuli were 500 full-color photographs of real-world bathrooms and kitchens (1000 photographs total), and 500 computer-rendered images of exemplars of kitchens (Fig. 1A; see Supplementary Fig. 1 for additional exemplars). Bathrooms and kitchens were selected for use because they are strongly associated with distinct sets of objects, as judged by the authors, and because, as indoor scenes, they have clearly identifiable bounding features (such as walls) that make relative judgments of spatial scale intuitive to naive observers. Real-world scenes were assigned to quintiles according to perceived size (hereafter referred to as “spaciousness”), as judged by web-based ratings made by a pool of paid raters recruited through Amazon Mechanical Turk (AMT). Rendered scenes were assigned to size quintiles according to their

Figure 1. Experimental stimuli and design. (A) Participants judged the spaciousness of images of exemplars of average-sized bathrooms, kitchens, and computer-rendered kitchens (top row, left to right). Judgments were preceded by adaptation to low- or high-spaciousness exemplars from the same categories (second and third rows, respectively) that had informative objects either unmasked or masked (left and right images in each cell, respectively). (B) Each adaptation block began with 800 adapting scenes (100 exemplars repeated 8 times, totaling 160 s) during which participants performed a one-back repetition detection task to ensure attention. Participants then rated the spaciousness of 30 test scenes. Test scenes always had all informative objects unmasked and were preceded by a 10-s top-up adaptation period containing 100 images from the initial adaptation period.
simulated floor areas. Scenes in the first and fifth quintiles (i.e., those that were very small and very large for their category, respectively) were selected for use as spatially extreme adapting scenes, and scenes in the middle three quintiles were designated for use as spatially average “test” scenes against which the effects of adaptation would be measured. Please see Supplementary Methods for details of the AMT scene rating procedure and methods of rendered scene generation.

Because we wished to understand how the objects in scenes might bias scenes’ encoded spatial properties, we generated a second version of each adapting scene in which informative objects were obscured. Informative objects were identified based on nominations made by each AMT rater of the 3 objects he or she most strongly associated with bathrooms and kitchens, with the 3 objects nominated most frequently for each category selected for use. Informative objects for bathrooms were toilets, sinks, and bath/showers; for kitchens, they were refrigerators, sinks, and stoves/ovens. Within each real-world scene, the spatial boundaries of as many as visible of the 3 appropriate objects were segmented with the LabelMe image annotation tool (Russell et al. 2008), and a masked version of each scene was generated by replacing those objects with scrambled biorthogonal 3.1 discrete wavelet transforms (Honey et al. 2008). On average, the number of informative objects masked in each scene was 2.89 for high-spaciousness bathrooms, 2.31 for low-spaciousness bathrooms, 2.98 for high-spaciousness kitchens, and 2.54 for low-spaciousness kitchens. Objects in rendered scenes were masked in the same way.

Procedure

Behavioral adaptation experiments were created and executed in MATLAB using the Psychophysics Toolbox (Brainard 1997). All scenes subtended 10° square when viewed from a chinrest positioned 57 cm from an LCD monitor. Each participant judged the spaciousness of test scenes in a series of adaptation blocks. Following the design of a previous study demonstrating negative aftereffects in judgments of scenes’ global properties after adaptation to scenes with extreme properties (Greene and Oliva 2010), each block began with 800 exemplars from either the high- or low-spaciousness quintiles of a single scene category, corresponding to 8 presentations of each of the 100 exemplars from the appropriate quintile, in a random order (Fig. 1B). Each adapting scene was shown for 100 ms and followed by a 100 ms gray screen, with exact timing yoked to the 60-Hz frame rate of the monitor. During adaptation sequences, participants were asked to perform a 1-back repetition detection task to maintain attention. Following the initial adaptation period, participants reported the subjective spaciousness of test scenes from either the same or a different category. Each test scene was shown for 100 ms followed by a 100-ms 1/f noise mask, after which participants rated the scene’s spaciousness according to the scheme outlined in the following paragraph. To maintain a consistent adaptation state, test scenes were separated by “top-up” adaptation periods consisting of 100 adapting scenes shown for 100 ms each without interruption over 10 s. Each participant rated 30 test scenes in each of 4 adaptation blocks corresponding to every combination of adapting scene spaciousness (high vs. low) and informative object masking state (all unmasked vs. all masked). Each block used a unique set of randomly selected test scenes, equated for average spaciousness across blocks. Each test scene was viewed by a participant only once.

In the first experiment undertaken, featuring adaptation to high- and low-spaciousness bathrooms, participants rated the spaciousness of test bathrooms on a 5-point scale (1 = “much less spacious than average,” 2 = “less spacious than average,” 3 = “about average spaciousness,” 4 = “more spacious than average,” 5 = “much more spacious than average”); ratings in all subsequent experiments were on a 4-point scale that was identical but for elimination of the middle “about average” option. This change was an attempt to improve the sensitivity of the rating scale to small changes in perceived spaciousness. Specifically, we were concerned that the “about average” response option in the 5-point scale was conceptually too broad, and therefore might be chosen by participants even for scenes which they felt were actually slightly above or below average in spaciousness. The 4-point scale theoretically avoided this problem by forcing participants to make a “more” versus “less” decision for every scene. Because the aim of the 4-point scale was to improve sensitivity to aftereffects that were already statistically significant in the initial bathroom experiment using the 5-point scale (see Results), repetition of that experiment with the 4-point scale was not necessary.

For reasons outlined in the Results, adaptation experiments were executed with 5 nonoverlapping groups of participants. For the first 3 groups, the design was exactly as described above: Participants rendered spaciousness judgments of scenes from a single category after adaptation to those from the same category. The categories used in these 3 experiments were bathrooms, kitchens, and computer-rendered kitchens; participants in each of these experiments are referred to hereafter as the “real bathroom” (participant n = 35), “real kitchen” (n = 17), and “rendered kitchen” (n = 30) groups, respectively. As described in the Results, the small size of the real kitchen group reflects early termination of that experiment when it became clear that the results were inconsistent with our hypothesis due to potential confounding factors in those photographs.

To understand whether adaptation effects were category specific, 2 additional groups of participants judged scenes from one category of real-world scenes (kitchens or bathrooms) after adaptation to scenes from the “other” category. The first of these groups were subjected to a variant of the main experiment design in which they rated kitchens after adaptation to bathrooms, and vice versa, but with objects always unmasked in all scenes. This experiment was designed to measure whether “basic” aftereffects (i.e., different test scene judgments after adaptation to high-vs. low-spaciousness scenes) could be observed between categories; participants in this experiment will be referred to as the “unmasked cross-category” group (n = 18). The small size of this participant group reflects the generally large magnitude of basic adaptation effects. Participants in the second “cross-category” group rated kitchens after adaptation to bathrooms that either had objects masked or unmasked. Participants in this experiment will be referred to as the “masked cross-category” group (n = 29).

Data Analysis

Owing to the evanescentness of aftereffects (Rhodes et al. 2007), we wished to exclude delayed spaciousness ratings. To do so, and to exclude trials in which participants responded implausibly quickly, all trials with reaction times (RTs) greater than 4 standard deviations above or below the mean RT for their participant group were excluded from analysis. We also employed a clustering algorithm to provide an unbiased means of filtering responses of inattentive participants. Details of both of these procedures can be found in Supplementary Methods. Note that these procedures reduced the number of participants in each group with corresponding data to final analyses. Final participant counts can be found in the Results.

Statistical Analysis

We used a series of permutation tests to assess the statistical significance of differences in judgments of test scene spaciousness among adapting scene types. For each comparison of interest between 2 types of adapting blocks (e.g., spaciousness ratings after adaptation to low-vs. high-spaciousness scenes), we randomly permuted the block labels of the 60 spaciousness judgments each participant made across the 2 types of adapting blocks. The difference between the average ratings from the 2 blocks was recomputed based on the permuted labels, and the average size of this difference across participants was stored. This permutation procedure was repeated 10,000 times, allowing us to construct the distribution of group-level rating differences to be expected under the null hypothesis that test scene spaciousness ratings were unaffected by adapting scenes. The P-value of the actual rating difference between the adaptation conditions being compared was the proportion of elements in the null distribution that exceeded the actual rating difference. Permutation testing was selected to avoid distributional assumptions of parametric tests.

Separate planned tests were performed to assess the significance of differences in aftereffects within a pair of adapting scene types. Each test involved its own independent set of label permutations. In the first test, labels were permuted between ratings following adaptation to high- and low-spaciousness unmasked adapting scenes in order to measure the strength of basic aftereffects that had been reported previously for spatial properties (Greene and Oliva 2010). In
the second test, labels were permuted between ratings following adap-
tation to high-spaciousness masked and unmasked adapting scenes in
order to measure the effect of object masking on aftereffects produced
by high-spaciousness scenes. Finally, in the third test, labels were per-
mutated between ratings following adaptation to low-spaciousness
masked and unmasked adapting scenes to measure the effect of object
masking on aftereffects produced by low-spaciousness scenes.
Because we had a clear hypothesis about the sign of the rating differ-
ence for each of these comparisons, $P$-values were determined from
one tail of null permutation distributions.

**fMRI Experiment**

**Participants**

Thirteen participants (12 females, aged 18–23 years) with normal or
corrected-to-normal visual acuity gave written informed consent in
compliance with procedures approved by the Boston College Institu-
tional Review Board. One participant with excessive motion artifacts
was excluded from analysis. Participants were paid $45.

**Stimuli**

Visual stimuli were real-world bathroom image exemplars assembled
for the behavioral experiments, except in gray- or blue-scale format
(Supplementary Fig. 3). Both masked and unmasked versions of these
images were used. Scenes subtended 9.3° of visual angle.

**Experimental Procedure**

Each stimulus event consisted of a bathroom image presented for
150 ms, followed by a white fixation cross for 1350 ms. Five types of
bathrooms were shown. Examples from the top and bottom spacious-
ness quintiles shown both with informative objects unmasked and
masked, and exemplars from the middle quintile with objects un-
masked. Participants indicated the color of the bathroom (gray or blue)
by button press when the fixation cross appeared. The 5 scene types
along with 3± null events were ordered according to third-order coun-
terbalanced de Bruijn sequences, a general class of pseudorandom se-
quences that provide the minimum length sequence needed to achieve
a desired depth of stimulus counterbalance for a condition set of arbi-
trary size (Aguirre et al. 2011; MacEvoy and Yang 2012). Each scan run
contained 36 repetitions of each stimulus type. Runs lasted 6 min and
18 s, including 15± fixation-only intervals attached to the end of each
run. Unique stimulus sequences were constructed for 6 scan runs for
each participant. Scan sessions also included 2 functional localizer scans
lasting 7 min 48 s each, during which subjects viewed blocks of color
pictures of scenes, faces, common objects, and scrambled objects
presented at a rate of 1.33 pictures per second (Epstein and Higgins

**MRI Acquisition**

All scan sessions were conducted at the Brown University MRI
Research Facility using a 3-T Siemens Trio scanner and a 32-channel
head coil. Structural $T_1^*$-weighted images for anatomical localization
were acquired using a 3D MPRAGE pulse sequences [time repetition
(TR) = 1620 ms, time echo (TE) = 3 ms, time to inversion (TD) = 950 ms,
voxel size $=0.9766 \times 0.9766 \times 1 \text{ mm}$, matrix size $=192 \times 256 \times 160$].
$T_2^*$-weighted scans sensitive to blood oxygenation level-dependent
contrasts were acquired using a gradient-echo, echo-planar pulse se-
quence (TR = 3000 ms, TE = 30 ms, voxel size $=3 \times 3 \times 3 \text{ mm}$, matrix
size $=64 \times 64 \times 45$). Visual stimuli were rear projected onto a screen at
the head end of the scanner bore and viewed through a mirror affixed
to the head coil. The entire projected field subtended $24° \times 18°$ at
1024 $\times$ 768 pixel resolution.

**fMRI Analysis**

Functional images were corrected for differences in slice timing by re-
sampling slices in time to match the first slice of each volume, re-
ligned with respect to the first image of the scan, and spatially
normalized to the Montreal Neurological Institute (MNI) template.
Volumes from experimental scans were analyzed with general linear
models (one for each scan run) implemented in SPM8 (http://www.fil.
ion.ucl.ac.uk/spm), including an empirically derived 1/f noise model,
filters that removed high and low temporal frequencies, and nuisance
regressors to account for global signal variations, between-scan signal
differences, and participant movements. Beta-value maps were ex-
tracted for each stimulus condition for each scan.

For each subject, a permutation test was used to identify those
voxels whose responses varied significantly among scene types and
would therefore be passed to multivoxel pattern analysis (MVPA) for
hypothesis testing. For each voxel, we stored the $F$ statistic from a
one-way analysis of variance (ANOVA) performed on beta values from
each of the 5 bathroom types, sampled across the 6 scans. This statistic
was compared with a null distribution of $F$ statistics computed from
10,000 within-scan permutations of condition labels, accumulated
across all voxels. A voxel was passed to subsequent analysis if its
unpermuted $F$ statistic exceeded the 95th percentile of the null
distribution. Selection based on a null distribution accumulated across
all voxels was a conservative approach, ensuring that only the voxels
with responses differing most consistently across conditions were
selected for further analysis. Note, however, that while this procedure
identified voxels whose responses varied among stimuli, it was not
biased toward identifying voxels with any particular ordinal relations-
ships among those responses. That is, a voxel could satisfy our selec-
tion criterion as easily with responses to stimuli labeled A, B, C, D, and
E that reliably fell in the order A $>$ B $>$ C $>$ D $>$ E as with responses that
were reliably ordered B $>$ C $>$ A $>$ E $>$ D, or any other order. Because
our hypotheses addressed ordinal relationships, as explained in the
following paragraph, our ANOVA-based feature selection procedure
did thus not amount to “peeking.” We required that at least 7 voxels be
selected from each region of interest (ROI; see below for definitions)
in each participant. This minimum was selected because it equaled the
number of voxels in each searchlight cluster used for whole-brain ana-
lyses, as described in a following section. Response vectors composed
of selected voxels were generated for each stimulus type and averaged
across scans, and pairwise Euclidean distances among all vectors were
computed for each participant.

**Statistical Analysis**

To test the hypothesis that informative objects bias patterns of neural
activity evoked by high- and low-spaciousness bathrooms, we used per-
mutation tests to assess the group-level significance of a series of contrasts
among pairwise Euclidean pattern distances extracted from each ROI.
First, we needed to ask whether patterns in each ROI were sensitive to
scenes’ actual spatial properties, without reference to any potential effects
of object. An ROI was considered sensitive to spatial properties if it
showed a significantly positive value for the distance contrast [(distance
from unmasked high-spaciousness scenes to unmasked low-spaciousness
scenes) minus (distance from unmasked high-spaciousness to average-
spaciousness) minus (distance from low-spaciousness to average-
spaciousness) minus (distance from unmasked low-spaciousness to
average-spaciousness)]. The significance of this contrast (i.e., the prob-
bility of being encountered by chance) was computed by comparing it to the
distribution of contrasts accumulated over 10,000 within-scan permutations
of the condition labels for unmasked high-spaciousness, unmasked low-
spaciousness, and average-spaciousness scenes. Second, to measure any
biasing effect of objects on patterns evoked by low-spaciousness scenes
relative to average-spaciousness scenes, we computed the contrast [(dis-
tance from masked low-spaciousness to average-spaciousness) minus
(distance from unmasked low-spaciousness to average-spaciousness)].
The significance of this contrast was computed by comparing it to the distri-
bution of contrasts accumulated over 10,000 within-scan permutations
of the condition labels for unmasked high-spaciousness, masked low-
spaciousness, and average-spaciousness scenes. Third, to measure any
biasing effect of objects on patterns evoked by high-spaciousness
scenes relative to average-spaciousness scenes, we computed the contrast
[(distance from masked high-spaciousness to average-
spaciousness) minus (distance from unmasked high-spaciousness to
average-spaciousness)]. The significance of this contrast was computed by
comparing it to the distribution of contrasts accumulated over 10,000 within-
scan permutations of the condition labels for unmasked high-spaciousness,
masked high-spaciousness, and average-spaciousness scenes. Last, to provide context for any of significant values
for the above contrast, we computed the contrast [(distance from masked
high-spaciousness to masked low-spaciousness) minus (distance from
unmasked high-spaciousness to unmasked low-spaciousness)]. The signifi-
cance of this contrast was computed by simultaneously permuting labels
between masked and unmasked high-spaciousness scenes and between
masked and unmasked low-spaciousness scenes. The utility of each of
these contrasts is explained further in the Results. Because we had clear

Downloaded from https://academic.oup.com/cercor/article-abstract/25/8/2267/312785 by guest on 23 March 2019
hypotheses about the sign of each contrast, one-tailed tests were applied, with values from unpermuted data considered significant if they were exceeded by fewer than 5% of permuted values. Our focus on specific distance contrasts obviated the need for cross-validation that is often employed with MvPA (Drucker and Aguirre 2009; Morgan et al. 2011).

Whole-brain searchlight pattern analysis was performed with 3 mm radius (7 voxels) searchlight centers on each voxel in the brain (Kriegeskorte et al. 2006). To measure any local effect of object visibility on scenes’ encoded spatial properties, Euclidean distances among patterns evoked by each stimulus at each searchlight position were used to compute the distance contrast (distance from masked high-spaciousness to masked low-spaciousness) minus (distance from unmasked high-spaciousness to unmasked low-spaciousness). The contrast value for each searchlight cluster was assigned to the voxel at its center. Resulting single-participant contrast volumes were passed to a second-level exact permutation test implemented with SnPM (http://www.neurolens.org) and custom MATLAB scripts to assess the group-level significance of regions showing large subject-averaged contrast values, which were consistent with a biasing effect of objects. First, voxelwise variance was smoothed with a 3-mm full-width at half-maximum (FWHM) Gaussian filter under the nonparametric assumption of smooth underlying variance in the searchlight volumes (Nichols and Holmes 2002). Smoothed variance maps were used to compute maps of pseudo t-values for each of the $2^{12}$ sign permutations of the 12 single-subject contrast volumes. The resulting distributions of pseudo t-values for each voxel were used to identify voxels in each permuted volume whose pseudo t-values were encountered with a probability <0.001, and the size of the largest 6-connected cluster of such voxels recorded for each permutation volume. Clusters identified in the same way from unpermuted volumes were considered significant if their sizes were exceeded by fewer than 5% of elements in the distribution of maximum cluster sizes accumulated across permuted volumes. The thresholded second-level volume was projected onto a surface-based representation of the MNI canonical brain with the SPM Surfrend toolbox (http://spmsurfrend.sourceforge.net), and then rendered in NeuroLens (http://www.neurolens.org).

Regions of Interest
All ROIs were defined from localizer scans using a recently described algorithmic approach applied to data from localizer scans (Julian et al. 2012). Briefly, for each contrast of interest (e.g., scenes > objects), a whole-brain group volume was created in which each voxel was tagged with the proportion of subjects in which that voxel showed an activation difference exceeding a threshold of $t = 1.6$. A 3-mm FWHM Gaussian filter was applied to this volume, followed by a watershed algorithm with an 8-connected part filter applied to each axial slice. The resulting volumes contained segmented parcels corresponding to activations shared between subjects. To reduce extraneous activations present in the segmented group volumes, parcels generated from the activations of fewer than 50% of subjects were removed. Individual subject ROIs associated with a given contrast were defined from the intersection between the shared activation volume and each subject’s contrast map thresholded at $t = 1.6$. This procedure was performed for the contrasts of scenes > objects [to identify the PPA, retrosplenial complex (RSC), and transverse occipital sulcus (TOS)], objects > scrambled objects [lateral occipital (LO) and posterior fusiform sulcus (pFs) subdivisions of the lateral occipital complex (LOC)], and scrambled objects > objects (early visual cortex). All voxels identified by the scenes > objects contrast that were inferior to the splenium of the corpus callosum were assigned to the PPA, and all superior voxels assigned to the RSC. An 11-voxel region of overlap between the group-defined candidate regions for the right PPA and right pFs was assigned to the PPA.

Results

Behavioral Experiments
To measure the impact of informative objects on scenes’ encoded spatial properties, participants were asked to rate the subjective spaciousness of test exemplars of bathrooms and kitchens that possessed spatial properties at or near the average for their category, after adaptation to exemplars which were much more spacious or much less spacious than their category average (see Materials and Methods). In the critical experimental manipulation, adapting scenes either had informative objects unmasked (i.e., fully visible) or masked. We reasoned that any effect of informative objects on adapting scenes’ encoded spatial properties should have been evident as a difference in the magnitude of aftereffects they produced when objects were masked versus unmasked. This adaptation-based approach was selected over the alternative of having participants directly rate the spatial properties of scenes with and without masked objects, because it avoided potential variability in participants’ interpretation of masks when rating scenes with masked objects. For instance, some participants could have interpreted masks as unspecified objects, while others might have interpreted them as empty space. Although this potential problem might have been addressed via appropriate instructions, it was more cleanly avoided using an adaptation approach in which participants were never forced to make judgments of manipulated scenes. The perceptual quantity of spaciousness was selected as the dependent measure because it is an easily understood concept that captures the scenes’ general spatial scales; we do not assert that spaciousness is a fundamental dimension along which scenes are encoded by the visual system and acknowledge that it likely draws upon a number of more basic spatial properties that have been characterized previously (Oliva and Schyns 2000; Oliva and Torralba 2001, 2006; McCotter et al. 2005).

Consistent with the susceptibility of scene spatial properties to adaptation (Greene and Oliva 2010), participants in the real bathroom group ($n = 34$) judged average bathrooms to be significantly less spacious after adaptation to high-spaciousness bathrooms than after adaptation to low-spaciousness bathrooms, all with objects unmasked (Fig. 2.4, vertical difference between data points on the left; one-tailed permutation test, $P = 0.0001$). The presence of this basic aftereffect demonstrates (1) that the perceptual quantity of spaciousness is subject to aftereffects similar to those described previously for individuated spatial properties of scenes, and (2) that aftereffects can be observed even within the spatial constraints of a single indoor scene category.

When informative objects in high-pole adapting bathrooms were masked, aftereffects were significantly enhanced: Spaciousness ratings of average bathrooms were significantly lower after adaptation to high-spaciousness bathrooms with masked objects than after adaptation to the same scenes with unmasked objects (Fig. 2.4, vertical difference between points connected by the lower line; $P = 0.018$). Based on the observation that negative aftereffects for high-level visual features generally increase with the perceptual distance between adapting and test stimuli (Webster and Maclin 1999; Leopold et al. 2001; Webster et al. 2004; Little et al. 2005), this indicates that large adapting bathrooms were encoded as more spacious when informative objects were masked versus when they were unmasked. Critically, this increase in encoded spaciousness did not simply reflect space “freed up” by object removal, as evident in the fact that the magnitude of the aftereffect produced by low-spaciousness bathrooms was also significantly enhanced by object masking (Fig. 2.4, vertical difference between points connected by the upper line; $P = 0.002$), and that this enhancement took the opposite sign. These results indicate that
large bathrooms were encoded as smaller and small bathrooms encoded as larger when informative objects were unmasked versus masked. In other words, adapting bathrooms at both spatial extremes were encoded as more similar to the average bathroom when their informative objects were visible.

Data from the real kitchen participant group (n = 16) showed a basic adaptation effect (ratings after adaptation to low-spaciousness unmasked adapting scenes > after adaptation to high-spaciousness unmasked adapting scenes = 0.0002). Furthermore, as with bathrooms, test kitchen spaciousness ratings after adaptation to low-spaciousness exemplars were significantly higher when adapting kitchens had objects masked than when they were unmasked (P = 0.013), indicating that object visibility biased small adapting scenes to be encoded as more spacious (i.e., more similar to average kitchens; Fig. 2B). Unlike adaptation with bathrooms, however, aftereffects produced by high-spaciousness kitchens were not enhanced when objects were masked. (The relatively small size of this participant group was a result of our decision to terminate data collection after it became clear that there was no trend whatsoever toward enhanced aftereffects with object masking in high-spaciousness adapting scenes.)

One explanation for the absence of aftereffect enhancement with high-spaciousness kitchens is that the extra space in large kitchens allowed them to accommodate a greater number of objects carrying information about scene category than low-spaciousness kitchens, potentially blunting the impact of masking the limited set of informative objects we targeted for masking; this potential complication applied less to bathrooms because they were associated with fewer informative objects to begin with (Table 1). Consistent with this explanation, high-spaciousness kitchens contained significantly more of the objects in Table 1 than did high-spaciousness bathrooms (6.99 vs. 5.01 objects per scene on average; \( t_{(196)} = 11.84, P < 0.0001 \)). Moreover, we observed that large kitchens often contained many objects that, while not appearing on the list in Table 1, may still have been associated with kitchens (e.g., kitchen counter stools). Large bathrooms did not appear to collect extra potentially informative objects in a similar way.

To avoid this potential confound, we repeated the kitchen adaptation experiment with a group of participants (n = 25) who viewed computer-rendered kitchens in which spatial parameters and object contents could be exactly and independently specified. As with real kitchens, average-sized rendered kitchens were susceptible to basic aftereffects (P = 0.001).

<table>
<thead>
<tr>
<th>Table 1</th>
<th>List of all objects nominated by online raters as associated with bathrooms and kitchens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bathrooms</td>
<td>Kitchens</td>
</tr>
<tr>
<td>Bathtub</td>
<td>Cabinet</td>
</tr>
<tr>
<td>Mirror</td>
<td>Countertop</td>
</tr>
<tr>
<td>Shower</td>
<td>Dish</td>
</tr>
<tr>
<td>Sink</td>
<td>Dishwasher</td>
</tr>
<tr>
<td>Toilet</td>
<td>Microwave</td>
</tr>
<tr>
<td>Towel</td>
<td>Oven</td>
</tr>
<tr>
<td>Vanity</td>
<td>Refrigerator</td>
</tr>
<tr>
<td></td>
<td>Sink</td>
</tr>
<tr>
<td></td>
<td>Stove</td>
</tr>
<tr>
<td></td>
<td>Table</td>
</tr>
<tr>
<td></td>
<td>Utensil</td>
</tr>
</tbody>
</table>

Figure 2. Behavioral results. (A) Average-spaciousness test bathrooms were judged significantly smaller after adaptation to high-spaciousness bathrooms than after adaptation to low-spaciousness bathrooms (black vs. gray filled squares). Spaciousness ratings after adaptation to high-spaciousness bathrooms were significantly lower when adapting scenes’ informative objects were masked. Object masking in low-spaciousness adapting scenes produced the opposite effect. (B) Real kitchens were similarly subject to basic (i.e., high vs. low) aftereffects, although a significant impact of informative objects was only present in aftereffects produced by low-spaciousness exemplars. (C) Bidirectional enhancement of aftereffects was restored in computer-rendered kitchens, which allowed exact specification of object contents. (D) Bidirectional enhancement was also evident in ratings of kitchens after adaptation to extreme bathrooms. Error bars are SEM; †P = 0.065; *P < 0.05; **P < 0.01; ***P < 0.001.
objects were masked (Fig. 2C; ratings after adaptation to masked high-spaciousness scenes < after adaptation to unmasked high-spaciousness scenes, \( P = 0.025 \); ratings after adaptation to masked low-spaciousness scenes > after adaptation to unmasked low-spaciousness scenes, \( P = 0.002 \)), indicating that both low- and high-spaciousness adapting kitchens were encoded as more similar to average kitchens when informative objects were unmasked.

To understand whether aftereffects required that adapting and test scenes belong to the same category, participants in the unmasked cross-category group (\( n = 17 \)) rated the spaciousness of scenes from one category after adaptation to extreme unmasked exemplars from the other. Adaptation to real-world bathrooms induced significant aftereffects in real-world kitchens and vice versa (kitchen ratings after adaptation to high-spaciousness bathrooms < after adaptation to low-spaciousness bathrooms, \( P = 0.0001 \); bathroom ratings after adaptation to high-spaciousness kitchens < after adaptation to low-spaciousness kitchens, \( P = 0.012 \)). These cross-category aftereffects were enhanced by object masking (Fig. 2D), as demonstrated in the masked cross-category participant group (\( n = 24 \)), who showed significantly greater aftereffects in ratings of real average-spaciousness kitchens after adaptation to low-spaciousness bathrooms with objects masked (\( P = 0.003 \)), and marginally greater aftereffects after adaptation to high-spaciousness bathrooms with objects masked (\( P = 0.065 \)). Because of the absence of an effect of object masking on aftereffects in the real kitchen participant group, we did not examine the impact of object masking state on aftereffects produced by real-world kitchen adapters on bathrooms.

**fMRI Experiment**

Our behavioral results indicate that the presence of objects strongly associated with a particular scene category produces a “centripetal bias” in the encoded spatial properties of scenes containing them. To understand where in the visual system this bias arises, we used fMRI to record activity patterns evoked by exemplars of each of the 5 types of bathrooms used in the behavioral experiments: high- and low-spaciousness exemplars, both with and without informative objects masked, plus average-spaciousness exemplars with objects unmasked. Note that this experiment sought to directly measure neural responses to scenes varying in spaciousness and masking state rather than any neural signature of aftereffects those scenes produced. This direct approach was feasible because no perceptual judgments of scene spatial properties were required of subjects in the scanner, who solely judged whether scenes were shaded in gray or blue. Only bathrooms were used because they were associated with a more reliable centripetal bias in the perceptual experiments.

Our analysis focused on the PPA, within which activity patterns have been shown to track the spatial properties of scenes (Kravitz et al. 2011; Park et al. 2011; Harel et al. 2012). Consistent with these previous studies, distances among activity patterns evoked in the right PPA by high-, low-, and average-spaciousness bathrooms, all with objects unmasked, qualitatively matched differences among their spatial properties: The average distance between patterns evoked by high- and low-spaciousness rooms was significantly greater than that of distances between each of those patterns and patterns evoked by average-spaciousness rooms (Fig. 3A, permutation test, \( P = 0.028 \)). This basic sensitivity to spatial properties was not significant in the left PPA, consistent with previous results suggesting greater sensitivity to spatial properties in the right PPA (Wagner et al. 1998; Kirchhoff et al. 2000; Epstein et al. 2003).

Patterns evoked in the right PPA by both high- and low-spaciousness bathrooms were significantly closer to that evoked by average-spaciousness bathrooms when informative objects in the 2 extreme scenes were unmasked versus when they were masked (Fig. 3B, C; \( P < 0.0001 \) for each). This combination of pattern distance differences matches the combination of distance differences among encoded spatial properties that was revealed by our adaptation results, in which scenes with objects unmasked were encoded as more similar to their category average than that with objects masked. However, it was possible that the greater similarity of patterns evoked by unmasked extreme scenes to those evoked by average-spaciousness scenes might have resulted from object masking state per se, rather than an effect of masking on encoded spatial properties. To assess whether this was the case, we also compared similarities of patterns evoked by high- and low-spaciousness exemplars that had objects unmasked with those evoked by the same scenes when they had objects masked. If the greater similarity of patterns evoked by average-spaciousness scenes to those evoked by unmasked high- and low-spaciousness scenes were simply an outcome of object masking by itself, patterns evoked by high- and low-spaciousness scenes should have been equally similar to each other when objects were masked and when they were unmasked, that is, when masking state was controlled. Instead, we find that patterns evoked by high- and low-spaciousness scenes were significantly more similar to each other when objects were unmasked than when masked (Fig. 3D, \( P = 0.0022 \)). Thus, patterns evoked by spatially extreme exemplars with objects unmasked were not only more similar to those evoked by average exemplars, but also more similar to those evoked by scenes at the opposite spatial pole. This combination exactly matches the relationships among encoded spatial properties inferred from our perceptual experiments.

Although these results were encouraging, it was possible that the greater distances between patterns evoked by extreme scenes with objects masked arose from differences in cognitive processes related to object masking. Therefore, there was still a risk that the greater similarity of patterns evoked by unmasked extreme scenes to those evoked by average-spaciousness scenes reflected a direct effect of object masking state, rather than an effect of objects on encoded spatial properties. To achieve a more direct comparison between our behavioral results and PPA activity patterns, we used multidimensional scaling (MDS) to visualize and isolate PPA pattern dimensions that specifically corresponded to scenes’ spatial properties. Matrices of pairwise Euclidean distances among the 5 scene-evoked patterns from the right PPA (Fig. 4A) were averaged across participants and passed to MDS, which produced as output the coordinates of patterns along the set of orthogonal dimensions that best accounted for the full suite of pairwise distances.

The positions of right PPA patterns expressed in terms of the first 2 dimensions returned by MDS are shown in Figure 4B. Taken together, these 2 dimensions accounted for the majority of total pairwise pattern distance, and individually accounted for similar shares 36.0% and 30.7% for the first and second dimensions, respectively; the remaining 2 dimensions each accounted for substantially less distance (17.5% and 15.8% for the
Figure 3. Analysis of pattern distances in the right PPA. (A) Average Euclidean distances between patterns evoked by high- and low-spaciousness bathrooms with unmasked objects were significantly greater than the average of distances between each of those extremes and the pattern evoked by average-spaciousness bathrooms. (B–D) Pattern distances satisfying the predictions made from behavioral results, demonstrating a centripetal object bias in the right PPA. The combination of these contrasts was not found in other ROIs. Distance data in each panel correspond to comparisons between patterns denoted by same-shaded arrows in the left half of each panel. Error bars are SEM. *P < 0.05; **P < 0.01; ***P < 0.001.

Figure 4. Visualization of relationships among scene-evoked patterns in the right PPA. (A) Matrix of Euclidean distances among bathroom-evoked patterns, averaged across participants. (B) Corresponding positions of patterns in two-dimensional space returned by MDS. Dimensions 1 and 2 capture 36.0% and 30.7%, respectively, of total between-pattern distance. Dashed contours are bootstrap 95% confidence ellipses for pattern coordinates, based on 10,000 resamples. Positions of patterns along the second (vertical) dimension qualitatively match relative encoded spaciousness of scenes indicated by behavioral experiments.
by the presence of informative objects, which bias encoded spatial properties. Critically coordinates of masked high- and low-spaciousness scenes along the second dimension are further from those of the average scene than are coordinates for their unmasked counterparts, exactly consistent with the centripetal bias we observed in our behavioral results. To assess the probability of observing this order by chance, we performed MDS on new distance matrices that were computed after random within-subject label swaps between activity patterns elicited by masked high-spaciousness scenes and by unmasked high-spaciousness scenes, and simultaneously between activation patterns elicited by masked low-spaciousness scenes and by unmasked low-spaciousness scenes. Across 10,000 sets of swaps, there was a probability of 0.019 of observing an MDS output dimension which (1) correctly ordered all 5 scenes in terms of their ground-truth spaciousness (as described in the previous paragraph) and (2) showed a mask-dependent increase in average distance from average- to extreme-spaciousness exemplars that were at least as large as the increase along the second dimension of Figure 4B. (At least one dimension that correctly ordered coordinates was observed for every swap; in the event that 2 such dimensions were returned, only the dimension accounting for the greater portion of pattern distance was considered.) These analyses indicate that patterns evoked in the PPA by extreme bathrooms with objects unmasked were more similar to those evoked by average bathrooms specifically along PPA pattern dimensions encoding scenes’ spatial properties. No other ROI possessed a profile of pattern similarity consistent with the perceptual experiments. Data from all other ROIs, including RSC, TOS, and LOC, can be found in Supplementary Figures 5–15.

Finally, we used a searchlight analysis to identify any regions outside our selected ROIs in which relationships among scene-evoked patterns were consistent with a centripetal bias by informative objects. Consistent with our ROI analysis, in the occipitotemporal cortex, only voxel clusters corresponding to the anterior portions of PPA showed evidence of centripetal bias (Fig. 5A). Evidence for centripetal bias was also found in a single right hemispheric frontal lobe cluster (Fig. 5B).

Discussion
We find that scenes’ encoded spatial properties are influenced by the presence of informative objects, which bias encoded properties toward the average of each scene’s category. This centripetal bias was evident both perceptually and in activity patterns in the PPA, a region that has been linked to processing of scenes’ spatial properties. Because scenes’ actual objective spatial properties are to some extent determined by the objects within them, it would not have been very surprising to find that the addition of objects exerted a negative effect on scenes’ encoded spatial properties. Critically, however, we found that the presence of informative objects led to both high-spaciousness scenes being encoded as smaller and low-spaciousness scenes being encoded as larger. This contingent directionality indicates that the presence of objects influenced scenes’ encoded spatial properties above and beyond what would be expected from objects’ simple occupancy of space.

Potential Explanations for Centripetal Bias
Perhaps the simplest explanation for the centripetal bias is feedback from conscious scene category decisions. Assuming that such decisions can drive scenes’ encoded spatial properties toward the average of the adjudged category, and further that the presence of objects in scenes leads to more accurate recognition, it stands to reason that a greater proportion of adapting scenes would have been encoded as spatially average scenes.
when informative objects were present. While this can theoretically explain the centripetal bias, it fails practically on several counts. First, category decisions were never required in either the perceptual or fMRI experiments, and a majority of participants (including all in the fMRI study) viewed scenes from a single category, leaving no impetus for even latent categorization. Second, even though participants in the masked cross-category group did see scenes from both categories and therefore might have had greater inclination to categorize scenes, their average centripetal bias was no greater. Third, and most importantly, the design of our fMRI experiment, in which scenes with and without objects masked were interleaved, meant that the category of masked scenes was always obvious; this applied to our perceptual experiments as well, albeit on a coarser time scale. Thus, even if participants persisted in categorization in the absence of any external motivation to do so, or if such categorization was automatic, it is highly unlikely that categorization accuracy would have differed appreciably between scenes with and without objects masked. Note that this does not challenge our designation of masked objects as “informative,” as this designation was based on the frequency of their association with a scene category rather than their impact on categorization in this experiment, although such an impact has been demonstrated previously for these exact object categories (MacEvoy and Epstein 2011). It simply means that, in this particular experimental context, those objects provided no more information about scene category than was already available from other cues.

Two other potential explanations for our results lie in differential attentional demands potentially placed by masked and unmasked scenes. Under both explanations, object masking led to greater attention to scenes’ spatial properties and a consequent improvement in the accuracy of their encoding, although in completely opposite ways. In the first, object masking potentially “freed” attentional resources ordinarily attracted by objects to be deployed to adapting scenes’ spatial properties, which were therefore encoded with greater fidelity to scenes’ true extreme spatial scales than when objects were unmasked. To accept this explanation, however, one must adopt the general view that codes for spatial properties are inherently less precise when objects are present. Considering that almost all real-world scenes contain objects, this imprecision would be maladaptive, and therefore unlikely. Under the second attention-based explanation, object masking drew greater attention to masked adapting scenes as wholes as observers struggled to identify masked objects, with the outcome again that the encoded values of these scenes’ spatial properties were more faithful to their extreme natures than when objects were unmasked. While this explanation cannot be directly discounted, the rapid timing of adapting sequences is likely to have dampened any efforts by participants to decipher masked objects, particularly given the ongoing repetition detection task. One way to avoid this potential problem could have been to simply excise, rather than mask, informative objects in scenes. Doing so, however, would have introduced other differences between masked and unmasked adapting scenes, including alterations to scenes’ low-level statistical properties. We wished to avoid such differences for the sake of our subsequent fMRI experiment, given evidence that PPA is sensitive to them (Rajimehr et al. 2011).

Moving beyond attention, it is very difficult to explain the centripetal bias as an outcome of some “direct” cognitive effect of object masking (i.e., an effect not mediated by some change in encoded spatial properties). Although it is possible and perhaps even likely that cognitive processes related to object recognition were differentially activated by masked and unmasked adapting scenes, this difference is unlikely to explain our results, for 2 reasons. First, it is unlikely that purely object-related cognitive differences would have influenced aftereffects exerted on the perceived spatial properties of test scenes. Secondly, even if they were able to exert such an influence, it is even less likely that they could have produced the bidirectional enhancements in aftereffects we observed. This is because any object-based differences in cognitive processes between masked and unmasked adapting scenes would have been identical for both high- and low-spaciousness adapters, and as a consequence, any potential contamination of spatial codes should therefore have likewise been identical. This conflicts with our observation that object masking exerted opposite effects on the encoded spatial properties of high- and low-spaciousness adapting scenes.

Finally, it is unlikely that the differing strength of aftereffects we observed with masked and unmasked adapters reflects the fact that test scenes matched the masking state of unmasked adapters but not of masked adapters. We acknowledge that, in general, the susceptibility of a test stimulus to aftereffects is dependent on the degree to which it matches the adapting stimulus along nonadapted dimensions. For instance, face-specific aftereffects are stronger when the adapting stimuli and judged stimuli occupy the same retinal location (Zimmer and Kovács 2011), and motion aftereffects are strongest when adapting and judged stimuli possess the same spatial frequency (Anstis et al. 1998). However, if we liken the masking states of adapting scenes in our study to different retinal positions or spatial frequencies, aftereffects should have been stronger for unmasked scenes because they matched the masking state of test scenes. This, of course, is the opposite of what we observed.

**Crosstalk Theory**

Rather than an outcome of decision feedback or attention, we propose instead that the centripetal bias reflects a form of here-tofore undescribed crosstalk between object- and spatial property-encoding pathways. In this theory, objects associated with a given scene category contribute a “normalizing” signal to codes for spatial properties, bringing potentially highly excursive encoded values into closer register with those typical of the scene category the objects are associated with. In contrast to the feedback account rejected above, in this framework, the centripetal influence of objects precedes scene recognition. Moreover, we propose that the purpose of this influence is to assist scene recognition by easing potential conflicts between scene category judgments derived from object contents and spatial properties.

For an example of how this might work, let us return to the task of deciding whether a room in an unfamiliar house is a bathroom, perhaps after being told that both a bathroom and a kitchen (but no other room type) can be found along a hallway one is walking down. These room categories differ both in their typical object contents (Table 1) and in their average spatial properties (Fig. 6A). As such, upon viewing the first encountered room, it can be expected that hypotheses about its identity could be generated from both its object contents and
its spatial properties. (We use the term “hypothesis” here to avoid any mechanistic implications attached to “schema” or “context frame.”) Let us assume that this room happens to be an inordinately large bathroom. Owing to the high degree of overlap between real-world distributions of the sizes of bathrooms and kitchens, it is quite possible that this room’s extreme spatial properties may place it on the “kitchen” side of a neutral spatial property-based category criterion. This would generate a spatial property-based hypothesis that conflicts with the hypothesis generated from its object contents, which we will assume leave no doubt about the room’s category. A final judgment of the room’s identity therefore requires some means of reconciling these competing hypotheses. To do so would presumably require consideration of a variety of factors to determine the appropriate weight that should be given to each hypothesis, a process that might take time and offer added opportunities for error.

We propose that crosstalk aids scene recognition by reducing the frequency with which this reconciliation process is required. By driving the encoded spatial properties of the very large bathroom toward those of the average bathroom, the centripetal bias we observed reduces the probability that the hypothesis of scene identity derived from those properties will conflict with the hypothesis derived from the scene’s object contents. Assessed across encounters with many scenes, we propose that centripetal bias narrows the distributions of each scene category’s encoded spatial properties, reducing the degree of overlap they would show if space were encoded veritically, and consequently decreasing the proportion of scenes on the “wrong” side of neutral spatial property criteria between categories (Fig. 6B). We expect that the resulting harmonization of category hypotheses derived from encoded spatial properties with those derived from objects would improve the speed and accuracy of categorization.

Based on this theory, we expect that the degree of centripetal bias produced by an object should bear some relationship to the strength of its association with a specific scene category; that is, that the identities of the masked objects in both our behavioral and fMRI experiments mattered. Although our study did not directly test this relationship, the alternative that all objects are equipotent in inducing centripetal bias is virtually impossible to reconcile with the bidirectional bias we observed. Consider a completely empty room with a floor area intermediate between the average floor areas of 2 room categories generally differing reliably in size. The addition of an object with no association to any particular scene category will, by definition, add no information about the identity of the scene, leaving the direction of any potential induced bias unspecified: Should the bias be toward the smaller or the larger scene category? With the target of bias undefined, such an object cannot produce any bias, seemingly negating the possibility that all objects are equipotent in producing the bias. We therefore consider it extremely likely that the centripetal bias we observed depended on the identities of those objects which varied in masking state.

We acknowledge, however, that our results do not tell us whether the ability of objects to bias scenes’ encoded spatial properties derives from objects’ statistical associations with specific base-level scene categories, such as “bathroom” versus “kitchen”, or with scenes grouped at some higher taxonomic level, such as “indoor scenes” versus “outdoor scenes.” In other words, while our results are consistent with our hypothesis that objects bias encoded spatial properties toward the average values of bathrooms or kitchens, they leave open the possibility that objects biased encoded properties toward those of the average indoor room. This ambiguity exists because the high- and low-spaciousness adapting scenes we used from each scene category were likely extreme enough that they bracketed the average spatial properties of both categories, and quite possibly those of the average across all indoor scene categories. However, while we cannot identify with certainty the level of scene specificity at which the centripetal bias operated, it seems that a bias which targeted the spatial properties of base-level categories would be more adaptive than one which targeted the average properties of a higher spatial property-based categorizations that con-
taxonomic cluster, such as indoor scenes. This is because while a bias targeting the average indoor room would benefit indoor/outdoor scene distinctions, it would simultaneously harm distinctions among base-level categories of indoor or outdoor scenes by compressing the range of encoded spatial properties of all categories within each group toward a single point. In contrast, a bias that targeted base-level scene categories, and therefore aided distinctions among them, would be at worst neutral with respect to the high taxonomic distinctions, such as indoor versus outdoor. Ultimately, additional experiments are necessary to clarify this issue. We emphasize, however, that the uncertainty we highlight does not challenge our crosstalk interpretation of the centripetal bias but merely raises questions about the level of scene distinctions that might benefit from it.

Although we favor the idea that the centripetal bias targets base-level scene categories, our crosstalk theory does not predict that all such categories will be equally susceptible. Instead, assuming equally strong object associations, the magnitude of centripetal bias should vary with the strength of scene categories' associations with any particular set of spatial properties. Specifically, we predict that the centripetal bias should be stronger for indoor scenes (such as the bathrooms and kitchens used here), which tend to occupy a relatively narrow range of real-world sizes, than for outdoor scenes. Given this, we do not interpret the fact that objects “controlled” spatial properties in this experiment to indicate that objects enjoy a general position of superiority over spatial properties during processing of all scenes. Thus, an important future test of our crosstalk hypothesis will be to show not only that the centripetal bias exists beyond the narrow range of scene types used in the present study, but also that it fails predictably for scene categories not strongly associated with any particular spatial scale.

We wish to emphasize that our crosstalk theory is not simply a restatement of the idea that objects activate scene schemata or context frames storing information about the features, including spatial properties, typically associated with each scene category (Bar 2004). Instead, we conceive of crosstalk as a direct translation of object information into spatial property codes, independent of (although perhaps coincident with) schemata activation. Support for this view comes from the fact, already mentioned, that the identities of masked adapting scenes were likely to have been quite obvious to participants, whether due to remaining identifying features, or for nearly half of our participant pool, previous exposure to adapting scenes in their unmasked forms. This makes it likely that scene schemata were equivalently activated regardless of adapting scenes' masking states. The persistence of centripetal bias in light of this suggests that object processing pathways enjoy direct access to spatial property codes that is distinct from their capacity to activate scene schemata. Similarly, it suggests that the influence of objects on encoded spatial properties arises from the objects' visual features, rather than their identities, since we expect that the latter also remained fairly firmly instantiated from context even during masked blocks.

Our crosstalk theory thus holds that informative objects benefit scene categorization in 2 ways: by directly activating schemata of their associated scenes and by biasing encoded spatial properties to reduce conflicts with properties associated with those schemata. This view makes the testable prediction that the presence of informative objects should aid performance on a binary scene discrimination task more under conditions that allow objects to produce a centripetal bias, such as when scene exemplars possess spatial properties departing from their category averages, versus when they do not, such as when exemplars from at least one category already match the spatial properties typical of their category. The competing view that the centripetal bias reflects feedback from object-activated schemata predicts no such difference. We expect, therefore, that future experiments will be able to clarify whether our feedforward crosstalk explanation of centripetal bias is correct.

**Role of Parahippocampal Cortex**

Matching our behavioral results, activity patterns evoked in the right PPA by scenes at each spatial extreme were more similar to patterns associated with the opposite extreme when objects were unmasked versus when they were masked. This correspondence to our behavioral results was not observed in any other ROI. Although PPA has been shown to be sensitive to low-level properties of stimuli, such as spatial frequency (Rajmehr et al. 2011; Zeidman et al. 2012) and texture (Cant and Goodale 2007; Cant and Xu 2012), response differences between unmasked and masked scenes are unlikely to have arisen from differences in low-level properties. Any influence of object masking on these properties should have taken the same sign for both high- and low-spaciousness scenes, whereas the influence of objects along the PPA space-coding dimension operated in opposite directions depending on scene spaciousness. While PPA activity has been shown previously to relate to spatial properties of scenes (Kravitz et al. 2011; Park et al. 2011; Harel et al. 2012) and to human judgments of scene category (Peelen et al. 2009; Walther et al. 2009), the present study joins a very small group demonstrating that PPA activity tracks scenes' encoded spatial properties even when those properties depart from physical reality (Park et al. 2007; Chadwick et al. 2013).

Viewed in the framework of our crosstalk theory, our results suggest that PPA, at least in the right hemisphere, is the brain area in which encoded spatial properties of scenes are brought into alignment with expectations derived from scenes' object contents. Our assignment to the PPA of this role as a junction point between codes for objects and spatial properties is consistent with its recent characterization (Harel et al. 2012) as the midpoint in a hierarchy of scene processing regions which ranges from the purely object-sensitive LOC to the purely space-sensitive RSC. Moreover, our results offer an alternative perspective on the contentious issue of the origin of object-evoked activity in the PPA (Aminoff et al. 2007; MacEvoy and Epstein 2009; Harel et al. 2012; Troiani et al. 2012), which has alternately been explained in terms of either object-triggered spatial representations (Epstein and Ward 2010; Mullally and Maguire 2011, 2013) or contextual associations among objects (Bar and Aminoff 2003; Aminoff et al. 2007, 2013). Our results suggest that the presence of object-evoked activity in the PPA also reflects the object information necessary for centripetal bias to take place. Indeed, as our crosstalk theory is based on associations between spatial properties and nonspatial information (i.e., object identity), the role we ascribe to the PPA as the effector of centripetal bias appears consistent with both the context- and layout-centered views of its function in scene processing.
Neither our ROI-level nor searchlight analyses showed a similar effect of objects on encoded spatial properties in the left PPA. This laterality can be separated into 2 distinctions between the right and left PPA. First, unlike patterns from the right PPA, patterns from the left PPA failed to pass even the basic test of distinguishing significantly among unmasked scenes on the basis of spatial properties: Pattern distances between high- and low-spaciousness unmasked exemplars were not significantly greater than pattern distances between those exemplars and the average-spaciousness exemplars. The reason for this failure is not clear, although some research has suggested that left parahippocampal cortex may have a relatively reduced capacity for spatial processing (Wagner et al. 1998; Kirchhoff et al. 2000; Epstein et al. 2003; Stevens et al. 2012), which may have been less apparent in previous MVPA studies of PPA spatial sensitivity that used scenes spanning a much greater range of spatial properties than those we used (Kravitz et al. 2011). Second, we observed no significant effect of object masking on relationships among left PPA activity patterns. This potentially reflects the demonstrated greater sensitivity of right parahippocampal cortex to the specific visual contents of scenes, contrasting with a greater capacity for abstraction in the left parahippocampal cortex (Koutstaal et al. 2001; Xu et al. 2007; Stevens et al. 2012).

While the medial temporal cluster identified by our searchlight analysis fell within the boundaries of group-defined PPA, it is positioned markedly anteriorly in the parahippocampal cortex. Our results thus join a growing set of findings suggesting that PPA is differentiable along its rostrocaudal axis in terms of both response properties (Rajimehr et al. 2011; Nasr et al. 2013) and connectivity (Ballassano et al. 2013), and dovetails very closely with some. For instance, anterior PPA has been recently shown to be much less sensitive to objects than posterior PPA (Ballassano et al. 2013), and less sensitive to the high spatial frequencies that might convey information about objects (Rajimehr et al. 2011). While this might appear at first to conflict with our searchlight mapping showing the most prominent effect of objects in anterior PPA, it is important to remember that the searchlight analysis identified subregions whose patterns were biased by the presence of objects in scenes, not necessarily those which contained information about the identities of the objects. Furthermore, inasmuch as the centripetal bias would appear to be a rather high-level refinement of spatial property codes, it makes sense that it would be found in the anterior PPA, which shares a greater degree of connectivity with fronto-parietal networks than posterior PPA (Ballassano et al. 2013). It is noteworthy in this regard that the only other area which showed evidence of centripetal bias in our searchlight analysis was a cluster in prefrontal cortex. Whether this indicates some functional association with PPA is unclear, but to the extent that it might, we are inclined toward the view that it results from prefrontal mirroring of a centripetal bias that arises in the PPA, potentially reflecting the channel through which PPA spatial codes contribute to categorical decisions.

In summary, although scenes’ spatial properties and object contents are formally independent descriptors of scenes, both our behavioral and fMRI results show that this theoretical independence is not respected by the visual system. While it has long been appreciated that objects can influence judgments of scene category, the biasing influence of objects on encoded spatial properties that we observed has not been previously described nor explicitly predicted by scene recognition models. We propose that this bias reflects a system of object/spatial property crosstalk supporting generation of unified judgments of scene category by reducing potential categorization conflicts. Further perceptual and neuroimaging experiments will be necessary to understand the neuroanatomical basis of this phenomenon, and to explicitly test the hypothesis that it aids the accuracy and speed of scene recognition.

Supplementary Material
Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding
This work was funded by Boston College.

Notes
The authors thank Lauren Beebe, Chris Gagne, Emilie Josephs, and Molly LaPoint for assistance with stimulus generation, Zoe Yang for assistance with data collection, and Russell Epstein for comments on the manuscript. Conflict of Interest: None declared.

References

Cerebral Cortex August 2015, V 25 N 8 2279

