



Functional Context Affects Scene Processing

Elissa M. Aminoff¹  and Michael J. Tarr²

Abstract

■ Rapid visual perception is often viewed as a bottom-up process. Category-preferred neural regions are often characterized as automatic, default processing mechanisms for visual inputs of their categorical preference. To explore the sensitivity of such regions to top-down information, we examined three scene-preferring brain regions, the occipital place area (OPA), the parahippocampal place area (PPA), and the retrosplenial complex (RSC), and tested whether the processing of outdoor scenes is influenced by the functional contexts in which they are seen. Context was manipulated by presenting real-world landscape images as if being viewed through a window or within a picture frame—manipulations that do not affect scene content but do affect one's functional knowledge regarding the scene. This manipulation influences neural scene processing (as measured by

fMRI): The OPA and the PPA exhibited greater neural activity when participants viewed images as if through a window as compared with within a picture frame, whereas the RSC did not show this difference. In a separate behavioral experiment, functional context affected scene memory in predictable directions (boundary extension). Our interpretation is that the window context denotes three-dimensionality, therefore rendering the perceptual experience of viewing landscapes as more realistic. Conversely, the frame context denotes a 2-D image. As such, more spatially biased scene representations in the OPA and the PPA are influenced by differences in top-down, perceptual expectations generated from context. In contrast, more semantically biased scene representations in the RSC are likely to be less affected by top-down signals that carry information about the physical layout of a scene. ■

INTRODUCTION

Although rapid visual perception is often considered as a primarily bottom-up process, it is well established that the processing of visual input involves both bottom-up and top-down mechanisms (Kay & Yeatman, 2017; Fang, Boyaci, Kersten, & Murray, 2008; Lamme & Roelfsema, 2000; Felleman & Van Essen, 1991). For example, the responses of the scene-selective network of category-preferred brain regions are affected by top-down information regarding learned contextual associations (Bar & Aminoff, 2003). This network of regions, the parahippocampal place area (PPA)/lingual region (Epstein & Kanwisher, 1998), the retrosplenial complex (RSC; Maguire, 2001), and the occipital place area (OPA; also known as the transverse occipital sulcus; Dilks, Julian, Paunov, & Kanwisher, 2013), appears to represent a wide variety of scene characteristics (reviewed in Epstein & Baker, 2019). The list of scene-relevant properties includes spatial layout, three-dimensionality, landmark processing, navigability, environment orientation and retinotopic bias, scene boundaries, scene categories, objects within a scene, and the contextual associative nature of the scene (Lescroart & Gallant, 2019; Lowe, Rajsic, Gallivan, Ferber, & Cant, 2017; Baldassano, Fei-Fei, & Beck, 2016; Çukur, Huth, Nishimoto, & Gallant, 2016;

Julian, Ryan, Hamilton, & Epstein, 2016; Aminoff & Tarr, 2015; Marchette, Vass, Ryan, & Epstein, 2015; Park, Konkle, & Oliva, 2015; Silson, Chan, Reynolds, Kravitz, & Baker, 2015; Troiani, Stigliani, Smith, & Epstein, 2014; Harel, Kravitz, & Baker, 2013; Auger, Mullally, & Maguire, 2012; Nasr & Tootell, 2012; Henderson, Zhu, & Larson, 2011; Kravitz, Peng, & Baker, 2011; Park, Brady, Greene, & Oliva, 2011; Bar, Aminoff, & Schacter, 2008; Janzen & van Turennout, 2004; Levy, Hasson, Avidan, Hendler, & Malach, 2001).

One of the significant open questions regarding the representation of scene properties is how they come to be encoded; that is, to what extent are the associated neural responses driven by visual properties within scenes as opposed to nonperceptual high-level scene properties, such as learned functional properties¹ and semantics? We address this question by exploring whether prior experience and expectations modulate scene-selective neural activity.

We used fMRI to measure neural responses while participants viewed the otherwise identical outdoor scenes in two different contexts: in a window frame (“WIN” condition) or in a picture frame (“PIC” condition; Figure 1). We hypothesize that viewing scene images surrounded by a window invokes a more naturalistic context that is closer to the perceptual experience of real-world scene processing. More specifically, a window connotes that the scene is 3-D, navigable, and extends beyond the boundaries presented. In contrast, we hypothesize that viewing scene images surrounded by a picture frame

¹Fordham University, ²Carnegie Mellon University



Figure 1. Sample stimuli showing the same scenes in both the picture frame (PIC) and the window frame (WIN) conditions. See Methods for more information.

invokes a less realistic context in which the scene is viewed as a 2-D picture without extension beyond the frame; as a consequence, inferential scene properties such as spatial affordances are likely to be limited. Based on these assumptions, we predict that the perception of a scene image will vary based on the context in which the image is situated. Under the assumption that the network of scene-preferred brain regions (PPA, RSC, and OPA) subserves different computational functions, we also predict that these regions will respond differently from one another across the manipulation of scene context. Alternatively, if scene preference is purely a function of scene content, one should predict no differences in responses across these regions.

To further explore the effect of functional context, we examined how the picture frame versus window frame manipulation affects boundary extension—a well-documented distortion of scene memory (Intraub, 2010, 2014; Intraub & Richardson, 1989). Boundary extension has been discussed as a memory distortion directly related to scene representation—a phenomenon that is intertwined with the spatial affordances arising from the process of scene perception applied to picture viewing (Intraub, 2010, 2020). When we experience a real-world scene via either direct viewing or a picture, we are not just perceiving the scene as a finite entity but as a percept that continues beyond the edges of our perception. Thus, if we manipulate the functional context of scenes by presenting them explicitly in picture frames, we are limiting the spatial context necessary for scene understanding and boundary extension should be reduced. As such, we predicted greater boundary extension for window-framed scenes as compared with picture-framed scenes.

More broadly, the manipulation of functional context addresses the question of whether scene-preferred brain regions process category-relevant inputs in a primarily bottom-up manner or whether they are sensitive to top-down influences. At the same time, the pattern of neural modulation across different scene-preferred brain regions adds to our understanding of the different functional roles for each.

METHODS

fMRI Experiment

Participants

Eighteen individuals participated in this experiment; 17 were included in the analysis (mean age = 23.6 years, range = 18–30 years; 8 women, 9 men; 1 left-handed). One participant was removed from the analysis because of extremely poor performance, indicative of falling asleep (missing 22% of the repeated trials in a trivial 1-back task). All participants had normal or corrected-to-normal vision and were not taking any psychoactive medication. Written informed consent was obtained from all participants before testing in accordance with the procedures approved by the institutional review board of Carnegie Mellon University. Participants were financially compensated for their time.

Stimuli

The main experiment included 120 outdoor scenes, including both manmade outdoor scenes such as a garden patio, as well as natural landscapes such as a mountain range. A majority of the stimuli were found and obtained through Google Image Search. There were two versions of each scene: one within the context of a window frame and the other within the context of a picture frame (see Figure 1).

A pool of 13 window frames and 13 picture frames was used across the 120 scenes. Each scene presented within the frame subtended 5.5° of visual angle, and the average extent of the frames was 9° with 0.68° (WIN) and 0.61° (PIC) standard deviations across the different frame exemplars. The frames were set against a gray rectangular background that subtended 10° of visual angle; the remainder of the screen background was black.

In a post hoc analysis, the brightness, contrast, and spatial frequency were measured for all stimulus images. Images in the PIC and WIN conditions were found to be matched across contrast and spatial frequency. However, there was a difference in brightness with PIC images brighter on average than WIN images.

Stimuli in the localizer experiment included 60 scenes (outdoor and indoor, nonoverlapping with the stimuli used in the main experiment), 60 weak contextual objects (Bar & Aminoff, 2003), and 60 phase-scrambled scenes. Phase-scrambled scenes were generated by running a Fourier transform of each scene image, scrambling the phases, and then performing an inverse Fourier transform back into the pixel space. All stimuli were presented at a 5.5° visual angle against a gray background.

Procedure

During fMRI scanning, images were presented to the participants via 24-in. MR compatible LCD display (BOLDScreen, Cambridge Research Systems LTD.) located at the head of the bore and reflected through a head coil mirror to the participant. There were two functional runs in the WIN/PIC experiment. Functional scans used a blocked design alternating WIN blocks and PIC blocks with fixation in between. The order of the blocks was balanced both across and within participants. Each functional scan began and ended with 12 sec of a white fixation cross (“+”) presented against a black background. Images were presented for 750 msec, with a 250-msec ISI. Each block contained 10 unique images and two repeated images, for a total block duration of 12 sec. Each run consisted of six blocks per condition. There were 10 sec of fixation between task blocks. Participants performed a 1-back task where they pressed a button if the picture immediately repeated, two per block. Each run presented all 120 stimuli, 60 presented in the WIN condition, and 60 presented in the PIC condition. The second run presented all 120 stimuli again, but with the presentation condition (PIC or WIN) swapped. The condition in which a stimulus was presented first was balanced across participants.

Most participants had two functional localizer runs (two participants had only one run because of time constraints) to functionally define scene-preferred regions.² Localizer runs consisted of three conditions: scenes, objects, and phase-scrambled scenes. These runs began and ended with 12 sec of a black fixation cross (“+”) presented against a gray background. Each run had four blocks per condition. Images were presented for 800 msec, with 200-msec ISI, with the exception that the first stimulus in each block other than the first block was presented for 2800 msec. Each block contained 12 unique images with two repeated images, for a total block duration of 14 sec for the first block and 16 sec thereafter because of the longer presentation of the first stimulus. There were 10 sec of fixation between task blocks. Participants performed a 1-back task where they pressed a button if the picture immediately repeated, two per block. The localizer runs occurred after the WIN/PIC functional runs.

fMRI Data Acquisition

fMRI data were collected on a 3T Siemens Verio MR scanner at the Scientific Imaging and Brain Research Center at

Carnegie Mellon University using a 32-channel head coil. Functional images were acquired using a T2*-weighted echo-planar imaging multiband pulse sequence (69 slices aligned to the AC/PC, in-plane resolution 2 mm × 2 mm, 2 mm slice thickness, no gap, repetition time [TR] = 2000 msec, echo time [TE] = 30 msec, flip angle = 79°, multiband acceleration factor = 3, field of view = 192 mm, phase encoding direction A >> P, ascending acquisition). Number of acquisitions per run was 139 for the WIN/PIC runs and 162 for the scene localizer. High-resolution anatomical scans were acquired for each participant using a T1-weighted MPRAGE sequence (1 mm × 1 mm × 1 mm, 176 sagittal slices, TR = 2.3 sec, TE = 1.97 msec, flip angle = 9°, GRAPPA = 2, field of view = 256). A field-map scan was also acquired to correct for distortion effects using the same slice prescription as the EPI scans (69 slices aligned to the AC/PC, in-plane resolution 2 mm × 2 mm, 2 mm slice thickness, no gap, TR = 724 msec, TE1 = 5 msec, TE2 = 7.46 msec, flip angle = 70°, field of view = 192 mm, phase encoding direction A >> P, interleaved acquisition).

fMRI Data Analysis

All fMRI data were analyzed using SPM12 (www.fil.ion.ucl.ac.uk/spm/software/spm12/). All data were preprocessed to correct for motion and to unwarped for geometric distortions using the field-map scan acquired. Data were smoothed using an isotropic Gaussian kernel (FWHM = 4 mm). Only data used for the group average activation maps were normalized to the Montreal Neurological Institute template. Otherwise, data used were in native space (i.e., all ROI analyses). The data were analyzed as a block design using a general linear model and canonical hemodynamic response function. A high-pass filter using 128 sec was implemented. The six motion parameter estimates that output from realignment were used as additional nuisance regressors. An autoregressive model of order 1, AR(1), was used to account for the temporal correlations of the residuals. For the whole-brain analysis in the group average, the contrasts were passed to a second-level random-effects analysis that consisted of testing the contrast against zero using a voxel-wise single-sample *t* test. All group-averaged activity maps are examined through a whole-brain analysis using a false discovery rate correction of $q = .05$. For visualization purposes, these average maps were rendered onto a 3-D inflated brain using CARET (Van Essen et al., 2001).

All ROIs analyzed were defined and extracted at the individual level using the MarsBaR toolbox (marsbar.sourceforge.net/index.html) or in-house MATLAB (The MathWorks) scripts and analyzed in native space. Scene-preferred regions (PPA, RSC, and OPA) were functionally defined using the contrast of scenes greater than the combined conditions of objects and phase-scrambled scenes from the localizer runs. Typically, a threshold of family-wise error, $p < .001$, was used to define the set of voxels.

In a post hoc analysis, the effect of stimulus brightness was evaluated. To test whether stimulus brightness contributed to any of our observed effects, we measured the mean brightness across all images within a block (as presented to the individual participant during the fMRI run). Blocks within the same frame condition (PIC, WIN) were separated into the brighter blocks ($n = 6$) and the darker blocks ($n = 6$), thereby yielding four conditions: PIC Bright, PIC Dark, WIN Bright, and WIN Dark. Conditions were compared to determine whether the differences in the WIN and PIC conditions could be accounted for by image brightness.

Behavioral Experiment

Participants

Thirty-seven individuals participated in the behavioral experiment examining boundary extension. Data from 36 individuals were included in the analysis, one participant was removed because of a technical error related to which buttons were pressed. The participants were undergraduates at Fordham University who were either paid for their participation or received course credit (mean age = 20.0 years, $SD = 1.36$ years, range = 18–22 years; 28 women, 7 men; 4 left-handed). Written informed consent was obtained from all participants before testing in accordance with the procedures approved by the institutional review board of Fordham University.

Stimuli

The stimuli for this experiment were 200 unique scenes, which included the 120 scenes used in the fMRI experiment as well as an additional 80 outdoor scenes added to increase the total number of trials. As in the fMRI experiment, there were two formats for each scene: one within the context of a window frame (WIN) and the other in the context of a picture frame (PIC). The same pool of window frames and picture frames from the fMRI experiment was applied to the 80 new pictures. Pictures were divided into two groups of 100 scenes, Group A and Group B. Images were presented to the participants on a 27-in. iMac using Psychtoolbox (Brainard, 1997) and MATLAB.

Procedure

Participants were instructed to memorize all of the scenes presented in the experiment. In the study phase, a single scene image was presented on each trial, and participants judged whether there was water in the picture. Each trial was composed of a white fixation cross presented against a gray background for 250 msec, a scene presented for 250 msec, and a repeat of the fixation cross for 250 msec. Following the second fixation cross, participants viewed a response screen showing: “(b) Water (n) No Water.” Participants had up to 2500 msec to respond with the

appropriate key press (b or n). Immediately after the participant responded, the next trial started.

Trials were broken into blocks of 25 trials, between which participants were offered a break. Each block consisted of pictures from a single condition, either PIC or WIN. Condition order alternated, starting with the WIN condition. Group A stimuli were presented in the WIN condition, and Group B stimuli were presented in the PIC condition. After 200 trials—a total of eight blocks, four from each condition—participants’ memory for the scenes was tested.

In the test phase, a fixation cross was presented for 250 msec, followed by a picture of a scene shown during the study phase, except without a frame. Participants judged whether the scene was identical to the version they had seen at study (absent the frame), was zoomed in (i.e., closer) relative to the version they had seen at study, or was zoomed out (i.e., wider) relative to the version they had seen at study. Participants responded on a 5-point scale: *very close*, *close*, *same*, *wide*, and *very wide*. The response screen was self-paced. After participants judged the amount of “zoom,” they rated their confidence on a 3-point scale: *sure*, *pretty sure*, or *don’t remember picture*. This screen was self-paced as well. Trials were broken into blocks of 25 trials, and as before, each block consisted of pictures from a single condition, either PIC or WIN. All scenes presented in the test phase were actually shown with the “same” boundaries as presented in the study phase—that is, with no zoom in or out. Thus, the correct answer was always “same.”

After the 200 test trials, participants were presented with another 200 study and 200 test trials using the same 200 scenes, but appearing in the opposite condition at study as compared with the first study/test session. Here, Group A stimuli appeared in the PIC condition, and Group B stimuli appeared in the WIN condition. The condition order again alternated across blocks, but here, starting with the PIC condition. Although presentation order was randomized for both sessions, a technical bug resulted in the stimuli and order of conditions not being balanced across conditions. See Results for detailed analysis demonstrating that this error did not affect the results.

Responses at test were converted to an integer score from -2 to $+2$ (corresponding to *very close*, *close*, *same*, *wide*, and *very wide*), where positive values denote when participants perceived the scene at test to be “wider” than they remembered seeing it at study (i.e., boundary contraction), zero represents no change from study to test, and negative values denote when participants perceived the scene at test to be “closer” than they remembered seeing it at study (i.e., boundary extension). Scores were summed across all test trials separately for the WIN and PIC conditions. Responses with RTs exceeding 3 SD s from the participant’s mean were considered outliers and removed from the analysis. A t test (WIN/PIC) was performed on these summed scores. A second analysis was run based on the confidence of the participant. If the participant responded “Don’t remember picture,” that trial

was removed from the analysis to ensure any effects arose from the frame context manipulation and not a failure of memory.

RESULTS

fMRI Experiment

We hypothesized that the PIC versus the WIN context manipulation would give rise to different top-down driven inferences—reflected in responses in scene-preferred brain regions—about the nature of the viewed scene. Neural responses were measured using fMRI in a block design, and we performed a whole-brain analysis comparing the BOLD activity elicited by WIN versus PIC blocks. This comparison revealed no voxel responses with larger magnitudes for the PIC as compared with the WIN condition (false discovery rate threshold at $q = .05$). In contrast, there were many voxel responses of larger magnitude for the WIN as compared with the PIC condition. These voxels were located within the dorsal visual stream, within the occipital cortex, and within the parietal cortex, close to the inferior portion (Figure 2).

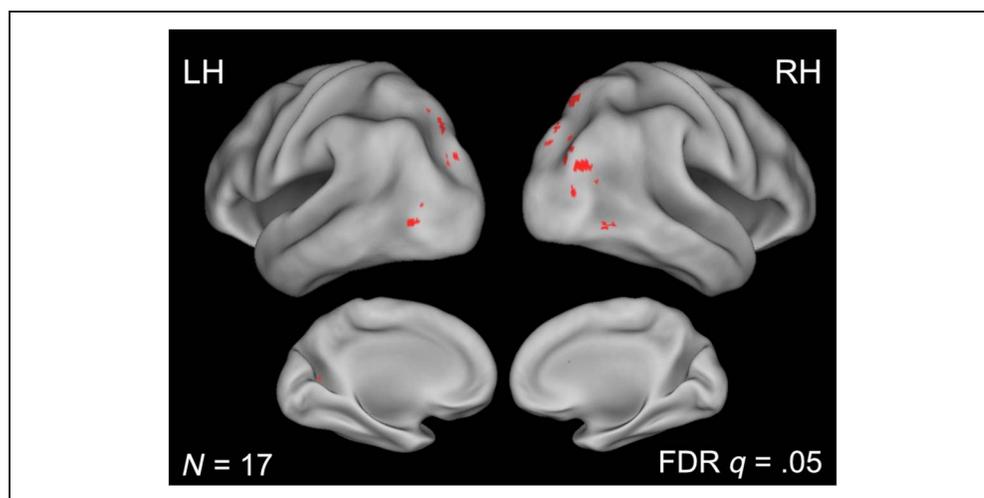
We next examined how our context manipulation affects different scene-preferred brain regions (Figure 3). An independent functional localizer was used to define ROIs commonly observed to be selective for scene processing—PPA, RSC, and OPA. An ANOVA with ROI \times Hemisphere \times Condition as factors revealed a significant main effect of Condition, with WIN eliciting more activity than PIC, $F(1, 16) = 11.83, p < .003, \eta_p^2 = .425$. There was also a main effect of ROI, $F(2, 32) = 85.02, p < 1.57 \times 10^{-13}, \eta_p^2 = .842$, with the PPA showing the highest magnitude response (2.3 parameter estimate) as compared with either the OPA (1.9 parameter estimate, $p < .001$ in planned comparisons) or the RSC (0.89 parameter estimate, $p < .0001$); the OPA response was also significantly higher than the RSC response ($p < .0001$). The effect of Hemisphere was significant, with the right hemisphere eliciting more activity than the left hemisphere, $F(1, 16) =$

19.07, $p < .0005, \eta_p^2 = .544$. There was also a significant interaction between ROI \times Condition, $F(2, 32) = 10.95, p < .0003, \eta_p^2 = .407$. Pairwise ROI \times Condition comparisons revealed that this interaction was driven by significant differences between both the PPA and OPA as compared with the RSC: PPA versus RSC, $F(1, 16) = 21.26, p < .0003, \eta_p^2 = .571$; OPA versus RSC, $F(1, 16) = 15.09, p < .001, \eta_p^2 = .485$. There was no significant effect when comparing the PPA to the OPA, $F(1, 16) = 0.080, p > .78, \eta_p^2 = .005$. No other interactions were significant.

To explore the effect of the context manipulation within each specific scene-preferred region, we ran separate ANOVAs for each ROI (Hemisphere \times Condition). In the PPA, there was a significant main effect of Condition, $F(1, 16) = 12.45, p < .003, \eta_p^2 = .438$, with WIN eliciting significantly more activity than PIC. There was also a significant difference in Hemisphere, $F(1, 16) = 17.72, p < .001, \eta_p^2 = .526$, with the right hemisphere showing more activity than the left hemisphere. The interaction was not significant ($p > .9$). In the OPA, there was a significant main effect of Condition, $F(1, 16) = 33.71, p < .00003, \eta_p^2 = .678$, with WIN eliciting significantly more activity than PIC. Neither the main effect of Hemisphere nor the Hemisphere \times Condition interaction were significant ($ps > .15$). In the RSC, there was no significant main effect of Condition ($p > .24$) nor any interaction between Hemisphere \times Condition. However, there was a main effect of Hemisphere, with the right-hemisphere response being greater than the left-hemisphere response, $F(1, 16) = 11.27, p < .004, \eta_p^2 = .413$.

Presentation order effects were explored by comparing Runs 1 and 2—where the same scene images appeared in different contexts. An ANOVA for each ROI was run with Hemisphere \times Condition \times Run as factors. Suggesting that order made no difference in neural responses, the main effect of Run was insignificant for each ROI ($p > .18, \eta_p^2 < .11$), as was the interaction between Condition \times Run ($p > .14, \eta_p^2 < .14$). The interaction of Hemisphere \times Run was not significant in the RSC ($p > .68, \eta_p^2 < .01$), was marginally significant for the PPA ($p < .07, \eta_p^2 < .19$), and was

Figure 2. Whole-brain analysis examining activity elicited for scenes in window frames (WIN) as compared with the activity for scenes in picture frames (PIC).



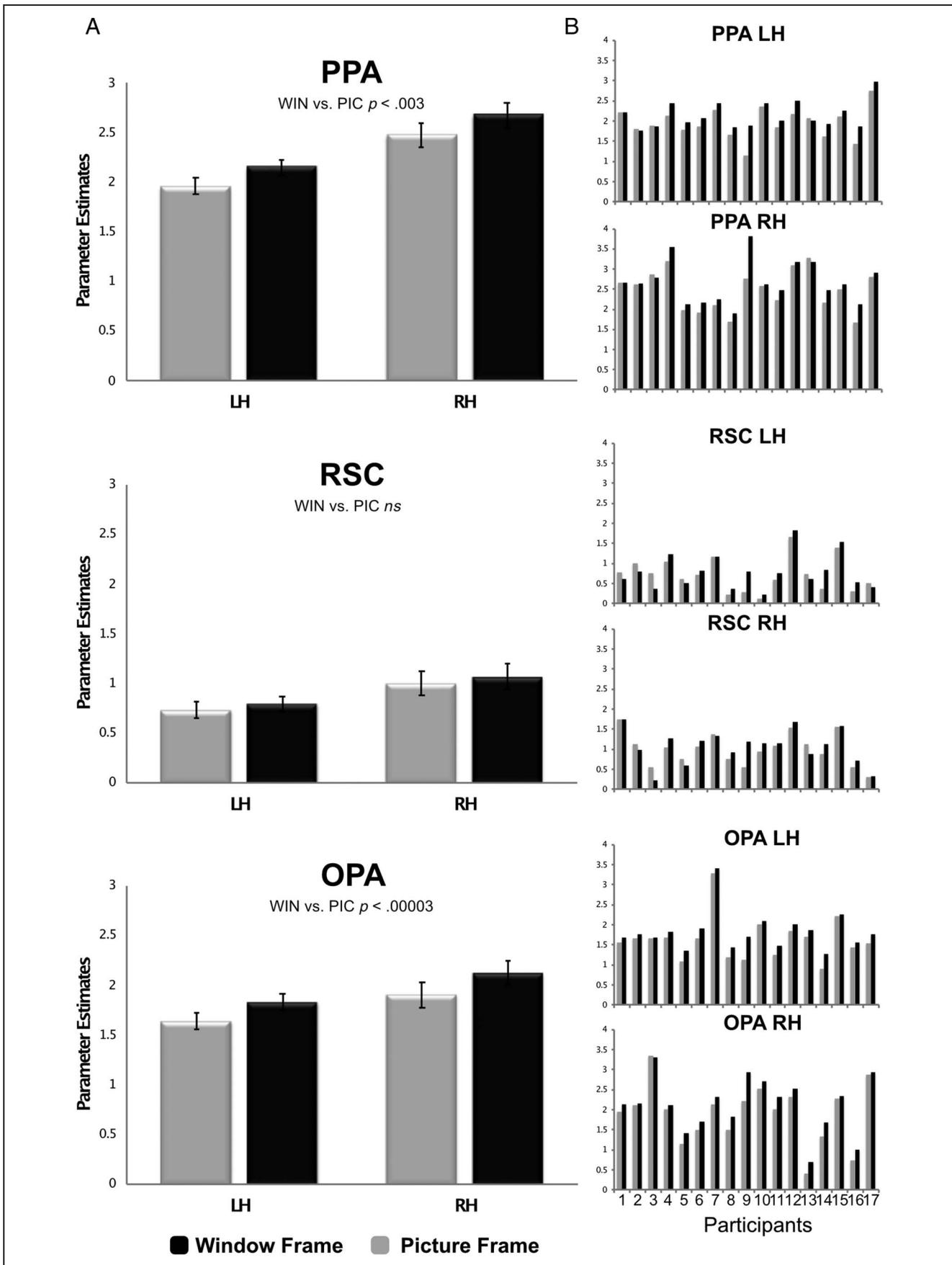


Figure 3. ROI analyses for both the group average (A) and individual participants (B). WIN condition = black; PIC condition = gray.

significant in the OPA ($p < .02$, $\eta_p^2 < .31$). The overall pattern does show greater activity in Run 1 as compared with Run 2, which is consistent with adaptation to the stimuli, regardless of condition. However, we found this effect to be modulated by hemisphere. In the PPA, the effect of adaptation was marginally greater in the left hemisphere than in the right hemisphere (Run 1 minus Run 2: left hemisphere 0.14, right hemisphere 0.05). In the OPA, adaptation was again observed in the left hemisphere (0.11); however, in the right hemisphere, there was slightly greater activity in Run 2 compared with Run 1, yielding the significant interaction (right hemisphere -0.02). The three-way interaction of Hemisphere \times Condition \times Run was not significant (PPA, $p < .94$, $\eta_p^2 < .0$; RSC, $p < .34$, $\eta_p^2 < .06$; OPA, $p < .07$, $\eta_p^2 < .2$).

A significant Hemisphere effect was found in a number of our analyses. However, our main manipulation of interest (WIN vs. PIC) did not interact with Hemisphere. However, our results do reflect a preference for scene processing in the right hemisphere—an effect that is difficult to compare to prior findings in that many studies examining scene selectivity collapse across hemispheres without statistical support. As such, the pervasiveness of this hemispheric effect is unknown. We suggest several reasons for observing a hemispheric difference in our study. First, the left hemisphere may preferentially process high spatial frequencies, whereas the right hemisphere may preferentially process low spatial frequencies (for a review, see Kauffmann, Ramanoël, & Peyrin, 2014). Low spatial frequencies have a unique role in the rapid processing of contextual and scene information (Greene & Oliva, 2009; Bar, 2004; Oliva & Torralba, 2001). Second, the right hemisphere may be biased toward perceptual properties of a scene, whereas the left hemisphere may be biased toward conceptual information (Stevens, Kahn, Wig, & Schacter, 2012; van der Ham, van Zandvoort, Frijns, Kappelle, & Postma, 2011). However, this difference would not seem to be able to account for why, in our study, scene processing recruits the right hemisphere preferentially, in that performing the 1-back task would seem to recruit both perceptual and conceptual information and that both levels of description are relevant to judging whether one image matches another.

A post hoc analysis was run to test whether differences in brightness accounted for the observed effects. When overall image brightness was considered as a separate factor, we failed to find any significant effect of brightness (PIC Bright = PIC Dark, WIN Bright = WIN Dark, $ps > .25$). Moreover, in 13 of the 17 participants, we were able to equate brightness across the PIC and WIN conditions, allowing us to directly compare the PIC and WIN conditions with equal average brightness for the images across the two conditions. Despite equivalent average brightness, we again found the predicted significant effect of context: left-hemisphere PPA, $t(12) = 2.40$, $p < .033$; left-hemisphere OPA, $t(12) = 3.54$, $p < .004$; right-hemisphere PPA, $t(12) = 2.69$, $p < .02$; right-hemisphere

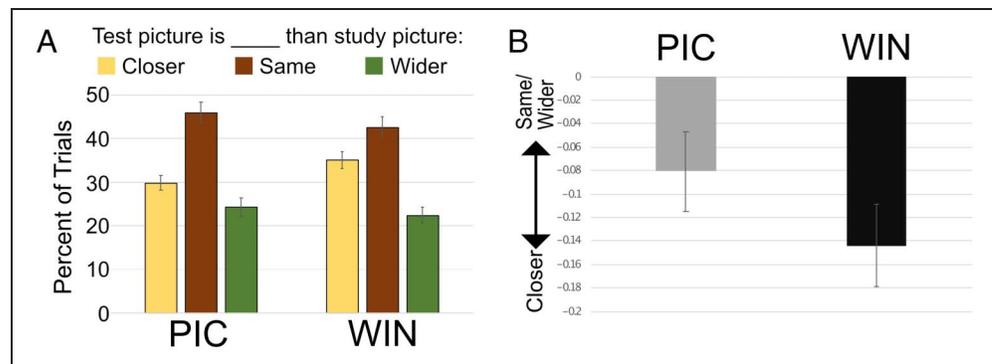
OPA, $t(12) = 4.17$, $p < .001$; left- and right-hemisphere RSC, *ns*). As such, we conclude that differences in low-level properties do not underlie our contextual interpretation of the observed differences between conditions.

Behavioral Experiment

Our neuroimaging results suggest that window frames render scene images more “scene-like”—that is, perceived as more realistic. But what does “more realistic” entail? Viewing a scene in a window frame versus a picture frame affects the functional context and thus the associated spatial affordances. More specifically, a scene in a picture frame is understood in the functional context of “what is in the picture is what is important,” whereas a scene in a window is understood to be only a part of the overall scene. For example, when we view only part of a real-world scene (e.g., the position of a bed in a bedroom), we know to turn our head to perceive and interpret additional features of the scene (e.g., the location of the closet). Under this view, we predict that differences found in the neural representations of the WIN and PIC scene conditions should also manifest in behavioral measures of scene perception because of these differences in functional context. In particular, boundary extension is a phenomenon where observers remember scenes with wider boundaries (i.e., more zoomed out) than what was originally experienced (Intraub, 2014; Intraub & Richardson, 1989). The boundary extension phenomenon is held to be specific to scene memory (for an alternative account, see Bainbridge & Baker, 2020). Moreover, there is evidence that boundary extension manipulations also recruit the PPA (Chadwick, Mullally, & Maguire, 2013; Park, Intraub, Yi, Widders, & Chun, 2007). As such, we do see consistency across boundary extension studies and our fMRI experiment in that PPA appears to correlate with BE and the observed significant recruitment of the PPA for our frame manipulation. Here, on the basis of the assumed differences between the window and picture frame contexts, we hypothesized a larger boundary extension effect for scenes presented in windows than for scenes presented in picture frames. This context manipulation—the same as used in our fMRI experiment—was included during the study phase of this experiment. During the subsequent test phase, the same scenes were presented without any frame, and participants’ memory was probed via reports as to whether each scene was identical (minus the frame) to its presentation at study, zoomed in (i.e., closer), or zoomed out (i.e., wider).

Across both study contexts, participants remembered the scene at test as being closer than what was actually presented at study (i.e., boundary extension; 32% of the trials) more often than the scene at test being farther than at study (i.e., boundary contraction; 23% of the trials)—a significant difference, $t(35) = 3.3$, $p < .002$. Relevant to our hypothesis, participants more often remembered that scenes in the WIN condition were closer at test relative

Figure 4. Boundary extension results. (A) Percentage of trials at test the participants thought the test image was closer, the same as, or wider than the study image. (B) The average converted bias scores where negative denotes that responses were biased to remember the test image as closer than what was actually presented at study.



to scenes in the PIC condition (35% vs. 30% of test trials; Figure 4). To measure this bias in scene memory, we computed an average based on the integer values assigned to each response (see Methods): The bias score for the WIN condition was -0.14 , whereas the bias score for the PIC condition was -0.08 (Figure 4). This difference in memory bias indicates that participants were more likely to remember the WIN scenes as wider compared with the PIC scenes, $t(35) = 2.85, p < .007$. We also examined the bias removing any trials in which the participants responded “Don’t remember picture” in their confidence judgment. Again, we observed a difference in memory bias: The bias score for the WIN condition was -0.15 , whereas the bias score for the PIC condition was $-0.09, t(35) = 2.96, p < .006$. These results support our prediction that scenes in a window frame context will elicit a greater boundary extension effect—consistent with the greater scene-selective neural responses observed in our fMRI study.

Presentation order effects were explored by comparing the two study/test sessions where the same scene images appeared in counterbalanced contexts. The main effect of Session was not significant, $F(1, 35) = 1.159, p = .289, \eta_p^2 = .032$; the main effect of Condition was significant (PIC or WIN), $F(1, 35) = 8.808, p < .007, \eta_p^2 = .188$; and there was a significant interaction, $F(1, 35) = 14.23, p < .001, \eta_p^2 = .289$. This interaction reflects similar boundary extension across conditions in the first session (WIN = $-.13, PIC = -.14$), whereas in the second session, there was stronger boundary extension for the WIN condition (WIN = $-.16, PIC = -.02$). We believe that this session interaction may be a consequence of a counterbalancing error—an issue that we further address next.

As mentioned in Methods, a technical error meant that the stimuli were not balanced across sessions or participants. Scenes were split into two static groups (A and B) across all participants. Group A was always shown first in the WIN condition, and Group B was always presented first in the PIC condition. To examine whether this contributed to the observed interactions, we performed an item analysis to investigate whether specific scenes consistently elicited greater boundary extension regardless of condition. Or critically, whether the “same scene” elicits greater boundary extension in the WIN condition as compared with the PIC condition. In this item analysis, we replicated

the overall effect of boundary extension across all stimuli and all conditions, mean = $-.11, t(199) = -4.15, p < .00005$, as well as a greater boundary extension effect for each scene in the WIN condition as compared with the PIC condition (WIN = $-.14, PIC = -.08, t(199) = 2.969, p < .003$). To rule out an effect driven by specific scenes, we compared the boundary extension of Group B—presented in the second session in the WIN condition—with Group A. When collapsing across the PIC and WIN conditions, both Groups A and B showed an overall boundary extension effect (A = $-.08, B = -.15$; no significant difference), $t(99) = 1.438, p = .15$, indicating that our observed context manipulation effects were not the result of any imbalance in which scenes appeared in which condition, but rather the result of the manipulation itself. However, Group B did elicit greater overall boundary extension (even in the PIC condition, although, critically, still greater for the WIN condition), which may have reduced the difference between PIC and WIN observed in the first presentation, yielding the significant interaction with session mentioned above. Overall, the item analysis provides further evidence that functional context affects how scenes are processed and perceived.

DISCUSSION

Rapid scene understanding is often construed as a feedforward process in which category-preferred neural substrates are mandatorily recruited. At the same time, there is clear evidence for high-level properties influencing scene perception (Biederman, Mezzanotte, & Rabinowitz, 1982; Biederman, 1981). We built on the idea of high-level knowledge influencing scene processing by asking whether the functional context in which a given scene is viewed (as opposed to the scene content in and of itself) affects scene perception. To address this question, we examined whether there is a difference in scene-selective neural responses when viewing a scene through a window as compared with in a picture frame. We found that two scene-preferring regions of the brain, the OPA and the PPA, respond differently when otherwise identical scenes are viewed in these two contexts. Consistent with the conception of these brain regions supporting real-world scene understanding, the more ecologically valid context,

through a window, elicited stronger neural responses as compared with the more artificial context, in a picture frame. These results support the proposal that high-level, top-down knowledge—even extraneous to the scene content—influences scene processing. We posit that this effect arises as a result of the window context triggering a set of task-related expectations with respect to scenes that modulate the manner in which the visual system processes incoming scene information.

Why should the context specified by the frame affect how we process scenes? In both conditions, each scene is a 2-D picture that participants are viewing on a screen. It seems highly unlikely that participants perceive the window-framed picture as if it were a real scene being viewed through a window (e.g., eventually seeing something move in the scenery). At the same time, statistical inference plays an important role in perception, and a variety of associations may automatically come into play because they are coupled with specific features (i.e., window frames). In our present experiment, we are capitalizing on such statistical regularities—in this case, those that give rise to specific functional contexts and spatial affordances. For example, previous studies have demonstrated differences in neural adaptation between the processing of 2-D pictures and 3-D real-world objects (Snow et al., 2011). However, Snow et al.'s (2011) study directly compared physical stimuli and pictorial stimuli—as such, there may be a variety of low- and mid-level visual cues, along with high-level inferences, that differed between their two presentation conditions. In contrast, the only differences between our presentation conditions would be carried by the frames rather than the images themselves (which were identical). Although it is possible—particularly in light of the differences in processing seen in Snow et al.'s study—that real-world stimuli would have prompted different results, the differences we observe in our presentation conditions must arise from either low-level image differences in the frames or high-level inferences about the frames that impact the processing of the contained scenes. We have tried to rule out the former and suggest that the latter is our preferred explanation. In this light, we argue that further research with physical stimuli may be needed to better characterize differences between perceiving 2-D and 3-D scenes (Snow et al., 2011, used object, not scene, stimuli). We do note that one way to address this issue is to examine whether our presentation manipulation has a behavioral effect, which would lend credence to the ecological validity of the manipulation—a question we address in the next section.

To better understand the functional impact of this neural processing difference, we examined how viewing scenes in windows and picture frames affects scene memory. More specifically, we explored whether boundary extension, a memory phenomenon associated with scene processing in which observers tend to remember scenes as wider than as actually presented, would be modulated

by functional context. We predicted that boundary extension would be greater for those scenes presented in window frames relative to those scenes presented in picture frames because of the more ecologically valid context afforded by windows. Our results were consistent with this prediction, demonstrating stronger boundary extension for scenes appearing in a window. Overall, we find support for the view that the functional context in which we view scenes can alter the perceived realism and the spatial cognitive affordances of those scenes (e.g., the multisource model; Intraub, 2010), thereby influencing the manner in which they are perceptually processed—an effect seen in both the magnitude of scene-preferred neural responses and the level of distortion of scene memories.

More broadly, scene-selective brain regions and mental processes are not simply responding to inputs that fall within their preferred domain. Instead, scene-preferred responses reflect some interplay between bottom-up and top-down information, including the associations/expectations that observers have formed about visual categories over their lifetimes. We posit that the responses of other category-preferred regions similarly reflect both feedforward and feedback processing (e.g., Hebart, Bankson, Harel, Baker, & Cichy, 2018; Brandman & Peelen, 2017; Vaziri-Pashkam & Xu, 2017; Çukur et al., 2016; Kaiser, Oosterhof, & Peelen, 2016; Kok, Brouwer, van Gerven, & de Lange, 2013; Yi & Chun, 2005).

We next turn to ask why the OPA and the PPA, but not the RSC, are sensitive to functional context. How might we account for higher neural responses for the window frame context as compared with the picture frame context for these two regions? Recent reports indicate that scene selectivity within the OPA reflects the processing of spatial properties. For example, the OPA was found to preferentially process scene boundaries and geometry relative to other properties such as landmarks (Julian et al., 2016). The OPA has also been found to process not just spatial information per se but spatial information that carries associative content (i.e., explicit coding of spatial relations within a scene and their relevance to a broader context; Aminoff & Tarr, 2015). Under this view, spatial properties such as boundaries not only help define a scene as a scene but also provide task-relevant information as to how an observer might navigate within their perceived environment. Reinforcing this claim, the OPA has also been associated with the position of the observer within an environment (Sulpizio, Committeri, Lambrey, Berthoz, & Galati, 2013) and with navigational affordances—information about where one can and cannot move in a local environment (Bonner & Epstein, 2017).

At an even finer grain, there is evidence that the OPA is not a singular functional area but is actually composed of at least two distinct functional regions: the OPA and the caudal inferior parietal lobule (cIPL). Baldassano, Esteva, Fei-Fei, and Beck (2016) argue that the OPA is tied to perceptual systems, whereas the cIPL is tied to

memory systems. Although our functional ROIs did not distinguish between the OPA and cIPL, our whole-brain analysis suggests that higher responses for the window frame context were localized to more dorsal regions that may include or overlap with the cIPL. We posit that the activation observed in these regions may be related to expectations arising from top-down information derived from memories of viewing scenes through windows. Such expectations facilitate task-related scene processing by biasing the observer to scene properties relevant to the local environment, for example, navigational affordances or scene boundaries. Supporting this view, in our behavioral experiment, we observed a boundary extension effect—remembering scene images with wider boundaries than were originally presented—when scene images were placed within a window frame. One possibility is that the perception and representation of scenes with wider boundaries may account for some of the differential activity we observe within the OPA.

As with the OPA, we observed that a second scene-preferred region, the PPA, is also sensitive to functional context. The PPA is sensitive to high-level associative scene content (Marchette et al., 2015; Aminoff & Tarr, 2015; Mégevand et al., 2014; Aminoff, Kveraga, & Bar, 2013; Diana, Yonelinas, & Ranganath, 2012; Troiani et al., 2014; Cant & Goodale, 2011; Peters, Daum, Gizewski, Forsting, & Suchan, 2009; Rauchs et al., 2008). We speculate that the larger neural responses observed for the window frame context reflect stronger associations arising from the more realistic nature of the experience. That is, scenes viewed through windows are more likely to be perceived as “real” scenes and therefore more likely to prompt the kinds of associations one experiences in day-to-day life. In contrast, scenes viewed within picture frames are understood to be depictions of scenes and less likely to be perceived as real. To the extent that the PPA is involved in bringing associative content, including associations, experiences, and expectations, to bear in scene perception, the more likely it is that the PPA will be engaged to a greater extent for the window frame context.

One caution is that, in our whole-brain analysis, the PPA did not demonstrate significant differential activity across context conditions. One possibility is that this lack of an effect may be a consequence of individual differences as to where within the PPA any differential activity was elicited. The PPA processes information differentially based on type of information; spatial information is biased to posterior regions, whereas nonspatial information is biased to anterior regions (Baldassano, Esteva, et al., 2016; Aminoff & Tarr, 2015; Aminoff, Gronau, & Bar, 2007). Across individuals, the difference between context conditions may be driven more by differences in the perception of the spatial properties of the scene and therefore recruit more posterior regions of the PPA, whereas in other individuals, the difference may be driven more by functional properties and semantics of the scene (e.g., viewing a

picture vs. being within the scene) and recruit more anterior regions of the PPA.

Finally, another scene-preferred region, the RSC, did not show any effects of our context manipulation. The RSC is believed to process nonperceptual aspects of scenes that are involved in defining higher-order properties such as strong contextual objects (Aminoff & Tarr, 2015; Bar & Aminoff, 2003); landmarks (e.g., Auger et al., 2012); or abstract, content-related episodic and autobiographical scene memories (Baldassano, Esteva, et al., 2016; Aminoff, Schacter, & Bar, 2008; Addis, Wong, & Schacter, 2007). Reinforcing the idea that the RSC is involved in more abstract aspects of scene processing, RSC responses to scenes are typically tolerant of shallow manipulations of the stimulus (Mao, Kandler, McNaughton, & Bonin, 2017). Similarly, the RSC generalizes across multiple views (e.g., Park & Chun, 2009), including indoor and outdoor views of specific places (Marchette et al., 2015). Such findings suggest that the RSC processes scenes abstracted away from their physical properties, that is, in terms of scene content and how this content relates to high-level properties of scenes encoded in memory. Given that our context manipulation focused on task-relevant inferences regarding scene structure, but not scene content, the lack of an effect of functional context in the RSC is consistent with this characterization. That is, irrespective of how one might interact with a scene, its high-level identity remains constant.

In summary, we demonstrate that top-down information modulates both the way the OPA and the PPA process and represent scenes and how observers remember scenes. In contrast, the RSC appears to be independent of this process, encoding a high-level representation of scene content that is not influenced by presentation context. Such results add to our understanding of the different roles of the OPA, PPA, and RSC in scene processing. More generally, our results demonstrate that responses in category-preferred brain regions do not arise solely from the processing of inputs within their preferential domains, but rather integrate high-level knowledge into their processing. Both feedforward and feedback pathways appear to play an important role in categorical perception and, in particular, in the specific neural substrates that support scene understanding.

Acknowledgments

We thank Alyssa Shannon for her work in the boundary extension experiment.

Reprint requests should be sent to Elissa M. Aminoff, Department of Psychology, Fordham University, Dealy Hall 332, 441 E. Fordham Rd., Bronx, NY 10458, or via e-mail: eaminoff@fordham.edu.

Author Contributions

Elissa M. Aminoff: Conceptualization; Data curation; Formal analysis; Writing—Original draft; Writing—Review & editing.

Michael J. Tarr: Conceptualization; Formal analysis; Writing—Original draft; Writing—Review & editing.

Funding Information

Elissa M. Aminoff, National Science Foundation (<http://dx.doi.org/10.13039/100000001>), grant number: 1439237.

Diversity in Citation Practices

A retrospective analysis of the citations in every article published in this journal from 2010 to 2020 has revealed a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were $M(\text{an})/M = .408$, $W(\text{oman})/M = .335$, $M/W = .108$, and $W/W = .149$, the comparable proportions for the articles that these authorship teams cited were $M/M = .579$, $W/M = .243$, $M/W = .102$, and $W/W = .076$ (Fulvio et al., *JoCN*, 33:1, pp. 3–7). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

Notes

1. “Functional properties” denotes high-level knowledge of how a visual stimulus is used and how it interacts with the environment (including other objects and people).
2. The participants of this study were also part of a study discussed in Yang, Tarr, Kass, and Aminoff (2019), and thus, the localizer data used here is common with the localizer data described in that paper.

REFERENCES

- Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45, 1363–1377. **DOI:** <https://doi.org/10.1016/j.neuropsychologia.2006.10.016>, **PMID:** 17126370, **PMCID:** PMC1894691
- Aminoff, E. M., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex*, 17, 1493–1503. **DOI:** <https://doi.org/10.1093/cercor/bhl078>, **PMID:** 16990438
- Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences*, 17, 379–390. **DOI:** <https://doi.org/10.1016/j.tics.2013.06.009>, **PMID:** 23850264, **PMCID:** PMC3786097
- Aminoff, E. M., Schacter, D. L., & Bar, M. (2008). The cortical underpinnings of context-based memory distortion. *Journal of Cognitive Neuroscience*, 20, 2226–2237. **DOI:** <https://doi.org/10.1162/jocn.2008.20156>, **PMID:** 18457503, **PMCID:** PMC3786095
- Aminoff, E. M., & Tarr, M. J. (2015). Associative processing is inherent in scene perception. *PLoS One*, 10, e0128840. **DOI:** <https://doi.org/10.1371/journal.pone.0128840>, **PMID:** 26070142, **PMCID:** PMC4467091
- Auger, S. D., Mullally, S. L., & Maguire, E. A. (2012). Retrosplenial cortex codes for permanent landmarks. *PLoS One*, 7, e43620. **DOI:** <https://doi.org/10.1371/journal.pone.0043620>, **PMID:** 22912894, **PMCID:** PMC3422332
- Bainbridge, W. A., & Baker, C. I. (2020). Boundaries extend and contract in scene memory depending on image properties. *Current Biology*, 30, 537–543. **DOI:** <https://doi.org/10.1016/j.cub.2019.12.004>, **PMID:** 31983637, **PMCID:** PMC7187786
- Baldassano, C., Esteva, A., Fei-Fei, L., & Beck, D. M. (2016). Two distinct scene-processing networks connecting vision and memory. *eNeuro*, 3, ENEURO.0178-16.2016. **DOI:** <https://doi.org/10.1523/ENEURO.0178-16.2016>, **PMID:** 27822493, **PMCID:** PMC5075944
- Baldassano, C., Fei-Fei, L., & Beck, D. M. (2016). Pinpointing the peripheral bias in neural scene-processing networks during natural viewing. *Journal of Vision*, 16, 9. **DOI:** <https://doi.org/10.1167/16.2.9>, **PMID:** 27187606
- Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5, 617–629. **DOI:** <https://doi.org/10.1038/nrn1476>, **PMID:** 15263892
- Bar, M., & Aminoff, E. M. (2003). Cortical analysis of visual context. *Neuron*, 38, 347–358. **DOI:** [https://doi.org/10.1016/S0896-6273\(03\)00167-3](https://doi.org/10.1016/S0896-6273(03)00167-3), **PMID:** 12718867
- Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: The parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *Journal of Neuroscience*, 28, 8539–8544. **DOI:** <https://doi.org/10.1523/JNEUROSCI.0987-08.2008>, **PMID:** 18716212, **PMCID:** PMC2707255
- Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–253). Hillsdale, NJ: Erlbaum. **DOI:** <https://doi.org/10.4324/9781315512372-8>
- Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relation violations. *Cognitive Psychology*, 14, 143–177. **DOI:** [https://doi.org/10.1016/0010-0285\(82\)90007-X](https://doi.org/10.1016/0010-0285(82)90007-X), **PMID:** 7083801
- Bonner, M. F., & Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences, U.S.A.*, 114, 4793–4798. **DOI:** <https://doi.org/10.1073/pnas.1618228114>, **PMID:** 28416669, **PMCID:** PMC5422815
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436. **DOI:** <https://doi.org/10.1163/156856897X00357>, **PMID:** 9176952
- Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *Journal of Neuroscience*, 37, 7700–7710. **DOI:** <https://doi.org/10.1523/JNEUROSCI.0582-17.2017>, **PMID:** 28687603, **PMCID:** PMC6596648
- Cant, J. S., & Goodale, M. A. (2011). Scratching beneath the surface: New insights into the functional properties of the lateral occipital area and parahippocampal place area. *Journal of Neuroscience*, 31, 8248–8258. **DOI:** <https://doi.org/10.1523/JNEUROSCI.6113-10.2011>, **PMID:** 21632946, **PMCID:** PMC6622867
- Chadwick, M. J., Mullally, S. L., & Maguire, E. A. (2013). The hippocampus extrapolates beyond the view in scenes: An fMRI study of boundary extension. *Cortex*, 49, 2067–2079. **DOI:** <https://doi.org/10.1016/j.cortex.2012.11.010>, **PMID:** 23276398, **PMCID:** PMC3764338
- Çukur, T., Huth, A. G., Nishimoto, S., & Gallant, J. L. (2016). Functional subdomains within scene-selective cortex: Parahippocampal place area, retrosplenial complex, and occipital place area. *Journal of Neuroscience*, 36, 10257–10273. **DOI:** <https://doi.org/10.1523/JNEUROSCI.4033-14.2016>, **PMID:** 27707964, **PMCID:** PMC5050324

- Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2012). Adaptation to cognitive context and item information in the medial temporal lobes. *Neuropsychologia*, *50*, 3062–3069. **DOI:** <https://doi.org/10.1016/j.neuropsychologia.2012.07.035>, **PMID:** 22846335, **PMCID:** PMC3483447
- Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience*, *33*, 1331–1336. **DOI:** <https://doi.org/10.1523/JNEUROSCI.4081-12.2013>, **PMID:** 23345209, **PMCID:** PMC3711611
- Epstein, R. A., & Baker, C. I. (2019). Scene perception in the human brain. *Annual Review of Vision Science*, *5*, 373–397. **DOI:** <https://doi.org/10.1146/annurev-vision-091718-014809>, **PMID:** 31226012, **PMCID:** PMC6989029
- Epstein, R. A., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601. **DOI:** <https://doi.org/10.1038/33402>, **PMID:** 9560155
- Fang, F., Boyaci, H., Kersten, D., & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. *Current Biology*, *18*, 1707–1712. **DOI:** <https://doi.org/10.1016/j.cub.2008.09.025>, **PMID:** 18993076, **PMCID:** PMC2638992
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47. **DOI:** <https://doi.org/10.1093/cercor/1.1.1>, **PMID:** 1822724
- Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*, 137–176. **DOI:** <https://doi.org/10.1016/j.cogpsych.2008.06.001>, **PMID:** 18762289, **PMCID:** PMC2759758
- Harel, A., Kravitz, D. J., & Baker, C. I. (2013). Deconstructing visual scenes in cortex: Gradients of object and spatial layout information. *Cerebral Cortex*, *23*, 947–957. **DOI:** <https://doi.org/10.1093/cercor/bhs091>, **PMID:** 22473894, **PMCID:** PMC3593580
- Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I., & Cichy, R. M. (2018). The representational dynamics of task and object processing in humans. *eLife*, *7*, e32816. **DOI:** <https://doi.org/10.7554/eLife.32816>, **PMID:** 29384473, **PMCID:** PMC5811210
- Henderson, J. M., Zhu, D. C., & Larson, C. L. (2011). Functions of parahippocampal place area and retrosplenial cortex in real-world scene analysis: An fMRI study. *Visual Cognition*, *19*, 910–927. **DOI:** <https://doi.org/10.1080/13506285.2011.596852>
- Intraub, H. (2010). Rethinking scene perception: A multisource model. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 52, pp. 231–264). Burlington, VT: Academic Press. **DOI:** [https://doi.org/10.1016/S0079-7421\(10\)52006-1](https://doi.org/10.1016/S0079-7421(10)52006-1)
- Intraub, H. (2014). Visual scene representation: A spatial-cognitive perspective. In K. Kveraga & M. Bar (Eds.), *Scene vision: Making sense of what we see* (pp. 5–26). Cambridge, MA: MIT Press. **DOI:** <https://doi.org/10.7551/mitpress/9780262027854.003.0001>
- Intraub, H. (2020). Searching for boundary extension. *Current Biology*, *30*, R1463–R1464. **DOI:** <https://doi.org/10.1016/j.cub.2020.10.031>, **PMID:** 33352122
- Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 179–187. **DOI:** <https://doi.org/10.1037/0278-7393.15.2.179>
- Janzen, G., & van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nature Neuroscience*, *7*, 673–677. **DOI:** <https://doi.org/10.1038/nn1257>, **PMID:** 15146191
- Julian, J. B., Ryan, J., Hamilton, R. H., & Epstein, R. A. (2016). The occipital place area is causally involved in representing environmental boundaries during navigation. *Current Biology*, *26*, 1104–1109. **DOI:** <https://doi.org/10.1016/j.cub.2016.02.066>, **PMID:** 27020742, **PMCID:** PMC5565511
- Kaiser, D., Oosterhof, N. N., & Peelen, M. V. (2016). The neural dynamics of attentional selection in natural scenes. *Journal of Neuroscience*, *36*, 10522–10528. **DOI:** <https://doi.org/10.1523/JNEUROSCI.1385-16.2016>, **PMID:** 27733605, **PMCID:** PMC6601932
- Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*, 37. **DOI:** <https://doi.org/10.3389/fnint.2014.00037>, **PMID:** 24847226, **PMCID:** PMC4019851
- Kay, K. N., & Yeatman, J. D. (2017). Bottom-up and top-down computations in word- and face-selective cortex. *eLife*, *6*, e22341. **DOI:** <https://doi.org/10.7554/eLife.22341>, **PMID:** 28226243, **PMCID:** PMC5358981
- Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience*, *33*, 16275–16284. **DOI:** <https://doi.org/10.1523/JNEUROSCI.0742-13.2013>, **PMID:** 24107959, **PMCID:** PMC6618350
- Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-world scene representations in high-level visual cortex: It's the spaces more than the places. *Journal of Neuroscience*, *31*, 7322–7333. **DOI:** <https://doi.org/10.1523/JNEUROSCI.4588-10.2011>, **PMID:** 21593316, **PMCID:** PMC3115537
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*, 571–579. **DOI:** [https://doi.org/10.1016/S0166-2236\(00\)01657-X](https://doi.org/10.1016/S0166-2236(00)01657-X)
- Lescroart, M. D., & Gallant, J. L. (2019). Human scene-selective areas represent 3D configurations of surfaces. *Neuron*, *101*, 178–192. **DOI:** <https://doi.org/10.1016/j.neuron.2018.11.004>, **PMID:** 30497771
- Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience*, *4*, 533–539. **DOI:** <https://doi.org/10.1038/87490>, **PMID:** 11319563
- Lowe, M. X., Rajsic, J., Gallivan, J. P., Ferber, S., & Cant, J. S. (2017). Neural representation of geometry and surface properties in object and scene perception. *Neuroimage*, *157*, 586–597. **DOI:** <https://doi.org/10.1016/j.neuroimage.2017.06.043>, **PMID:** 28647484
- Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology*, *42*, 225–238. **DOI:** <https://doi.org/10.1111/1467-9450.00233>, **PMID:** 11501737
- Mao, D., Kandler, S., McNaughton, B. L., & Bonin, V. (2017). Sparse orthogonal population representation of spatial context in the retrosplenial cortex. *Nature Communications*, *8*, 243. **DOI:** <https://doi.org/10.1038/s41467-017-00180-9>, **PMID:** 28811461, **PMCID:** PMC5557927
- Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2015). Outside looking in: Landmark generalization in the human navigational system. *Journal of Neuroscience*, *35*, 14896–14908. **DOI:** <https://doi.org/10.1523/JNEUROSCI.2270-15.2015>, **PMID:** 26538658, **PMCID:** PMC4635136
- Mégevand, P., Groppe, D. M., Goldfinger, M. S., Hwang, S. T., Kingsley, P. B., Davidesco, I., et al. (2014). Seeing scenes: Topographic visual hallucinations evoked by direct electrical stimulation of the parahippocampal place area. *Journal of Neuroscience*, *34*, 5399–5405. **DOI:** <https://doi.org/10.1523/JNEUROSCI.5202-13.2014>, **PMID:** 24741031, **PMCID:** PMC6608225
- Nasr, S., & Tootell, R. B. H. (2012). A cardinal orientation bias in scene-selective visual cortex. *Journal of Neuroscience*, *32*, 14921–14926. **DOI:** <https://doi.org/10.1523/JNEUROSCI.2036-12.2012>, **PMID:** 23100415, **PMCID:** PMC3495613

- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175. **DOI:** <https://doi.org/10.1023/A:1011139631724>
- Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience*, *31*, 1333–1340. **DOI:** <https://doi.org/10.1523/JNEUROSCI.3885-10.2011>, **PMID:** 21273418, **PMCID:** PMC6623596
- Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage*, *47*, 1747–1756. **DOI:** <https://doi.org/10.1016/j.neuroimage.2009.04.058>, **PMID:** 19398014, **PMCID:** PMC2753672
- Park, S., Intraub, H., Yi, D.-J., Widders, D., & Chun, M. M. (2007). Beyond the edges of a view: Boundary extension in human scene-selective visual cortex. *Neuron*, *54*, 335–342. **DOI:** <https://doi.org/10.1016/j.neuron.2007.04.006>, **PMID:** 17442252
- Park, S., Konkle, T., & Oliva, A. (2015). Parametric coding of the size and clutter of natural scenes in the human brain. *Cerebral Cortex*, *25*, 1792–1805. **DOI:** <https://doi.org/10.1093/cercor/bht418>, **PMID:** 24436318, **PMCID:** PMC4459284
- Peters, J., Daum, I., Gizewski, E., Forsting, M., & Suchan, B. (2009). Associations evoked during memory encoding recruit the context-network. *Hippocampus*, *19*, 141–151. **DOI:** <https://doi.org/10.1002/hipo.20490>, **PMID:** 18777560
- Rauchs, G., Orban, P., Balteau, E., Schmidt, C., Degueldre, C., Luxen, A., et al. (2008). Partially segregated neural networks for spatial and contextual memory in virtual navigation. *Hippocampus*, *18*, 503–518. **DOI:** <https://doi.org/10.1002/hipo.20411>, **PMID:** 18240326
- Silson, E. H., Chan, A. W.-Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *Journal of Neuroscience*, *35*, 11921–11935. **DOI:** <https://doi.org/10.1523/JNEUROSCI.0137-15.2015>, **PMID:** 26311774, **PMCID:** PMC4549403
- Snow, J. C., Pettypiece, C. E., McAdam, T. D., McLean, A. D., Stroman, P. W., Goodale, M. A., et al. (2011). Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects. *Scientific Reports*, *1*, 130. **DOI:** <https://doi.org/10.1038/srep00130>, **PMID:** 22355647, **PMCID:** PMC3216611
- Stevens, W. D., Kahn, I., Wig, G. S., & Schacter, D. L. (2012). Hemispheric asymmetry of visual scene processing in the human brain: Evidence from repetition priming and intrinsic activity. *Cerebral Cortex*, *22*, 1935–1949. **DOI:** <https://doi.org/10.1093/cercor/bhr273>, **PMID:** 21968568, **PMCID:** PMC3388897
- Sulpizio, V., Committeri, G., Lambrey, S., Berthoz, A., & Galati, G. (2013). Selective role of lingual/parahippocampal gyrus and retrosplenial complex in spatial memory across viewpoint changes relative to the environmental reference frame. *Behavioral Brain Research*, *242*, 62–75. **DOI:** <https://doi.org/10.1016/j.bbr.2012.12.031>, **PMID:** 23274842
- Troiani, V., Stigliani, A., Smith, M. E., & Epstein, R. A. (2014). Multiple object properties drive scene-selective regions. *Cerebral Cortex*, *24*, 883–897. **DOI:** <https://doi.org/10.1093/cercor/bhs364>, **PMID:** 23211209, **PMCID:** PMC3948490
- van der Ham, I. J. M., van Zandvoort, M. J. E., Frijns, C. J. M., Kappelle, L. J., & Postma, A. (2011). Hemispheric differences in spatial relation processing in a scene perception task: A neuropsychological study. *Neuropsychologia*, *49*, 999–1005. **DOI:** <https://doi.org/10.1016/j.neuropsychologia.2011.02.024>, **PMID:** 21356223
- Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., & Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *Journal of the American Medical Informatics Association*, *8*, 443–459. **DOI:** <https://doi.org/10.1136/jamia.2001.0080443>, **PMID:** 11522765, **PMCID:** PMC131042
- Vaziri-Pashkam, M., & Xu, Y. (2017). Goal-directed visual processing differentially impacts human ventral and dorsal visual representations. *Journal of Neuroscience*, *37*, 8767–8782. **DOI:** <https://doi.org/10.1523/JNEUROSCI.3392-16.2017>, **PMID:** 28821655, **PMCID:** PMC5588467
- Yang, Y., Tarr, M. J., Kass, R. E., & Aminoff, E. M. (2019). Exploring spatiotemporal neural dynamics of the human visual cortex. *Human Brain Mapping*, *40*, 4213–4238. **DOI:** <https://doi.org/10.1002/hbm.24697>, **PMID:** 31231899, **PMCID:** PMC6865718
- Yi, D.-J., & Chun, M. M. (2005). Attentional modulation of learning-related repetition attenuation effects in human parahippocampal cortex. *Journal of Neuroscience*, *25*, 3593–3600. **DOI:** <https://doi.org/10.1523/JNEUROSCI.4677-04.2005>, **PMID:** 15814790, **PMCID:** PMC6725381