Parametric Coding of the Size and Clutter of Natural Scenes in the Human Brain

Soojin Park[1,], Talia Konkle[2] and Aude Oliva[3]

[1]Department of Cognitive Science, Johns Hopkins University, Baltimore, MD 21218, USA, [2]Department of Psychology, Harvard University, Cambridge, MA 02138, USA and [3]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Address correspondence to Soojin Park, Department of Cognitive Science, The Johns Hopkins University, 237 Krieger Hall, 3400 North Charles Street, Baltimore, MD 21218, USA. Email: park@cogsci.jhu.edu
Soojin Park and Talia Konkle equally contributed to the manuscript.

Estimating the size of a space and its degree of clutter are effortless and ubiquitous tasks of moving agents in a natural environment. Here, we examine how regions along the occipital–temporal lobe respond to pictures of indoor real-world scenes that parametrically vary in their physical "size" (the spatial extent of a space bounded by walls) and functional "clutter" (the organization and quantity of objects that fill up the space). Using a linear regression model on multivoxel pattern activity across regions of interest, we find evidence that both properties of size and clutter are represented in the patterns of parahippocampal cortex, while the retrosplenial cortex activity patterns are predominantly sensitive to the size of a space, rather than the degree of clutter. Parametric whole-brain analyses confirmed these results. Importantly, this size and clutter information was represented in a way that generalized across different semantic categories. These data provide support for a property-based representation of spaces, distributed across multiple scene-selective regions of the cerebral cortex.

Keywords: fMRI, multivoxel pattern analysis, scene and space perception, the parahippocampal place area, the retrosplenial cortex

Introduction

It is well accepted that real-world scenes, akin to objects, have a rich taxonomy of categorical structure and that each scene exemplar can be labeled with semantic category descriptors, for example, a pantry or a stadium (Tversky and Hemmenway 1983; Xiao et al. 2010). However, real-world scenes, like objects (Konkle and Oliva 2012) can also be described in terms of geometrical properties—e.g., a stadium is a large space whereas a pantry is a small space (Oliva and Torralba 2001; Greene and Oliva 2009a, b; Ross and Oliva 2010; Kravitz, Peng, et al. 2011; Park et al. 2011). What are the meaningful geometric properties of a scene that are effortless and ubiquitous for the brain to compute?

Psychophysical research has shown that, at the very beginning of a glance, the visual system has information about how large the space depicted in a 2D image is, whether the space is indoor or outdoor, enclosed or open, full or empty, navigable or not, or made up of natural or manufactured objects (e.g., Fei-Fei et al. 2007; Joubert et al. 2007; Greene and Oliva 2009a, b, 2010). In other words, when we step into a new environment or glance at a novel scene picture, the "size" of the space (the spatial extent bounded by walls) and its functional "clutter" (the organization and quantity of objects that fill up the space) are general properties of all scenes that are immediately accessible and can constrain our action or navigation in the environment (e.g., Hermer-Vasquez et al. 2001; Learmonth

et al. 2002). However, while much has been learned about the scene properties we extract in a glance and how these relate to navigational capacities, our understanding about how and where the brain represents properties of real-world scenes remains sparse.

A well-known network of scene-processing regions, including the parahippocampal place area (PPA) and the retrosplenial complex (RSC), are characterized by their greater response to visually presented scenes than that to objects, but their exact role and interactions during scene processing is still an active topic (Epstein et al. 2003; Epstein 2008 for review; Park and Chun 2009). Several neuroimaging studies have shown that the PPA is sensitive to the semantic category of a scene (Naselaris et al. 2009; Walther et al. 2009; Morgan et al. 2011). A recent study provides insight as to what features may underlie this scene–category information, showing successful scene–category prediction based on an encoding model that represents scenes as mixtures of object probabilities (Stansbury et al. 2013). Interestingly, scene–category information was found not only in the well-known nodes of the scene network (including PPA, RSC), but also across the extended occipitotemporal cortex. These results provide support for an object-based representation of scene–category information, which is supported by a broad expanse of neural regions extending beyond the classic scene-processing regions.

However, recent neuroimaging studies have just begun to examine the hypothesis that scenes might also be represented based on global properties describing the spatial geometry and featural content of the scene independent of the objects. For example, the scene property of "openness" is represented in the PPA while the scene property of "naturalness" is represented in the lateral occipital cortex (LOC) (Kravitz, Peng, et al. 2011; Park et al. 2011). These studies show that for example a street and a forest have similar patterns of activation in the PPA as long as they have similar spatial layout, even though they are from different semantic categories (Kravitz et al. 2011; Park et al. 2011). Further, both the PPA and RSC (but not LOC) show sensitivity to the "spatial layout" of a scene, while the PPA and LOC (but not RSC) show sensitivity to the presence of an object inside the scene (Harel et al. 2013). Thus, there is also emerging evidence for a property-based representation of scene information, distributed primarily within the major nodes of the scene network.

In the current work, we examined the neural representation of 2 unexplored properties of real-world scenes, their "physical size" and their "functional clutter." Are scenes that are similar in their depicted physical size, or their level of clutter, represented similarity in the brain, even if they are from different semantic categories? And, if so, in what regions of the brain are

these representations apparent? Further, these properties of size and clutter do not exist as discrete distinctions in the real world (e.g., small space vs. large space; or empty vs. full scene), but rather lie on a continuous scale. Thus, we also asked, do neural scene regions represent size and clutter information in a way that reflects these properties as continuous dimensions? To answer these questions, we used a powerful method to probe the representational content of a region by examining whether the neural regions show "parametric variation" based on these properties, with analysis methods that require generalization over multiple semantic categories.

We approach these questions in 2 experiments. In Experiment 1, we look for evidence of a neural representation of the size of a space, testing whether scenes of small sizes (e.g., closet, shower, or pantry) are represented differently from scenes of medium and large sizes (e.g., shopping mall, concert hall, or sports stadium; illustrated in Fig. 1). In Experiment 2, we consider the neural representation of size and clutter properties together, by building a 2D parametric design. Specifically, 36 categories of real-world scenes were selected to lie in a 2D space with size and clutter as orthogonal axes (illustrated in Fig. 2). With this design, scene categories can be grouped depending on which property the analysis focuses on: for example "closets" are similar to "showers" but different from "stadiums" on the size property, whereas closets are similar to stadiums and different from showers on the clutter property. This allows us to isolate and compare the variation of size and clutter, holding the stimulus set constant, in order to explore how regions across the cerebral cortex are sensitive to these scene properties.

Materials and Methods

Subjects
Thirteen participants (6 females; ages: 19–35 years) in Experiment 1 and 12 participants (9 females; ages: 20–27 years) in Experiment 2 were recruited from the MIT and Cambridge, MA community. One participant of Experiment 1 was excluded from the analyses due to excessive head movement (over 8 mm across runs). All participants had normal or corrected-to-normal vision. Informed consent was obtained, and the study protocol was approved by the Institutional Review Board of the Massachusetts Institute of Technology.

Visual Stimuli and Experimental Design
In Experiment 1, scene categories were chosen to cover the full magnitude of physical size of indoor environments (Fig. 1). The stimulus set was organized in 6 size levels, roughly following a logarithmic scale based on the number of people the space may contain: from a small space that would contain 1 to 2 people (level 1), to a large space that could hold a thousand people (level 6). Each level contained images from 3 different scene categories (for a total of 18 scene categories, see the list in the caption of Fig. 1). There were 16 image exemplars per category. In the functional neuroimaging experiment, images were presented in blocks of 16 s each, which was followed by 10-s fixation periods. Within a block, each image was displayed for 800 ms, followed by a 200-ms blank. Across 4 runs, 12 blocks per size level were presented, with each category block repeating 4 times in the experiment.

In Experiment 2, a stimulus set was constructed to orthogonally vary the physical size (6 levels) and the functional clutter (6 levels) of real-world scenes (Fig. 2). Thirty-six different scene categories were chosen such that each category fit into a cell of this 6 × 6 stimulus grid, Physical size followed the same scale as in Experiment 1. Functional clutter was broadly constructed as the layout and quantity of components (including objects, walls, people) that fill the scene, in a

natural way. Its levels ran from a completely empty space (level 1) to a highly cluttered or full space (level 6). Importantly, as shown in Figure 2, each size or clutter level was represented, respectively, by all the levels along the other property in a fully crossed design, making the 2 properties independent and orthogonal of each other for data analyses. There were 12 different image exemplars for each of the 36 categories, presented in blocks of 16 s per category. Within a block, each scene was displayed for 1 s, followed by a 330-ms blank. Across 6 runs, there were 24 blocks per level (of either size or clutter), with each category block repeating 4 times in the experiment. Only 22 images of 432 images (5%) used in Experiment 2 overlapped with those used in Experiment 1.

To validate the size and clutter levels in this stimulus set, we obtained behavioral ratings of the size and clutter of each image using Amazon's Mechanical Turk service (See Supplementary Fig. 4). The results confirmed that 1) the average category ratings were highly correlated with the expected size and clutter levels ($r = 0.98$, $r = 0.97$, P's < 0.01); 2) that, within a category, the items were consistently rated at the expected level for both size and clutter (average item standard deviation = 0.30 (size), 0.36 (clutter), which is less than half of a level on the 6-point scale); and 3) that there was no relationship between the size levels and clutter ratings or vice versa ($r = 0.0$, $r = -0.3$, P's > 0.2), confirming the orthogonal nature of the design.

To take into account differences in the low-level image statistics in the Experiment 2 stimuli set, we also created an "equalized set" that consists of a subset of the 36 categories with roughly equal spectral energy in the image categories on average across the levels of scene size (small to large). To do this, a power spectrum analysis was performed on each image to calculate the quantity of energy across the range of spatial frequencies. In the original set, the average spectral energy for each category above 10 cycles/image ranged from 27% to 39% across size levels, and, in the equalized set, this range was reduced to 29–33%. The equalized set included 24 categories, with 4 categories for each size level. This equalized set was used in the whole-brain analysis of Experiment 2 to specifically localize regions of the brain that responded to different levels of size, beyond the spectral energy differences.

In both experiments, colored photographs were 500-by-500 pixels resolution (9° × 9° of visual angle) and were normalized to have a mean pixel value averaging 127 (on a 0–255 scale). Images were presented in the scanner using a Hitachi (CP-X1200 series) projector through a rear-projection screen. Participants performed a one-back repetition detection task to maintain attention.

MRI Acquisition and Preprocessing
Imaging data were acquired with a 3T Siemens fMRI scanner with 32-channel phased-array head coil (Siemens) at the Martinos Center at the McGovern Institute for Brain Research at MIT. Anatomical images were acquired using a high-resolution (1 × 1 × 1 mm voxel) MPRAGE structural scan. Functional images were acquired with a gradient echo-planar T_2* sequence (TR, 2 s; TE 30; 33 axial 3 mm slices with no gap; acquired parallel to the anterior commissure–posterior commissure line).

Functional data were preprocessed using Brain Voyager QX software (Brain Innovation, Maastricht, Netherlands). Preprocessing included slice scan-time correction, linear trend removal, and 3D motion correction. For multivariate pattern analysis, no additional spatial or temporal smoothing was performed and data were analyzed in each individual's ACPC space. For whole-brain group parametric analysis, all data were smoothed with Gaussian kernel with 4 mm FWHM and were transformed to a Talairach brain. For retinotopic analysis, the cortical surface of each subject was reconstructed from the high-resolution T_1-weighted anatomical scan, acquired with a 3D MPRAGE protocol. These 3D brains were inflated using BV surface module, and the obtained retinotopic functional maps were superimposed on the surface rendered cortex.

Regions of Interest
Regions of interest (ROIs) were defined for each participant using independent localizers. A localizer run presented blocks of scenes, faces,

Figure 1. Examples of scenes used in Experiment 1, varying in the size of the space from small to large, on a 6-point scale (from left to right). Scene categories representing each size levels are level 1 = closet, shower, pantry; level 2 = bathroom, pilothouse, music studio; level 3 = bedroom, hair salon, classroom; level 4 = gym, lobby, church; level 5 = warehouse, hangar, airport terminal; level 6 = shopping mall, concert hall, sports stadium.
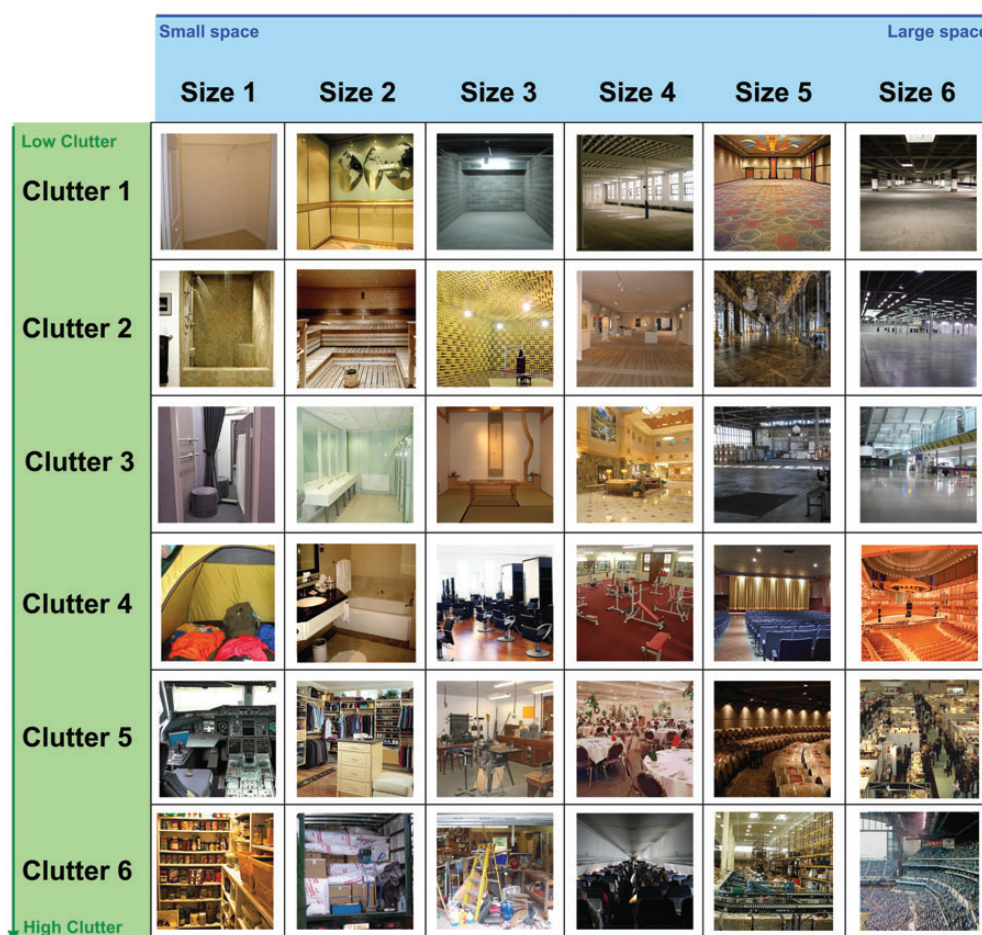

Figure 2. Examples of 36 scene categories varying in 2 orthogonalized properties used in Experiment 2: The size of the scene varied from small to large size on 6-point scale (size levels 1–6 from left to right); the amount of clutter varied from low clutter to high clutter on 6-point scale (clutter levels 1–6 from top to bottom). Scene categories representing each size and clutter level are (organized by size level – clutter level; column–row): 1–1 = closet, 1–2 = shower, 1–3 = fitting room, 1–4 = tent, 1–5 = cockpit, 1–6 = pantry; 2–1 = elevator, 2–2 = sauna, 2–3 = restroom, 2–4 = bathroom, 2–5 = walk-in closet, 2–6 = moving truck; 3–1 = garage, 3–2 = anechoic chamber, 3–3 = Japanese room, 3–4 = hair salon, 3–5 = workshop, 3–6 = messy garage; 4–1 = loft, 4–2 = art gallery, 4–3 = lobby, 4–4 = gym, 4–5 = banquet hall, 4–6 = inside airplane; 5–1 = ball room, 5–2 = palace hall, 5–3 = hangar, 5–4 = movie theater, 5–5 = barrel storage, 5–6 = warehouse; 6–1 = parking garage, 6–2 = industrial warehouse, 6–3 = airport terminal, 6–4 = concert hall, 6–5 = exhibition hall, 6–6 = arena.

objects, and scrambled objects to define the PPA and RSC (Scenes–Objects) and LOC (Objects-Scrambled objects), and FFA (Faces–Scenes). There were 4 blocks per each stimuli type, with 20 images per each block. Each block was presented for 16 s with 10-s rest periods in between. Within each block, each stimulus was displayed for 600 ms, followed by 200 ms blank. The order of blocks was randomized. Participants performed a one-back repetition detection task. A retinotopic localizer presented vertical and horizontal visual field meridians to delineate borders of retinotopic areas (Sereno et al. 1995; Spiridon and Kanwisher 2002). Triangular wedges of black and white checkerboards were presented either vertically (upper or lower vertical meridians) or horizontally (left or right horizontal meridians) in 12 s blocks, alternated with 12 s blanks. There were 5 blocks per each condition. Participants were instructed to fixate on a small central fixation dot. Area V1 was defined based on the contrast between vertical and horizontal meridian activations.

Univariate Analysis: ROI-Based Correlation Analysis

We tested whether the average activity within ROIs parametrically varied as the scene size or clutter level changed. A measure of the overall activity in the ROI for each size (or clutter) level was obtained, by first estimating the response magnitude for each scene category using a ROI GLM. Then, we computed the correlation between the true size/clutter levels and the magnitude of the beta weights for each scene category. This correlation was computed for each participant. To test for significance across participants, the correlation coefficients (r) for each subject were transformed using Fisher's z' transformation to represent normally distributed variable z', and then entered into a t-test.

Univariate Analysis: Whole-Brain Analysis with Parametric Regressors

To find brain regions that are modulated parametrically along a scene property (e.g., a parametric increase in response with increasing size levels), we performed an exploratory random-effects group analysis. Parametric predictors for the time course were modeled as a boxcar function for each block, with the amplitude equal to the size level (or clutter level) from 1 to 6, normalized to have a zero-mean, and subsequently convolved with the hemodynamic response function. Main effect predictors for the time course were modeled with a boxcar function for each block, convolved with the hemodynamic response. The response for each voxel was modeled with these regressors using a GLM. To localize regions with a parametric response, we performed a random-effects conjunction analysis with the main and parametric predictors. Without the conjunction analysis, voxels that are simply activated by some or all of the conditions without any clear parametric modulation will still have much of their variance accounted for by a single non-normalized parametric regressor. By using a conjunction analysis between the main effect and the de-meaned parametric predictor, we avoid the partial colinearity between the main effect and parametric predictors and isolate truly modulatory regions.

Multivariate Analysis with ROIs: Data Format

For all multivariate pattern analysis, we obtained patterns of activity across the voxels of an ROI for each presentation of a scene category using the following procedure. The MRI signal intensity from each voxel within a ROI across all time points was transformed into z-scores by run. This helps mitigate overall differences in fMRI signal across different ROIs and across runs and sessions (Kamitani and Tong 2005). The z-scored signal intensity was then extracted for each stimulus block, spanning 16 s (8 TRs) with a 4 s (2 TR) offset to account for the hemodynamic delay of the BOLD response. These time points were averaged to generate a pattern across voxels (within an ROI) for each stimulus block. Each category was presented 4 times in the experiment, thus this design yielded 144 multivoxel patterns (36 categories × 4 repetitions) used for the parametric pattern analysis.

Multivariate Analysis: Parametric Pattern Analysis

To examine if the patterns in an ROI contain parametric information about the size and clutter levels of scenes, we conducted a regression analysis. This analysis was only performed on Experiment 2, due to lack of power for multivariate analysis in Experiment 1. We used a regression analysis rather than discrete classification to take advantage of the parametric nature of the data: this method solves for a set of weights on each voxel in a ROI, such that any new scene pattern of activity, multiplied by the weights, predicts level of that scene's size on a continuous scale. In other words, this regression analysis learns a set of weights from an input feature vector (e.g., the response of a voxel to each of the stimulus blocks) and an outcome variable (e.g., the size levels across each of the stimulus blocks). This analysis differs from the assumptions in classification analyses (e.g., SVM), where each scene size level would be treated categorically and classification performance requires setting up multiple pairwise linear classifiers (see Supplementary Fig. 1 for the standard SVM analysis).

In standard linear regression, weights on each feature vector are adjusted to minimize the squared error between the predicted label and the correct label (the first term in eq. 1). Here, we used ridge regression, which is the same as standard regression, but also includes a regularization term that biases it to find a solution that also minimizes the sum of the squared feature weights (the second term in eq. 1). This is the simplest regression technique that can cope with underdetermined regression problems for example where the number of voxels far exceeds the number of training blocks. Ridge regression calculates the weights in B to minimize the following equation:

$$||\mathbf{y} - \mathbf{XB}||^2 + \lambda ||\mathbf{B}||^2 \qquad (1)$$

Here, the $\mathbf{y}$ vector reflects the actual levels of size (or clutter) for each scene pattern in the training set (120×1, where 120 patterns come from 30 training categories and 4 patterns per category); $\mathbf{X}$ is the matrix with the multivoxel patterns of activity for each scene category ($120 \times N$), where N is the number of voxels in the ROI; $\mathbf{B}$ is the model, characterized as a vector of betas, or weights, on each voxel ($N \times 1$); and λ is the ridge parameter which determines the impact of the regularization term (scalar value 1×1). In the current analysis, λ was set at 5% of the number of voxels, which ensures that lambda exerts a similar force on ROIs of different sizes (Hoerl and Kennard 1970; Hastie et al. 2001; Newman and Norman 2010).

Train–Test Procedure

We used a train–test procedure that requires generalization over scene category for successful performance. That is, a regression model was fit using training data that contained data from 5 of 6 scene categories for each size (or clutter) level (120 training patterns). That model was then tested with data from the remaining 6 scene categories (24 test patterns), where the model predicted the level of scene size (or clutter) of each of the test patterns. We conducted a 6-fold validation procedure. For each iteration, the 6 categories that were withheld for the testing phase were always from the same level on the orthogonal property. In the first train–test iteration, one scene category was withheld from each size level (1–6), all with clutter level = 1. On the next train–test iteration, a different 6 categories were withheld from each size level (1–6), all with clutter level = 2. Simulations verified that this orthogonal hold-out method leads to unbiased estimates of the true size and clutter parameters, whereas other hold-out methods such as a latin-square leave-out procedure or a random sampling leave-out procedure do not (see Supplementary Methods).

In this leave-out procedure, at each iteration, the test categories always spanned the full range of size (or clutter) levels from 1 to 6. Accordingly, performance was assessed by calculating the correlation between the actual size (or clutter) level and the predicted level (which could be any real-valued number) of the 24 test patterns. Additionally, we calculated percent accuracy by rounding the predicted level to the nearest integer between 1 and 6, and computed the proportion of correctly predicted labels. Chance performance in this analysis for random guessing is 1/6, and is reported in the supplementary methods. These measures were calculated for each iteration. The overall performance was computed as the average correlation and percent correct across the 6 iterations for each subject and each ROI. The correlation measure takes advantage of the continuous nature of the predicted levels, and will show a higher correlation if the predicted level is closer in magnitude even if the rounded predicted level is incorrect (e.g., predicting level 4 size rather than level 5 size). The accuracy measure only captures whether the predicted level was the same as the actual level and does not take into account near misses. Additional analyses verified that several alternate model validation schemes (e.g., assessing the goodness-of-fit between the predicted and actual values using only 6 points averaged across repetitions of category, or assessing the slope of the fit between predicted and actual levels using either averaging scheme) yielded the same overall pattern of results.

To show that these effects were not biased by the size of the ROI, we conducted the ridge regression analysis using an equal number of voxels across ROIs, by randomly selecting 200 voxels for each ROI 50 times and averaging. The results remained the same whether we selected 200 voxels or used all the voxels in the ROI, so we report the results using the entire ROI (Cox and Savoy 2003; Pereira et al. 2009; Park et al. 2011).

Results

Experiment 1: Parametric Variation to Scene Size

In Experiment 1, we first searched for neural regions that show sensitivity to the size of the space depicted in the scene in a way that does not rely on any category-specific features. Observers performed a one-back task while viewing images of scenes, presented in blocks by semantic category (e.g., kitchens, dining rooms, pantries). Each observer saw 18 scene categories spanning 6 levels of scene size. We first approach the data with a targeted ROI-based analysis to ask whether our ROIs show any increase or decrease with changes in the scene size. Then, we analyze the whole brain to see if any regions outside our ROIs show a parametric change with scene size.

ROI Analysis: Average Activity Modulation by Size

We first examined how the average BOLD activity in our ROIs changed as a function of the size of the space. In each independently localized ROI, we extracted an average BOLD activity for each scene category, and tested a linear relationship between size levels and the magnitude of the response (Fig. 3B). The activity within both the PPA and RSC showed a
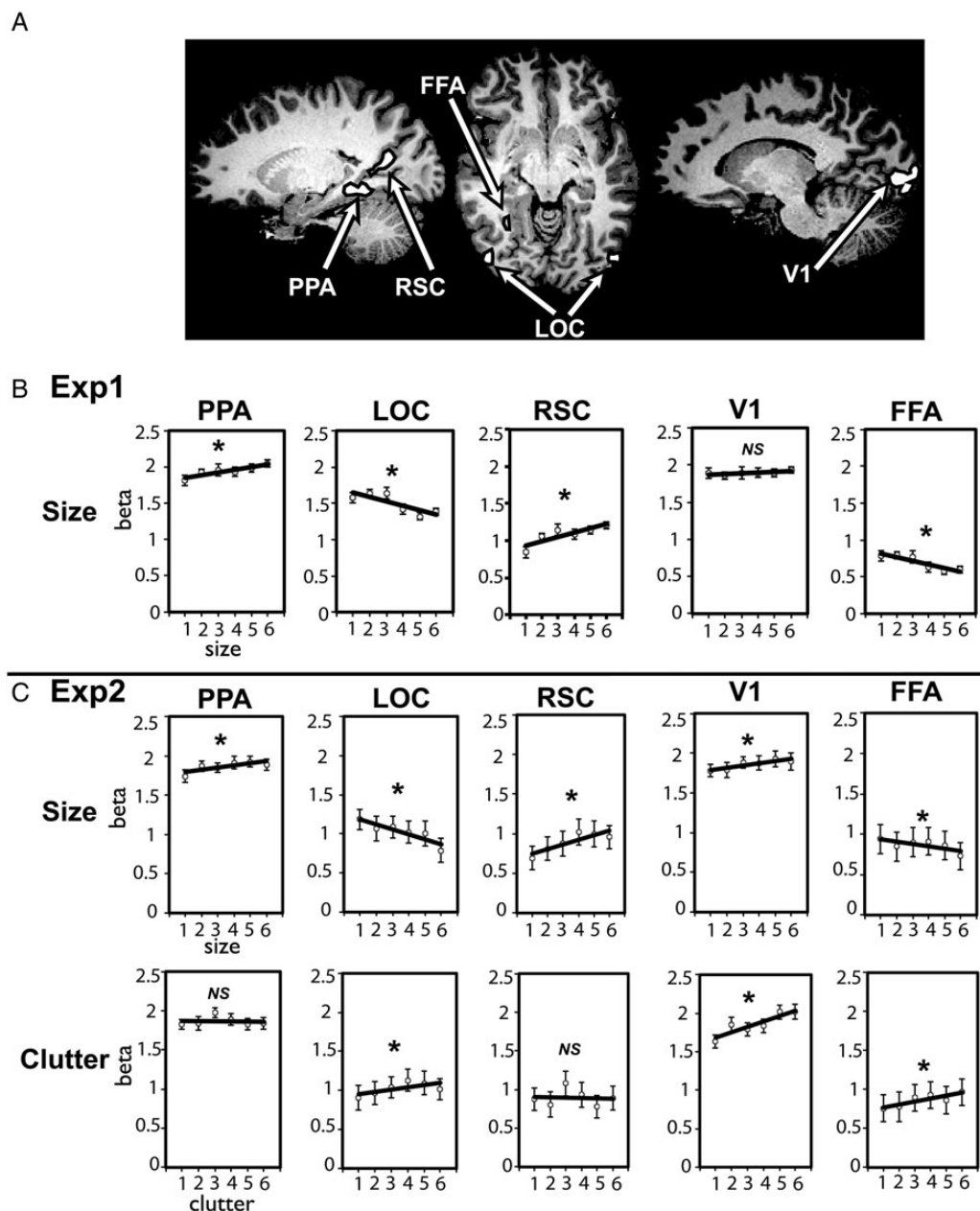
Figure 3. (A) Regions of interest shown on a representative participant's brain. (B) Experiment 1: Average beta weights are shown for the size of the space (1–6) in the PPA, RSC, LOC, FFA, and V1. The PPA and RSC showed a significant increase of activity as the size of scenes increased; LOC and FFA showed a significant decrease of activity as scene size increased. (C) Experiment 2: Average beta weights are shown for the size of the space (1–6) and for the levels of clutter (1–6) in the PPA, RSC, LOC, FFA, and V1. The PPA, RSC, and V1 showed a significant increase of activity as the size of scenes increased; LOC and FFA showed a significant decrease of activity as scene size increased. For clutter, there was no significant difference in the PPA and RSC activity as the amount of clutter increased or decreased. The LOC, FFA, and V1 showed a significant increase of activity as the amount clutter in a scene increased. Error bars reflect ±1 standard error of the mean.

significant increase in activity as the size of scenes increased (PPA: $r = 0.3$, $z' = 0.32$, $t_{(11)} = 5.7$, $P < 0.01$; RSC: $r = 0.3$, $z' = 0.32$, $t_{(10)} = 4.15$, $P < 0.01$). In other words, these regions showed a preference for large spaces such as shopping malls, concert halls, and sports stadiums, relative to smaller spaces such as closets, showers, and pantries. It is interesting to note that, while the PPA showed a significant increase of activity as scene size increased, the amount of increase was relatively small (12.6% increase from size levels 1–6; average beta for size level $1 = 1.81$; size level $6 = 2.0$) compared with RSC, which showed much bigger modulation (44% increase from size level 1–6; average beta for size level $1 = 0.83$; size level $6 = 1.2$).

In contrast, activity in LOC decreased as a function of scene size, showing a preference for small spaces over large spaces ($r = -0.33$, $z' = -0.39$, $t_{(10)} = -3.8$, $P < 0.01$). This region is well established to process object and shape information (Malach et al. 1995; Grill-Spector et al. 1998; Kourtzi and Kanwisher 2000; Eger et al. 2008; Vinberg and Grill-Spector 2008), which may become more apparent and defined as the scenes become smaller. FFA had low overall responses to these scene categories, but nonetheless showed modulations that patterned with LOC ($r = -0.24$, $z' = -0.26$, $t_{(11)} = -3.1$, $P < 0.01$). Finally, V1 showed no modulation of activity by the size of space ($r = 0.09$, $z' = 0.10$, $t_{(10)} = 1.19$, $P = 0.26$), suggesting that stimuli were relatively similar in their average spectral energy across the size levels.

This parametric design allows us to more directly assess the relevance of the size dimension than had we used only the 2 poles (e.g., small vs. big). For example, suppose that we only knew the PPA and RSC showed a higher response to large scenes than small scenes. This would present initial evidence that these regions are sensitive to the size of the space. However, if these regions had an even lower response to medium spaces than to small spaces, it suggests these regions may actually be responding to a property of all large scenes that is not related to size per se. Put another way, predicting a parametric response is a stronger test of what is explicitly coded in the magnitude of a region's response. In this case, the higher response to large scenes that we observe in the PPA and RSC is not likely to be driven by an unrelated property, and more likely to reflect scene size, when put in the context of a parametric modulation over levels of scene size.

Whole-Brain Analysis: Parametric Main Effects

To locate any regions beyond our targeted ROIs that show a parametric response to scene size, we next performed a whole-brain random-effects group analysis with parametric regressors (see Materials and Methods). Two different parametric models were fit: one predicting an increase of BOLD activity as the levels of scene size increased (Fig. 4A), and the other predicting a decrease of BOLD activity as the levels of scene size decreased (Fig. 4B).

Regions with a parametric increase in activity as scene size increased were right retrosplenial cortex (Tal coordinates: 13 −45 11) and left and right parahippocampal gyri (−14 −33 −4; 14 −33 −4), consistent with the location of the functionally localized ROIs. Additionally, the parahippocampal region extended more anteriorly in the right hemisphere (17 −27 −13). This result hints at an anterior specificity in the parahippocampal cortex—in most participants, the anterior part of the

parahippocampal cortex showed a stronger response with increasing scene size, with a locus sometimes outside of the functionally localized PPA. Regions that showed parametrically decreasing activity as the scene size increased were the left and right LOC (−37 −75 14; 49 −72 6), left posterior fusiform gyrus (−32 −42 −13), and left superior occipital cortex which corresponds to V3/V3A regions (−13 −96 20). Taken together, the results of this whole-brain analysis confirm that the well-known components of the scene network area are also the primary locations that show strong parametric sensitivity to scene size.

Split-ROI Analysis for the Anterior and Posterior PPA

Based on the anterior parahippocampal cortex activity found in the whole-brain analysis, we conducted an additional analysis in which we divided the PPA region in each participant into anterior and posterior halves (Fig. 5A). Note that the functionally defined PPA is localized to the posterior aspect of the parahippocampal gyrus, and thus the "anterior PPA" ROI does not exactly correspond to the "anterior parahippocampal" region that we observed in the whole-brain analysis. However, a number of studies have now indicated that the PPA may not be homogeneous, but may have functionally distinct subregions along the anterior and posterior axis (e.g., Bar and Aminoff 2003; Epstein 2008; Baldassano et al. 2013; Fairhall et al. 2013). Thus, any reliable differences between anterior versus posterior PPA provide further evidence that there are as-yet-uncharacterized subdivisions that exist within the PPA, or more generally, along the extent of the parahippocampal gyrus (e.g., Cant and Xu 2012).

While both the posterior and the anterior subdivision of the PPA show a significant increase in activity with scene size (Fig. 5B), the modulation in the anterior PPA was significantly stronger (left hemisphere: anterior vs. posterior PPA: $t_{(10)} = 2.6$, $P < 0.03$; right hemisphere: anterior vs. posterior PPA: $t_{(11)} = 2.9$, $P < 0.02$; Fig. 5B; left/right anterior/posterior subdivisions increase with size: all t's > 3.7, all P's < 0.01). On average, the anterior parts of the PPA had a 19.5% increase from size levels 1–6, while the posterior parts of the PPA only showed a 9.3% increase from size levels 1–6. As a control, and to test for the generality of this anterior locus, we also examined the anterior versus posterior difference within RSC, but did not find any difference across the anterior and posterior parts (left hemisphere: anterior vs. posterior RSC: $t_{(9)} = -0.95$, $P = 0.37$; right hemisphere: anterior vs. posterior RSC: $t_{(11)} = 1.6$, $P = 0.14$). These split-PPA analyses converge well with our whole-brain analysis results, demonstrating that the more anterior aspects of the parahippocampal cortex have a stronger parametric response to increasing scene size.

Experiment 2: Parametric Variation to Size and Clutter

In Experiment 2, we further examined how the physical size of a scene was represented across the brain, while independently manipulating another scene property–"functional clutter." Different sized spaces can be more or less filled, raising potential covariance between the size of the scene and the degree of clutter. By manipulating both the size and clutter properties, this experiment allows us to examine multiple scene properties simultaneously, and also provides the opportunity to replicate Experiment 1, using different stimuli sets and different participants.
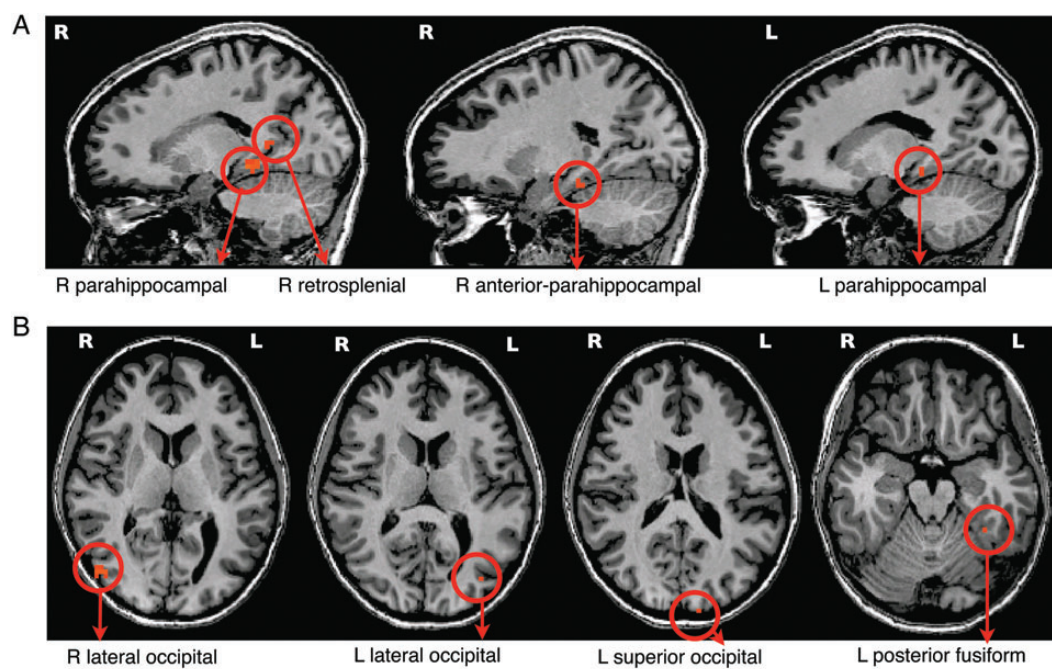
Figure 4. The figure shows regions from parametric whole-brain analysis (random-effects analysis, $P < 0.005$, cluster threshold >54 mm^3) in Experiment 1. Names of regions are marked with arrows below each figure. (A) Regions showing parametrically higher activity as the size of scenes increase; (B) Regions showing parametrically higher activity as the size of scenes decrease; In the table, names of regions showing parametric modulation of activity with Talairach coordinates for peak voxels within each regions, peak and average T values, number of voxels, and the count of subjects who showed the same regions in individual whole-brain analyses.

Observers performed a one-back task while viewing images of scenes presented in blocks by semantic category (e.g., kitchens, dining rooms, pantries). Each observer saw 36 scene categories that spanned 6 size levels and 6 clutter levels. Only 5% of the images in Experiment 1 overlapped with the images in Experiment 2; thus, any convergence between the 2 experiments suggests that the results are unlikely to be driven by any peculiarities of the stimulus sets. Further, with this expanded and more powerful design, we also introduce a new parametric multivoxel pattern analysis, to go beyond the main effects and examine whether the finer scale patterns within a region can predict the size of the space or the level of clutter.

ROI Analysis: Average Activity Modulation by Size and Clutter
In each independently localized ROI, we extracted the average BOLD activity for each scene category. First, we tested a linear relationship between the size levels and the magnitude of the response (Fig. 4C). The PPA and RSC were increasingly more active for scenes with larger sizes (PPA: $r = 0.19$, $z' = 0.20$, $t_{(11)} = 3.9$, $P < 0.01$, 8.5% increase from size level 1 to level 6; RSC: $r = 0.25$, $z' = 0.27$, $t_{(10)} = 4.97$, $P < 0.01$, 38.8% increase from size level 1 to level 6). In contrast, LOC and FFA showed the opposite pattern, and were more active for scenes with smaller sizes (LOC: 33.4% decrease from size level 1 to size level 6, $r = -0.26$, $z' = -0.28$, $t_{(9)} = -3.99$, $P < 0.01$; FFA: 22.7%
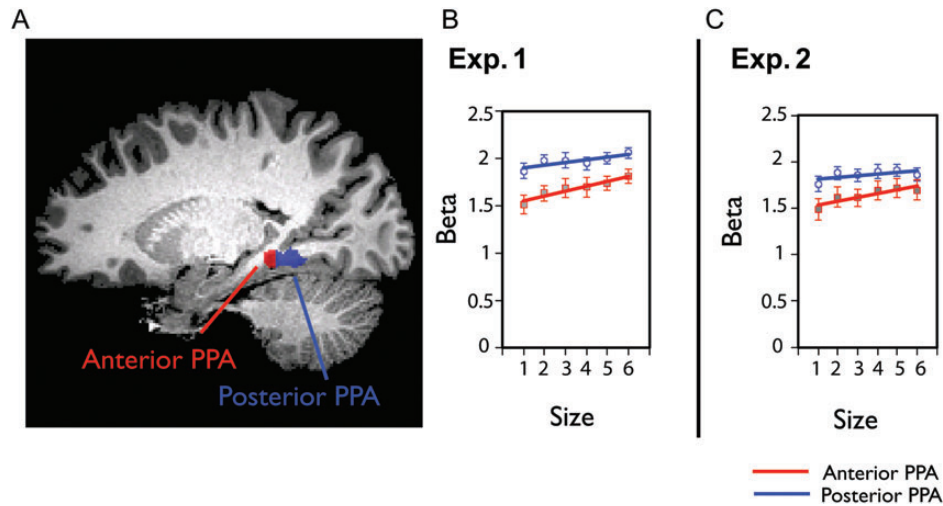
Figure 5. (*A*) An example of the anterior and posterior PPA of one representative subject is shown. The anterior and posterior PPA were defined by a split-half analysis within each individual's functionally localized PPA. (*B*) The average beta weights for the size of space in the anterior and posterior halves of the PPA in Experiment 1. The anterior and posterior halves are indicated by red (anterior) or blue (posterior) lines. The PPA shows greater increase of activity for anterior subdivisions compared with posterior subdivision. (*C*) The anterior–posterior PPA analysis in Experiment 2 showed the same result. Error bars reflect ±1 standard error of the mean.

decrease from size level 1 to size level 6, $r = -0.14$, $z' = -0.14$, $t_{(11)} = -2.94$, $P < 0.05$). We also observed that V1 showed an increase in activity as the scene size increased ($r = 0.19$, $z' = 0.20$, $t_{(11)} = 3.26$, $P < 0.01$). With the exception of V1, these data nicely replicate the results obtained in Experiment 1.

Examining the relationship between the degree of clutter and the overall BOLD response yielded a different pattern of results. Neither the PPA nor RSC showed a significant modulation to clutter (all t's < -0.37, $P > 0.5$; Fig. 3C). In contrast, LOC showed an increase of activity as the amount of clutter increased, indicating that LOC activity is greater as more objects fill the space ($r = 0.13$, $z' = 0.13$, $t_{(9)} = 3.25$, $P < 0.01$). As might be expected, V1 also showed an increase of activity as the amount of clutter increased ($r = 0.43$, $z' = 0.47$, $t_{(11)} = 9.62$, $P < 0.001$). Comparing size and clutter across the PPA and RSC directly, the RSC showed a marginally stronger parametric variation to size versus clutter, compared with PPA ($F_{1,10} = 3.3$, $P = .09$).

Note that size and clutter effects are calculated with the same set of brain data, only grouped differently by size or clutter properties. Thus, these data demonstrate that there are systematic physical properties that are shared across scene categories—here size and clutter, which drive the overall responsiveness of these scene-processing regions. Further, each of the regions has a different pattern of sensitivity to these properties, suggesting that size and clutter information is not isolated to single nodes in the scene network but instead is heterogeneously distributed across these regions.

*ROI Analysis: Parametric Multivoxel Pattern Analysis*

While any change in the overall responsiveness of an ROI shows a broad-scale sensitivity to that scene property, there may also be more fine-grained information about the size of the space or the degree of clutter contained in the patterns of the ROI. To examine this possibility, we developed a novel multivoxel pattern analysis that tests for a parametric response in the patterns of an ROI. This method can take advantage of a heterogeneous population in which for example some voxels have decreasing activity with scene size, while others have increasing activity with scene size, and where some voxels may

be more strongly modulated than others. All of these voxels are informative for predicting the size level of a scene, and this can be capitalized on in a parametric pattern analysis.

Patterns of activity were extracted for each presented block of each scene category, for each participant, and for each ROI. Using a subset of the patterns as training data (Fig. 6A), we used linear regression to fit a model (which is a set of weights on each voxel) that best maps the multivoxel scene patterns to their corresponding size or clutter levels (Fig. 6B). Given that, for any regression, the number of voxels in a given ROI is much more than the number of scene patterns, we included a regularization term in the regression (ridge regression, see Materials and Methods). Next, we tested the model by trying to predict the level of size or clutter, on a real-valued scale, for an independent set of scenes (Fig. 6C). These test scenes were always from different semantic categories than were used to train the model. Performance was assessed as the correlation between the actual size of the scene categories and the predicted size from the model. This correlation measure takes advantage of the continuous nature of the predicted levels, and will show a higher correlation if the predicted level is closer in magnitude even if the rounded predicted level is incorrect (e.g., predicting level 4 size rather than level 5 size).

The results of the ridge regression are shown in Figure 7 (see also Supplementary Table 1). We observed that the patterns in the PPA were able to predict both the size and clutter properties of the scene categories, but were significantly better at predicting the size of the scene ($t_{(11)} = 2.34$, $P < 0.05$). This effect was even more pronounced in the RSC: the patterns of the RSC were much more sensitive to the size of the space than the degree of clutter ($t_{(10)} = 8.27$, $P < 0.001$). Comparing our scene-selective ROIs directly confirmed that these 2 regions had significantly different sensitivity to the 2 properties, where RSC showed a significantly larger size versus clutter difference than the PPA ($F_{1,43} = 15.1$, $P = 0.003$). Patterns in the lateral occipital complex were able to predict both size and clutter levels, with a trend toward better performance on the clutter property though this did not reach significance ($t_{(9)} = 1.26$, $P > 0.2$).

A ROI Regression Procedure: Divide scene categories into train/test sets
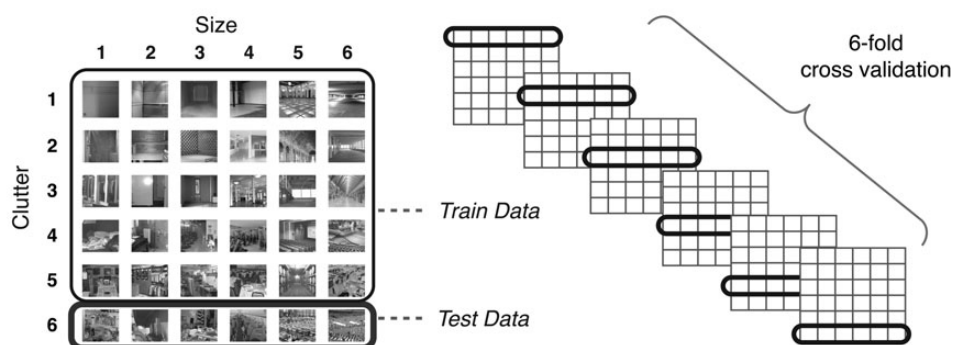
B Train: Use the patterns from 30 scene categories (train data) and the actual size levels to fit model (weight vector)
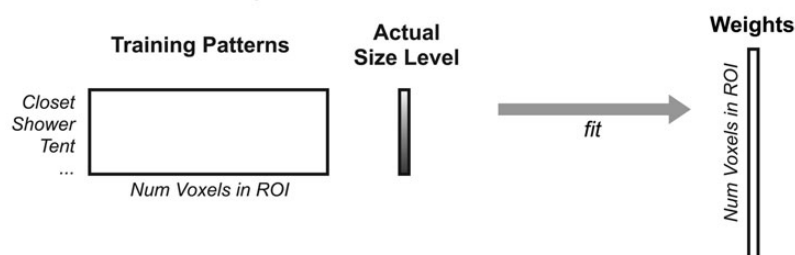
C Test: Use model to predict the size levels of the 6 remaining scene categories (test data), and correlate with the actual size levels to assess model performance
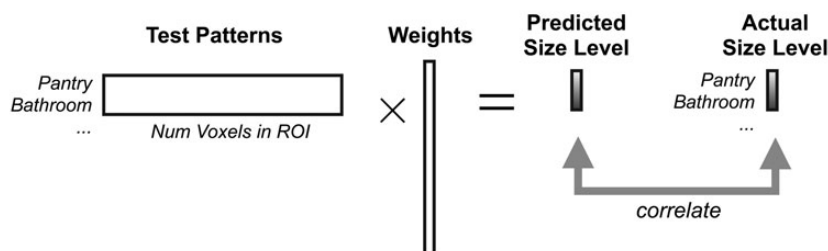

Figure 6. ROI regression procedure. (*A*) To test for a parametric pattern representation of a size (and clutter), we first divided the data into a training set with 5 of 6 categories per size level (30 categories total), and a test set with the remaining 6 categories, which all shared the same clutter level. This method of dividing data into train and test sets is necessary to have an unbiased estimate of size prediction independent of clutter (and vice versa). A 6-fold validation procedure was conducted which iteratively left out 6 scene categories that spanned the range of sizes and shared a clutter level. Analogous procedures were used for clutter-regression models. (*B*) In the training phase, the patterns from 30 training categories and their actual size levels were used to fit a model (set of weights on each voxel). (*C*) During the testing phase, the test patterns were multiplied by the weight vector to generate predicted size levels (which could take on any real value). Performance was assessed by computing the correlation between the actual and predicted size levels, averaged over all iterations, and was aggregated across subjects for each ROI.

Interestingly, we also observed very high performance in early visual areas, with V1 performance as high as the PPA ($F_{(1,43)} = 0.6$, $P = 0.447$) and better than LOC ($F_{(1,39)} = 22.2$, $P = 0.001$). Computational modeling approaches have shown that the size of a scene and its clutter level can be estimated from a set of orientation and spatial frequency measurements across the image (Oliva and Torralba 2001; Rosenholtz et al. 2007), and area V1 is known to have a high resolution of representation of orientation and spatial frequency across the visual field (DeYoe and Van Essen 1988; Tootell et al. 1998; Boynton and Finney 2003; Van Essen 2003; Murray et al. 2006). Thus, this result suggests that V1 patterns as measured by the BOLD signal are of a sufficient resolution and reliability to allow for successful multivoxel predictions of size and clutter levels (see also Naselaris et al. 2009).

The FFA showed very low overall performance on both size and clutter properties, with no difference between size and clutter ($t_{(11)} = 0.3$, $P = 0.8$), but surprisingly, the overall performance was slightly above chance in both properties (t's > 3.75; P's < 0.01). Thus, as a control, we examined a noncortex region in the skull, to verify that the regression procedure is unbiased. Indeed, in the skull ROI, there was no significant correlation between predicted and actual size or clutter levels (size: mean $r = 0.05$, $t_{(11)} = 1.14$, $P = 0.277$ (n.s.), clutter: mean $r = 0.07$, $t_{(11)} = 1.76$, $P = 0.107$ (n.s.), nor was percent correct significantly different from chance (16.67%) for either of these properties (size: mean pc = 16.4%, $t_{(11)} = -0.22$, $P = 0.831$, clutter: mean pc = 17.8%, $t_{(11)} = 0.95$, $P = 0.364$). This suggests that FFA, which does contain reliable responses to nonface
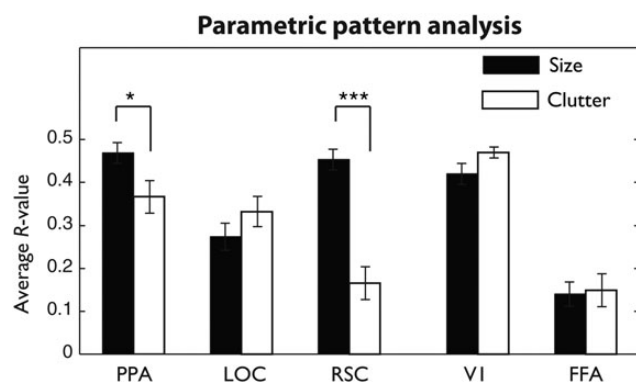
Parametric pattern analysis

Figure 7. Correlation between multivoxel model predictions and actual size or clutter levels, averaged over train–test iterations for each subject and each region of interest. Error bars reflect ±1 within-subject standard error of the mean (*$P < 0.05$, ***$P < 0.001$).

objects, also has some sensitivity to low-level features of scenes that vary across size and clutter (see also Stansbury et al. 2013).

To what extent are these parametric pattern analyses relying on the overall magnitude modulations of the ROI? To address this, we re-analyzed our data after mean-centering each pattern, thereby eliminating any overall main effects, and the same results were obtained (see Supplementary Analysis and Supplementary Fig. 2). Thus, the multivoxel pattern results are not simply reducible to the modulations in the overall response.

Finally, we also performed a classification analysis using a support-vector machine classifier. This analysis does not take into account the parametric relationship among the levels during training. Nonetheless, the errors the classifier makes tend to be similar to the true size level. This analysis is reported in the Supplementary Analysis and Supplementary Figure 1.

Whole-Brain Parametric Analysis

As in Experiment 1, we conducted a whole-brain analysis to locate any regions beyond our targeted ROIs that showed a parametric response to scene size or clutter. The results are shown in Figure 8. Areas that showed increased BOLD activity with increasing scene size again included the right retrosplenial cortex (Talairach coordinates: 19 −54 12), left and right parahippocampal gyri (−19 −32 −7; 14 −30 −7). Areas with the opposite trend, showing higher overall BOLD activity as scene size decreased, included the right LOC (38 −78 8) and right posterior fusiform gyrus (25 −42 −10). These results are consistent with main effects observed in the overall modulation of the targeted ROIs, and also nicely replicate the results of Experiment 1.

We also found a left and right medial occipital cortex corresponding to V1 (−11 −92 0; 28 −93 5), which showed a systematic increase with scene size. This is inconsistent with the results of Experiment 1, for which V1 activity was equal across the scene levels. To test whether this finding of V1 was due to spectral energy differences across size levels, we created an equalized image set based on a subset of the scene categories and re-analyzed the data (See Materials and Methods). With the equalized stimulus set, the effect in V1 was no longer present, and only the right retrosplenial cortex (19 −54 17) and bilateral anterior parahippocampal gyrus (−19 −39 −6; 14 −33 −7)

showed a parametric increase with scene size. This analysis provides an explanation for why the increase in area V1 was found in Experiment 2 but not in Experiment 1, and also generally supports the intuitive relationship between overall spectral energy in an image and the overall responsiveness of V1.

Turning to the clutter property, areas which increased with clutter included the left and right posterior fusiform gyri (−40 −66 −13; 37 −53 −14), as well as several more low-level visual areas around V1, V2 (−17 −95 −1; 14 −87 3) and even the lateral geniculate nucleus of the thalamus (LGN; −23 −27 5; 20 −27 3). No areas were found to increase in activity as clutter levels decreased. While the posterior fusiform gyri are likely related to object processing as this is the ventral surface analog to area LOC (e.g., Schwarzlose et al. 2008), the increase of activity in the early visual areas likely reflects the general increase in spectral energy at high spatial frequencies. Indeed, while for the scene size property we could create a equalized image set by simply removing a few of the object categories per size level, there was no similar way to equalize spectral energy in across clutter levels: High clutter is strongly correlated with power at high spatial frequencies.

Split-ROI Analysis for the Anterior and Posterior PPA

Finally, we examined whether the anterior PPA had a stronger modulation to size than the posterior PPA, as was found in Experiment 1. Consistent with our previous results, in both left and right PPA, the modulation by scene size was significantly greater in the anterior subdivision than posterior subdivision (left anterior vs. posterior PPA: $t_{(11)} = 2.5$, $P < 0.03$; right anterior vs. posterior PPA: $t_{(11)} = 3.1$, $P < 0.02$) (Fig. 5C, Experiment 2). On average, the anterior parts of the PPA had 14.6% increase from size levels 1 to 6, while the posterior parts of the PPA only showed a 5.7% increase from size levels 1 to 6. Further, in Experiment 1, all 4 subdivisions (left/right anterior/posterior PPA) maintained an overall modulation to scene size. In Experiment 2, however, after splitting the left PPA, the posterior aspect no longer showed a modulation by scene size ($r = 0.07$, $z' = 0.07$, $t_{(11)} = 1.1$, $P = 0.31$). Thus, in a new set of participants and with new stimuli, we replicated the result that the anterior PPA has an increased overall sensitivity to the size of a scene relative to the posterior PPA.

Because the anterior PPA showed a stronger modulation to scene size in its overall response, we also examined whether the anterior PPA had a loss of sensitivity to clutter across its multivoxel patterns, analogous to RSC. The results showed this was not the case: anterior PPA and posterior PPA showed similar predictive correlations for size and clutter levels as the full PPA (see Supplementary Fig. 3). Given that pattern analyses are more sensitive to fine-grained information, these results indicate that while anterior PPA is more strongly modulated by scene size than posterior PPA at a large scale, its fine-grained patterns maintain their sensitivity to clutter as well as size.

Discussion

Here, we examined whether there is evidence for neural coding of 2 basic scene properties: physical size (how large an enclosed space is) and functional clutter (the organization and quantity of objects that fill up the space). To approach this, we measured responses to different scene categories that parametrically varied along these particular scene properties, and we
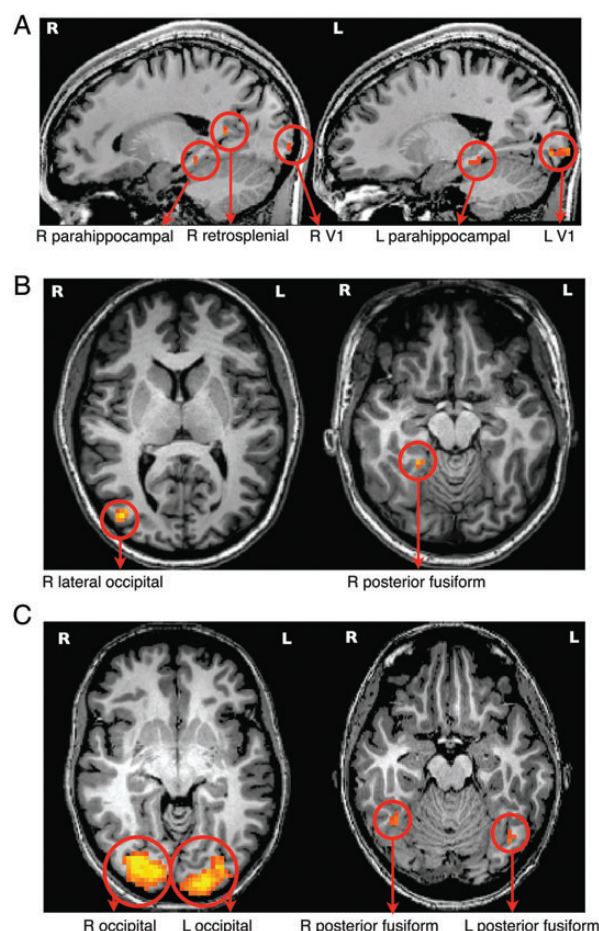
Figure 8. The figure shows regions from parametric whole-brain analysis (random-effects analysis, $P < 0.001$, cluster threshold >54 mm^3) for Experiment 2. Names of regions are marked with arrows below each figure. (*A*) Regions showing parametrically increasing activity as the size of scenes increase; (*B*) Regions showing parametrically increasing activity as the size of scenes decrease; (*C*) Regions showing parametrically increasing activity as the amount of clutter in scenes increases. (*D*) The table indicates the names of regions showing parametric modulation of activity with the Talairach coordinates for the peak voxel within each region, the magnitude of the peak *T* value and average the *T* values, the number of voxels, and the count of subjects who showed the same regions in individual whole-brain analyses.

examined whether grouping them by these properties led to systematic variation in the overall responsiveness of a region. We found that the PPA and RSC, both major components of the visual scene-processing network, showed a modulation in their overall response to scene size, but differed in the information present in their fine-grained patterns. PPA showed strong pattern sensitivity to both size and clutter, while RSC showed much stronger pattern sensitivity to size. LOC showed a complementary pattern of results, with a larger overall response for smaller and more cluttered scenes, but also with fine-grained patterns that could predict both clutter and size information. Importantly, by the logic of the design and analyses, these findings cannot be attributed to category-specific features in any of these regions. Broadly, these results expand our understanding of the distinctive roles of the regions in this network during natural scene analysis, and provide support for a property-based representation of scene information.

Parahippocampal Sensitivity to Both Spatial Boundaries and Content Properties

Here, we found that the PPA has a high response to all scene categories, but there was a reliable modulation by size (and not clutter), with larger scene categories (e.g., stadiums, arenas)

also eliciting a larger response. Interestingly, the patterns of the PPA were informative not only for the size of the space, a property describing the spatial boundary of a scene (see also Kravitz et al. 2011; Park et al. 2011), but also for the degree of clutter, a property describing the contents of the space (Oliva and Torralba 2001; Park et al. 2011). Traditionally, the PPA has been labeled as a "spatial layout analyzer," showing little modulation to object properties (Epstein and Kanwisher 1998; Epstein et al. 1999; Kravitz et al. 2011; Park et al. 2011). However, our understanding of the response properties of PPA have been updated by a number of recent studies that indicate sensitivity to objects (Mullally and Maguire 2011; Auger et al. 2012; Konkle and Oliva 2012; Troiani et al. 2012; Harel et al. 2013).

For example, in a related study to the present one, using a limited but controlled set of artificially generated simple scenes and single objects, Harel et al. (2013) also found that the PPA response is modulated not only by the layout of the scene, but also by the presence or absence of an object, and even by the specific identity of the object. Further, PPA has also been shown to respond parametrically to the real-world size of the object (Mullally and Maguire 2011; Troiani et al. 2012; Konkle and Oliva 2012). For example, Konkle and Oliva

(2012) found that large real-world objects such as cars activate the PPA, while small real-world objects such as keys activate lateral–occipital and inferior-temporal regions. These findings suggest that the size of both objects and scene enclosures in the world is a fundamental property of visually evoked neural responses, with a large-scale organization for small and large objects across the cortex, and with parametric sensitivity to different sizes of scenes. It will be important in future work to identify and explore the distinct and common properties of large objects, landmarks, and scenes to fully characterize the PPA response.

To add more complexity and richness to the PPA representation, a number of studies have shown that this region is also responsive to statistical summaries of multiple objects (Cant and Xu 2012) and even has preferences for particular low-level visual image features such as high spatial frequency and vertical/horizontal contours (Nasr et al. 2011; Rajimehr et al. 2011). Thus, a current challenge to our understanding of this scene-preferring region is whether the information in the PPA pertaining to objects and clutter in the space is better characterized as a more low-level statistical summary of featural content (e.g., Oliva and Torralba 2001; Alvarez and Oliva 2008; Cant and Xu 2012) or as a high-level object-based representation that explicitly represents object identity (e.g., Stansbury et al. 2013), or as some combination of both. Importantly, considering our findings with previous work on the responsiveness of the PPA, there is a strong convergence with the emerging view that PPA is not just a spatial layout analyzer, but is jointly sensitive to both the spatial boundaries properties and the content/textural properties of a scene view.

Anterior-to-Posterior Organization of the Parahippocampal Cortex

We also observed, in 2 independent experiments, that the more anterior aspect of the parahippocampal cortex showed a stronger overall response modulation to scene size than the posterior aspect. While this posterior/anterior difference was subtle, it hints at a functional division along the parahippocampal gyrus that echoes other recent findings (Bar and Aminoff 2003; Arcaro et al. 2009; Baldassano et al. 2013). For example, using a functional connectivity analysis, Baldassano et al. recently argued for such a division: The more anterior aspect of the PPA correlated more strongly at rest with the RSC, and the most posterior aspect of the PPA correlated more strongly at rest with LOC (Kravitz, Saleem, et al. 2011; Baldassano et al. 2013). This fits nicely with our findings, in which anterior PPA and the RSC showed more similar response patterns in their overall sensitivity to size. However, we might also have expected to find that posterior PPA would show a greater sensitivity to clutter than anterior PPA, and this was not observed. Thus, while there is not yet a simple method for functionally dissociating a posterior and anterior aspect of the PPA, the present results add to the mounting evidence for the existence of a functional subdivision within the PPA.

Retrosplenial Complex as a Geometric Analyzer

The present results here showed a very clear dissociation between the response properties of the PPA and those of RSC. Specifically, RSC showed, both in the overall response and in the patterns, a clear sensitivity to scene size with a markedly low sensitivity to functional clutter. The RSC is known to have sensitivity to scene layout and perspective (Kravitz, Peng, et al. 2011; Harel et al. 2013), with slight modulation to object properties strongly related to landmark or navigation (Auger et al. 2012; Harel et al. 2013; Troiani et al. 2013). Relatedly, RSC has also been linked to the behavioral phenomenon of boundary extension, in which the mental representation of a scene is larger than its physical percept—an illusion about the perceived size of a space (Park et al. 2007). Finally, the RSC is also situated on the medial surface, which places it far from the well-known object responsive regions on the lateral surface, and near to the transition zone between the PPA along the ventral stream, the far periphery of early visual areas, and medial dorsal stream regions (Kobayashi and Amaral 2003; Kravitz, Saleem, et al. 2011). This positioning with respect to other regions, along with its observed response properties, broadly supports the currently accepted view that RSC is involved in linking environmental spaces (Epstein 2008). Intuitively, knowing the size of a space is likely to be an important component for integrating a view into the larger environment, while the degree of clutter is not; thus, the current results showing RSC's strong sensitivity to size align with this interpretation. As such, our results echo and extend our understanding of RSC as a region that reflects the geometric properties of a space rather than the contents inside it.

A Property-Based Framework for Scene Processing

We know scene-processing regions have some information about categories (Walther et al. 2009), but what are the critical features that humans use to make such categorizations? Objects, or the co-occurrence of objects, are one proposed representational framework (where scenes have some probability of containing a fridge, a table, a tree, etc.; for a review, Oliva and Torralba 2007; Bar et al. 2008; MacEvoy and Epstein 2011; Greene 2013). A complementary representational approach is to consider a scene according to its global properties (where each scene has some degree of openness, clutter, size, perspective). Several proposals have been developed using both behavioral and computational modeling approaches to understand what these global properties are, how they might be extracted from natural scene statistics, and how they can support semantic categorization of scenes (Oliva and Torralba 2001, 2006; Greene and Oliva 2009a, 2010; Xiao et al. 2010; Kadar and Ben-Shahar 2012). Here, we show that size and clutter are 2 such global properties that may have an explicit neural coding, adding to a list that includes spatial boundaries (open/closed), and content-based properties (urban/natural) (Baker et al. 2011; Park et al. 2011). Importantly, both object-based and property-based representations are extracted at the early stage of scene processing to facilitate everyday tasks like scene recognition and way-finding. In future work, it will be informative to explore how the different tasks of scene categorization and way-finding (navigation) operate over representations that are more object-based or global property-based, and how these are accomplished by the scene-processing network of the brain.

Conclusion

In summary, the current study shows that information about two meaningful geometric properties of scenes—size and clutter—are explicitly coded in scene-selective cortical regions. By using different analytical approaches including regression

on multivoxel patterns, whole-brain random-effects analyses, and split-ROI analyses, we conclude that dimensions of size and clutter properties are parametrically coded in the brain, and these representations are "property-based" and flexible enough to generalize to different semantic categories. In particular, we propose a specialized role of RSC in representing physical size of space, independent of the amount of clutter in a scene. We suggest different sensitivity of anterior and posterior subdivisions of the PPA, adding a further support to recent studies that propose the PPA as a nonuniform region with anterior–posterior subfunctions. Broadly, we suggest that a property-based representation of size and clutter may support our rapid scene recognition and navigation in real-world environment.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxford journals.org/

Funding

Notes

References

Alvarez GA, Oliva A. 2008. The representation of simple ensemble visual features outside the focus of attention. Psychol Sci. 19 (4):392–398.

Arcaro MJ, McMains S, Singer B, Kastner S. 2009. Retinotopic organization of human ventral visual cortex. J Neurosci. 29:10638–10652.

Auger SD, Mullally SL, Maguire EA. 2012. Retrosplenial cortex codes for permanent landmarks. PLoS One. 7(8):e43620.

Baldassano C, Beck DM, Fei-Fei L. 2013. Differential connectivity within the parahippocampal place area. Neuroimage. 75:228–237.

Bar M, Aminoff E. 2003. Cortical analysis of visual context. Neuron. 38:347–358.

Bar M, Aminoff E, Schacter DL. 2008. Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. J Neurosci. 28:8539–8544.

Boynton GM, Finney EM. 2003. Orientation-specific adaptation in human visual cortex. J Neurosci. 23:8781–8787.

Cant JS, Xu Y. 2012. Object ensemble processing in human anterior-medial ventral visual cortex. J Neurosci. 32:7685–7700.

Cox DD, Savory RL. 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage. 19:261–270.

DeYoe EA, Van Essen DC. 1988. Concurrent processing streams in monkey visual cortex. Trends Neurosci. 11:219–226.

Eger E, Ashburner J, Haynes J, Dolan RJ, Rees G. 2008. fMRI activity patterns in human LOC carry information about object exemplars within category. J Cogn Neurosci. 20:356–370.

Epstein R, Graham KS, Downing PE. 2003. Viewpoint-specific scene representations in human parahippocampal cortex. Neuron. 37:865–876.

Epstein R, Harris A, Stanley D, Kanwisher N. 1999. The parahippocampal place area: recognition, navigation, or encoding? Neuron. 23:115–125.

Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. Nature. 392(6676):598–601.

Epstein RA. 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn Sci. 12:388–396.

Fairhall SL, Caramazza A. 2013. Category selective neural substrates for person- and place-related concepts. Cortex, doi: 10.1093/cercor/bht039.

Fei-Fei L, Iyer A, Koch C, Perona P. 2007. What do we perceive in a glance of a real-world scene? J Vis. 7(1):10, 1. –29.

Greene MR. 2013. Statistics of high-level scene context. Front Percept Sci. 4:777.

Greene MR, Oliva A. 2009b. The briefest of glances: the time course of natural scene understanding. Psychol Sci. 20:464–472.

Greene MR, Oliva A. 2010. High-level aftereffects to global scene property. J Exp Psychol Hum Percept Perform. 36(6):1430–1442.

Greene MR, Oliva A. 2009a. Recognition of natural scenes from global properties: seeing the forest without representing the trees. Cogn Psychol. 58(2):137–176.

Grill-Spector K, Kushnir T, Edelman S, Itzchak Y, Malach R. 1998. Cue-invariant activation in object-related areas of the human occipital lobe. Neuron. 21(1):191–202.

Harel A, Kravitz DJ, Baker CI. 2013. Deconstructing visual scenes in cortex: gradients of object and spatial layout information. Cereb Cortex. 23(4):947–947.

Hastie T, Tibshirani R, Friedman J. 2001. The elements of statistical learning. New York: Springer.

Hermer-Vasquez L, Moffet A, Munkholm P. 2001. Language, space, and the development of cognitive flexibility in the case of humans: the case of two spatial memory task. Cognition. 79:263–299.

Hoerl AE, Kennard RW. 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 12:69–68.

Joubert OR, Rousselet GA, Fize D, Fabre-Thorpe M. 2007. Processing scene context: fast categorization and object interference. Vis Res. 47:3286–3297.

Kadar I, Ben-Shahar O. 2012. Small sample scene categorization from perceptual relations. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR, Providence, Rhode Island, USA, December 2012.).

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. Nat Neurosci. 8:679–685.

Kobayashi Y, Amaral DG. 2003. Macaque monkey retrosplenial cortex: II. cortical afferents. J Comp Neurol. 466:48–79.

Konkle T, Oliva 2012. A real-world size organization of object responses in occipito-temporal cortex. Neuron. 74(6):1114–1124.

Kourtzi Z, Kanwisher N. 2000. Cortical regions involved in processing object shape. J Neurosci. 20(9):3310–3318.

Kravitz D, Peng CS, Baker CI. 2011. Real-world scene representations in high-level visual cortex—its the spaces not the places. J Neurosci. 31(20):7322–7333.

Kravitz DJ, Saleem KS, Baker CI, Mishkin M. 2011. A new neural framework for visuospatial processing. Nat Rev Neurosci. 12:217–230.

Learmonth A, Nadel L, Newcombe NS. 2002. Children's use of landmarks: implications for modularity theory. Psychol Sci. 13:337–341.

MacEvoy SP, Epstein RA. 2011. Constructing scenes from objects in human occipitotemporal cortex. Nat Neurosci. 14(10):1323–1329.

Malach R, Reppas JB, Benson R, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RBH. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci USA. 92:8135–8139.

Morgan LK, MacEvoy SP, Aguirre GK, Epstein RA. 2011. Distances between real-world locations are represented in the human hippocampus. J Neurosci. 31(4):1238–1245.

Mullally SL, Maguire EA. 2011. A new role for the parahippocampal cortex in representing space. J Neurosci. 31(20):7441–7449.

Murray SO, Olman CA, Kersten D. 2006. Spatially specific fMRI repetition effects in human visual cortex. J Neurophysiol. 95:2439–2445.

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL. 2009. Bayesian reconstruction of natural images from human brain activity. Neuron. 53:902–915.

Nasr S, Liu N, Devaney KJ, Yue X, Rajimehr R, Ungerleider LG, Tootell RBH. 2011. Scene-selective cortical regions in human and nonhuman primates. J Neurosci. 31:13771–13785.

Newman EL, Norman KA. 2010. Moderate excitation leads to weakening of perceptual representations. Cereb Cortex. 20(11):2760–2770.

Oliva A, Torralba A. 2006. Building the gist of a scene: the role of global image features in recognition. Prog Brain Res. 155:23–36.

Oliva A, Torralba A. 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vis. 42:145–175.

Oliva A, Torralba A. 2007. The role of context in object recognition. Trends Cogn Sci. 11(12):520–527.

Park S, Brady TF, Greene MR, Oliva A. 2011. Disentangling scene content from its spatial boundary: complementary roles for the PPA and LOC in representing real-world scenes. J Neurosci. 31 (4):1333–1340.

Park S, Chun MM. 2009. Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in scene perception. Neuroimage. 47:1747–1756.

Park S, Intraub H, Yi D-J, Widder D, Chun MM. 2007. Beyond the edges of a view: boundary extension in human scene-selective visual cortex. Neuron. 54:335–342.

Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage. 45:S199–S200.

Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RBH. 2011. The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. PLoS Biol. 9: e1000608.

Rosenholtz R, Li Y, Nakano L. 2007. Measuring visual clutter. J Vis. 7 (2):17, 1. –22.

Ross MG, Oliva A. 2010. Estimating perception of scene layout properties from global image features. J Vis. 10(1):2, 1. –25.

Schwarzlose RF, Swisher JD, Dang S, Kanwisher N. 2008. The distribution of category and location information across object-selective regions in human visual cortex. Proc Natl Acad Sci USA. 105 (11):4447.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Beliveau JW, Brady TJ, Rosen BR, Tootell RB. 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science. 268(5212):889–893.

Spiridon M, Kanwisher N. 2002. How distributed is visual category information in human occipital-temporal cortex? An fMRI study. Neuron. 35(6):1157–1165.

Stansbury DE, Naselaris T, Gallant JL. 2013. Natural scene statistics account for the representation of scene categories in human visual cortex. Neuron. 79, 5, 1025. –1034.

Tootell RB, Hadjikhani NK, Vanduffel W, Liu AK, Mendola JD, Sereno MI, Dale AM. 1998. Functional analysis of primary visual cortex (V1) in humans. Proc Natl AcadSci USA. 95:811–817.

Troiani V, Stigliani A, Smith ME, Epstein RA. 2012. Multiple object properties drive scene-selective regions. Cereb Cortex, doi: 10.1093/cercor/bhs364.

Tversky B, Hemenway K. 1983. Categories of scenes. Cogn Psychol. 15:121–149.

Van Essen DC. 2003. Organization of visual areas in macaque and human cerebral cortex. In: Chalupa L, Werner J, editors. The visual neuroseicnes. Cambridge (MA): MIT Press. 507–521.

Vinberg J, Grill-Spector K. 2008. Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. J Neurophysiol. 99(3):1380–1393.

Walther DB, Caddigan E, Fei-Fei L, Beck DM. 2009. Natural scene categories revealed in distributed patterns of activity in the human brain. J Neurosci. 29(34):10573–10581.

Xiao J, Hays J, Ehinger K, Oliva A, Torralba A. 2010. SUN database: large-scale scene recognition from abbey to zoo. Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, June 2010. pp. 3485–3492.

SUPPLEMENTAL DATA

Parametric coding of the size and clutter of natural scenes in the human brain

Soojin Park, Talia Konkle, & Aude Oliva

<u>**Table of contents:**</u>

Supplementary Analysis: SVM classification and confusion matrix for all ROIs
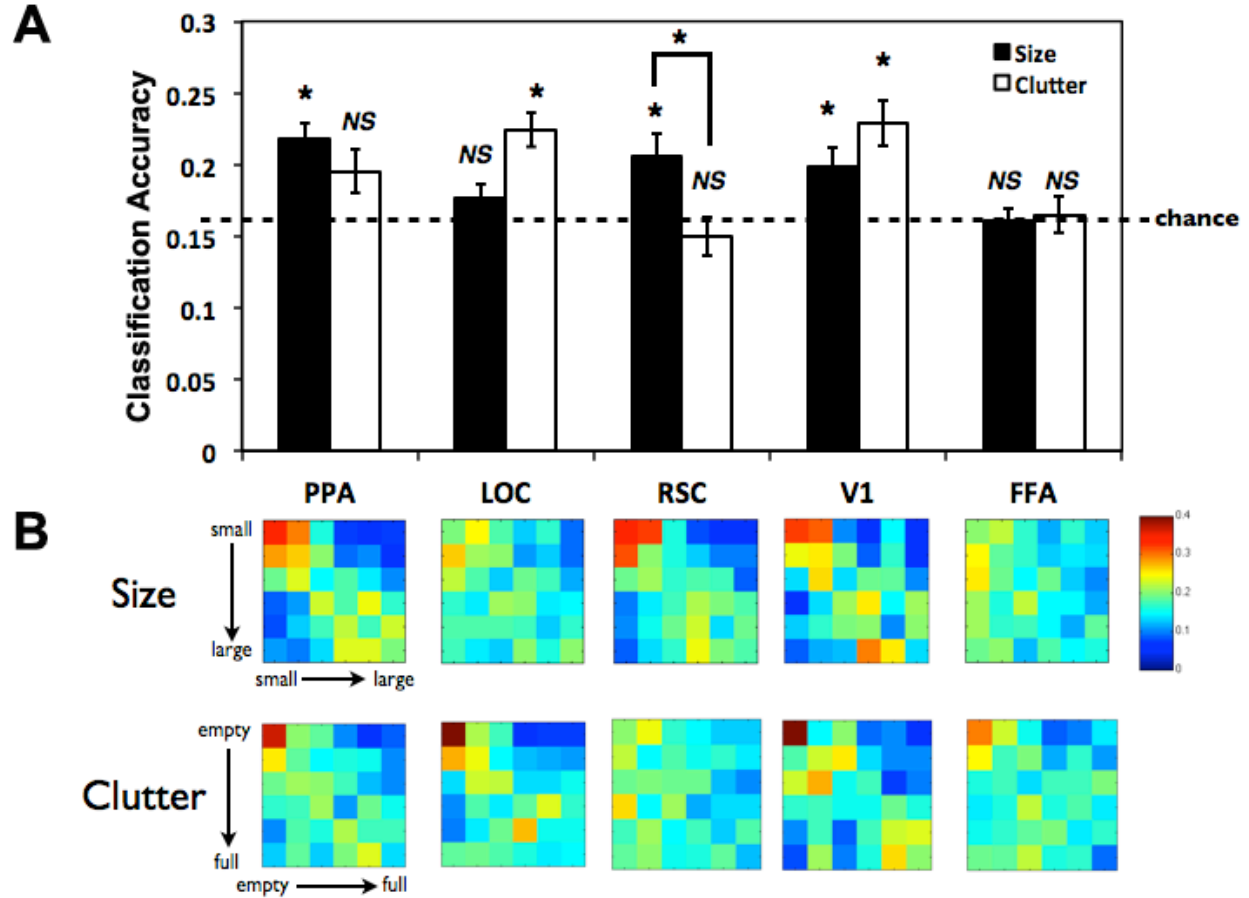

Figure S1. A. Standard SVM classification accuracy for the size and clutter levels. In addition to the ridge regression analysis, we also performed a standard multivariate pattern classification analysis using linear support vector machine (SVM) classifier using LIBSVM. For each block, we computed the average pattern of activity across the voxels. We used a "leave-one-entire-category-out" cross validation in which we used different sets of categories representing the same size of a scene as the training and test samples. For example, an SVM classifier was trained to classify the six size levels using all categories but one, and then was tested on the remaining category. This was repeated across all samples so that each sample of the dataset played a role in training and testing. The procedure was repeated for the clutter dimension as well. Percent correct classification for each subject and each ROI was calculated as the average performance over the cross-validation iterations.

When the classifier was tested for the size of a scene, the PPA, RSC, and V1 could classify the size of a novel scene category above chance (PPA: t(11) = 4.47, *p* < .01; RSC: t(10) = 2.49, *p* < .05), while LOC could not (t(9) = 1.13, *p* > .28). Such accurate classification requires the generalization of representation across semantic categories, since the classifier was always tested on a novel semantic category. On the contrary, when the classifier was tested for the amount of clutter with the same exact sets of scenes, LOC performed significantly above chance (t(9) = 4.84, *p* <.01), while the PPA only marginally classified the clutter level (t(11) = 1.85, *p* =

2

.09) and RSC could not classify clutter levels at all (t(10) = -1.271, *p* = .23). V1 could classify the clutter levels correctly (t(11) = 3.88, *p* < .01). Interestingly, RSC showed specificity for the representation of scene size and not clutter (paired t-test t(10) = 2.89, *p* < .02), while LOC showed a specificity for the representation of clutter and not size (paired t-test t(9) = -2.83, *p* < .05). There was a significant interaction across the RSC and LOC for the size and clutter (F(1,8)=17.49, *p* < .01).

B. Confusion matrices of the size and clutter classifier. The confusion matrix in the PPA and RSC for size shows a 'fuzzy diagonal', which suggests that confusion errors in the classifier were not random, and instead reflect a sensitivity to scene size (e.g., the classifier made more errors between adjacent levels, creating a fuzzy diagonal in the confusion matrix).

Supplementary Analysis: Parametric pattern analysis overall accuracy results

ROI	Size		Clutter		Size vs Clutter
	%Correct (sem)	difference from chance (chance = 16.7%)	%Correct (sem)	difference from chance (chance = 16.7%)	
PPA	25.2% (1.2)	t(11)=7.4, p<0.001*	22.5% (1.2)	t(11)=4.3, p<0.005*	t(11)=1.9, p=0.09 ~
LOC	20.1% (1.1)	t(9)=3.1, p<0.05*	20.9% (1.1)	t(9)=5.8, p<0.001*	t(9)=-0.8, p=0.46
RSC	26.3% (0.7)	t(10)=14.4, p<0.001*	17.9% (0.7)	t(10)=1.1, p=0.28	t(10)=8.9, p<0.001*
V1	23.1% (1.0)	t(10)=6.3, p<0.001*	24.3% (1.0)	t(10)=5.4, p<0.001*	t(10)=-0.6, p=0.53
FFA	17.8% (1.0)	t(11)=1.1, p=0.29	19.6% (1.0)	t(11)=2.5, p<0.05*	t(11)=-1.1, p=0.29

Table S1. Parametric pattern analysis overall accuracy results. In addition to the correlation between the predicted size/ clutter levels and the actual size/clutter levels, we also computed overall percent correct, in which we rounded the predicted level and compared it to the actual level. Chance in this method is 1/6 or 16.7%. The overall accuracy and standard error of the mean for each ROI is shown in Table S1. The table also shows the statistical tests between each ROI and chance, as well as the t-tests between performance on size versus clutter dimension.

Overall percent accuracies were on the range of 17.8% to 26.3%. We observed that all areas performed above chance classification for the size dimension except for the FFA; and all regions performed above chance classification for the clutter dimension except RSC. The RSC region was the only region to show significantly different classification performance between size and clutter, with a trend for this in PPA.

Supplementary Analysis: Permutation Analysis

To measure the probability of obtaining high correlations between predicted and actual levels of size or clutter by chance, we conducted a permutation analysis. For each subject, for a given ROI, we shuffled the scene labels and computed the average r-value following the 6-fold validation procedure. We repeated this procedure 100x to obtain a distribution of r-values. A p-value was obtained by fisher-z transforming the distribution of r-values obtained from 100 shuffled iterations along with the actual r-value for size (or clutter), and computing the z-score of the actual size or clutter regression coefficient with respect to the shuffled distribution. This procedure assumes a normal distribution of transformed r-values, which was appropriate given the shuffled distributions we observed; a non-parametric statistical procedure would require more samples than was computationally feasible. This analysis was performed for the PPA, RSC, and the FFA.

Comparing the r-value obtained when the labels are not shuffled allows us to estimate whether the regression pattern analysis was significantly different from a shuffled baseline in each ROI of individual participant. We found that, in the PPA, 11 of 12 participants showed a significant correlation for size and 9 of 12 participants showed a significant correlation for clutter. In the RSC, 11 of 11 participants showed a significant correlation for size and 2 of 12 participants showed a significant correlation for clutter. Interestingly, in the FFA, 2 of 12 showed a significant correlation for size and 3 of 12 for clutter. Importantly, across all cases across subjects and ROIs, the mean of the shuffled distribution was centered on zero. Together, these results indicate that (i) a correlation of zero is expected by chance, and (ii) at the single subject level, the regression model performance largely follows the results found presented at the group level.

Supplementary Analysis: Mean-Centered Parametric Pattern Analysis

To eliminate the influence of main effects on multi-voxel analysis, we mean-centered each pattern for each scene category in each ROI (so that the average beta was zero across the voxels in the region for each scene category pattern). When we re-analyzed these zero-mean centered patterns in ridge regression analysis, we found the same overall pattern of results. That is, the patterns of the PPA were able to predict both the size and clutter properties, but were significantly better at predicting the size of the scene ($t(11) = 2.76$, $p<.05$); the patterns of the RSC were much more sensitive to the size than clutter ($t(10)=6.48$, $p<.001$). There was a significant interaction across the two ROIs, where RSC showed a significantly larger size vs. clutter difference than the PPA ($F(1,43) = 11.3$, $p=.007$). These direct replications of the results using mean-centered multi-voxel patterns show that results that we obtained in the main manuscript can not be reduced to the main effect.

Comparing the original and mean-centered results directly revealed that the original ROI data showed very slightly but reliably higher correlations for size, but not clutter, dimensions. Size paired t-test: PPA vs PPA-MeanCentered: $t(11)=2.44$, $p=0.033$ *; Clutter paired t-test: PPA vs PPA-MeanCentered: $t(11)=1.99$, $p=0.072$ (n.s.); Size paired t-test: RSC vs RSC-MeanCentered $t(10)=3.15$, $p=0.010$ *; Clutter paired t-test: RSC vs RSC-MeanCentered: $t(10)=-1.21$, $p=0.255$ (n.s.). These likely reflect the contribution of the main effects of the ROI overall response to size.
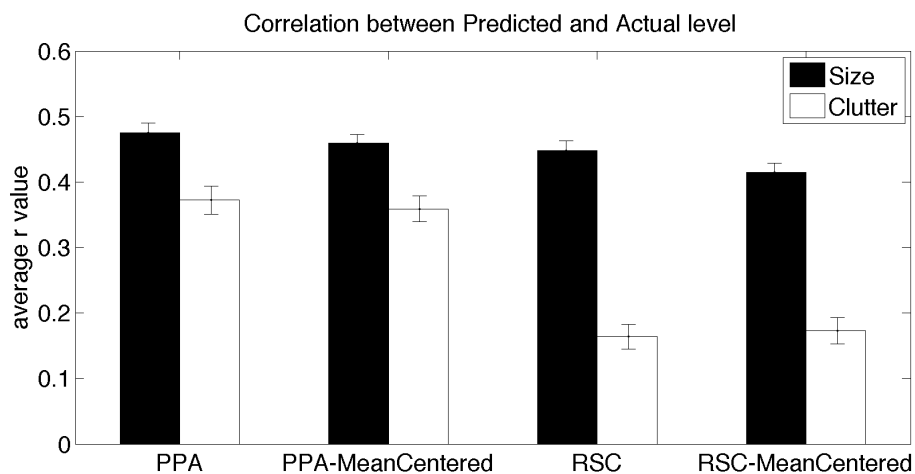

Figure S2. Original and mean-centered multi-voxel regression results in the RSC and PPA. Regions are along the x-axis and the average correlation for size and clutter dimensions across participants is plotted on the y-axis. Error bars reflect +/- 1 within-subject standard error of the mean.

Supplementary Analysis: Anterior and posterior PPA pattern analysis

We examined the degree of size and clutter information present in the multi-voxel patterns of the anterior and posterior subdivisions of the PPA. The anterior and posterior aspect of the PPA showed similar correlations with size and clutter levels as the full ROI (Figure S3). Overall, there was no main effect of anterior/posterior regions ($F(1,47)=0.5$, $p=0.503$), nor was there a main effect of size vs clutter ($F(1,47)=3.0$, $p=0.112$). However both subregions show a trend for higher correlations for the size dimension. Further, there was no interaction between anterior/posterior and size/clutter dimension ($F(1,47)=1.1$, $p=0.325$).
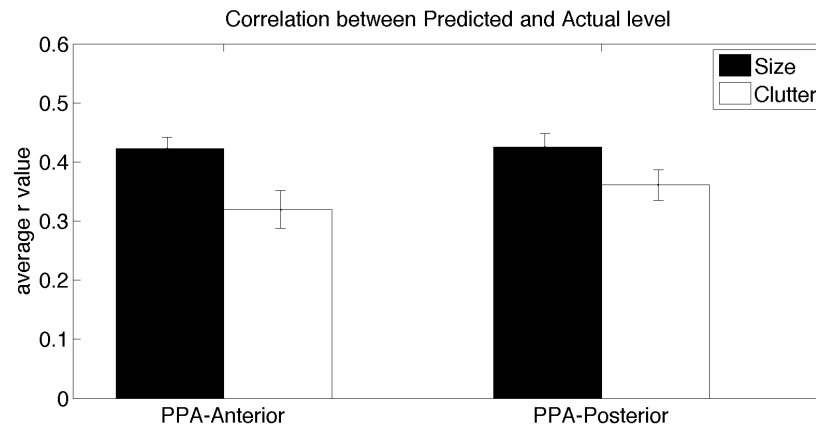

Figure S3. Split-ROI parametric pattern analysis. The PPA was divided into an anterior and posterior section (x-axis) and the average correlation for size and clutter dimensions across participants is plotted on the y-axis. Error bars reflect +/- 1 within-subject standard error of the mean.

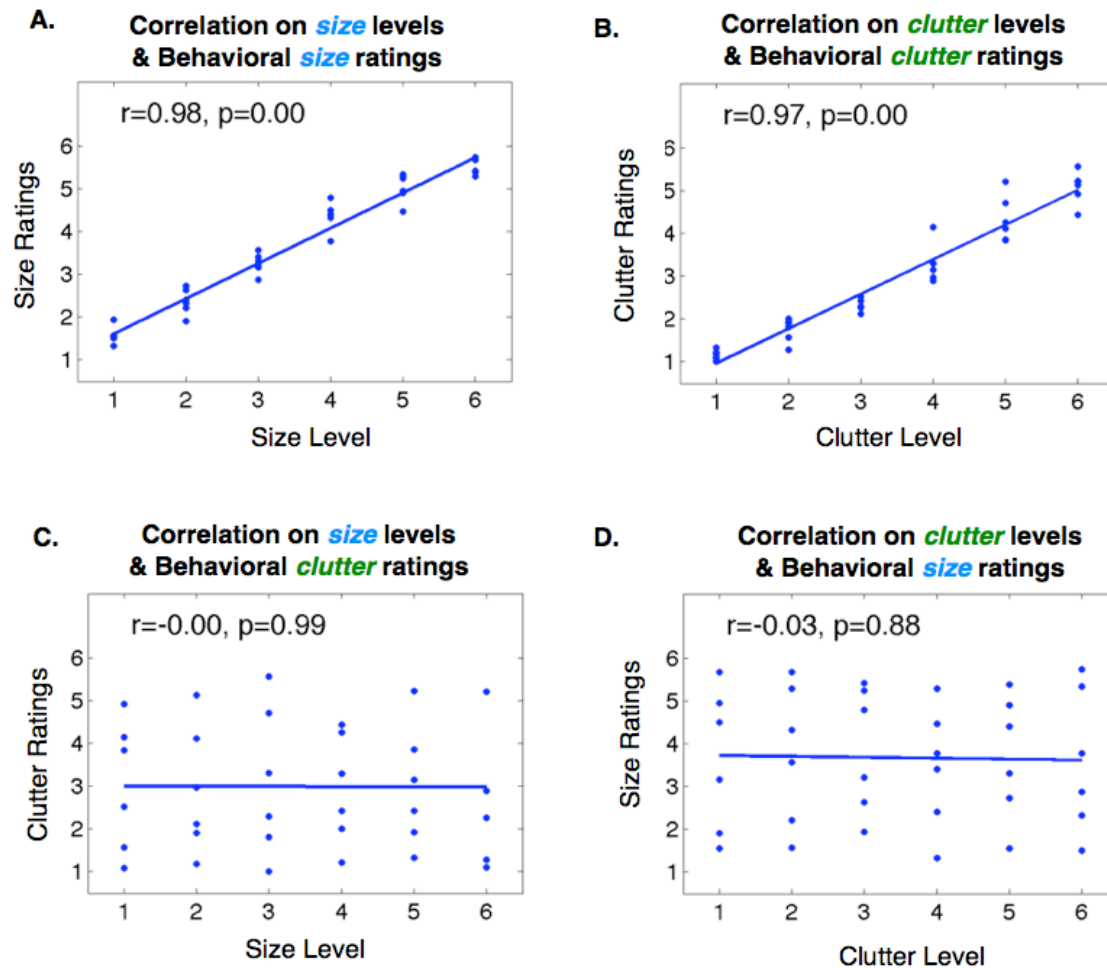# Supplementary Methods: Stimulus set validation experiment

Figure S4. Correlation on pre-selected size and clutter levels and behavioral ratings. A. Correlation between size levels and size behavioral ratings. **B.** Correlation between clutter levels and clutter behavioral ratings. **C & D.** Comparison between size levels and clutter behavioral ratings, and vice versa

In order to validate the size and clutter levels of the stimulus set used in Experiment 2, we obtained 20 ratings for the 432 images (36 categories * 12 exemplars) on both the size and clutter dimensions. Participants were recruited through Amazon Mechanical Turk. Each trial, an image was shown on the screen. In the size experiment, the instructions were to judge the size of the space on a scale of 1 to 6, where 1=very small and 6=very big. In the clutter experiment, the instructions were to judge the clutter in the scene on a scale from 1-6, where 1=empty and 6=fully cluttered. In total, we collected 20 judgments for both size and clutter dimensions per each of 432 individual images.

We compared the pre-selected size and clutter values from the stimulus set ("levels") with size and clutter average ratings from the behavioral experiments ("ratings"). We observed that there was a strong correlation between the pre-selected levels and the behavioral ratings at the

categorical level, for both size (r=.98, p<0.01) and clutter (r=0.97, p<0.01) dimensions (Figure S4 A & B). Consistent with the orthogonal nature of the design, there was no relationship between the size levels and clutter ratings or vice versa (r=0.0, r=-0.3; Figure S4 C & D), and the size ratings and the clutter ratings were not correlated with each other (r=-0.3, all p's > .2). Finally, we found that the ratings were quite consistent across the images within each category: the item standard deviation was 0.30 for size and 0.32 for clutter, averaged over the 36 scene categories. In other words, across the items within each category, the standard deviation of the rankings was less that half of a level on the 6-point scale.

Supplementary Methods: Simulation analysis for ridge regression

We performed a simulation to verify that our hold-out method in the ridge regression train/test procedure leads to unbiased estimates of true size and clutter relationships. We considered how different leave-out methods were able to recover underlying structure, considering several possible relationships between size and clutter. Here we show the results for 3 such data sets: when the data contained a perfect parametric modulation of one dimension and no modulation of the other, vice versa, or perfect modulations of both dimensions. We simplified the dimensionality of the data, considering a 1-voxel ROI. The different leave-out methods were:

(1) *random* item per level: for each size level, leave out one of the 6 categories randomly, ignoring clutter levels

(2) *latin* square leave-out: for each size level, leave out one of the 6 categories so that, across all 6 levels of size, one of each of the 6 levels of clutter has been left out.

(3) *diagonal* leave out: for each size level, leave out the scene with the corresponding level of clutter (e.g. for size level 1, leave out category with clutter level 1, etc.)

(4) *opposite diagonal* leave out: for each size level, leave out the scene with the "opposite" level of clutter (e.g. for size level 1, leave out category with clutter level 6, etc.)

(5) *orthogonal*-leave-out: for each size level, leave out a specific level of clutter and repeat this six times. (e.g. on iteration 1 leave out the categories with clutter level 1; on iteration 2, leave out the categories with clutter level 2, etc.)

For each data set and each leave-out method, we (i) simulated a noisy voxel based on the true parametric relationship across size and clutter dimensions, (ii) fit a regression model (a weight on the voxel) to predict the training levels given the voxel's response; (iii) computed the predicted levels of the held-out test set based on this model, (iv) correlated the predicted levels with the actual levels, (v) and summarized the results as the average fisher-z transformed correlation over 1000 iterations.

Figure S5 summarizes the results. We found that only the *orthogonal*-leave-out procedure returned unbiased estimates of the size and clutter dimensions. Random category leave out methods correctly estimated the informative dimension, but showed systematic negative correlations for the non-informative dimension (rather than zero correlations, as would be expected from the structure of the data). The same was true for a latin-square leave out procedure and a diagonal leave-out procedure. These simulation results hold over a range of noise levels in the voxel's response. Based on these simulations, we used the *orthogonal*-leave-out procedure.
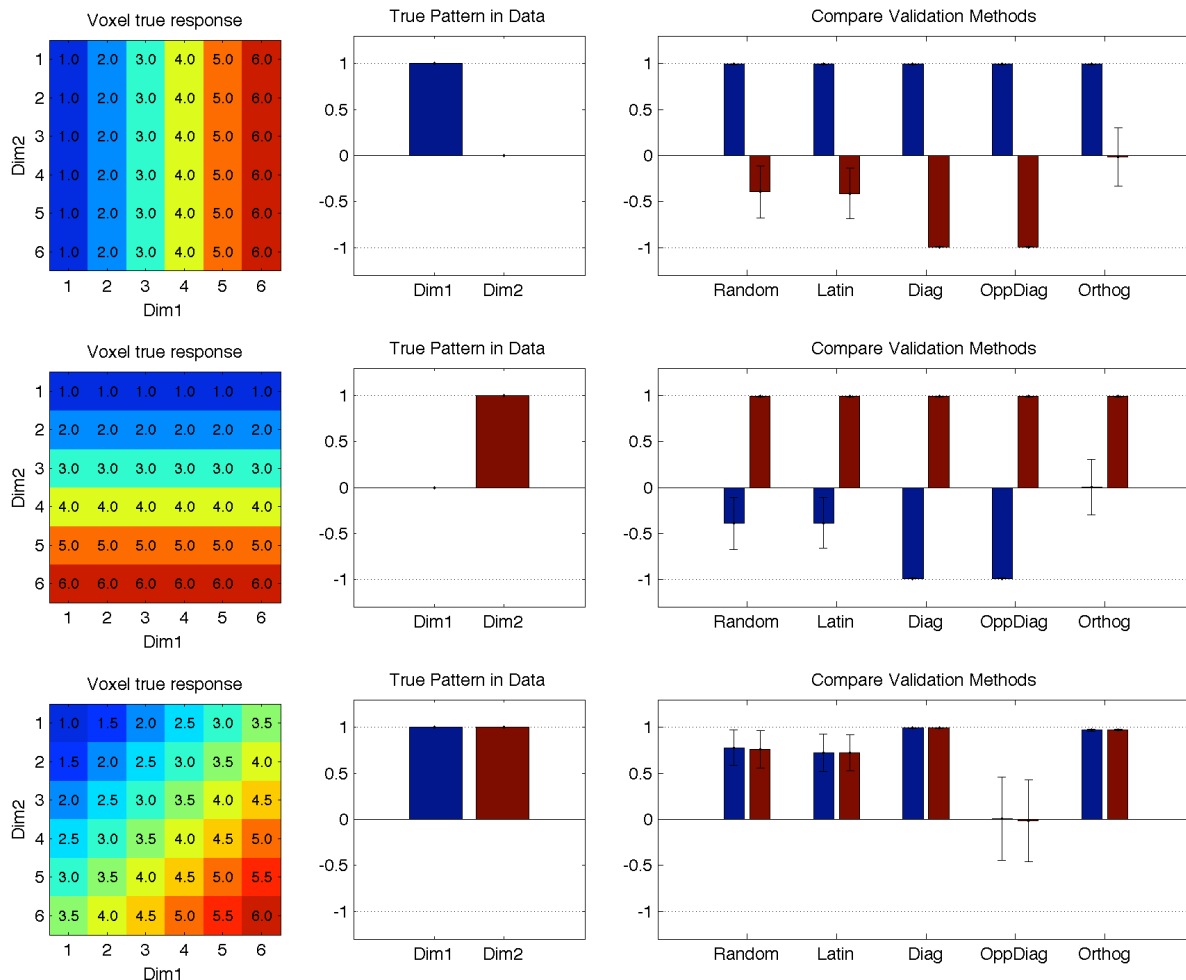
Figure S5. Summary of simulation results. Left Column: 3 example voxel data patterns. Middle Column: Voxel's true correlation with dimension1 or dimension2. Right Column: Mean correlations between the predicted and actual levels on the hold out data, averaged over 1000 iterations, for each of the hold-out validation procedures. Error bars reflect ± 1 standard deviation over iterations.