Vision Research 227 (2025) 108539

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Task-modulated neural responses in scene-selective regions of the human brain☆

Aysu Nur Koc [a,b,*], Burcu A. Urgen [b,c,e], Yasemin Afacan [b,d,e]

[a] Department of Psychology, Justus Liebig University Giessen, Giessen, Germany
[b] Interdisciplinary Neuroscience Program, Bilkent University, Ankara, Turkey
[c] Department of Psychology, Bilkent University, Ankara, Turkey
[d] Department of Interior Architecture and Environmental Design, Bilkent University, Ankara, Turkey
[e] Aysel Sabuncu Brain Research Center and National Magnetic Resonance Imaging Center, Bilkent University, Ankara, Turkey

ARTICLE INFO

Keywords:
Scene perception
PPA
RSC
OPA
MVPA
RSA

ABSTRACT

The study of scene perception is crucial to the understanding of how one interprets and interacts with their environment, and how the environment impacts various cognitive functions. The literature so far has mainly focused on the impact of low-level and categorical properties of scenes and how they are represented in the scene-selective regions in the brain, PPA, RSC, and OPA. However, higher-level scene perception and the impact of behavioral goals is a developing research area. Moreover, the selection of the stimuli has not been systematic and mainly focused on outdoor environments. In this fMRI experiment, we adopted multiple behavioral tasks, selected real-life indoor stimuli with a systematic categorization approach, and used various multivariate analysis techniques to explain the neural modulation of scene perception in the scene-selective regions of the human brain. Participants (N = 21) performed categorization and approach-avoidance tasks during fMRI scans while they were viewing scenes from built environment categories based on different affordances ((i)access and (ii)circulation elements, (iii)restrooms and (iv)eating/seating areas). ROI-based classification analysis revealed that the OPA was significantly successful in decoding scene category regardless of the task, and that the task condition affected category decoding performances of all the scene-selective regions. Model-based representational similarity analysis (RSA) revealed that the activity patterns in scene-selective regions are best explained by task. These results contribute to the literature by extending the task and stimulus content of scene perception research, and uncovering the impact of behavioral goals on the scene-selective regions of the brain.

1. Introduction

A scene is often defined as a view of an environment with a spatial structure one can act within (R. Epstein et al., 1999; R. Epstein, 2005; Henderson & Hollingworth, 1999). The human brain can process incredibly large amounts of data regarding the environment one is situated within in a matter of milliseconds, which is often attributed to various evolutionary mechanisms, such as identifying one's surroundings, finding shelter, food, and protection from threats (Kaplan, 1992). Providing the background and the context for virtually any cognitive process, the study of scene perception can be crucial for fully understanding the human brain. It can contribute to the study of many topics, such as attention, learning, memory, and social behavior (R. A. Epstein & Baker, 2019). Accordingly, this significance of scenes has inspired a large body of research investigating the behavioral and neural bases of scene perception and their implications.

Three main regions were discovered to be selective to scenes in the human brain compared to other categorical stimuli. The parahippocampal place area (PPA) is located in the posterior parahippocampal gyrus and is highly sensitive to scenes; it shows increased activation to stimuli with spatial layout, including photos, drawings, and even spaces built with Lego blocks (R. Epstein et al., 1999; R. Epstein & Kanwisher, 1998; Walther et al., 2011). PPA was demonstrated to be sensitive to spatial, categorical, and semantic properties of scenes that help us recognize scenes and distinguish between them (Persichetti & Dilks, 2018, 2019). In contrast, it was not emphasized in processes

☆ This article is part of a special issue entitled: 'Scene Processing' published in Vision Research.

* Corresponding authors at: Interdisciplinary Neuroscience Program, Bilkent University, Ankara, Turkey.
E-mail addresses: aysu.n.koc@uni-giessen.de (A.N. Koc), burcu.urgen@bilkent.edu.tr (B.A. Urgen), yasemine@bilkent.edu.tr (Y. Afacan).

https://doi.org/10.1016/j.visres.2024.108539
Received 13 December 2023; Received in revised form 29 October 2024; Accepted 20 December 2024
Available online 28 December 2024

related to navigation, planning, or memory regarding scenes.

The retrosplenial cortex (RSC) is located in the posterior cingulate region (Brodmann's 29 and 30). In addition to its scene-selectivity, it is emphasized in processes such as learning, navigation, and episodic memory regarding scenes (Vann et al., 2009). In contrast, it may not be as successful as the other regions regarding scene category sensitivity (Persichetti & Dilks, 2019). Instead, it is viewed as an association hub integrating a wide range of information regarding one's environment (Alexander et al., 2022), necessary for higher-level complex processes such as map-based navigation and route-planning (Dilks et al., 2022), supported by its sensitivity to heading direction, reference frames, boundaries, landmarks (Alexander et al., 2022; Auger et al., 2012; Dilks et al., 2011; Stacho & Manahan-Vaughan, 2022; Troiani et al., 2014), and role in reconstructing places by prediction or by retrieval from memory (R. A. Epstein, Higgins, et al., 2007; R. A. Epstein, Parker, et al., 2007).

Lastly, the occipital place area (OPA) is located near the transverse occipital sulcus, and its activation is modulated by egocentric distance, first-person perspective motion, direction, obstacles, and borders in a scene (Dilks et al., 2022; Henriksson et al., 2019; Kamps et al., 2016). Considering these findings, the OPA is thought to have a major role in perceiving and navigating the immediate scene one is situated within (called *visually guided navigation*) rather than complex navigation referring to memory or experience (Dilks et al., 2013; Silson et al., 2015). Its disruption can cause problems in scene categorization (Ganaden et al., 2013); however, this is thought to be related to the OPA's sensitivity to border and layout that could differentiate scenes and not to semantic processing like the PPA. In addition to its role in visually-guided navigation, the OPA is suggested to be an information source for higher processing of scenes in the PPA and RSC (Dilks et al., 2013).

These scene-selective regions and the underlying mechanisms of scene perception have been studied using various stimuli and tasks, contributing to our current knowledge. Scene perception research often employs categorical stimuli. Common categories of scene stimuli include various outdoor categories such as mountains, fields, and beaches (M. Greene & Oliva, 2009; Johnson & Johnson, 2014), indoors and outdoors (M. R. Greene & Hansen, 2020; Henderson et al., 2007; Rousselet et al., 2005), natural and manmade environments (Gopnarayan et al., 2022; Groen et al., 2013), environments that vary by their openness and closedness(Guo et al., 2012; Kravitz et al., 2011), or by other distinctions including properties such as scale, spatial frequency, and color (Kaping et al., 2007; Oliva & Schyns, 2000; Oliva & Torralba, 2001). These categorical stimuli are often presented to the participants with an accompanying task, such as categorization to measure their understanding of the stimuli, or tasks that aim to keep the participants focused on the screen, such as n-back paradigms.

Although past research uncovered so much about scene processing dynamics in the human brain, there are several gaps in knowledge, especially regarding higher-level scene perception. First, the field can benefit from the use of various behavioral tasks that are related to goals one can have regarding particular scenes and as they are limited and commonly to tasks that mainly aim to keep the participant focused although this has been changing in the recent years. The effect of task on brain activity patterns while participants look at the same stimuli has been shown for other category-sensitive brain regions (Harel et al., 2014) and given the growing interest in the role of tasks in visual neuroscience (Kay et al., 2023), scene perception study can benefit from in-depth examination of task demands and resulting changes in neural activity. Second, the approach to choosing stimulus categories has not been systematic, and the stimulus content has been limited. Categories of stimuli used in past research are primarily outdoor scenes, and categorical distinctions are decided based seemingly on common sense or statistical image features.

Further, built indoor spaces in which we spend 80 to 90 % of our daily lives (Cholowsky et al., 2023; Klepeis et al., 2001) that are made up of highly functional parts suitable for incorporating various behavioral goals are often overlooked, and can greatly contribute to the field. Moreover, the stimuli representing these categories are often professional high-quality images, previously referred to by Epstein et al. as "holiday snap-shot perception" (R. A. Epstein & Baker, 2019) or the complete opposite: images stripped off of their visual properties, reducing them to a singular aspect a particular study focuses on, neither of which reflecting real-life scenes we interact with (Groen et al., 2017). That real-life tasks and stimuli can contribute greatly to the scene literature for a more comprehensive and high-level study of scenes both behaviorally and neurally has been pointed out multiple times in the past (R. A. Epstein & Baker, 2019; Groen et al., 2017; Malcolm et al., 2016).

In this study, we aim to address these issues by examining how different behavioral goals modulate the neural processing of built environment categories in the scene-selective regions of the human brain. To this end, we employed two behavioral tasks in an fMRI experiment while participants viewed indoor scenes: one is a categorization task to measure the semantic processing of categories and monitor the participants' understanding of categories, and the other is an approach-avoidance task measuring initial actions taken by participants regarding a scene. We studied these tasks in less studied indoor environments, comprising separate functional units suitable for studying behavioral goals. To systematically decide on indoor categories, we adapted a categorization method from the architecture literature, defining built-environment categories based on their elements that afford distinct actions and serve distinct functions. Finally, we selected real-life, ordinary scene stimuli from a database, to keep them as close to real-life as possible in an MRI setting. We then used multivariate pattern analyses (MVPA) techniques, decoding and representational similarity analysis (RSA) to examine how tasks modulate the neural representation of scenes in the scene-selective regions of the human brain.

2. Methods

2.1. Participants

28 participants volunteered for this fMRI experiment. After data quality checks explained in section 2.6, the final analyses included 21 participants (11 males, ages 18–31, M = 23.7, SD = 3.25). All participants had normal or corrected-to-normal visual acuity, reported no neurological disorders, and did not use any related medication. The study was approved by the human research Ethics Committee of Bilkent University in line with the Declaration of Helsinki. Each participant filled out a prescreening form for MR safety and gave written consent before the experiment. After the experiment, they were briefed upon request and compensated for their time, with course credit when applicable and 100TL.

2.2. Stimuli and apparatus

Stimulus categories were chosen based on the Universal Design: Users-Built Environments Model by Froyen (2012), which evaluates built environments from various aspects and provides an in-depth classification. The model dissects environments based on the impairments and activities of users, the physical aspects such as lighting and ergonomics, and the physical concrete elements. The *elements* section of the model separates environments into physical parts: approach (the surrounding), access (entrances), horizontal circulation (e.g., corridors), vertical circulation (e.g., stairs), social interaction (e.g., canteens), rest (waiting and seating areas), food and drinks (food courts), and sanitary facilities (restrooms). Considering the objectives of this study, we selected only the elements section of this categorization, as these were less subjective, and they categorized environments based on more concrete and visually distinct elements. However, since some of these categories had shared features, we grouped them and created a simplified,

easier-to-understand categorization approach. The final set of experimental stimuli consisted of 32 different images chosen from the Scene Understanding (SUN) database (Xiao et al., 2010, 2016). There were 8 images of access elements to buildings (4 entrances, 4 exit points), 8 images of circulation elements (2 stairs, 2 escalators, 2 corridors, and 2 elevators), 8 images of sanitary facilities (4 bathroom stalls, 4 sink areas), and 8 images of seating areas (4 eating areas, 4 seating areas). Simplified categories and the corresponding stimuli used in the experiment can be seen in Fig. 1.

The experiment and the stimuli were programmed to be displayed in a mirrored fashion to be reflected on an MRI-compatible screen (detailed in section 2.6), and each image had a 600x450 pixel resolution. The visual properties of images were not manipulated to keep the stimuli more realistic and ordinary, but these properties were addressed in the analyses.

2.3. Behavioral tasks

Participants performed 2 different tasks during the experiment. One of these was a categorization task, where participants responded by indicating the main category of the presented stimulus (architectural or functional). The categories required considering different actions a scene affords, along with understanding its semantic associations, so this task aimed to engage participants with various high-level processes needed to comprehend a scene wholly.

The other was an approach-avoidance task, where they simply indicated whether they would like to enter the presented environment. We address this task as an "approach-avoidance task"; however, it must be emphasized that it does not precisely represent the nature of such tasks found in the literature. Approach-avoidance tasks often involve distinctly separated stimuli based on a positive–negative dynamic related to valence (Korn & Elliot, 2015). Therefore, it is more sensible to consider this task a subjective "enter-or-not" decision since the stimuli were not chosen based on valence. Instead, we aim to observe how an immediate decision about acting in an environment is made, that is, to enter or not, based on initial perception. Previous research in neuroarchitecture literature shows that approach-avoidance decisions are not entirely explained by beauty or otherwise valence-based responses and that these tasks can measure initial sensorimotor processing regarding an environment (Coburn et al., 2020). Participants were also instructed with this in mind: they were asked to respond quickly based on their initial opinions. They were not expected to be consistent across runs for the same stimuli. Since there were no correct answers, this task was not analyzed or interpreted like a traditional approach-avoidance task and is better considered unique to this study.

2.4. Experimental design

The experiment consisted of 8 runs (~9 min each), containing 2 behavioral task blocks (categorization and approach-avoidance) with their order counterbalanced across runs (ABBABAAB). With one trial per stimulus, each task consisted of 32 trials. In each run and task, participants viewed the same set of 32 stimuli in randomized order. Trials began with a 2000 ms presentation of the stimulus, followed by a 2000 ms response period, and ended with a randomized variable interstimulus interval (ISI) that ranged between 3000–4000 ms, averaging at 3500 ms. Each task block started with a 10-second instruction screen indicating the task and the corresponding buttons and ended with a 10-second rest. A black fixation cross was presented at the center of the screen throughout the experiment, which only turned white during the response period. The procedure of a single run is shown in Fig. 2.

The categories were explained to the participants in detail, and they also performed a shorter version of these tasks on a laptop as a practice task before going into the MRI scanner using different images selected based on the same criteria as the experiment. Architectural elements comprising access points and circulation elements were explained to the participants as parts of buildings that allow us to access and circulate within an environment, and functional elements comprising restrooms and eating/seating areas were explained as those that serve basic human needs and comprise relatively stationary activities. Participants were instructed to focus on the fixation throughout the experiment. The responses were collected using a fiber optic response box.

2.5. Functional localizer

In addition to the experimental session, a separate functional

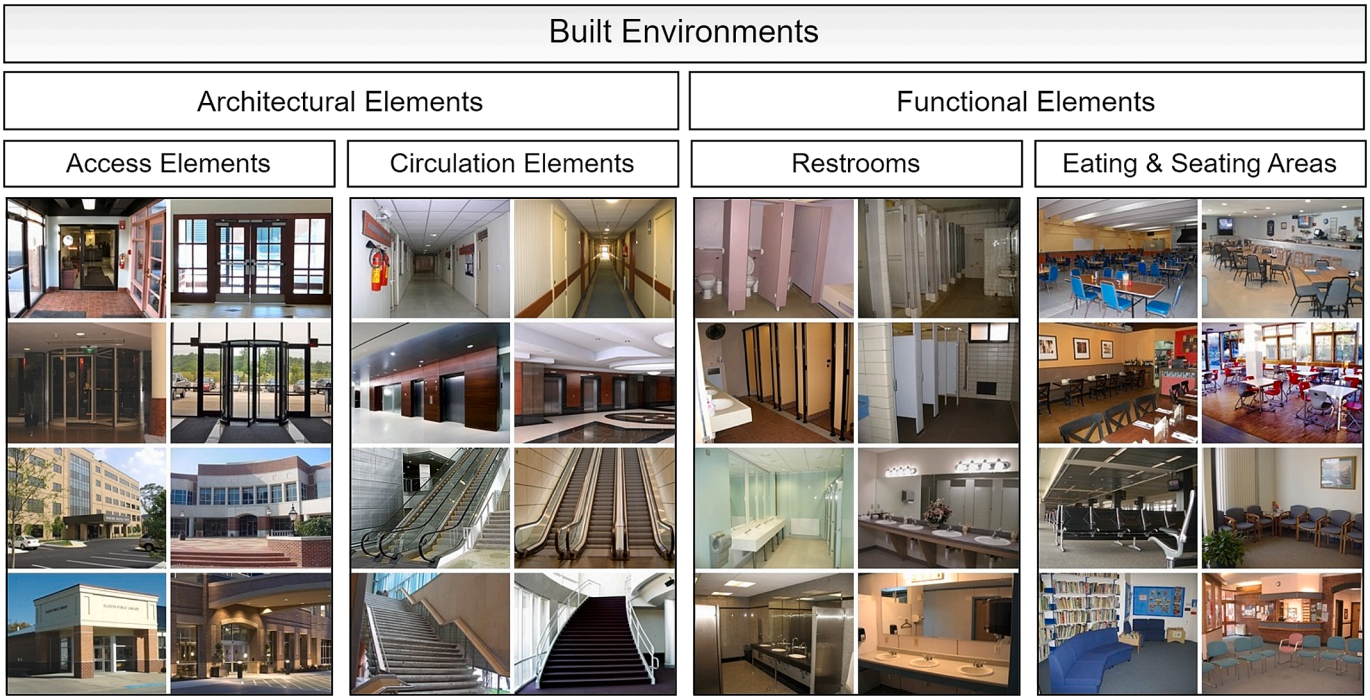

Fig. 1. Simplified stimulus categories and the corresponding stimuli.

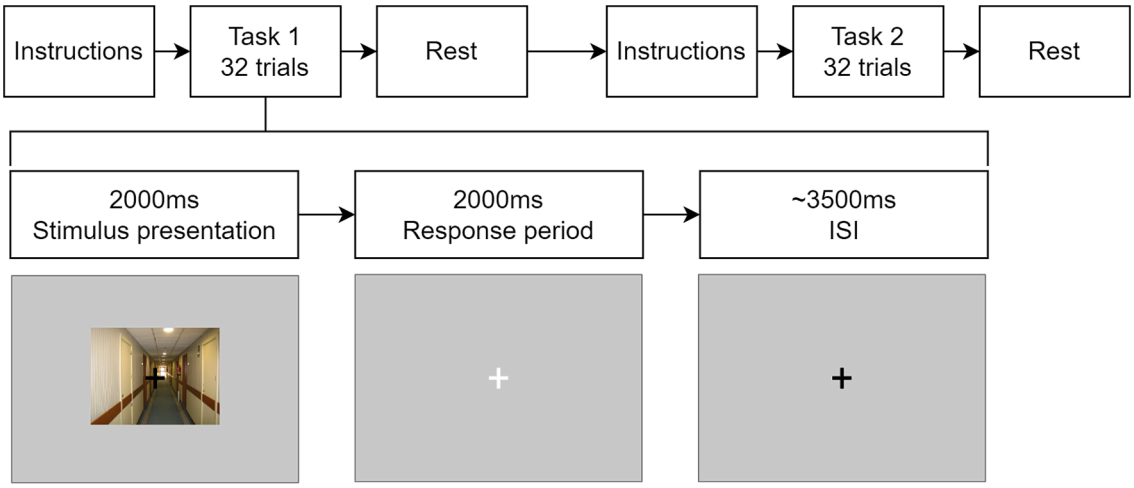

Fig. 2. Experimental procedure of a run. Each run consists of two task blocks, each including 32 trials with 2000 ms stimulus presentation, 2000 ms response period, and ~ 3500 ms ISI.

localizer session was conducted to delineate scene-selective regions of interest (ROIs) across our participants. This localizer was adapted from the code made available by Meissner et al. (2019) and was shown to activate scene-selective regions in multiple age groups successfully. This session consisted of 4 short runs (~3.5 min) of scene, object, and fixation blocks, in a pseudo-randomized order in each run, where related categorical images were presented rapidly in separate blocks while participants performed a one-back repetition task. This allowed us to have multiple blocks of neural activity to use for a scene > object contrast and delineate the scene-selective regions in our participant at the group level.

2.6. FMRI data acquisition and preprocessing

MRI data was collected using a 3 T Siemens Trio scanner (Magnetom Trio, Siemens AG, Erlangen, Germany) with a 32-channel head coil, at the National Magnetic Resonance Research Center (UMRAM), İhsan Doğramacı Bilkent University. Participants viewed visual stimuli presented on an MRI-safe LCD screen (1920x1080px, 125x70cm, vertical refresh rate = 60 Hz, TROYKA MED, İstanbul) through a mirror mounted on the head coil with a total view distance of ~ 168 cm. The sessions started with a short localizer scan to check if the head was positioned correctly, followed by a T1-weighted structural scan. The main experimental session included 8 functional runs, each lasting about 9 min. The high resolution T1-weighted structural images were obtained with a standard protocol (TR = 2600 ms, TE = 2.92 ms, flip angle = 12°, FoV read = 256 mm, FoV phase = 87.5 %, 176 slices, voxel size = 1x1x1mm3). During the functional runs, 263 functional volumes were obtained using gradient-echo planar imaging (TR = 2000 ms, TE = 22 ms, flip angle = 90°, 64x64 matrix, FoV read = 192 mm, 43 slices with a thickness of 2.5 mm, voxel size = 3x3x2.5 mm3). For the localizer session for ROI definition, the fMRI data acquisition parameters were the same as the main experiment, only this time, we obtained 4 functional runs (~3.5 min/104 volumes).

Anatomical and functional fMRI data was converted to BIDS format (Gorgolewski et al., 2016) and later preprocessed using fMRIPrep 21.0.1 (Esteban, Blair, et al., 2018; Esteban, Markiewicz, et al., 2018) using the default steps. Structural (T1w) images were corrected for intensity non-uniformity and skull stripped. Brain tissue segmentation was performed, and brain surfaces were reconstructed. Volume-based spatial normalization to the ICBM 152 Nonlinear Asymmetrical template (2009c) was performed. For the functional images (both the main experiment and the functional localizer), a reference volume and its skull-stripped version were created, and head-motion parameters were calculated based on the reference. Slice-time correction was applied to BOLD runs and then resampled onto the native space after head-motion estimation transforms were applied. The BOLD reference was co-registered to the T1w reference with six degrees of freedom. Confounding time-series were calculated for framewise displacement, demeaned variance (DVARS), global signals, and physiological regressors to be used for component-based noise correction. High-pass filtering was applied to the pre-processed BOLD time-series with a 128 s cut-off. BOLD time-series were also resampled into the same standard space as the structural scans. Detailed preprocessing documentation outputted by the software itself can be found in the supplementary material.

Based on the head-motion output of this procedure, runs with spikes of motion of more than 1.5 mm and total relative motion of more than 3 mm were defined and removed from the analysis. Participants with less than 5 suitable runs remaining and those who did not attend the localizer session were completely omitted from the analyses. Therefore, the final analyses were performed on 21 participants in total, with complete data (8 runs) from 17 participants, 7 runs from 1, 6 runs from 2, and 5 runs from 1 participant. This sample size is smaller than we aimed for after losing 7 participants, but still comparable to contemporary scene perception and fMRI research (e.g. Peer & Epstein, 2021; Szucs & Ioannidis, 2020).

2.7. Data analyses

2.7.1. Behavioral analyses

The accuracy of responses was calculated for the categorization task. Since there were no correct responses for the approach-avoidance task, only the distribution of responses was examined. Reaction times were compared across tasks using a *t*-test, and across categories using a repeated measures ANOVA.

2.7.2. Univariate analyses and internal consistency

Univariate analyses were performed using the Statistical Parametric Mapping software package (SPM12, Wellcome Trust Centre for Neuroimaging, University College London, UK) implemented under MATLAB (The Mathworks Inc., Natick, MA). For the following classification/decoding analyses, the general linear model (GLM) was calculated using a design matrix consisting of categories separately coded under two tasks, fixation, ISI, and rest periods in the experimental procedure and 6 motion regressors. These regressors were then convolved with the default canonical hemodynamic response function of the software. Contrast images were created by grouping trials for each category under each task (4 categories during the categorization task and 4 during the

approach-avoidance task) versus fixation periods. Another set of univariate analyses were conducted for the RSA, using the same protocol, except this time by coding each stimulus under each task as separate regressors. Then, 64 contrast images, grouping trials of each stimulus under each task separately, were created. For both of these multivariate analyses, GLMs were calculated using unsmoothed images.

In order to confirm the internal consistency of our data, we have conducted a split-half analysis. We conducted the same first-level analysis as the data prepared for decoding analysis, but this time we have conducted it separately for odd vs even runs of each participant, as it is common practice (Luking et al., 2017). For those participants with excluded runs, the new run order was considered. We used the ROI masks that we created (section 2.7.3 & 3.3) to extract beta values at the group level. Pearson correlation coefficients (r) were calculated for the two task vs fixation contrasts, as well as the Spearman-Brown (SB) coefficients ($2r/(1 + r)$).

2.7.3. Functional localizer analysis and definition of regions of interest (ROIs)

Functional localizer scans were analyzed similarly using SPM12. However, this time preprocessed functional scans were smoothed with a 6 mm full width at half maximum (FWHM) kernel. The GLM was calculated using a design matrix consisting of 11 regressors, including scene, object, and fixation blocks, as well as the rest of the intervals in the experimental procedure and 6 motion regressors. Scenes > objects contrast images were created for each participant to define scene-selective regions. One sample *t*-test was performed at the group level using these contrast images with default SPM parameters.

To decide on the maximum extent of our ROIs, we used the scene parcels from a localizer study by Julian et al. focusing on the definition of category-sensitive brain regions with localizer experiments on a large participant group (Julian et al., 2012). Group-level activation maps were overlaid by these parcels, and by adjusting the threshold, the surviving 50-voxel contiguous clusters were selected in each hemisphere at the sites indicated for the PPA, the RSC, and the OPA. We chose this approach to avoid very large or very small (sometimes non-existent) ROIs and to keep them around same size. However, there exist many approaches to ROI selection, as we further discuss in detail in section 4.3. For comparing two of the common approaches, a separate ROI selection procedure was performed- this time using a fixed threshold and selecting all the remaining voxels without a limit (see Supplementary Material). Group-level ROIs were defined using the MarsBar toolbox (Brett et al., 2010).

2.7.4. ROI-based multivariate pattern analysis (MVPA)

An ROI-based classification analysis was performed to observe classification performance for tasks and categories in these ROIs. First, we performed GLM analyses on the unsmoothed experimental functional scans using SPM12.

For the classification analysis, we used The Decoding Toolbox (TDT − version 3.999F) (Hebart et al., 2015), implementing a linear support vector machine (LibSVM) (Chang & Lin, 2011) with a fixed cost parameter $c = 1$. Classifiers were trained with a leave-one-run-out procedure using runwise beta images of participants and repeated multiple times to test for each run while training with the rest ($k = 8$ for 17, $k = 7$ for 1, $k = 6$ for 2, and $k = 5$ for 1 participant), separately for each region. This procedure was performed 4 times per ROI, one decoding for categories (labels: 1,2,3,4,1,2,3,4), decoding for the task (1,1,1,1,2,2,2,2), for categories during the categorization task (1,2,3,4,0,0,0,0), and lastly for categories during the approach-avoidance task (0,0,0,0,1,2,3,4). For each classification, accuracy values were computed for participants individually.

Additionally, a permutation procedure was applied using the same toolbox. Permutation analyses perform the same steps in calculating classification performance a specified number of times, shuffling the data and labels randomly at each repetition. This procedure creates an experimental baseline for comparing actual decoding results to, by calculating a distribution of responses that could be expected by chance (null hypothesis) instead of using a theoretical chance level (Valente et al., 2021). Then, the resulting permuted means are compared to the observed values, and compared to a minimum p-value calculated based on the number of permutations ($p_{min} = 1/n + 1$) and a significance cut-off ($p_{max} = 0.05$). After initial decoding output was computed, same data and label configurations were used input subject by subject to create 1000 permutations for each condition separately and the results were used to compute group means (chance levels). The number of repetitions was selected based on both past research and computational constraints (Edgington, 1969; Nichols & Holmes, 2002). Later, built-in functions that use Fisher's exact test were applied for statistical analyses for category, task, and task dependent category decoding performances. Finally, one-way Analyses of Variance (ANOVA) was performed separately to analyze differences in category and task decoding performances across ROIs, and a 2(task)x3(ROI) repeated measures ANOVA was performed for the categorization performance of ROIs across tasks. In cases of significant main effects, post-hoc tests were conducted and GG (Greenhouse-Geisser) and Bonferroni corrected p-values are reported where appropriate.

2.7.5. Representational similarity analysis (RSA)

To examine the changes in activation patterns for each condition across ROIs compared to tasks, categories, and visual image properties, we performed a model-based representational similarity analysis (RSA) using the RSA Toolbox (Nili et al., 2014). RSA allows us to explore how information is represented in the brain by comparing patterns of neural activity to different measurements or conditions, such as stimuli, tasks, and hypothetical models. By calculating representational dissimilarity matrices (RDMs) of neural activity in different brain regions and comparing them to matrices of other measurements or models, it helps discern the underlying structure of neural activities (Kriegeskorte, 2008).

First, we created 4 different 64x64 candidate models that could potentially explain the changes in activation patterns in the form of RDMs (Fig. 3). The first was a task model, indicating that the dissimilarity values would be low for all conditions during the same task and high during the other. The second was a category model with low dissimilarity values between items of the same category and higher dissimilarity between items of other categories independent of the task. The third was a visual model created using Gabor filters on the stimuli to account for the visual differences among the stimuli, as scene processing is affected and sometimes explained by low-level properties (Groen et al., 2017; Kauffmann et al., 2014). A Gabor bank was defined with 3 wavelengths (4,8,16 pixels/cycle) and 4 orientations (0,45,90,135°). Gabor filters of each combination were applied to the stimuli, and the resulting magnitude values per pixel were used to calculate filter power (by squaring the values). These matrices corresponding to each pixel in each stimulus were then converted into arrays and combined for each of the 32 stimuli. Later, the obtained 32 arrays were compared to one another (corr2 in MATLAB). The obtained similarity values were subtracted from 1 to calculate dissimilarity values and placed in a 32x32 matrix. This matrix was repeated 2x2 to create a 64x64 model. The fourth and final model was a random one for control.

Thresholded contrast images (spmT) created for each stimulus (32) under each task (2) versus fixation periods and the 3 ROI masks were used as the input for the RSA Toolbox. Averaged 64x64 RDMs were computed across 21 subjects for each ROI. Then, via Kendall's τ, a one-sided Wilcoxon signed-rank random-effects test was performed between the neural RDMs and the candidate models explained above. By estimating a hypothetical model with maximum average correlation to the reference RDM via Kendall's τ, the upper and lower bounds for the noise ceiling were calculated, indicating the range a model must reach to be considered a *true model* fully explaining the variance in the data. Lastly, by calculating Spearman's ρ, pairwise correlations of all the neural and

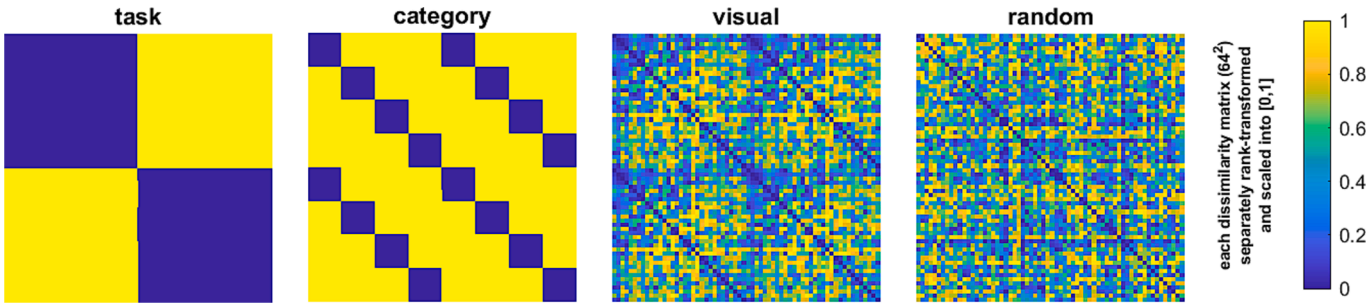

Fig. 3. Candidate model RDMs created for the RSA.

model RDMs were compared. Both for ROI to candidate model comparisons and overall RDM comparisons, multidimensional scaling (MDS) plots were also generated. Matrix values of both model and neural RDMs were rank-transformed into [0,1] and color-coded accordingly.

3. Results

3.1. Behavioral results

Participants responded accurately 97.013 % (SD = 3.846) of the time during the categorization task, indicating the categorization method was well understood. The reaction times were significantly different between the tasks $t(20) = -2.762$, $p = 0.012$, with a small effect size (cohen's d = −0.227).

A 2x3 repeated measures ANOVA between 4 categories revealed a significant effect ($F(3,60) = 10.343$, $p_{GG} < 0.000$) and was followed with post-hoc comparisons using the Bonferroni correction. The following paired t-tests revealed that there were significant differences in reaction time between the restrooms category and the others. Restrooms had significantly shorter reaction times compared to access points ($t(20) = -3.538$, $p_{bonf} = 0.012$), circulation elements ($t(20) = -4.846$, $p_{bonf} < 0.001$), and eating/seating areas ($t(20) = -4.017$, $p_{bonf} = 0.004$). There were no significant differences in reaction time between access and circulation categories ($t(20) = -2.189$, $p_{bonf} = 0.244$), access and eating/seating categories ($t(20) = -0.423$, $p_{bonf} = 1$), or circulation and eating/seating categories ($t(20) = 1.752$, $p_{bonf} = 0.571$).

The distribution of answers for the approach-avoidance task by stimulus and participant can be seen in the supplementary material (Supplementary Figs. 3 & 4).

3.2. Internal consistency results

Split-half reliability was calculated using odd vs even runs to determine internal consistency. Beta values extracted from each ROI in each hemisphere at the group level indicate good level of internal consistency, with generally high correlations and good SB coefficient values (SB > 0.8), with slightly lower but acceptable results for left RSC (SB < 0.7) (see Table 1).

Table 1
Split-half reliability results. Results are significant at $p < 0.000$.

ROI	Categorization Pearson's r	SB	Approach-Avoidance Pearson's r	SB
R PPA	0.955	0.977	0.982	0.991
L PPA	0.981	0.990	0.979	0.989
R RSC	0.810	0.895	0.914	0.955
L RSC	0.613	0.762	0.749	0.856
R OPA	0.963	0.981	0.954	0.977
L OPA	0.961	0.980	0.983	0.992

3.3. Functional localizer results and ROI definition

Statistical information regarding the ROIs is listed in Table 2. Peak locations were comparable to past literature (e.g. Marchette et al., 2015; Ramanoël et al., 2019). We obtained three 100-voxel ROIs by adjusting the threshold to obtain 50-voxel clusters in each hemisphere for each region, and combining these clusters in bilateral masks (Fig. 4).

3.4. ROI-based MVPA results (decoding)

For each region, ROI-based classification and permutation analyses were performed. Observed classification performances were then compared to permutated mean chance levels using Fisher's exact test. These analyses revealed several significant effects (Fig. 5).

For overall categorization performance, OPA showed a significant result ($p = 0.002$, chance = 25.051) whereas PPA ($p = 0.134$, chance = 25.074) and RSC ($p = 0.476$, chance = 24.984) did not show above-chance decoding performances. When examined for task-dependent categorization performance, there were also distinct results between ROIs. During the categorization task, both PPA ($p = 0.022$, chance = 25.097) and OPA (0.026, chance = 25.023) were significantly above chance for decoding categories, while RSC was not statistically significant ($p = 0.334$, chance = 25.017). During the approach-avoidance task, however, only RSC activity was successful at decoding categories ($p = 0.049$, chance = 25.008), while PPA ($p = 0.188$, chance = 25.026) and OPA ($p = 0.054$, chance = 25.042) were not different from chance. Finally, for task decoding performance, RSC was significantly above chance level ($p = 0.002$, chance = 24.936), whereas PPA ($p = 0.184$, chance = 24.998) and OPA ($p = 0.182$, chance = 24.981) were not.

For category decoding performance, one-way ANOVA indicated significant results ($F(2,80) = 4.69$, $p = 0.013$), and thus was followed by a multiple comparisons test. Paired t-tests were applied between ROIs for category decoding, indicating a significant effect, $t(20) = -2.508$, $p_{bonf} = 0.014$. There were no significant effects between the PPA and the RSC($t(20) = 0.939$, $p_{bonf} = 1$), nor between the PPA and the OPA ($t(20) = 0.127$, $p_{bonf} = 0.091$). Results show that the OPA (M = 29.677, SD = 4.796) had significantly higher decoding performance compared to the RSC (M = 25.819, SD = 3.743). The repeated measures ANOVA on task-dependent category decoding performances across ROIs did not reveal any significant effects for neither task ($F(1,20) < 0.000$, $p_{GG} = 0.98$), ROI ($F(2,40) = 0.133$, $p_{GG} = 0.872$) or their interaction ($F(2,40) = 0.632$, $p_{GG} = 0.534$). Similarly, the one-way ANOVA performed on task decoding across ROIs did not reveal a significant effect ($F(2,60) = 0.520$, $p = 0.598$).

3.5. RSA results

We performed a model-based RSA defining responses to each stimulus under each task as separate conditions to examine the activity patterns in scene-selective regions compared to various candidate models. The resulting 64x64 neural RDMs and the MDS plots (metric stress minimized) can be seen in Fig. 6.

Table 2
Coordinates and peaks of the ROI clusters, ~50 voxels.

ROI	Right Hemisphere Cluster Threshold	Peak Threshold	Peak Coordinates	Left Hemisphere Cluster Threshold	Peak Threshold	Peak Coordinates
PPA	T = 9.2	T = 19.71	26—46 −6	T = 9	T = 15.42	−24—48 −6
					T = 9.41	−22—36 −14
RSC	T = 8.5	T = 14.11	20—54 14	T = 6.5	T = 11.27	−18—58 9
					T = 8.43	−16—48 4
					T = 5.43	−6—42 2
OPA	T = 6.06	T = 8.25	32—78 19	T = 3.95	T = 6.57	−30—84 22
		T = 8.07	32—84 26		T = 5.05	−34—82 32
		T = 6.72	38—78 26			

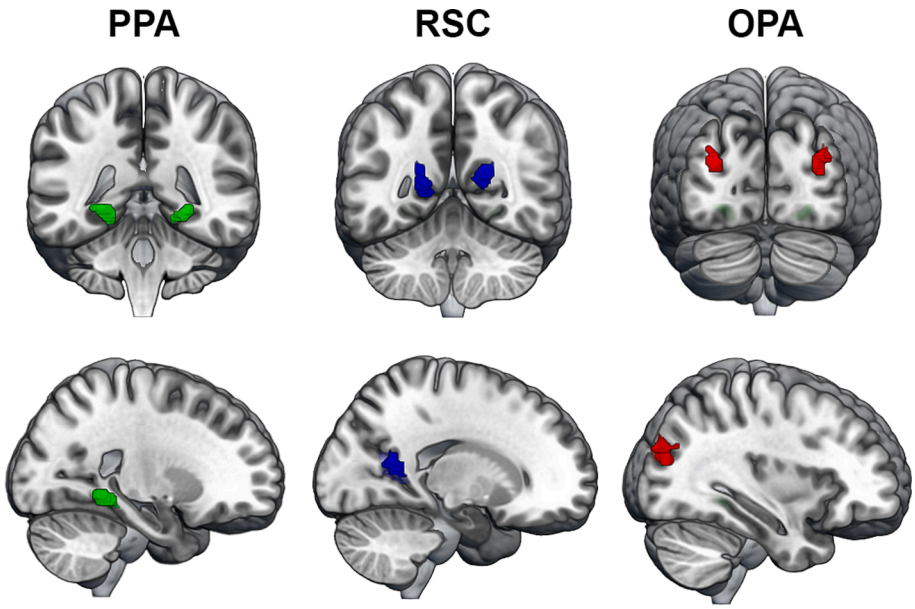

Fig. 4. 100-voxel bilateral ROI masks, obtained by the group-level results of the localizer session.

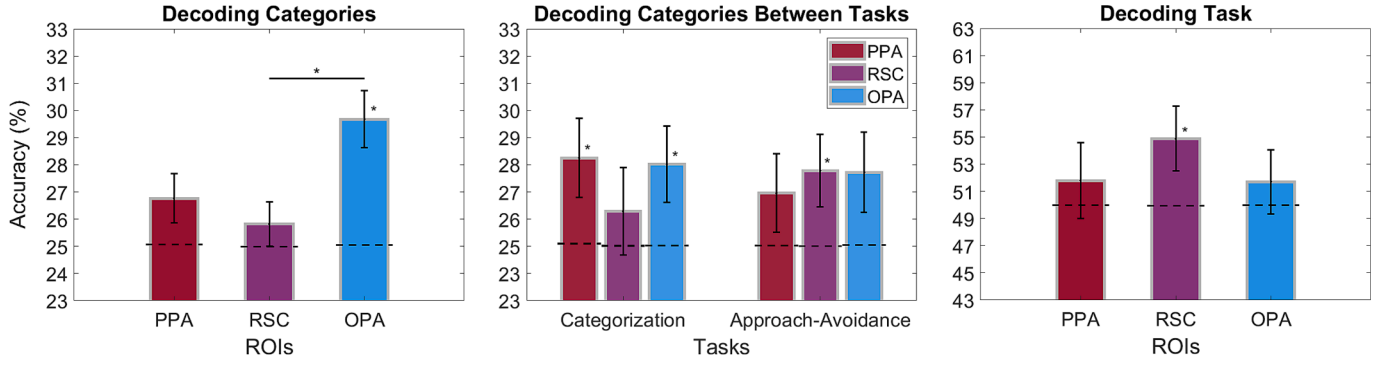

Fig. 5. MVPA decoding results by ROI. Dashed lines indicate the chance levels, based on permutation test means for each comparison. Significant decoding performances are indicated by asterisks right above the corresponding bar. Significant differences between ROIs are at $p < 0.05$, Bonferroni corrected. Error bars represent ± 1 standard error.

Precisely, Kendall's τ values for the model comparisons were as follows. For PPA: task ($\tau = 0.184$, $p < 0.000$), category ($\tau = -0.002$, $p = 0.939$), visual ($\tau = 0.001$, $p = 0.305$), random ($\tau = 0.008$, $p = 0.144$). The task model was significantly similar to the PPA but did not reach within the bounds of the noise ceiling (0.250, 0.314). The findings were similar for the RSC: task ($\tau = 0.219$, $p < 0.000$), category ($\tau = -0.000$, $p = 0.659$), visual ($\tau = 0.005$, $p = 0.102$), and random ($\tau = 0.002$, $p = 0.341$). The task model was also significantly similar to the RSC but did not reach the noise ceiling (0.274,0.332). Finally, the same trend was observed for the OPA: task ($\tau = 0.166$, $p < 0.000$), category ($\tau = 0.001$, $p = 0.293$), visual ($\tau = -0.018$, $p = 0.997$), random ($\tau = -0.001$, $p = 0.541$). The task model was again significant, still not reaching the noise ceiling (0.221, 0.289) (see Fig. 7).

4. Discussion

This study used classification and RSA techniques to examine the effect of built environment categories and behavioral tasks on scene-selective regions in the human brain. The findings indicate that indoor categories were successfully decoded in the OPA in general. In contrast,

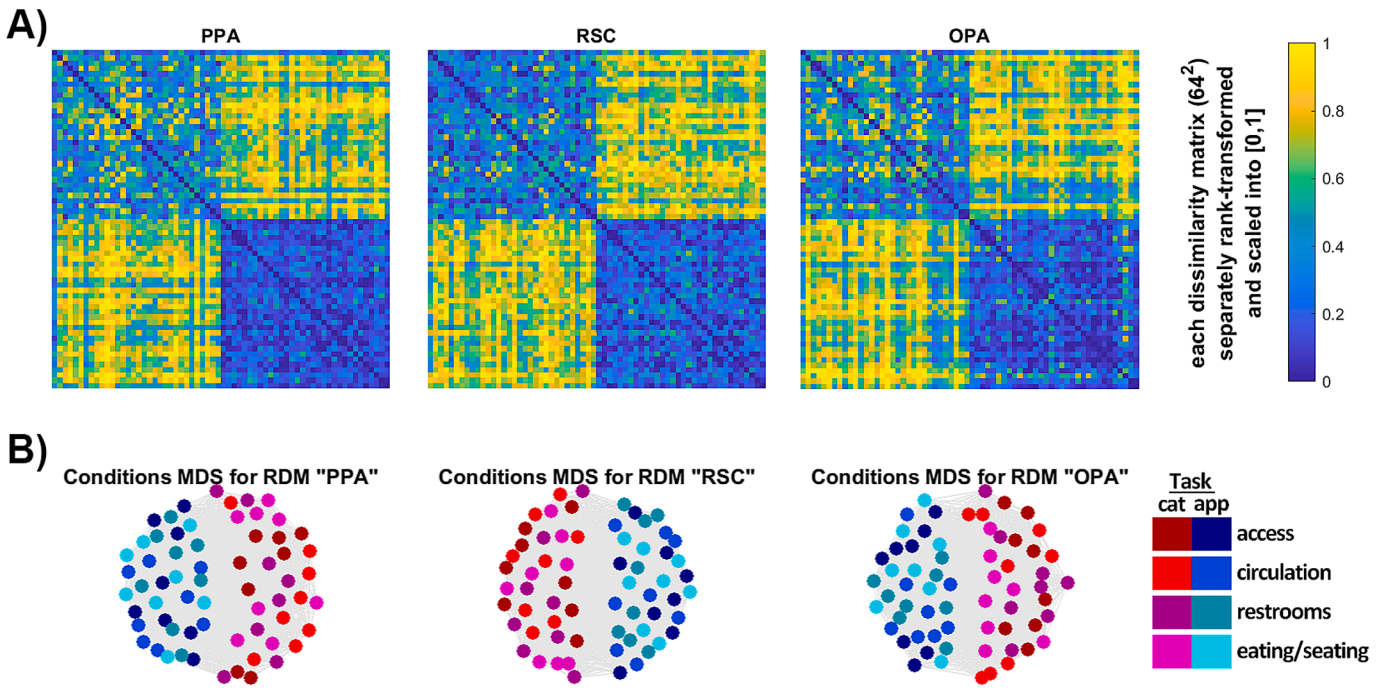

Fig. 6. RSA results. (A) Neural RDMs for each ROI. Stimuli are represented by individual cells in the matrix, ordered by categories and tasks. (B) Conditions MDS plots for each ROI. The correlational distances and dissimilarities between color-coded stimuli for each task are visualized.

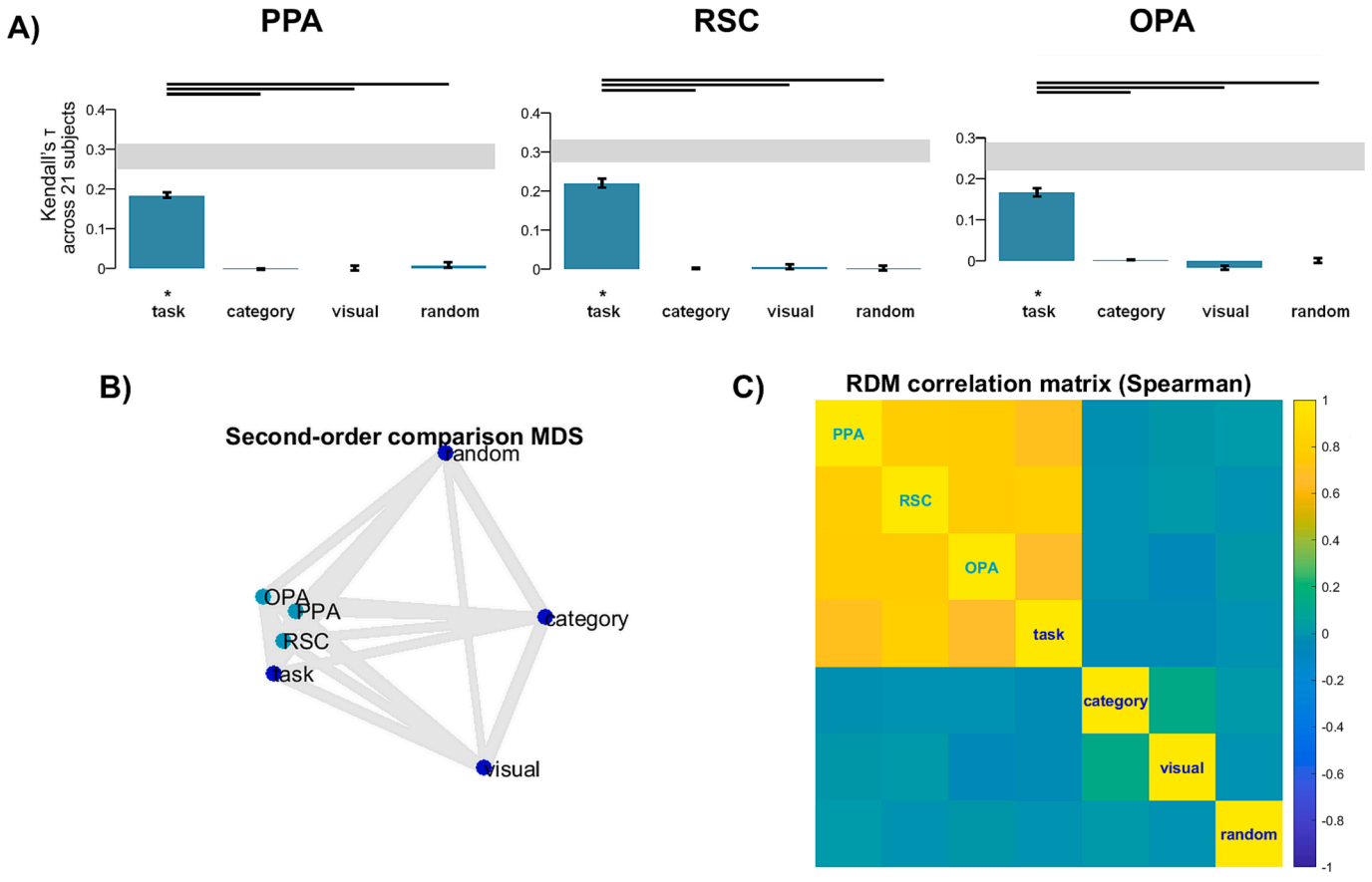

Fig. 7. RDM-candidate model comparison results. (A) Kendall's τ comparisons of ROIs to candidate models. The noise ceiling for a hypothetical true model is indicated with the gray bands over the graphs. The significant results are indicated with an asterisk at $p < 0.01$. Error bars represent the standard errors of average RDM correlations. (B) Second-order comparison MDS, indicating the spatial distances between neural and candidate RDMS. (C) Spearman correlations of all the RDMS.

the PPA and the OPA were significantly successful when we only looked at their performances during the categorization task, and the RSC was significantly successful only during the approach-avoidance task. For decoding the task, only the RSC was significantly successful. When we looked at the activation patterns to the categorical stimuli under each task and compared them to candidate task, category, visual, and random models, we observed that the task model was significantly similar to the actual neural activation patterns in all the ROIs. In contrast, other models were not at all significant. Despite being strongly correlated to the neural RDMs for all ROIs, the task model did not reach the noise ceiling, indicating that there are other factors a potential true model would incorporate that our models did not account for.

4.1. Built-environment categories are decoded best in the OPA, and the task affects decoding performance in all scene-selective regions

When we looked at the overall categorization performance of the ROIs, we see that although above chance in all the ROIs, it was only significantly high in the OPA. Also, the OPA was significantly more successful than RSC regarding category decoding performance. Although one would expect the PPA to have a significant category decoding performance based on the previous literature as it is pronounced in processing semantic information and identifying our surroundings (Dilks et al., 2022; Persichetti & Dilks, 2019), it was not observed in this analysis when we used data from both tasks in the classification process. This could result from a task effect or ROI selection, which will be discussed later. However, despite not being considered the most prominent in category discrimination, the overwhelming success of the OPA activity in decoding built environment categories can be explained by its proposed role in a particular type of navigation and how it may have been reflected in the criteria of our categorization approach, rather than it processing categorical information. Our categories of access points, circulation elements, restrooms, and eating/seating areas are essentially defined by their functions as part of a building, and they differ significantly by affordances. The access points include wide entrance areas, including the surroundings of buildings and doors one can walk through, and circulation elements consist of halls and stairs that primarily serve movement within a building. On the other hand, restrooms and eating/seating areas are less about movement and navigation, and these areas are less open and contain more borders, corners, and obstacles. The OPA may have been better at classifying these categories due to its involvement in *visually-guided navigation*, which concerns the immediate space visible to the subject, based on detecting paths and avoiding any obstacles by its sensitivity to borders, obstacles, and the openness of a scene (Dilks et al., 2022; Patai & Spiers, 2017; Persichetti & Dilks, 2018), and also its processing of navigational affordances even when they are not task-relevant (Bonner & Epstein, 2017, 2018).

Interestingly, when we analyzed the two tasks separately, the category decoding performances of these ROIs changed. Both the PPA and the OPA were significantly above the chance level during the categorization task when the participants actively considered the categorical properties of the stimuli. As explained before, the PPA is expected to have good decoding performance for categories; however, in this case, it was only apparent when the task was related, indicating an effect of behavioral goals. On the other hand, only the RSC was significantly above the chance level for category decoding during the approach-avoidance task, as well as for task decoding. These could be related to the nature of the task, possibly engaging multiple processes such as considering oneself in relation to the scene (egocentric reference frames) and spatial memory, including retrieval of previously learned scenes from memory (Alexander et al., 2022; Miller et al., 2014; Mitchell et al., 2018; Vann et al., 2009) as well as RSC's multiple roles in cognition as described earlier. One explanation concerning spatial memory could be the repeated exposure of participants to the same stimuli during the experiment. Indeed, although participants were instructed to give approach-avoidance responses based on their initial opinions and not worry about being consistent throughout, some participants later indicated trying to remember their previous responses in subsequent runs. Another explanation may come from a possible role of the RSC in approach-avoidance processes, which, to the best of our knowledge, is unexplored in humans but is supported by animal studies on approach-avoidance learning (Todd et al., 2019). More research is necessary to clarify the RSC's role in related processes.

Overall, the decoding results indicate that the category decoding performances of scene-selective regions are modulated by the behavioral goals of the observers.

4.2. Activation patterns in scene-selective regions are best explained by the task model

The model-based RSA results indicated the Task model best explained the activation patterns in scene-selective regions. Still, the similarity did not reach the noise ceiling, which a true model that explains the variation perfectly would. This means that, although the task model was significantly similar to neural activation patterns, it does not perfectly explain the data. We suggest that a combination of factors may better explain activation patterns in these regions than discrete models for task, category, and visual properties of the stimuli since brain regions do not necessarily process single, separate properties. Indeed, according to the literature, despite each being more pronounced in different functions, all three scene-selective regions are affected by various low- and high-level properties (Groen et al., 2017). Also, the functional differences among scene-selective regions that are demonstrated in the past literature, such as variations in category and visual property processing (Dilks et al., 2022; R. A. Epstein & Baker, 2019), were not observed with RSA, as all three regions resulted in highly similar RDMs. This may be due to the nature of the RSA and the way it probes the data as we observed these differences in the classification analyses, and it is natural for different types of analyses to uncover distinct aspects of the data. Therefore, comprehensive models, possibly separately for each ROI, are necessary to enlighten the differential nature of scene processing these regions.

Another possible reason we did not observe any ROI differences could be related to our method of ROI selection based on our localizer experiment. There exist different approaches in the literature, some more conservative, such as defining individual ROI masks and omitting participants who do not have clusters that reach a strict threshold. There also exist more liberal approaches, which claim the best practice to be rather lenient thresholding, with much lower cut-offs compared to our approach here (Kawabata Duncan & Devlin, 2011). As explained earlier, we used group-level activation maps to define ROIs and kept their sizes equal while not omitting any participants to be more representative of the general population. This caused us to be less stringent with the thresholds in some cases. As a result, we may have had ROIs that included surrounding regions in some participants with smaller ROIs while not covering the whole area for some other participants, while also introducing variability into the activity patterns we examined with these analyses. Both have advantages and disadvantages, possibly introducing different biases to the analyses, and there is no single correct way of defining ROIs using localizers (Friston et al., 2006; Saxe et al., 2006). Here, we chose not to be too strict by only looking at a few voxels with very high responses from just a few participants who have them. We tried to have a broader, generalizable approach without losing any valuable data.

4.3. ROI definition impacts decoding results in scene-selective regions

To further investigate if our ROI selection method had a remarkable impact on our results, we conducted separate decoding analyses and RSA by only changing the ROI masks and using the same exact remaining parameters. ROI selection and the complementary results can

be found in the Supplementary Material. In short, instead of having flexible thresholds granting us 50-voxel surviving clusters, we used a fixed threshold of $p = 0.05$, family-wise error (FWE) corrected. This resulted in us having larger ROIs for PPA and RSC, but a very small one for OPA. While we did not observe any considerable changes in RSA results, performing the decoding analyses with these ROIs, we were presented with quite a different picture. Unlike our main analyses, this time PPA and RSC were significantly successful in category decoding, whereas OPA did not. However, this time there were no significant differences between any regions, and overall decoding performances were higher. Category decoding examined separately under each task also resulted in differences. During the categorization task, PPA and RSC were significantly successful, and during the approach-avoidance task only PPA was significantly successful. In both cases, OPA showed near-chance levels of decoding performance. Finally, for task decoding, again PPA and RSC were significantly above chance, while OPA was around chance levels. While these results, especially for PPA and RSC with regards to category decoding, are more compatible with the past literature (Bilalić et al., 2019; Dilks et al., 2022; Walther et al., 2009, 2011), the very low and insignificant performance we observed with OPA could be due to the size and definition of the ROI. As a result of a fixed and very conservative threshold, we obtained a 26-voxel unilateral mask only (which is about 0.6 cm^3), whereas other regions totaled to much larger bilateral masks. The size and location of functional regions are not strictly defined, and vary individually, and it is not rare for individuals to have much smaller or non-existing corresponding regions in one hemisphere (commonly left) or have no significant clusters at all. On top of this, there also exist great individual differences in both anatomically and functionally defined structures (Liu, 2011). This points to a third possible approach to ROI selection, which is to define individual ROI masks and/or exclude participants that do not fit our expectations from the analysis completely. We did not follow up with this approach, since it greatly reduces the available data and does not produce generalizable results. However, considering functionally defined parcels change borders and activity across tasks and states, both at the individual and group level measured at various cluster sizes (Salehi et al., 2020), there does not seem to be a certain ideal way of functionally defining the brain in parts anyway.

What is revealed by such distinct results is not necessarily a single "correct" approach to ROI selection, but rather the nature of the examined processes and how they are localized or distributed across the brain and nearby structures. It can be interpreted comparing these two decoding results, then, that the PPA and the RSC process tasks and categories in a more distributed manner that are represented by much larger ROIs, and that the OPA and its immediate vicinity together process indoor scene categories well.

4.4. Limitations

Although we aimed to address important gaps in the literature, using a systematic categorization method, realistic stimuli, and multiple tasks to study the perception of built environments, this study also has certain limitations.

Instead of professional, retouched photos or strictly controlled, unrealistic images, we opted to choosing ordinary, unedited images from a database as our stimuli. This approach may have posed certain limitations since we cannot control the low-level variables. It is known that scene-selective regions are affected by low-level features of the stimuli, which may have added variability to the neural activity patterns we were investigating. However, low-level features are relatively well-studied compared to higher-level processes, and it has been communicated in the field that it is time to view scene processing as it is: a combination of multiple processes using realistic stimuli and tasks (Malcolm et al., 2016). Also, we attempted to address this limitation by adding a visual similarity model to the RSA, which did not result in any significant similarity to the neural activation patterns.

Another limitation is related to the content of our stimulus set. Since we followed the categorization method we chose while deciding on the stimuli and were also limited by the content of the database, some of the images may have affected the participants' behavioral responses and the resulting neural activation patterns. Various characteristics of the environment, such as geometry and aesthetics, can affect brain activity and approach-avoidance tendencies (Shemesh et al., 2021, 2022; Vartanian et al., 2013, 2015), and we indeed found that some stimuli elicited unbalanced responses (Supplementary Fig. 3). Confirming this, we actually saw significantly faster reaction times for the restroom condition, depicting bathrooms and sinks, which may have evoked negative emotions and prompted participants to respond quickly. This can be alleviated by performing a norming study first, and deciding on neutral stimuli if a similar task is to be used and the topic of study is the valence-related judgments. In addition, approach-avoidance tasks often result in variable responses even when the stimuli are controlled for valence due to individual tendencies, and again, we observed considerable individual differences in the distribution of approach-avoidance responses (Supplementary Fig. 4). In this case, however, we were not interested in the valence of the stimuli and only aimed to measure the initial processing by employing an approach-avoidance task, responses to which are not entirely explained by beauty or pleasantness judgments.

4.5. Future directions

This study was a step towards investigating task-driven brain activity differences in the scene-selective regions and future studies can expand our understanding of this topic in several directions. This study only used one branch of one particular categorization approach, and future research can make use of other aspects of this model, use other existing models, or develop new strategies to include various types of scene categories based on different criteria. In this study, we used multiple tasks. However, this was only a first step, and it is necessary to develop and use various ecologically valid behavioral tasks that represent real-life scene perception and our actions within scenes engaging a wide range of cognitive processes. Such in-depth and high-level study of scene perception will contribute to uncovering new information regarding the neural and behavioral mechanisms underlying these processes. In the long run, such approaches can provide a scientific basis for fields such as environmental psychology and neuroarchitecture to aid environmental and architectural design processes, resulting in improved spaces that both serve human needs and promote cognitive functions.

5. Conclusion

This study examined the task influence on the neural activity patterns of scene-selective regions to the scenes from built environment categories. We utilized multivariate analysis techniques such as decoding and RSA. We found that the OPA had overall the highest decoding performance for built environment categories, and task-specific category decoding performances were different in all of these regions, with the PPA and the OPA being significantly successful during the categorization task, and the RSC being significantly successful during the approach-avoidance task. Also, RSC was significantly successful at task decoding. Moreover, with RSA, we found that behavioral goals significantly explain the changes in the neural activity patterns in scene-selective regions while not reaching a true model level. We delved further into ROI definition practices and their impact, highlighting common approaches and their implications. This study is a first step in examining the task influence on scene perception in depth, contributing to our understanding of high-level scene processing.

CRediT authorship contribution statement

Aysu Nur Koc: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Burcu A. Urgen:**

Supervision, Methodology, Conceptualization. **Yasemin Afacan:** Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was funded by a TUBITAK (The Scientific and Technological Research Council of Turkiye) grant, under the 1002-A program (Project No: 122K706), awarded to Yasemin Afacan. The authors would like to thank Aslı Eroğlu and Ömer Rençbereli for their support in data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2024.108539.

Data availability

Data will be made available on request.

References

Alexander, A. S., Place, R., Starrett, M. J., Chrastil, E. R., & Nitz, D. A. (2022). Rethinking retrosplenial cortex: Perspectives and predictions. *Neuron, S0896627322010273*. https://doi.org/10.1016/j.neuron.2022.11.006

Auger, S. D., Mullally, S. L., & Maguire, E. A. (2012). Retrosplenial cortex codes for permanent landmarks. *PLoS ONE, 7*(8), Article e43620. https://doi.org/10.1371/journal.pone.0043620

Bilalić, M., Lindig, T., & Turella, L. (2019). Parsing rooms: The role of the PPA and RSC in perceiving object relations and spatial layout. *Brain Structure and Function, 224*(7), 2505–2524. https://doi.org/10.1007/s00429-019-01901-0

Bonner, M. F., & Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences of the United States of America, 114*(18), 4793–4798. https://doi.org/10.1073/pnas.1618228114

Bonner, M. F., & Epstein, R. A. (2018). Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. *PLoS Computational Biology, 14*(4). https://doi.org/10.1371/journal.pcbi.1006111

Brett, M., Anton, J.-L., Valabrègue, R., & Poline, J. B. (2010). Region of interest analysis using an SPM toolbox.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3), 1–27. https://doi.org/10.1145/1961189.1961199

Cholowsky, N. L., Chen, M. J., Selouani, G., Pett, S. C., Pearson, D. D., Danforth, J. M., Fenton, S., Rydz, E., Diteljan, M. J., Peters, C. E., & Goodarzi, A. A. (2023). Consequences of changing Canadian activity patterns since the COVID-19 pandemic include increased residential radon gas exposure for younger people. *Scientific Reports, 13*(1), 5735. https://doi.org/10.1038/s41598-023-32416-8

Coburn, A., Vartanian, O., Kenett, Y. N., Nadal, M., Hartung, F., Hayn-Leichsenring, G., Navarrete, G., González-Mora, J. L., & Chatterjee, A. (2020). Psychological and neural responses to architectural interiors. *Cortex, 126*, 217–241. https://doi.org/10.1016/j.cortex.2020.01.009

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *Journal of Neuroscience, 31*(31), 11305–11312. https://doi.org/10.1523/JNEUROSCI.1935-11.2011

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of Neuroscience, 33*(4), 1331–1336. https://doi.org/10.1523/JNEUROSCI.4081-12.2013

Dilks, D. D., Kamps, F. S., & Persichetti, A. S. (2022). Three cortical scene systems and their development. *Trends in Cognitive Sciences, 26*(2), 117–127. https://doi.org/10.1016/j.tics.2021.11.002

Edgington, E. S. (1969). Approximate randomization tests. *The Journal of Psychology, 72*(2), 143–149. https://doi.org/10.1080/00223980.1969.10543491

Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition, 12*(6), 954–978. https://doi.org/10.1080/13506280444000607

Epstein, R. A., & Baker, C. I. (2019). *Scene Perception in the Human Brain*. https://doi.org/10.1146/annurev-vision-091718

Epstein, R. A., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of Neurophysiology, 97*(5), 3670–3683. https://doi.org/10.1152/jn.00003.2007

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2007). Where am i now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *Journal of Neuroscience, 27*(23), 6141–6149. https://doi.org/10.1523/JNEUROSCI.0799-07.2007

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area. *Neuron, 23*(1), 115–125. https://doi.org/10.1016/S0896-6273(00)80758-8

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598–601. https://doi.org/10.1038/33402

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, M., James D. andGoncalves, DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIPrep. Software. doi: 10.5281/zenodo.852659.

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., & Gorgolewski, K. J. (2018). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*. https://doi.org/10.1038/s41592-018-0235-4

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *NeuroImage, 30*(4), 1077–1087. https://doi.org/10.1016/j.neuroimage.2005.08.012

Froyen, H. (2012). *Universal Design, a methodological approach: A pathway to human-friendly and elegant architecture*. Institute for Human Centered Design.

Ganaden, R. E., Mullin, C. R., & Steeves, J. K. E. (2013). Transcranial magnetic stimulation to the transverse occipital sulcus affects scene but not object processing. *Journal of Cognitive Neuroscience, 25*(6), 961–968. https://doi.org/10.1162/jocn_a_00372

Gopnarayan, M. N., Rathore, D. R., Bauer, F. S., Hillard, J., & Chawla, P. (2022). Differential representation of natural and manmade images in the human ventral visual stream [Preprint]. *Neuroscience*. https://doi.org/10.1101/2022.09.22.509086

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., & Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data, 3*(1), Article 160044. https://doi.org/10.1038/sdata.2016.44

Greene, M., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137–176. https://doi.org/10.1016/j.cogpsych.2008.06.001

Greene, M. R., & Hansen, B. C. (2020). Disentangling the independent contributions of visual and conceptual features to the spatiotemporal dynamics of scene categorization. *Journal of Neuroscience, 40*(27), 5283–5299. https://doi.org/10.1523/JNEUROSCI.2088-19.2020

Groen, I. I. A., Ghebreab, S., Prins, H., Lamme, V. A. F., & Scholte, H. S. (2013). From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category. *Journal of Neuroscience, 33*(48), 18814–18824. https://doi.org/10.1523/JNEUROSCI.3128-13.2013

Groen, I. I. A., Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1714). https://doi.org/10.1098/rstb.2016.0102

Guo, F., Preston, T., Giesbrecht, B., & Eckstein, M. P. (2012). Automatic neural coding of open and closed scenes in RSC and PPA during visual search. *Journal of Vision, 12*(9), 595. https://doi.org/10.1167/12.9.595

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences, 111*(10). https://doi.org/10.1073/pnas.1312567111

Hebart, M. N., Görgen, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics, 8*. https://doi.org/10.3389/fninf.2014.00088

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*(1), 243–271. https://doi.org/10.1146/annurev.psych.50.1.243

Henderson, J. M., Larson, C. L., & Zhu, D. C. (2007). Cortical activation to indoor versus outdoor scenes: An fMRI study. *Experimental Brain Research, 179*(1), 75–84. https://doi.org/10.1007/s00221-006-0766-2

Henriksson, L., Mur, M., & Kriegeskorte, N. (2019). Rapid invariant encoding of scene layout in human OPA. *Neuron, 103*(1), 161–171.e3. https://doi.org/10.1016/j.neuron.2019.04.014

Johnson, M. R., & Johnson, M. K. (2014). Decoding individual natural scene representations during perception and imagery. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00059

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage, 60*(4), 2357–2364. https://doi.org/10.1016/j.neuroimage.2012.02.055

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *NeuroImage, 132*, 417–424. https://doi.org/10.1016/j.neuroimage.2016.02.062

Kaping, D., Tzvetanov, T., & Treue, S. (2007). Adaptation to statistical properties of visual scenes biases rapid categorization. *Visual Cognition, 15*(1), 12–19. https://doi.org/10.1080/13506280600856660

Kaplan, S. (1992). Environmental preference in a knowledge-seeking, knowledge-using organism. In *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 581–598). Oxford University Press.

Kauffmann, L., RamanoÃ«l, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience, 8*. https://doi.org/10.3389/fnint.2014.00037

Kawabata Duncan, K. J., & Devlin, J. T. (2011). Improving the reliability of functional localizers. *NeuroImage, 57*(3), 1022–1030. https://doi.org/10.1016/j.neuroimage.2011.05.009

Kay, K., Bonnen, K., Denison, R. N., Arcaro, M. J., & Barack, D. L. (2023). Tasks and their role in visual neuroscience. *Neuron, 111*(11), 1697–1713. https://doi.org/10.1016/j.neuron.2023.03.022

Klepeis, N. E., Nelson, W. C., Ott, W. R., Robinson, J. P., Tsang, A. M., Switzer, P., Behar, J. V., Hern, S. C., & Engelmann, W. H. (2001). The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. *Journal of Exposure Science & Environmental Epidemiology, 11*(3), 231–252. https://doi.org/10.1038/sj.jea.7500165

Korn, R. M., & Elliot, A. J. (2015). Avoidance and approach motivation: a brief history. In International Encyclopedia of the Social & Behavioral Sciences (pp. 326–331). Elsevier. doi: 10.1016/B978-0-08-097086-8.26071-6.

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*(20), 7322–7333. https://doi.org/10.1523/JNEUROSCI.4588-10.2011

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/neuro.06.004.2008

Liu, T. (2011). A few thoughts on brain ROIs. *Brain Imaging and Behavior, 5*(3), 189–202. https://doi.org/10.1007/s11682-011-9123-6

Luking, K. R., Nelson, B. D., Infantolino, Z. P., Sauder, C. L., & Hajcak, G. (2017). Internal consistency of functional magnetic resonance imaging and electroencephalography measures of reward in late childhood and early adolescence. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2*(3), 289–297. https://doi.org/10.1016/j.bpsc.2016.12.004

Malcolm, G. L., Groen, I. I. A., & Baker, C. I. (2016). Making sense of real-world scenes. *Trends in Cognitive Sciences, 20*(11), 843–856. https://doi.org/10.1016/j.tics.2016.09.003

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(44), 14896–14908. doi: 10.1523/JNEUROSCI.2270-15.2015.

Meissner, T. W., Nordt, M., & Weigelt, S. (2019). Prolonged functional development of the parahippocampal place area and occipital place area. *NeuroImage, 191*, 104–115. https://doi.org/10.1016/j.neuroimage.2019.02.025

Miller, A. M. P., Vedder, L. C., Law, L. M., & Smith, D. M. (2014). Cues, context, and long-term memory: The role of the retrosplenial cortex in spatial cognition. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00586

Mitchell, A. S., Czajkowski, R., Zhang, N., Jeffery, K., & Nelson, A. J. D. (2018). Retrosplenial cortex and its role in spatial cognition. *Brain and Neuroscience Advances, 2*, Article 239821281875709. https://doi.org/10.1177/2398212818757098

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping, 15*(1), 1–25. https://doi.org/10.1002/hbm.1058

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLoS Computational Biology, 10*(4). https://doi.org/10.1371/journal.pcbi.1003553

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41*(2), 176–210. https://doi.org/10.1006/cogp.1999.0728

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.

Patai, E. Z., & Spiers, H. J. (2017). Human navigation: occipital place area detects potential paths in a scene. *Current Biology, 27*(12), R599–R600. https://doi.org/10.1016/j.cub.2017.05.012

Peer, M., & Epstein, R. A. (2021). The human brain uses spatial schemas to represent segmented environments. *Current Biology, 31*(21), 4677–4688.e8. https://doi.org/10.1016/j.cub.2021.08.012

Persichetti, A. S., & Dilks, D. D. (2018). Dissociable neural systems for recognizing places and navigating through them. *The Journal of Neuroscience, 38*(48), 10295–10304. https://doi.org/10.1523/JNEUROSCI.1200-18.2018

Persichetti, A. S., & Dilks, D. D. (2019). Distinct representations of spatial and categorical relationships across human scene-selective cortex. *Proceedings of the National Academy of Sciences, 116*(42), 21312–21317. https://doi.org/10.1073/pnas.1903057116

Ramanoël, S., York, E., Le Petit, M., Lagrené, K., Habas, C., & Arleo, A. (2019). Age-related differences in functional and structural connectivity in the spatial navigation brain network. *Frontiers in Neural Circuits, 13*, 69. https://doi.org/10.3389/fncir.2019.00069

Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes? *Visual Cognition, 12*(6), 852–877. https://doi.org/10.1080/13506280444000553

Salehi, M., Greene, A. S., Karbasi, A., Shen, X., Scheinost, D., & Constable, R. T. (2020). There is no single functional atlas even for a single individual: Functional parcel definitions change with task. *NeuroImage, 208*, Article 116366. https://doi.org/10.1016/j.neuroimage.2019.116366

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage, 30*(4), 1088–1096. https://doi.org/10.1016/j.neuroimage.2005.12.062

Shemesh, A., Leisman, G., Bar, M., & Grobman, Y. J. (2021). A neurocognitive study of the emotional impact of geometrical criteria of architectural space. *Architectural Science Review, 64*(4), 394–407. https://doi.org/10.1080/00038628.2021.1940827

Shemesh, A., Leisman, G., Bar, M., & Grobman, Y. J. (2022). The emotional influence of different geometries in virtual spaces: A neurocognitive examination. *Journal of Environmental Psychology, 81*, Article 101802. https://doi.org/10.1016/j.jenvp.2022.101802

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*(34), 11921–11935. https://doi.org/10.1523/JNEUROSCI.0137-15.2015

Stacho, M., & Manahan-Vaughan, D. (2022). Mechanistic flexibility of the retrosplenial cortex enables its contribution to spatial cognition. *Trends in Neurosciences, 45*(4), 284–296. https://doi.org/10.1016/j.tins.2022.01.007

Szucs, D., & Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990–2012) and of latest practices (2017–2018) in high-impact journals. *NeuroImage, 221*, Article 117164. https://doi.org/10.1016/j.neuroimage.2020.117164

Todd, T. P., Fournier, D. I., & Bucci, D. J. (2019). Retrosplenial cortex and its role in cue-specific learning and memory. *Neuroscience & Biobehavioral Reviews, 107*, 713–728. https://doi.org/10.1016/j.neubiorev.2019.04.016

Troiani, V., Stigliani, A., Smith, M. E., & Epstein, R. A. (2014). Multiple object properties drive scene-selective regions. *Cerebral Cortex, 24*(4), 883–897. https://doi.org/10.1093/cercor/bhs364

Valente, G., Castellanos, A. L., Hausfeld, L., De Martino, F., & Formisano, E. (2021). Cross-validation and permutations in MVPA: Validity of permutation strategies and power of cross-validation schemes. *NeuroImage, 238*, Article 118145. https://doi.org/10.1016/j.neuroimage.2021.118145

Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nature Reviews Neuroscience, 10*(11), 792–802. https://doi.org/10.1038/nrn2733

Vartanian, O., Navarrete, G., Chatterjee, A., Fich, L. B., Gonzalez-Mora, J. L., Leder, H., Modroño, C., Nadal, M., Rostrup, N., & Skov, M. (2015). Architectural design and the brain: Effects of ceiling height and perceived enclosure on beauty judgments and approach-avoidance decisions. *Journal of Environmental Psychology, 41*, 10–18. https://doi.org/10.1016/j.jenvp.2014.11.006

Vartanian, O., Navarrete, G., Chatterjee, A., Fich, L. B., Leder, H., Modroño, C., Nadal, M., Rostrup, N., & Skov, M. (2013). Impact of contour on aesthetic judgments and approach-avoidance decisions in architecture. Proceedings of the National Academy of Sciences, 110(supplement_2), 10446–10453. doi: 10.1073/pnas.1301227110.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *The Journal of Neuroscience, 29*(34), 10573–10581. https://doi.org/10.1523/JNEUROSCI.0559-09.2009

Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences, 108*(23), 9661–9666. https://doi.org/10.1073/pnas.1015666108

Xiao, J., Ehinger, K. A., Hays, J., Torralba, A., & Oliva, A. (2016). SUN database: exploring a large collection of scene categories. *International Journal of Computer Vision, 119*(1), 3–22. https://doi.org/10.1007/s11263-014-0748-y

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010*, 3485–3492. https://doi.org/10.1109/CVPR.2010.5539970