Neural codes of architectural styles

Heeyoung Choo[1,2], Jack Nasar[3], Bardia Nikrahei [3], & Dirk B. Walther[4]

Keywords:
Architectural styles; brain decoding; scene categorization; parahippocampal place area; fMRI; multi-voxel pattern analysis

Word count: 2485/2500

Correspondence:
Heeyoung Choo
405 N Mathews Ave
Urbana, IL 61801, United States
E-mail: hchoo@illinois.edu
Phone: +1-217-244-2354

[1] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois, 61801, United States
[2] Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, Illinois, 61820, United States
[3] Department of City and Regional Planning, The Ohio State University Columbus, Ohio, 43210, United States
[4] Department of Psychology, University of Toronto, Toronto, Ontario, M5S 3G3, Canada

Abstract

Images of iconic buildings, for example, the Empire State Building, instantly transport us to New York City. Despite the substantial impact of architectural design on people's visual experience of built environments, we know little about its neural representation in the human brain. We have found patterns of neural activity associated with specific architectural styles in a network of several high-level visual brain regions including the scene-selective parahippocampal place area (PPA). Surprisingly, this network, which is characterized by correlated error patterns, includes the fusiform face area. Accuracy of decoding architectural styles from the PPA was negatively correlated with expertise in architecture, indicating a shift from purely visual cues to the use of domain knowledge with increasing expertise. Our study showcases that neural representations of architectural styles in the human brain are driven not only by perceptual features but also by semantic and cultural facets, such as expertise for architectural styles.

149/150 words

As of 2014, more than half of the world's population resided in urban environments [1]. Architectural design has profound impact on people's preferences and productivity in such built environments [2]. Despite the ubiquity and importance of architecture for people's lives, it is so far unknown where and how architectural styles are represented in people's brains. Here we show that architectural styles are represented in distributed patterns of neural activity in several visually active brain regions in ventral temporal cortex but not in primary visual cortex.

In a functional magnetic resonance imaging (fMRI) scanner, 23 students in their final year at The Ohio State University (11 majoring in architecture, 12 majoring in psychology or neuroscience, one psychology major excluded due to excessive head motion) passively viewed blocks of images. Each block comprised four images from one of the following sixteen categories; (1) representative buildings of four architectural styles (Byzantine, Renaissance, Modern, and Deconstructive); (2) representative buildings designed by four famous architects (Le Corbusier, Antoni Gaudi, Frank Gehry, and Frank Lloyd-Wright); (3) four entry-level scene categories (mountains, pastures, highways, and playgrounds); and (4) photographs of faces of four different non-famous men (Fig. 1). Brain activity was recorded in 35 coronal slices, which covered approximately the posterior 70% of the brain. For each participant, several visually active regions of interest (ROI) were functionally localized: the parahippocampal place area (PPA), the occipital place area (OPA), the retrosplenial cortex (RSC), the lateral occipital complex (LOC), and the fusiform face area (FFA). Primary visual cortex (V1) was delineated based on anatomical atlases.

Following standard pre-processing, data from the image blocks were subjected to a multi-voxel pattern analysis (MVPA). For each of the four groups of stimuli, a linear support vector machine decoder was trained to discriminate between the activity patterns associated with each of the four sub-categories. The decoder was tested on independent data in a leave-one-run-out (LORO) cross validation. Separate decoders were trained and tested for each participant and each ROI. Accuracy was compared to chance (25%) at the group level using one-tailed t tests.

Results

Successful decoding of architectural categories from human visual cortex

Consistent with previous results [3, 4, 5] we could decode entry-level scene categories from all visually active ROIs (Fig. 1A). Furthermore, we could decode architectural styles from all five high-level visual brain regions, but not from V1 (Fig. 1B). In addition, it was possible to decode buildings by famous architects from brain activity in the PPA, the OPA, and the LOC, but not from V1, the RSC, or the FFA (Fig. 1C). Decoding of facial identity succeeded only in V1 and was not possible in any of the high-level ROIs, including the FFA. We also found statistically significant differences in average activity levels between sub-categories for the categorization conditions in a subset of the ROIs. However, such

differences were not sufficient to allow for 4-way decoding. The discrimination between sub-categories was only possibly by considering the *spatial pattern* of brain activity within ROIs.

Searchlight analysis of the scanned parts of the brain confirmed the ROI-based results. The searchlight map of decoding entry-level scene categories showed significant clusters at both occipital poles and calcarine gyri as well in bilateral lingual, fusiform, and parahippocampal gyri and bilateral transverse occipital sulci. On the other hand, the searchlight map of decoding architectural styles showed clusters encompassing bilateral fusiform gyri and transverse occipital sulci, but not the occipital poles and nearby areas. The searchlight map for decoding buildings by famous architects was similar to that of decoding architectural styles, with an additional small cluster on the right occipital pole. Table 1 provides a full list of significant clusters from each searchlight map. Analysis of the overlap of individual's searchlight maps with their ROIs is shown in Table S1.

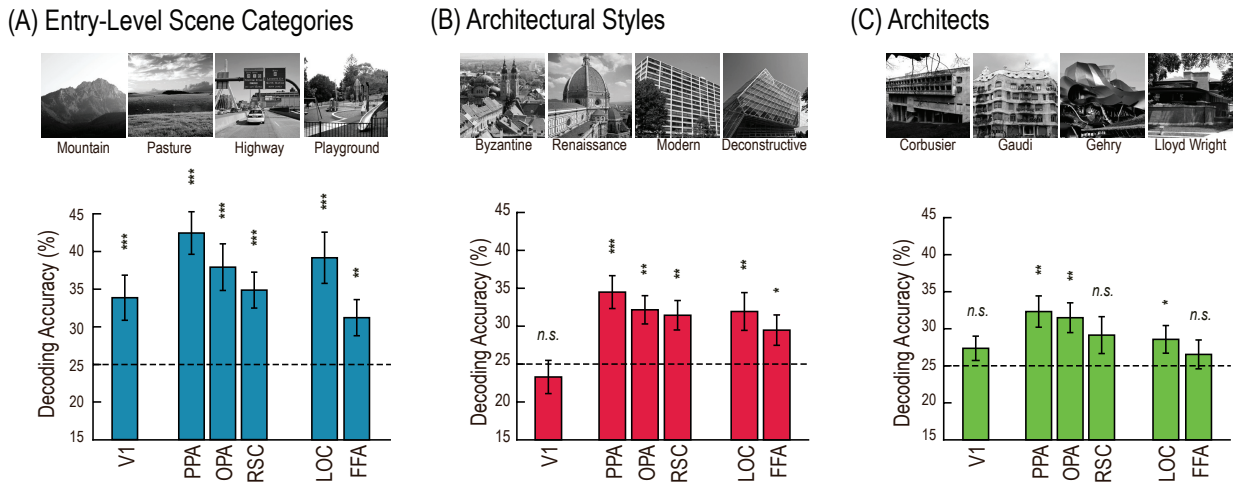

Figure 1. Example images and category decoding accuracy rates for three image types across the ROIs: (A) entry-level scene categories, (B) architectural style, and (C) architects (for differences in mean activity levels see Fig S1). Participants viewed 32 different examples per category in nine runs. Participants were asked to detect rare back-to-back repetitions of images to engage their attention. Four randomly chosen images from each category were repeated back-to-back. Here, we show one example per category for the three image types. Separate decoders were trained to discriminate between four categories per condition, separately for V1, PPA, OPA, RSC, LOC, and FFA. The decoders were tested on independent data from the same participants in a leave-one-run-out cross-validation procedure. We here report group-average decoding accuracy rates. Significance with respect to chance (25%) was assessed with one-sample t-tests (one-tailed). Error bars indicate *SEM*. $*p < .05$, $**p < .01$, $***p < .001$.

Table 1. Clusters identified in the searchlight analysis for the four categorization conditions (See Table S1 for the group average percent of overlap with the ROIs).

Decoding Condition	Peak				Volume (µl)	Description
	x	y	z	Accuracy (%)		
Scenes	40.0	85.5	13.2	43.4	167542	Bilateral occipital poles, calcarine gyri, fusiform gyri. lingual gyri, bilateral hippocampus, parahippocampal gyri, inferior occipital gyri and sulci, middle occipital gyri, superior occipital gyri, transverse occipital sulci, inferior parieto-angular gyri, superior parieto-occiptial sulci, cerebella
	-7.5	63.0	48.2	32.3	688	Right precuneus
	-42.5	30.5	-6.8	31.2	281	White matter between right hippocampus and right superior temporal sulcus
	62.5	43.0	25.8	30.2	203	Left inferior parietosupramarginal gyrus
Styles	-47.5	63.0	-9.2	34.1	10360	Right inferior occipital gyrus and sulcus, right occipito-temporal (lateral fusiform) gyrus, right medial occipito-temporal sulcus, right middle temporal gyrus, right inferior temporal gyrus, right middle temporal gyrus
	-30.0	85.5	20.8	33.6	8063	Right superior occipital sulcus, right transverse occipital sulcus, right middle occipital gurus, right occipito-temporal (lateral fusiform) gyrus, right lateral occipto-temporal sulcus
	47.5	63.0	-6.8	33.7	6672	Left inferior occipital gyrus and sulcus, left occipito-temporal (lateral fusiform) gyrus, left inferior temporal gyrus, left middle temporal gyrus.
	15.0	78.0	50.8	32.9	2813	Left superior occipital sulcus, left transverse occipital sulcus, left superior parietal gyrus, left precuneus
	42.5	85.5	15.8	32.2	2235	Left middle occipital gyrus
	7.5	90.5	8.2	33.9	1063	Left cuneus
	-10.0	48.0	5.8	32.3	906	Right posterior ventral cingulate gyrus
	-35.0	25.5	3.2	32.2	531	Right superior parietal gyrus
	15.0	93.0	33.2	30.5	500	Left superior occipital gyrus
	-12.5	93.0	18.2	31.1	391	Right superior occipital gyrus
	20.0	85.5	8.2	30.9	375	White matter between left middle occipital gyrus and left cuneus
	-50.0	55.5	40.8	31.7	266	Right inferior parieto-angular gyrus
Architects	22.5	100.5	-9.2	33.4	10032	Left occipital pole, left inferior occipital gyrus and sulcus, left middle occipital gyrus, left superior occipital gyrus
	-47.5	68.0	-6.8	32.5	5672	Right inferior occipital gyrus and sulcus, right middle occipital gyrus
	-15.0	93.0	5.8	31.2	1938	Right occipital pole
	-30.0	40.5	-6.8	32.3	1047	Right lateral occipito-temporal (fusiform) gyrus, right medial occipito-temporal (lingual) gyrus, right hippocampus, right parahippocampal gyrus
	22.5	73.0	-6.8	32.1	453	White matter near the left medial occipito-temporal gyrus and sulcus
	30.0	63.0	-6.8	31.2	344	White matter between left medial occipito-temporal (lingual) gyrus and left lateral occipito-temporal (fusiform) gyrus
	-20.0	25.5	58.2	31.2	344	White matter near the left precentral gyrus
Face	-2.5	93.0	8.2	41.0	27174	Bilateral occipital pole, calcarine gyri, cuneus, medial occipito-temporal (lingual) gyri, superior occipital gyri
	-32.5	50.5	8.2	31.2	469	Right superior parietal gyrus, right intraparietal sulcus, right transverse occipital sulcus

Analysis of error patterns

To explore the nature of the representation of architectural styles in visual cortex in more detail, we analyzed patterns of decoding errors. Decoding errors were recorded in confusion matrices, whose rows (*r*) indicate the ground truth of the presented category, and whose columns (*c*) represent predictions by the decoder. Individual cells (*r,c*) contain the proportion of blocks with category *r*, which were decoded as category *c*. Diagonal elements contain correct predictions, summarized as decoding accuracy in Fig. 1. Off-diagonal elements represent decoding errors. The patterns of decoding errors serve as a proxy for the nature of the neural representation of categories in a particular brain region. We computed the correlations of error patterns as a measure of the similarity between these neural representations across ROIs. Significance of error correlations was established non-parametrically against the null distribution of correlations obtained by jointly permuting the rows and columns of one of the confusion matrices. Only error correlations with none of the 24 permutations resulting in higher correlation than the correct ordering ($p < 0.0417$) were deemed significant.

In the case of entry-level scene categorization, we found significant correlations of error patterns between the three ROIs known to specialize in scene perception: the PPA, the RSC, and the OPA (Fig. 2A). We also found significant error correlation between the PPA and the LOC, which is likely due to the recruitment of LOC for the detection of diagnostic objects in scenes [6, 7]. Note that error patterns from the FFA did not correlate significantly with any of the other ROIs, even though we could decode entry-level scene categories from the FFA.

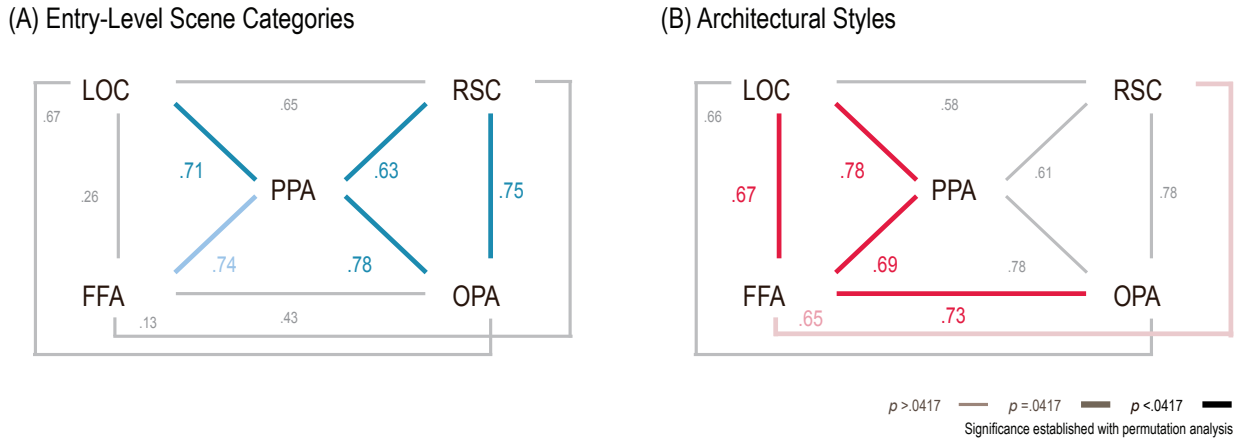

Figure 2. Group-average error pattern correlations and their significance between the PPA, OPA, RSC, LOC, and FFA for (A) entry-level scene categories (in blue) and (B) architectural styles (in red). Significance of error correlations was established by permutation analysis. Thick vivid lines indicate significant error pattern correlations. Thick faint lines indicates marginal correlations ($p = .047$; one permutation resulted in a correlation higher than the correct ordering). Thin gray lines indicate correlations failing to reach significance. See Figure S2 for the confusion matrices underlying this analysis.

For architectural styles, we found a different error correlation structure (Fig. 2B). For this more specialized, subordinate level categorization, we found statistically significant error correlations of the FFA with the LOC, the PPA, and the OPA, as well as between the PPA and the LOC. Note that the RSC no longer shows significant error correlations with any of the other ROIs. We here show for the first time that the FFA is recruited into the scene processing network for demanding subordinate-level scene categorization but not for simple entry-level categorization. This is consistent with the FFA's role in visual expertise as shown for object categories as varied as birds, cars, motorcycles or artificial "Greeble" objects [8], even though those results were shown for mean activity levels, whereas ours appear in the interpretation of patterns of brain activity

We did not find any statistically significant error correlations between ROIs for decoding architects, possibly due to the difficulty of decoding architects from brain activity in the first place. Given that facial identity could not be decoded from any of the high-level visual ROIs, we did not further pursue error correlations for the face identification condition.

The effect of expertise

The involvement of the FFA in the representation of categories for architectural styles suggests a role of expertise in the subordinate-level categorization of architectural styles, but not in entry-level scene categorization. However, unlike the typical scenario of subordinate-level visual categorization (i.e., Golden Retrievers vs. Chihuahua), accurate recognition of architectural styles or architects of the buildings is highly affected by non-visual factors. The distinction between architectural styles relies not only on visual consistency within a style but also on the historical, regional, and cultural context of buildings. Prior knowledge of a building's style may be an important factor in accurate classification, without requiring reference to the visual aspects of the building. How, then, does expertise for architecture affect the neural representation of architectural categories in visual cortex?

We measured expertise for architectural styles in a post-scan behavioral experiment employing the Vanderbilt Expertise Test [9]. During the behavioral experiment, participants were asked to identify which of three displayed images belonged to a given set of six target categories. Behavioral accuracy ranged from 20.0% to 100.0% with a mean of 72.5% (chance: 33.3%). Architecture students were more accurate than non-architecture students at a statistically significant level for architectural styles, $t(20) = 3.963$, $p < .001$ (architecture students: 77.1%, $SD = 8.9\%$, psychology and neuroscience students: 59.5%, $SD = 11.8\%$), and architects, $t(20) = 3.960$, $p < .001$ (architecture students: 72.5%, $SD = 12.0\%$, psychology and neuroscience students: 44.2%, $SD = 9.2\%$), but not for entry-level scene categories, $t(20) = .869$, $p = .395$ (architecture students: 98.2%, $SD = 2.9\%$, psychology and neuroscience students: 96.1%, $SD = 2.3\%$).

Comparison of decoding accuracy from neural data showed no between-group differences in any of the ROIs, except for a marginal effect for decoding architectural styles from the PPA, where architecture students showed lower decoding accuracy (30.1%) than did psychology and neuroscience students (34.3%), $t(10) = 2.038$, $p = .069$. To account for the full range of individual differences, we correlated each individual's behavioral accuracy with their MVPA decoding accuracy for architectural styles in the PPA. We found significant negative correlation between behavior and decoding accuracy ($r = -.56$, $p = .007$; Fig 3).

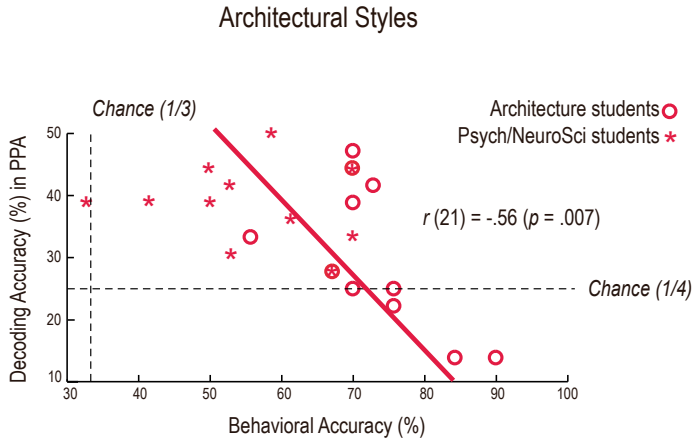

Figure 3. Correlation between behaviorally measured expertise and decoding accuracy from the PPA for architectural styles. Circles indicate data points of architecture students, and asterisks indicate data points of psychology and neuroscience students. Dashed lines mark chance performances for behavior (33%) and decoding (25%). See Figure S3 for details on the behavioral performances in the expertise experiment.

Discussion

We have shown for the first time that subordinate categories of buildings, architectural styles, are represented in the neural activity patterns of several high-level visual areas in human temporal cortex. It was even possible to decode the architects of buildings from neural activity elicited by images of the buildings in the PPA, the OPA, and the LOC. Unlike entry-level scene categories, architectural style and architects could not be decoded from activity in V1, indicating that the simple visual properties encoded in V1 are insufficient to discriminate between architectural styles. These findings suggest that the neural representations of architectural features rely on complex visual structure beyond simple feature statistics. For instance, byzantine architecture is characterized by symmetry in the global shape of buildings and a dome roof, whereas deconstructive architecture is well-known for its non-collinearity and fragmented global shape. Complex visual properties associated with architectural design elements have previously been suggested to contribute to successful cross-decoding between interior and exterior views of landmark buildings [10].

When discriminating between architectural styles, the fusiform face area, previously implicated in the preferential processing of faces [11] as well as visual expertise [8], is recruited as part of a network of regions that share similar error patterns. The FFA could be involved in the encoding of configural characteristics of buildings, such as Lloyd-Wright's signature horizontally elongated proportions. By contrast, entry-level categorization of scenes does not include the FFA in the same way, instead relying on a tight network of three scene-selective areas, the PPA, the RSC, and the OPA, as well as the LOC.

Categorizing a building by its architectural style or its designer involves not only detecting characteristic visual features, but also recruitment of semantic knowledge. Indeed, domain knowledge of architecture is likely to contribute to the neural representations of architectural styles. This was shown clearly by the negative correlation between behavioral expertise scores and individual decoding accuracy for architectural styles in the PPA. We presume that participants with more expertise in architecture relied more on their domain knowledge and less on the high-level visual features represented in the PPA when making judgments about architectural styles.

In summary, several high-level visual regions, but not the V1, contain decodable neural representations of architectural styles and architects of buildings. The FFA participates in a network of high-level visual areas characterized by similar error patterns, but only in the subordinate categorization of architectural styles and not in entry-level categorization of scenes. Furthermore, accuracy of decoding architectural styles from the PPA is negatively correlated with expertise in architecture, indicating a shift from purely visual cues to domain knowledge with increasing expertise. Our study showcases that neural correlates of human classification of visual categories are driven not only by perceptual features but also by semantic and cultural facets, such as expertise of architectural styles and architects of buildings. Most importantly, we have identified in the human visual system a neural representation of architecture, one of the predominant and longest-lasting artefacts of human culture.

Methods

Participants: Twenty-three healthy undergraduate students in their final year at The Ohio State University participated in the study for monetary compensation. We recruited eleven students from the Department of Architecture (2 females; l left-handed; age range = 21–27, M = 22.4, SD = 3.0), and twelve senior students majoring in psychology or neuroscience (3 females; 2 left-handed, age range = 21– 24, M = 21.8, SD = 0.9). Data from one psychology student were not included in the analysis due to excessive head motion during the scan.

fMRI Experiment: MRI images were recorded on a 3T Siemens MAGETOM Trio with a 12-channel head coil at the Center for Cognitive and Behavioral Brain Imaging at The Ohio State University. High-resolution anatomical images were obtained with a 3D-MPRAGE sequence with coronal slices covering

the whole brain; inversion time = 930 ms, repetition time (TR) = 1900 ms, echo time (TE) = 4.44 ms, flip angle = 9°, voxel size = 1 x 1 x 1 mm, matrix size = 224 x 256 x 160. Functional images were obtained with T2*-weighted echo-planar sequences with coronal slices covering approximately the posterior 70% of the brain: TR = 2000ms, TE = 28ms, flip angle = 72°, voxel size = 2.5 x 2.5 x 2.5 mm, matrix size = 90 x 100 x 35.

Participants viewed 512 grayscale photographs of four image types: (1) 32 images of representative buildings of each of four architectural styles: Byzantine, Renaissance, Modern, and Deconstructive; (2) 32 images of buildings designed by each of four well-known architects: Le Corbusier, Antoni Gaudi, Frank Gehry, and Frank Lloyd-Wright; (3) 32 scene images per each of four entry-level scene categories: mountains, pastures, highways, and playgrounds; (4) 32 face images per each of four different individuals [12]. Brightness and contrast were equalized across all images. Images were back-projected with a DLP projector (Christie DS+6K-M 3-chip SXGA+) onto a screen mounted in the back of the scanner bore and viewed through a mirror attached to the head coil. Images subtended approximately 12º x 12 º of visual angle. A fixation cross measuring 0.5º x 0.5º of visual angle was displayed at the center of the screen.

During each of nine runs, participants saw sixteen 8-second blocks of images. In each block, four photographs from a single category were each shown for 1800 ms, followed by a 200 ms gap. The order of images within a block and the order of blocks within a run were randomized in such a way that the four blocks belonging to the same stimulus type (entry-level scenes, styles, architects, faces) were shown back to back. A 12-sec fixation period was placed between blocks as well as at the beginning and the end of each run, resulting in a duration of 5 min 32sec per run. Occasionally, (approximately one out of eight blocks), an image was repeated back-to-back within a block. Participants were asked to press a button when they detected image repetitions.

FMRI data were motion corrected, spatially smoothed (2 mm full width at half maximum), and converted to percent signal change. We used a general linear model with only nuisance regressors to regress out effects of motion and scanner drift. Residuals corresponding to image blocks were extracted with a 4 s hemodynamic lag and averaged over the duration of each block. Block-average activity patterns within pre-defined ROIs was used for MVPA.

Author Contributions

Conceptualization, D. B. W, J. N., and C. H.; Methodology, H. C., D. B. W., B. N.; Investigation, H. C., and D. B. W.; Writing – Original Draft, H. C.; Review & Editing, D. B. W., and J. N.; Funding Acquisition, D. B. W.; Resources, D. B. W.; Supervision, D. B. W. and J. N.

References

1. United Nations, Department of Economic and Social Affairs, Population Division. (2014). World Urbanization Prospects: The 2014 Revision, Highlights (ST/ESA/SER.A/352)

2. Nasar, J. L. (1994). Urban design aesthetics the evaluative qualities of building exteriors. Environment and behavior, 26(3), 377-401.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. J. Neurosci. 29, 10573–10581.

3. 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. J. Neurosci. 29, 10573–10581.

4. 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. Proc. Natl. Acad. Sci. U. S. A. 108, 9661–9666.

5. Park, S., Brady, T. F., Greene, M. R. & Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the PPA and LOC in representing real-world scenes. J. Neurosci. 31(4), 1333-1340.

6. MacEvoy, S. P., & Epstein, R., A. (2011). Constructing scenes from objects in human occipitotemporal cortex. Nat. Neurosci. 14, 1323-1329.

7. Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the parahippocampal place area. Neuroimage, 75, 236-245.

8. Gauthier, I., Tarr, M. J., Anderson, A. W., Skularski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat. Neurosci, 2, 568-573.

9. McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012). The Vanderbilt expertise test reveals domain-general and domain-specific sex effects in object recognition. Vision Res. 69, 10-22.

10. Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2015). Outside looking in: Landmark generalization in the human navigational system. J. Neurosci. 35(44), 14896-14908.

11. Kanwisher, N., McDermott, J., & Chun, M. (1997) The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for the Perception of Faces. J. Neurosci. 17, 4302-4311.

12. Ho J., Lee K., & Kriegman, D. (2005). Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27(5), 684 - 698.

Supplemental Experimental Procedures

Regions of interest

High-level visual regions of interest (ROI) were defined functionally using separate localizer scans. Participants saw one to three runs (7 minutes and 12 seconds each) of blocks of images of faces, scenes, objects, and grid-scrambled objects while responding to image repetitions with a button press. Following motion correction, spatial smoothing (4 mm full width at half maximum Gaussian kernel) and normalization to percent signal change, localizer data were analyzed using a general linear model (3dDeconvolve in AFNI). ROIs were defined as contiguous clusters of voxels with significant contrasts ($q < 0.05$; corrected for multiple comparisons using false discovery rate) in the following comparisons: scenes > (faces, objects) for the parahippocampal place are (PPA), retrosplenial cortex (RSC), and the occipital place area (OPA) [1, 2]; faces > (scenes, objects) for the fusiform face area (FFA) [3]; and objects > scrambled objects for the lateral occipital complex (LOC) [4]. The PPA and RSC were successfully identified in all twenty-two participants. We could not find significant clusters corresponding to the OPA in five participants, the FFA in one participant, and the LOC in two participants. Group statistics of ROI-based results was performed only for the participants for whom we could identify the ROIs.

Primary visual cortex (V1) was defined on each participant's original cortical surface map using the automatic cortical parcellation provided by Freesurfer [5]. Surface-defined V1 was registered back to the volumetric brain separately for each hemisphere using AFNI.

Univariate analysis

We tested whether the four types of visual categories elicited different levels of mean activity in each of the ROIs. We conducted a mixed-effects analysis of variances (ANOVA) for each ROI separately, using participant group (Architecture vs. Psychology and Neuroscience students) as a between-subjects factor, and visual category (entry-level scene categories vs. architectural styles vs. architects vs. faces) as within-subjects factors. Since there was neither a main effect for group nor an interaction between group and visual category, we collapsed the data for the two groups. Results are show in Fig. S1A. Differences in mean activity between the three scene-type categories and faces were assessed using planned paired t-tests, separately for each ROI. Differences in mean activity among the subordinate categories for each of the four main categories were evaluated with one-way ANOVAs. Results are shown in Fig. S1B.

Searchlight analysis

We explored representations of image categories outside of the pre-defined ROIs with a searchlight analysis using the Searchmight Toolbox [6]. The size of the searchlight region was chosen as a 5x5x5 =125 voxel cube to approximate the average size of a unilateral PPA of the participants (159 voxels). The searchlight was centered on each voxel in turn [7], and decoding analysis with leave-one-run-out cross-validation was performed using the voxels within the searchlight regions. Decoding accuracies for the searchlight locations were recorded in a brain map, thresholded at $p < 0.01$ (one-tailed analytical p value), and corrected for multiple comparisons at the cluster level with a minimum cluster size determined separately for each participant, ranging from 4 to 8 voxels ($M = 4.8$, $SD = 0.9$). We evaluated the agreement between the searchlight analysis and the pre-defined ROIs as the fraction of voxels within each ROI that was found to be significantly above chance in the searchlight analysis (Table S1).

For group analysis, anatomical brain volumes of each of the participants were registered to the Montreal Neurological Institute (MNI) 152 template [8]. Searchlight accuracy maps were registered to MNI space using the parameters from the anatomical registration, followed by smoothing with a 2 mm full width at half maximum Gaussian filter. Significance of group-average decoding accuracy versus chance (25%) was assessed with a one-sample one-tailed t-test ($p < 0.01$), followed by cluster-level correction for multiple comparisons (minimum cluster size of 13 voxels, determined by α probability simulation).

Post-scan behavioral experiment

We measured participants' visual domain knowledge in a post-scan behavioral experiment similar to the Vanderbilt Expertise Test [9]. Domain knowledge for each image type was tested in four separate blocks. Each block consisted of three components: study, practice, and testing. During study, participants were introduced to six target categories. Example images for each of the six target categories were displayed on the screen with correct category labels: (1) entry-level scene categories: fountains, highways, mountains, pastures, skylines, and waterfalls; (2) architectural styles: Byzantine, Gothic, Renaissance, Modern, Postmodern, and Deconstructive; (3) buildings by famous architects: Peter Eisenman, Antoni Gaudi, Frank Gehry, Michael Graves, Le Corbusier, and Frank Lloyd-Wright; (4) faces: six non-famous individuals varied in gender and race. Following the study phase, participants experienced twelve practice trials. In these trials, three images (12° x 12° of visual angle each) were presented side by side. Participants were asked to indicate which of the three images belonged to a given target category by pressing one of the keys, "1," "2," or "3." During practice, one of the three images was always drawn from the set of studied examples. The images were presented until the participant made a response, and feedback was provided by displaying the word "CORRECT" or "INCORRECT." Study exemplars were shown again halfway through practice and at the beginning of the subsequent test phase. For the 35 test trials, 24 new grayscale images from the target categories and 48 new grayscale foil images from different categories were used. Structure of the test trials was the same as practice, except that participants no longer received feedback. The entire experiment lasted approximately 30 min.

We confirmed that architecture students had higher expertise for architectural styles and buildings by famous architects in an analysis of variance (ANOVA) of average accuracy rates, using participant group as a between-subjects factor, and visual category (entry-level scene categories vs. architectural styles vs. architects vs. faces) as a within-subjects factor. Furthermore, planned comparisons between the two groups were conducted for each of the four visual categories. We also conducted the same analyses on average reaction times (RT). Results are shown in Fig. S4.

References

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

2. Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. J. Neurosci. 33(4), 1331-1336.

3. Kanwisher, N., McDermott, J., & Chun, M. (1997) The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for the Perception of Faces. J. Neurosci. 17, 4302-4311.

4. Grill-Spector, K., Kushnir,T., Hendler, T., Edelman, S,, Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum. Brain. Mapp. 6, 316-328.

5. Hinds, O. P., Rajendran, N., Polimeni, J. R., Augustinack, J. C., Wiggins, G., Wald, L. L., Rosas, H. D., Potthast, A., Schwartz, E. L.. & Fischi, B. (2008). Accurate prediction of V1 location from cortical folds in a surface coordinate system. Neuroimage, 39(4), 1585-1599.

6. Pereira, F. & Botvinick, M. (2011). Information mapping with pattern classifiers: a comparative study. Neuroimage, 56, 476-496.

7. Kriegeskorte, N., Goebel R, Bandettini, P. (2006). Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A., 103, 3863–3868.

8. Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. Neuroimage, 54(1), 313-327.

9. McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012). The Vanderbilt expertise test reveals domain-general and domain-specific sex effects in object recognition. Vision Res., 69, 10-22.

Supplemental Data

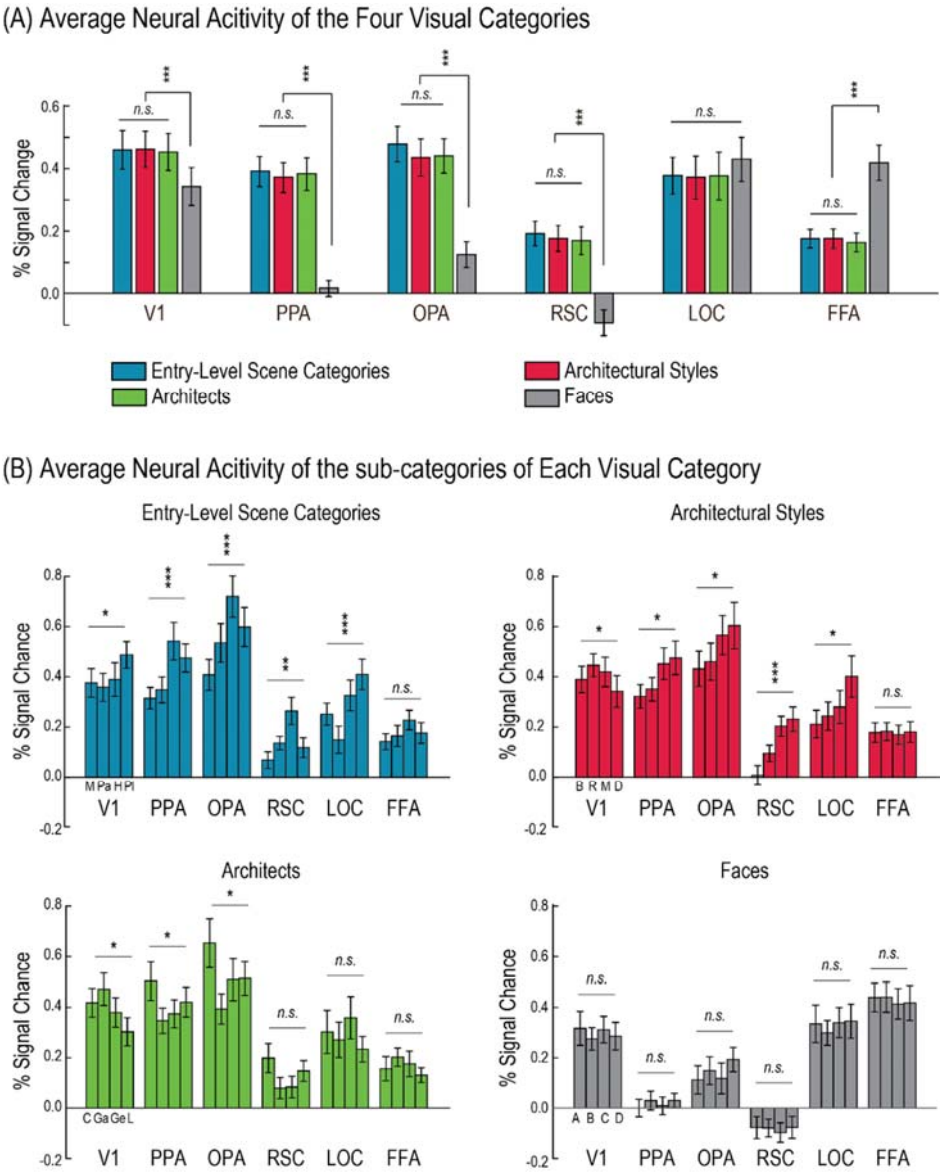

Figure S1. The effects of four visual categories and their sub-categories on group average neural activity levels: entry-level scene categories (blue), architectural styles (red), architects (green), and faces (gray). Error bars indicate standard error of the mean *(SEM)*. (A) Average neural activation for the four visual categories. As expected, the PPA, OPA, and RSC showed higher mean activation for scenes, architectural styles, and architects than faces. By contrast, the FFA showed higher activation for faces than for the other visual categories. We found no main effect of visual category in the LOC. (B) Average neural activation for subordinate categories: (M)ountains, (Pa)strues, (H)ighways, and (Pl)ayground for entry-level scene categories; (B)yzantine, (R)enaissance, (M)odern, and (D)econstructive for architectural styles; Le (C)orbusier, (Ga)udi, (Ge)hry, and (L)loyde-Wright for architects; and (A, B, C, and D) for faces of non-famous individuals. $*p < .05$, $**p < .01$, $***p < .001$. Related to Figure 1.

Table S1. Percent of ROI-voxels with significant decoding accuracy in the searchlight analysis. Numbers shown are averages over participants with *SEM*s shown in parentheses. The searchlight map for decoding entry-level scene categories showed the largest amount of overlap with all ROIs. Overlap was smaller for styles, architects and faces. The searchlight map for decoding face identity showed the largest overlap with V1. Related to Table 1.

ROI	N	Entry-level Scenes		Styles		Architects		Faces	
V1	22	13.0	(3.4)	2.9	(0.8)	3.7	(1.4)	9.4	(1.9)
PPA	22	32.9	(5.6)	12.5	(2.8)	7.3	(1.8)	5.8	(2.2)
OPA	17	33.4	(7.6)	15.4	(5.6)	7.7	(2.0)	4.2	(2.2)
RSC	22	20.2	(4.7)	6.6	(2.2)	9.2	(3.1)	5.3	(2.2)
LOC	20	27.1	(4.7)	11.3	(2.7)	8.2	(2.7)	2.2	(0.9)
FFA	21	12.5	(2.1)	6.9	(2.1)	6.4	(1,9)	4.9	(1,4)

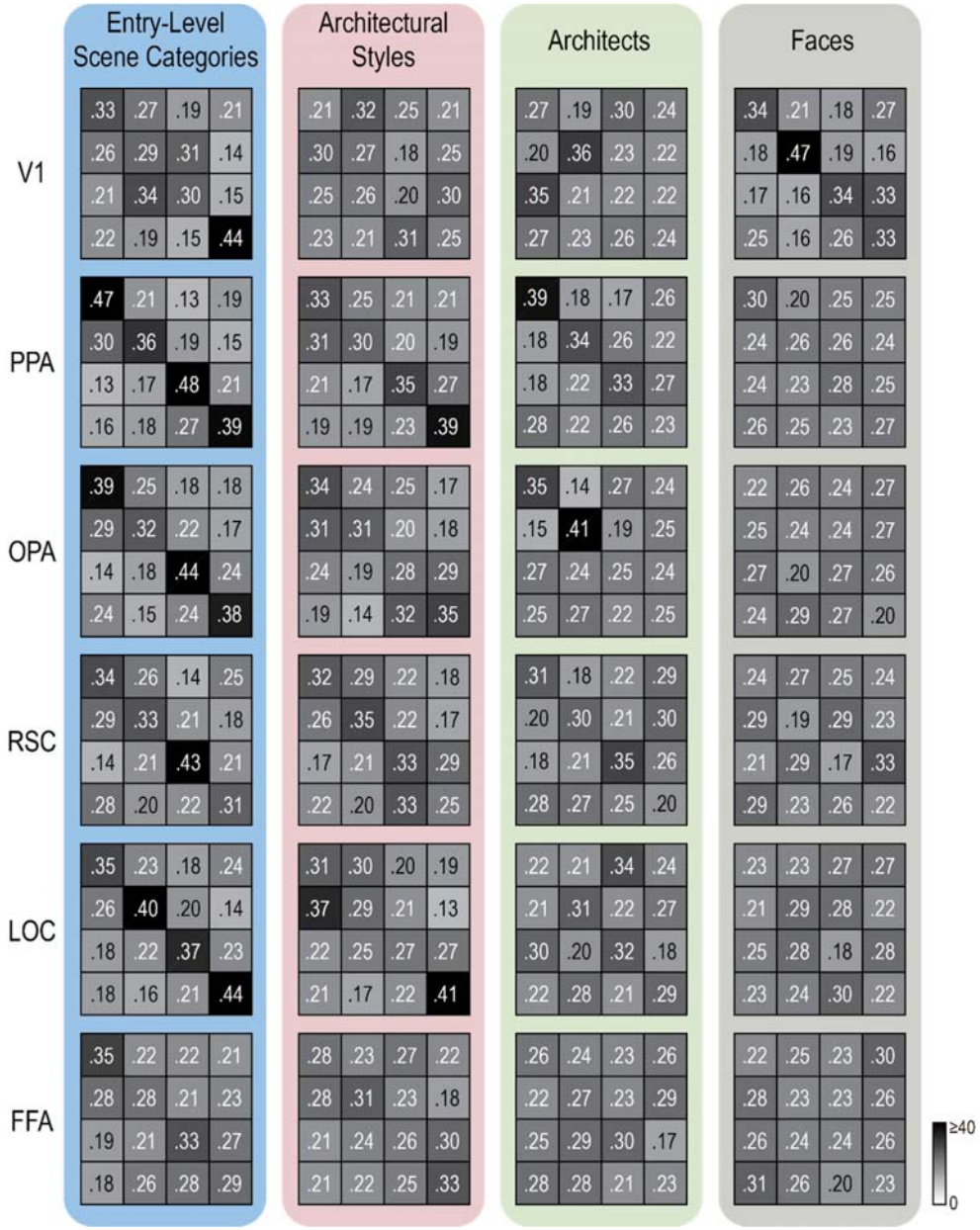

Figure S2. Group-average confusion matrices for decoding from the PPA, OPA, RSC, LOC, and FFA for the four visual categories: entry-level scene categories (blue panel), architectural styles (red panel), architects (Green panel), and faces (gray panel). For each confusion matrix, rows (*r*) indicate the ground truth of the presented category, and columns (*c*) represent predictions by the decoder. Individual cells (*r*,*c*) contain the proportion of blocks with category *r*, which were decoded as category *c*. Labels of subordinate categories (from top to bottom and left to right) are: entry-level scene categories: mountains, pastures, highways, and playgrounds; architectural styles: Byzantine, Renaissance, Modern, and Deconstructive; architects: Le Corbusier, Antoni Gaudi, Frank Gehry, and Frank Lloyd-Wright; and faces of four non-famous men, A, B, C, and D. Related to Figure 2.

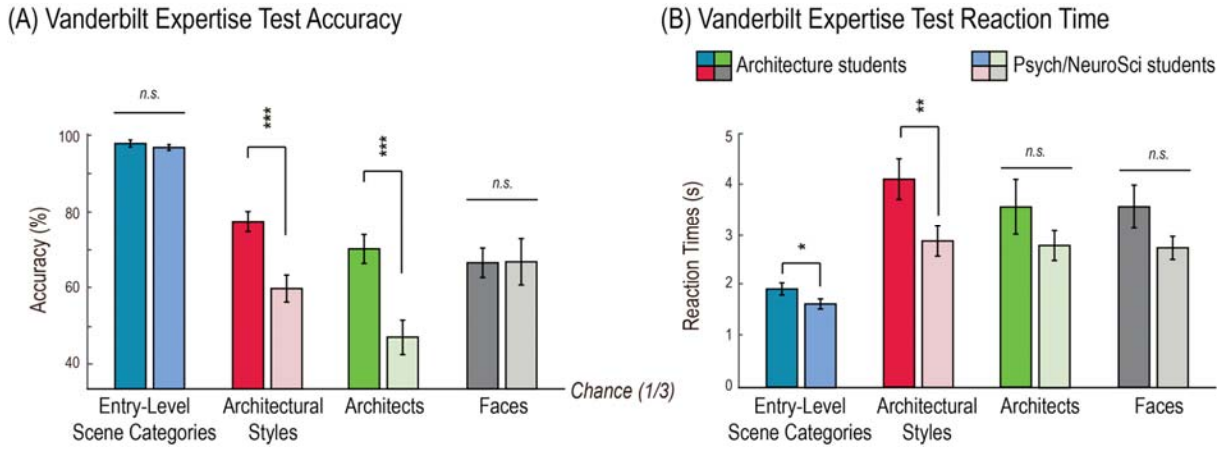

Figure S3. Group-average accuracy rates and reaction times for the four categorization tasks. Dark-colored bars indicate behavioral performances of the eleven architecture students, and bright-colored bars indicate performances of the eleven psychology and neuroscience students. Different hues indicate the four visual categories: entry-level scene categories in blue, architectural styles in red, architects in green, and face identities in gray. Error bars show *SEMs*. (A) The ANOVA for accuracy showed significant main effects of group and visual category, as well as significant interaction between the two factors, showing that the architecture students were more accurate at categorizing architectural styles and architects than psychology and neuroscience students. As expected, such group differences were not found for categorizing entry-level scene categories and face identities. (B) The same ANOVA for reaction times showed significant main effects of group and visual category, but no significant interaction between them, suggesting that architecture students were slower for all types of categorization tasks. Participants showed the fasted reaction times for entry-level scene categorization compared to the other subordinate-level categorization tasks. Planned comparisons between the two groups per visual category confirmed these findings, as their significance is shown above the bar graphs separately for each visual category. *$p < .05$, **$p < .01$, ***$p < .001$. Related to Figure 3.