Visual search for object categories is predicted by the representational architecture of high-level visual cortex

Articles in PresS. J Neurophysiol (November 2, 2016). doi:10.1152/jn.00569.2016

Visual search for object categories is predicted by the representational architecture of high-level visual cortex

Michael A. Cohen[1*], George A. Alvarez[2], Ken Nakayama[2], Talia Konkle[2]

[1] McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

[2] Department of Psychology, Harvard University

*Corresponding author

Running title: Neural organization constrains visual cognition.

Corresponding author information: Michael A Cohen, 77 Massachusetts Avenue 46-4141, Cambridge MA, 02139, michaelthecohen@gmail.com

Abstract

Visual search is a ubiquitous visual behavior, and efficient search is essential for survival. Different cognitive models have explained the speed and accuracy of search based either on the dynamics of attention or on similarity of item representations. Here, we examined the extent to which performance on a visual search task can be predicted from the stable representational architecture of the visual system, independent of attentional dynamics. Participants performed a visual search task with 28 conditions reflecting different pairs of categories (e.g., searching for a face amongst cars, body amongst hammers, etc.). The time it took participants to find the target item varied as a function of category combination. In a separate group of participants, we measured the neural responses to these object categories when items were presented in isolation. Using representational similarity analysis, we then examined whether the similarity of neural responses across different subdivisions of the visual system had the requisite structure needed to predict visual search performance. Overall, we found strong brain/behavior correlations across most of the higher-level visual system, including both the ventral and dorsal pathways when considering both macro-scale sectors as well as smaller meso-scale regions. These results suggest that visual search for real-world object categories is well predicted by the stable, task-independent architecture of the visual system.

New and noteworthy

Here, we ask which neural regions have neural response patterns that correlate with behavioral performance in a visual processing task. We found that the representational structure across all of high-level visual cortex has the requisite structure to predict behavior. Furthermore, when directly comparing different neural regions, we found that they all had highly similar category-level representational structures. These results point to a ubiquitous and uniform representational structure in high-level visual cortex underlying visual object processing.

Introduction

We spend a lot of time looking for things: From searching for our car in a parking lot to finding a cup of coffee on a cluttered desk, we continuously carry out one search task after another. However, not all visual search tasks are created equal: sometimes what we are looking for jumps out at us, while other times we have difficulty finding something that is right in front of our eyes. What makes some search tasks easy and others difficult?

Most prominent cognitive models of visual search focus on factors that determine how attention is deployed, such as bottom-up stimulus saliency or top-down attentional guidance (Itti and Koch, 2000; Nakayama and Martini, 2011; Treisman and Gelade, 1980; Wolfe and Horowitz, 2004). Accordingly, the majority of neural models of visual search primarily focus on the mechanisms supporting attention in the parietal and frontal cortices (Kastner and Ungerleider, 2000; Treue, 2003; Buschman and Miller, 2007; Eimer, 2014) or examine how attention dynamically alters neural representations within occipitotemporal cortex when there are multiple items present in the display (Chelazzi et al., 1993; Luck et al., 1997; Reddy & Kanwisher, 2007; Peelen and Kastner, 2011; Seidl et al., 2012). Other cognitive work that does not focus on attention has explored how representational factors contribute to visual search speeds, such as target-distractor similarity (Duncan and Humphreys, 1989). These representational factors might reflect competition between stable neural representations (Scalf et al., 2013; Cohen, et al., 2014; Cohen et al., 2015;), rather than the flexibility of attentional operations. Here, we extend this representational framework by relating visual search for real-world object categories to the representational architecture of the visual system.

Prior work relating perceptual similarity and neural similarity can provide some insight into the likely relationship between visual search and the architecture of the visual system. In particular, a number of previous studies have explored how explicit judgments of the similarity between items are correlated with measures of neural similarity in various regions across the ventral stream (Edelman et al., 1998; Haushofer et al., 2008; Op de Beeck et al., 2008; Peelen & Caramazza, 2012; Carlson et al., 2014; Bracci et al., 2015; Bracci & Op de Beeck, 2016; Jozwik et al., 2016). For example, Bracci and Op de Beeck (2016) asked how the similarity structures found within a variety of regions across the visual system reflected shape/perceptual similarity, category similarity, or some combination. This study, along with several others, found a transition from perceptual/shape representations in posterior visual regions (Edelman et al., 1998; Op de Beeck et al., 2008) to more semantic/category representations in anterior visual regions (Connolly et al., 2012; Bracci et al., 2015; Bracci and Op de Beeck, 2016; Prokolova et al., 2016), with both kinds of structure coexisting independently in many of the regions. Thus, to the extent that visual search is correlated with perceptual and/or semantic similarity, then we are likely to find that visual search correlates with at least some of regions of the visual system.

Expanding on this previous work, here we explore an extensive set of neural subdivisions of the visual system at different spatial scales to determine which neural substrates have the requisite representational structure to predict visual search behavior for real-world object categories. The broadest division of the visual system we considered is between the

ventral and dorsal visual pathways, canonically involved in processing “what” and “where/how” information about objects, respectively (Mishkin and Ungerleider, 1982; Goodale and Milner, 1992). This classic characterization predicts a relationship between search behavior and representational similarity in ventral stream responses, but not dorsal stream responses. Consistent with this assumption, most prior work examining neural and perceptual similarity has not explored the dorsal stream. Further, the ventral visual hierarchy progresses from representing relatively simple features in early areas to increasingly complex shape and category information in higher-level visual areas (DiCarlo et al., 2012; Bracci et al., 2016). This hierarchy predicts that representational similarity for higher-level object information will be reflected in the neural representations of higher-level visual areas. However, this relationship may be complicated by the fact that the ventral stream contains regions that respond selectively to some object categories (e.g., faces, bodies, and scenes; Kanwisher, 2010). Given these distinctive regions, it is possible that the visual search speeds between different pairs of categories can only be predicted by pooling neural responses across large sectors of the ventral stream that encompass multiple category-selective sub-regions.

To examine the relationships between visual search and neural representation, we first measured the speed of visual search behavior between high-level object categories (e.g., the time it takes to find a car amongst faces, or a body amongst hammers). Given the combinatorial considerations, we purposefully selected 8 object categories that spanned major known representational factors of the visual system (e.g., animacy, size, faces, bodies, and scenes; Kanwisher, 2010; Huth et al., 2012, Konkle and Caramazza, 2013), yielding 28 possible object category pairs. We then recruited a new group of participants and measured neural responses using functional neuroimaging (fMRI) while they viewed images of these object categories presented in isolation while performing a simple vigilance task. By obtaining neural response patterns to each category independently, this design enables us to test the hypothesis that competition between pairs of categories in visual search is correlated with the stable representational architecture of the visual system. Finally, we used a representational similarity approach to relate behavioral measures to a variety of subdivisions of the visual system (Kriegeskorte et al., 2008a).

In addition to comparing visual search performance with neural measures, we had a third group of observers perform similarity ratings, explicit judging the similarity between each pair of categories. These data enabled us to examine whether visual search shows correlations with neural similarity above and beyond what we would expect from explicit similarity judgments. If so, it would suggest that the similarity that implicitly limits visual search is at least partially distinct from the similarity that is captured by explicit similarity ratings.

Materials and Methods

Participants: Sixteen observers (ages 18-35, 9 females) participated in the visual search behavioral experiments. Another group of sixteen observers (ages 23-34, 10 females) participated in the similarity ratings experiment. Six participants (ages 20-34, 3 females), including author M.A.C., participated in the neuroimaging experiment. Both of these sample

sizes were determined based on pilot studies estimating the reliability of the behavioral and neural measures. All participants had normal or corrected-to-normal vision. All participants gave informed consent according to procedures approved by the Institutional Review Board at Harvard University.

Stimuli: Eight stimulus categories were used for both the behavioral and fMRI experiments: bodies, buildings, cars, cats, chairs, faces, hammers, and phones. Thirty exemplars within each category were selected so that there would be high within-category diversity (e.g. images of items taken from different angles, in different positions, etc.) (**Figure 1a**). In addition, all images were matched for average luminance, contrast, and spectral energy across the entire image using the SHINE toolbox (Willenbockel et al., 2010). The selection and processing of stimuli was done to minimize the extent to which visual search could depend on low-level features that differentiate the stimulus categories.

Behavioral experiment materials: The experiment was run on a 24-inch Apple iMac computer (1920 x 1200 pixels, 60 Hz) created and controlled with MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). Participants sat approximately 57 cm away from the display, where 1 cm on the display corresponds to 1° of visual angle.

Visual search paradigm: Participants performed a visual search task in which a target from one category had to be detected amongst distractors from another (e.g., a face amongst cars). On each trial, 8 images were presented in a circular arrangement, 11.5° from a central fixation point (**Figure 1b**). The same 8 locations were used throughout the experiment. Each image was viewed through a circular aperture (radius = ~3.25°) with Gaussian blurred edges.

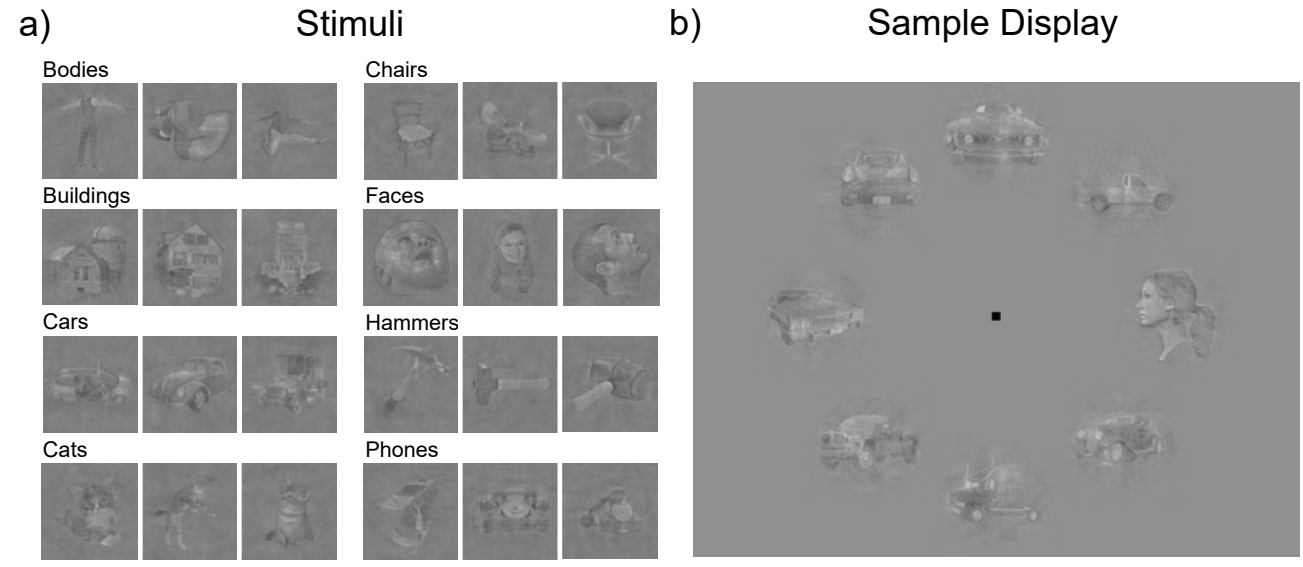


Figure 1. a) Examples of stimuli from each of the 8 categories. b) Sample display of a target present trial from the visual search task. On this trial, one target item (e.g., a face) was shown amongst seven distracting items from another category (e.g., cars).

The visual search task was comprised of eight blocks. Each block had 10 practice trials and 112 experimental trials and corresponded to a particular target category (e.g., “In this block, look for a face”). The order in which the blocks were presented was counterbalanced across subjects using a balanced Latin square design (Williams, 1949). Within each block, a target was shown on half the trials at a randomly chosen location, and participants reported whether an item from the target category was present or absent. Responses were given via button presses on the keyboard with visual feedback given immediately. Critically, on each trial, the distractors were chosen from only one of the remaining categories (e.g., a face amongst seven distractor cars). Thus, each target-present display tested a particular category pairing. Across every block, each of the seven distracting categories appeared equally often. For each participant, trials in which the response times were less than 300ms or greater than three standard deviations from the mean were excluded.

Neuroimaging paradigm: Functional neuroimaging (fMRI) was used to obtain whole-brain neural response patterns for these object categories using a standard blocked design. Each participant completed 4 experimental runs, 1 meridian mapping run used to localize early visual cortex (V1-V3), and 2 localizer runs used for defining all other regions of interest. All data was collected with a 3T Siemens Trio scanner at the Harvard University Center for Brain Sciences. Structural data were obtained in 176 axial slices with 1x1x1 mm voxel resolution, TR=2200 ms. Functional blood oxygenation level-dependent (BOLD) data were obtained using a gradient-echo echo-planar pulse sequence (33 axial slices parallel to the anterior commissure-posterior commissure line; 70x70 matrix; FoV=256x256 mm; 3.1x3.1x3.1 mm voxel resolution; Gap thickness = 0.62 mm; TR=2000 ms; TE=60 ms; flip angle = 90 degrees). A 32-channel phased-array head coil was used. Stimuli were generated using the Psychophysics toolbox for MATLAB and displayed with an LCD projector onto a screen in the scanner that subjects viewed via a mirror attached to the head coil.

Neuroimaging experimental runs: Experimental runs were part of a larger project within our laboratory. Thus, more categories were presented than were ultimately used for this project. While being scanned, participants viewed images from nine categories: bodies, buildings, cars, cars, chairs, faces, fish, hammers, and phones. Fish were not presented in the visual search experiment (see above) and fMRI data from fish were not analyzed for this study. Stimuli were presented in a rapid block design with each 4s block corresponding to one category. Six images were shown in each block for 667ms/item. Images were presented in isolation and participants were instructed to maintain fixation on a central cross and perform a vigilance task, pressing a button indicating when a red circle appeared around one of the images. The red circle appeared on 40% of blocks randomly on image 2, 3, 4, or 5 of that block. In each run, there were 90 total blocks with 10 blocks per category. All runs started and ended with an 6s fixation block, and further periods of fixation that could last 2, 4, or 6 seconds were interleaved between stimulus blocks, constrained so that each run totaled 492s. The order of the stimulus blocks and the sequencing and duration of the fixation periods was determined using Optseq (http://surfer.nmr.mgh.harvard.edu/optseq/).

Localizer runs: Localizer Runs: Participants performed a one-back repetition detection task with blocks of faces, bodies, scenes, objects, and scrambled objects. Stimuli in these runs were different from those in the experimental runs. Each run consisted of 10 stimulus blocks of 16s, with intervening 12s blank periods. Each category was presented twice per run, with the order of the stimulus blocks counterbalanced in a mirror reverse manner (e.g. face, body, scene, object, scrambled, scrambled, objects, scene, body, face). Within a block, each item was presented for 1s followed by a 330ms blank. Additionally, these localizer runs contained an orthogonal motion manipulation: In half of the blocks, the items were presented statically at fixation. In the remaining half of the blocks, items moved from the center of the screen towards either one of the four quadrants or along the horizontal and vertical meridians at 2.05°/s. Each category was presented in a moving and stationary block.

Meridian mapping: Meridian map runs: Participants were instructed to maintain fixation and were shown blocks of flickering black and white checkerboard wedge stimuli, oriented along either the vertical or horizontal meridian (Sereno et al., 1995; Wandell, 1999). The apex of each wedge was at fixation and the base extended to 8° in the periphery, with a width of 4.42°. The checkerboard pattern flickered at 8 Hz. The run consisted of 4 vertical meridian and 4 horizontal meridian blocks. Each stimulus block was 12 s with a 12 s intervening blank period. The orientation of the stimuli (vertical vs. horizontal) alternated from one block to the other.

Analysis procedures: All fMRI data was processed using Brain Voyager QX software (Brain Innovation, Mastricht, Netherlands). Preprocessing steps included 3D motion correction, slice scan-time correction, linear trend removal, temporal high-pass filtering (0.01 Hz cutoff), spatial smoothing (4mm FWHM Kernel), and transformation into Talairach space. Statistical analyses were based on the general linear model. All GLM analyses included box-car regressors for each stimulus block convolved with a gamma-function to approximate the idealized hemodynamic response. For each experimental protocol, separate GLMs were computed for each participant, yielding beta maps for each condition for each subject.

Defining neural sectors: Sectors were defined in each participant using the following procedure. Using the localizer runs, a set of visually active voxels was defined based on the contrast of [Faces + Bodies + Scenes + Objects + Scrambled Objects] vs. Rest (FDR<0.05, cluster threshold 150 contiguous 1x1x1 voxels) within a gray matter mask. To divide these visually-responsive voxels into sectors, the early visual sector included all active voxels within V1, V2, and V3, which were defined by hand on an inflated surface representation based on the horizontal vs. vertical contrasts of the meridian mapping experiment. The occipitotemporal and occipitoparietal sectors were then defined as all remaining active voxels (outside of early visual), where the division between the dorsal and ventral streams was drawn by hand in each participant based on anatomical landmarks and the spatial profile of active voxels along the surface. Finally, the occipitotemporal sector was divided into ventral and lateral sectors by hand using anatomical landmarks, specifically the occipitotemporal sulcus.

To define category selective regions, we computed standard contrasts for face selectivity (faces>[bodies scenes objects]), scene selectivity (scenes>[bodies faces objects]), and body

selectivity (bodies>[faces scenes objects]) based on independent localizer runs. For object-selective areas, the contrast of objects>scrambled was used. Category selective regions included FFA (faces), PPA (scenes), EBA (bodies), and LO (objects). In each participant, face-, scene-, body-, and object-selective regions were defined using a semi-automated procedure that selects all significant voxels within a 9 mm radius spherical ROI around the weighted center of category selective clusters (Peelen and Downing, 2005), where the cluster is selected based on proximity to the typical anatomical location of each region based on a meta analysis. This is a standard procedure, yielding appropriately sized category-specific ROIs. All ROIs for all participants were verified by eye and adjusted if necessary. Any voxels that fell in more than one ROI were manually inspected and assigned to one particular ROI, ensuring that there was no overlap between these ROIs.

Representational similarity analysis: To explore the relationship between behavioral and neural measures, we used representational similarity analysis to compute a series of brain/behavior correlations targeting different neural subdivisions within the visual hierarchy. This required constructing a "representational geometry" from both behavioral and neural measures that could then be directly compared to one another (Kriegeskorte and Kievit, 2013). A representational geometry is the set of pairwise similarity measures across a given set of stimuli. To construct our behavioral representational geometry, our similarity measure was the average reaction time it took to find a target from one category amongst distractors from another. For each category pairing, we averaged across which category was the target (i.e., the face/hammer pairing reflects trials in which observers searched for a face amongst hammers and a hammer amongst faces) to get one reaction time value per pairing. To construct our neural representational geometry, our similarity measure was the correlation across all voxels in a particular neural region between all possible category pairings (i.e., the correlation between faces and cars in ventral occipitotemporal cortex). These geometries were then correlated with one another to produce brain/behavior correlations in all neural regions.

Statistical analysis of brain/behavior correlations: The statistical significance of these correlations was assessed using both (i) group-level permutation analyses and (ii) linear mixed-effects (LME) analyses. The first method is standard, but reflects fixed effects of both behavioral and neural measures, which is well suited for certain cases (e.g., in animal studies where the number of participants is small, e.g. Kriegeskorte et al., 2008b). The second method uses a linear mixed-effects model to estimate the brain/behavior correlation with random effects across both behavioral and neural measures (Barr et al., 2013; Winter, 2013). This analysis takes into account the mixed nature of our design (i.e., that there are within-subject behavioral measures, within-subject neural measures, and between-subject comparisons). Unlike the group-level analysis, these mixed-effects models enable us to determine whether brain/behavior correlations generalize across both behavioral and neural participants.

For the permutation analyses, the condition labels of the data of each individual fMRI and behavioral participant were shuffled and then averaged together to make new, group-level similarity matrices. The correlation between these matrices was computed and Fisher z-transformed, and this procedure was repeated 10,000 times, resulting in a distribution of

correlation values. A given correlation was considered significant if it fell within the top 5% of values in this distribution. For visualization purposes, all figures show the relationship between the group-average neural measures and group-average behavioral measures, with statistics from the group-level permutation analysis.

For linear mixed effects modeling, we modeled the Fisher z-transformed correlation values between all possible fMRI and behavioral participants as a function of neural region. This included random effects analyses of neuroimaging and behavioral participants on both the slope term and intercept of the model, which was the maximal random-effects structure justified by the current design (Barr et al., 2013). All modeling was implemented using the R packages languageR (Baayen, 2009) and lme4 (Bates and Maechler, 2009). To determine if correlations were statistically significant, likelihood ratio tests were performed in which we compared a model with a given brain region as a fixed-effect to another model without it, but that was otherwise identical. P-values were estimated using a normal approximation of the t-statistic and a correlation was considered significant if $p<0.05$ (Barr et al., 2013).

Reliability analysis. The reliability of each group-level neural geometry was estimated using the following procedure: For each participant, the data were divided into odd and even runs, and two independent neural geometries were estimated. Group-level geometries were calculated by averaging across participants, yielding separate odd and even-run representational similarity matrices. These matrices were correlated to estimate the split-half reliability, and then adjusted using the Spearman-Brown formula to estimate the reliability of the full data set (Spearman, 1910; Brown, 1910). This procedure was carried out separately for each neural region.

To determine whether differences in the strength of brain/behavior correlations across neural regions can be explained by differences in reliability, we adjusted the brain behavior correlations. To do this, we used the correction for attenuation formula: the observed brain/behavior correlation from a given neural region divided by the square root of the reliability of that region (Nunnally, 1994).

Searchlight analysis. To examine the relationship between neural structure and behavioral performance within and beyond our selected large-scale sectors, we conducted a searchlight analysis (Kriegeskorte et al., 2006). For each subject, we iteratively moved a sphere of voxels (3 voxel radius) across all locations within a subject-specific gray matter mask. At each location, we measured the response pattern elicited by each of the eight stimulus categories. Responses to each category were correlated with one another, allowing us to measure the representational geometry at each searchlight location. That structure was then correlated with the behavioral measurements, resulting in an *r*-value for each sphere at every location.

One-item search task: The same 16 behavioral participants also completed a one-item search task. Since there was only one item on the display at a time, whatever effects we find cannot be due to competition between items or limitation of attention and must be due to the similarity of a target template held in memory and the item being considered on the display. On

every trial, participants were shown one item on the screen and had to say whether or not that item belonged to a particular category (i.e., “Yes or no, is this item a face?”). The item was viewed through a circular aperture (radius ~3.25°) with Gaussian blurred edges, and positioned with some random jitter so that the image fell within ~7° from fixation.

The task was comprised of eight blocks that contained 10 practice and 126 experimental trials. Each block corresponded to a particular object category (e.g. “In this block, report whether the item is a face”). The order in which the blocks were presented was counterbalanced across subjects using a balanced Latin square design (Williams, 1949). Within each block, the target item was present on half of the trials, and participants reported whether the item belonged to the target category with a yes/no button response. Critically, when the item was not from the target category, it was from one of the remaining seven distractor categories. For the purposes of this study, the reaction to reject a distractor was taken as our primary dependent measure. For example, the average time to reject a car as not-a-face was taken as a measure of car/face similarity. For each participant, trials in which the response times were less than 300ms or greater than three standard deviations from the mean were excluded.

Explicit similarity ratings task: A new group of 16 participants who did not perform the previous search tasks were recruited to provide similarity ratings on the eight stimulus categories. To familiarize participants with the full stimulus set, they were presented with Apple Keynote slides that had each stimulus from a given category on a particular slide (i.e., “Here is one slide with all 30 of the faces” etc.). Participants were allowed to study them for as long as they wanted before beginning the task. We then showed participants pairs of category names (e.g., “Faces and Cars”) and instructed them to “Please rate these categories according to their similarity on a 0-100 scale with 0 indicating that these categories are not at all similar and 100 indicated that these categories are highly similar.” Participants then rated the similarity of every category pairing. The order in which they were cued to rate a particular category pairing was randomized across participants, and participants were allowed to review and change the ratings they provided at any time. Additionally, the stimulus slides remained available throughout the experiment in case they wanted to re-examine the images.

Partial-Correlation Analysis: In addition to computing standard brain/behavior correlations, we also computed several partial correlations in which we measure the relationship between a given behavioral task (e.g., visual search) and a neural region (e.g., ventral occipitotemporal cortex) while factoring out the contribution of another behavioral task (e.g., the similarity ratings task). The formula for determining the partial correlation between two variables (X and Y), while controlling for a third variable (z) is:

$$\frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

Animacy category-model analysis: To determine the extent to which our behavioral and neural measures were driven by the distinction between the animate categories (i.e., faces,

bodies, and cats) and inanimate categories (i.e., buildings, cars, chairs, hammers, and phone), we constructed an animacy category-model following the procedures of Khaligh-Razavi et al., 2014. In this case, when one category was animate (e.g., faces) and another category was inanimate (e.g., buildings), that category pairing was designated as a 0. When both categories were animate (e.g., faces and bodies), that category pairing was designated as a 1. When both categories were inanimate (e.g., cars and chairs), that category pairing was designated as a 2. To measure the strengths of the correlations between this model and our behavioral and neural measures, we computed Spearman's rank correlation coefficient. Statistical significance of these correlations was determined by group-level permutation analyses.

Results

Visual Search Task

Observers completed a visual search task, and their reaction time to find a target from one category amongst distractors from another category was measured for all 28 combinations of our 8 selected categories (**Figure 1a**). We examined reaction times for all target-present trials in which the participant responded correctly (with accuracies at 92% or higher across participants). As anticipated, some combinations of categories were faster than others, ranging from reaction times of 874ms to 1,352ms ($F(1,27)=14.85$, $p<0.001$) (**Figure 2**). Seven of the nine fastest conditions contained a pairing with the face category, consistent with previous results showing an advantage for faces in visual search (Crouzet et al., 2010). However, even excluding faces, the remaining category pairings showed significant differences, ranging from 1,019ms to 1,352ms ($F(1,20)=7.71$, $p<0.001$). Thus, there was significant variability in search times for different pairs of categories, and the full set of 28 reaction times were used to construct a behavioral representational similarity matrix. The key question of interest is whether representational similarity within different areas of the visual system can predict this graded relationship of visual search reaction times across pairs of categories.

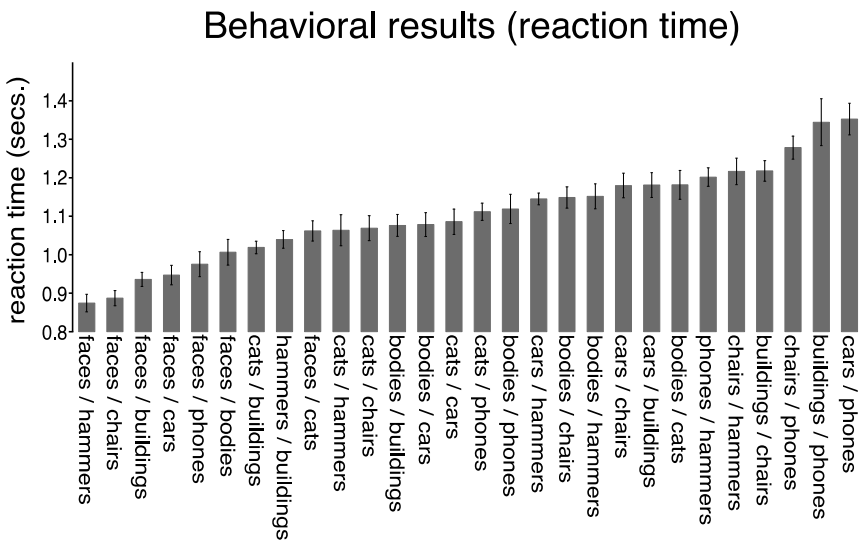


Figure 2. Reaction time results from correct target present responses for all possible category pairings. Reaction times are plotted on the y-axis, with each category pairing plotted on the x-axis. Error bars reflect within-subject standard error of the mean. The actual reaction time values are reported in the Appendix.

We made an a priori decision to focus on correct target-present responses because target-absent trials can depend on observer-specific stopping rules and decision processes (Chun and Wolfe, 1996). However, we found that target-absent trials showed a similar overall representational geometry: the correlation between target-absent and target-present reaction times was high (r=0.92, p<0.001).

Major Visual System Subdivisions

Armed with this pattern of visual search speeds, we examined whether this similarity structure is evident across major neural divisions of the visual system. Macro-scale neural sectors were defined to capture the major divisions of the visual processing hierarchy, including early visual cortex, ventral stream, and dorsal stream regions (**Figure 3a**). In addition, we further divided the ventral pathway into ventral and lateral occipitotemporal sectors, given the mirrored representation of response selectivities across occipitotemporal cortex (Taylor and Downing, 2011; Konkle and Caramazza, 2013), and the fact that some evidence suggests that the ventral surface is more closely linked to perceptual similarity than the lateral surface (Haushofer et al., 2008).

The relationship between the neural and behavioral similarity measures at the group level is plotted for each sector in **Figure 3b**. We observed strong brain/behavior correlations across the ventral stream: ventral occipitotemporal (r=0.77*;* Permutation analysis: p<0.001; Mixed-effects model parameter estimate=0.64, t=11.81, p<0.001), lateral occipitotemporal (r=0.65; p<0.001; parameter estimate=0.49, t=6.78, p<0.001). We also observed significant correlations in the dorsal stream sector: occipitoparietal (r=0.49; p<0.01; parameter estimate=0.26, t=3.15, p<0.05). Taken together, these analyses revealed a robust relationship

between the functional architecture of each high-level division of the visual processing stream and visual search speeds.

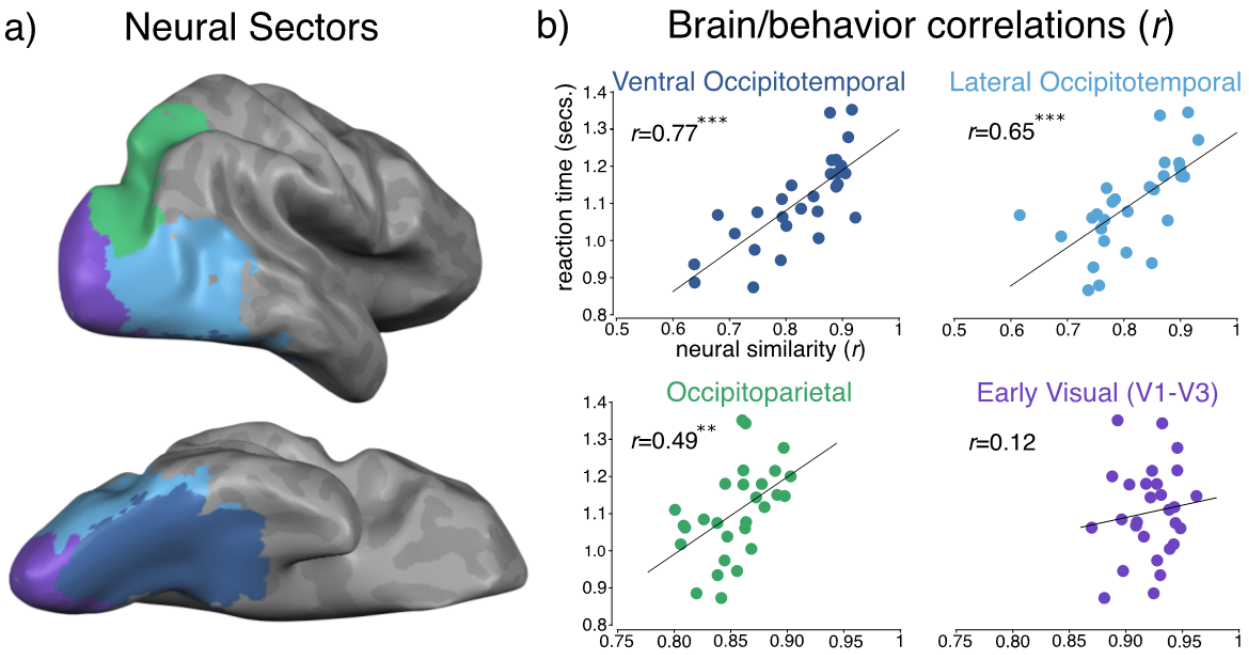


Figure 3: a) Visualization of four macro-scale sectors from a representative participant. b) Group level brain/behavior correlations in each sector. Each dot corresponds to a particular category pairing (e.g., faces and hammers). Times for each category pairing are plotted on the y-axis, while the similarity of the neural responses for those category pairings are plotted on the x-axis. Note the change in scales of the x-axes between the occipitotemporal sectors and the occipitoparietal/early visual sectors. **p<0.01, ***p<0.001.

In contrast, the correlation in early visual cortex (V1-V3) was not significant (r=0.12; Permutation analysis: p=0.26; parameter estimate=0.09, t=0.97, p=0.33). This low correlation was expected, given that the object exemplars in our stimulus set were purposefully selected to have high retinotopic variation and were matched for average luminance, contrast, and spectral energy. These steps were taken to isolate a category-level representation, which is reflected in the fact that all pairs of object categories had very similar neural patterns in early visual cortex (range r=0.87—0.96).

Comparing the brain/behavior correlations across sectors, the ventral occipitotemporal cortex showed the strongest relationship to behavior relative to all other sectors (LME parameter estimates <-0.15, t<-2.03, P<0.05 in all cases), consistent with Haushofer (2008). Further, the correlation in lateral occipitotemporal cortex was greater than those in occipitoparietal and early visual cortex (parameter estimates <-0.23, t<-3.74, P<0.001 in both cases), and the correlation in occipitoparietal cortex was marginally greater than that in early visual cortex (parameter estimate=-0.16, t=-1.82, P=0.06).

However, comparisons between brain regions should be interpreted with caution, as differences in the strength of the brain/behavior correlation could be driven by noisier neural

measures in some regions. Overall, the neural sectors had high reliability, ranging from r=0.56 in occipitoparietal cortex to r=0.95 ventral occipitotemporal cortex (**Figure 4a**). After adjusting for neural reliability, the differences in brain/behavior correlation across ventral and parietal cortex were reduced, while the correlation between early visual cortex and behavior remained low (**Figure 4b**). Thus, we did not observe strong evidence for major dissociations amongst the ventral and dorsal stream sectors in their relationship to these behavioral data.

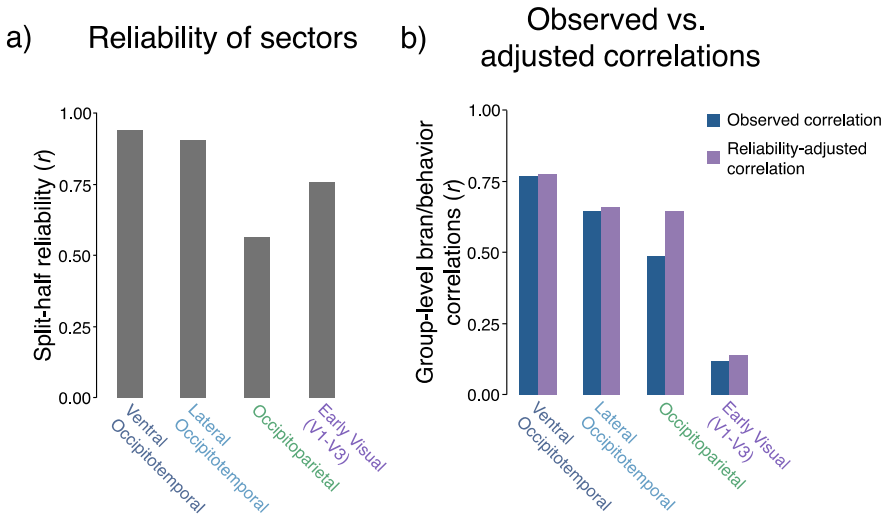


Figure 4. a) Group-level split-half reliability of the representational structures of the macro-scale sectors. The average reliability estimates (y-axis) for the group-level fMRI data is plotted for each of eight neural regions (x-axis). b) Group-level brain/behavior correlations within the macro-scale sectors. The blue bars represent the observed brain/behavior correlations in these regions. Purple bars represent the adjusted group-level correlations after correcting for attenuation using the split-half reliability of the neural regions.

Searchlight Analysis

Does this brain/behavior relationship only hold when considering neural responses pooled across large swaths of cortex, or can such a relationship be observed at a smaller spatial scale? To address this question, we conducted a searchlight analysis (Kriegeskorte et al., 2006), in which brain/behavior correlations were computed for all 3mm-radius spherical brain regions across the cortical surface. The resulting brain/behavior correlation map is shown in **Figure 5**. These results show that the graded relationship between visual categories that was observed in search behavior is also strongly correlated with the neural similarity patterns in meso-scale regions throughout the visual system except for early visual cortex.

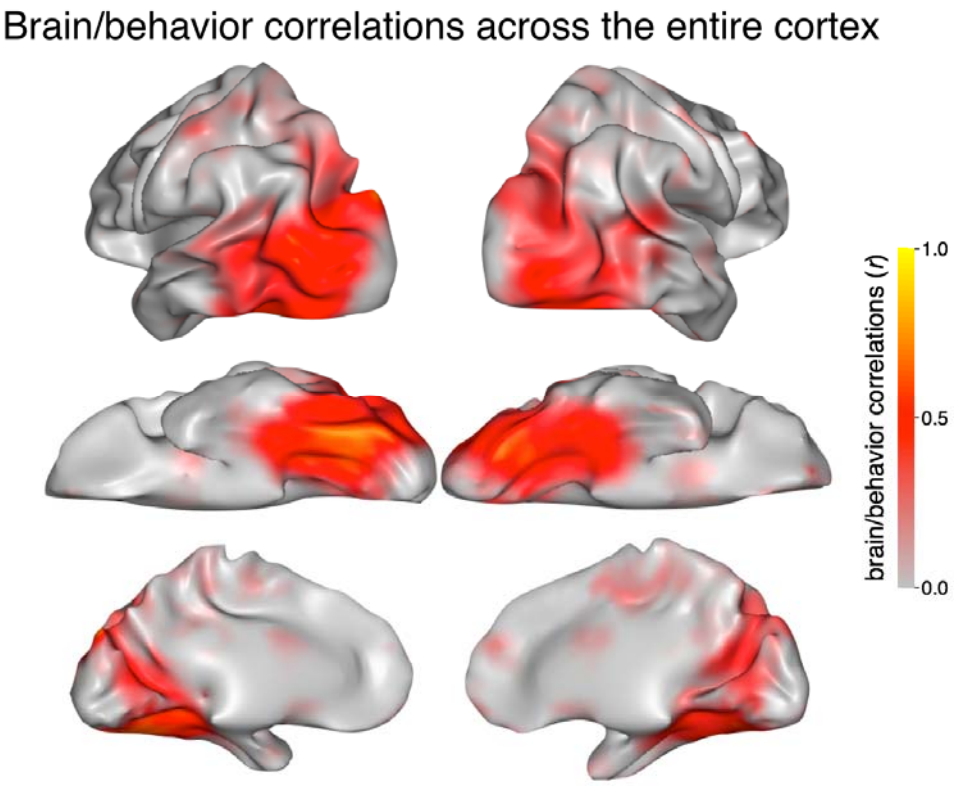


Figure 5. Searchlight analysis across the entire cortex with brain/behavior correlations computed at every location. The strength of the correlation is indicated at each location along the cortical surface.

Delving into this result, we took a targeted look at several classic regions of interest defined from a separate localizer. The fusiform face area (FFA), extrastriate body area (EBA), parahippocampal place area (PPA), and object preferential lateral occipital (LO) were defined in each individual subject. The neural similarities were then extracted and compared with the visual search data. As implied by the searchlight analysis, each of these category-selective regions had neural response patterns that correlated with visual search speeds (**Figure 6a**): FFA ($p<0.001$; parameter estimate=0.53, $t=6.09$, $p<0.001$), PPA ($p<0.001$; parameter estimate=0.52, $t=6.01$, $p<0.001$), EBA ($p<0.01$; parameter estimate=0.34, $t=4.18$, $p<0.001$), and LO ($p<0.01$; parameter estimate=0.40, $t=9.50$, $p<0.001$).

Critically, we found that this brain/behavior relationship did not depend on the category pairings that included preferred category: After removing all category pairings with faces from our analysis in FFA, a significant correlation with search speeds was still evident ($r=0.67$, $p<0.001$; parameter estimate=0.39, $t=6.87$, $p<0.001$), and also remained when excluding both faces and cats, which have faces ($r=0.61$, $p<0.001$; parameter estimate=0.32, $t=3.87$, $p<0.001$). Similarly, the brain/behavior relationship was evident in PPA when all pairings containing buildings were removed ($r=0.64$, $p<0.001$; parameter estimate=0.51, $t=7.23$, $p<0.001$), and in EBA when bodies were removed ($r=0.67$, $p<0.001$; parameter estimate=0.57, $t=4.49$, $p<0.001$) (**Figure 6b**). These results show that the preferred categories of these regions are not driving the brain/behavior correlations.

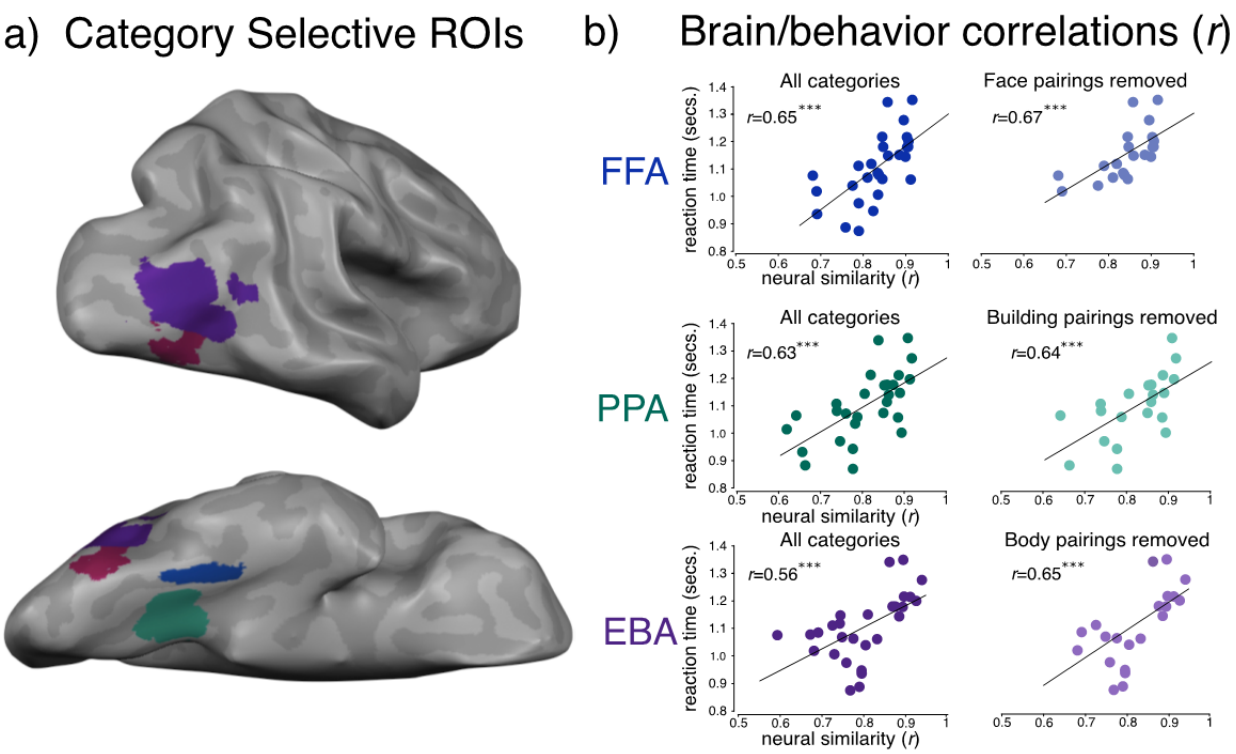


Figure 6. a) Visualization of four category selective regions from a representative participant. b) Group level brain/behavior correlations in each region. Each dot corresponds to a particular category pairing (e.g., faces and hammers). Reaction times for each pairing are plotted on the y-axis, while the similarity of the neural responses for those category pairings are plotted on the x-axis, separately for FFA (top), PPA (middle), and EBA (lower). The left panels show the brain/behavior correlation with all category pairs included. The right panels show the brain/behavior correlation with selected pairings removed. ***p<0.001.

Brain/behavior correlations across search asymmetries

Up until this point, we have measured the behavioral similarity between any two categories in a way that ignores any potential search asymmetries (e.g.,, the face and building pairing is computed as the average of search times for a face amongst buildings and a building amongst faces). The choice to collapse over target categories was made a priori based on power considerations for obtaining reliable behavioral measures, and was further supported by a post-hoc analysis showing that only 1 of the 28 pairs of categories had a significant asymmetry after Bonferonni correcting for multiple comparisons: a face was found amongst bodies faster than a body was found amongst faces (RT=900ms, SEM=56ms vs RT=1,110ms, SEM=44ms, respectively, p<0.001)**.** However, for completeness, we calculated the brain/behavioral correlation using search times separated by target category (56 points rather than 28 points). As shown in **Figure 7**, the pattern of results is the same when breaking down search times by target category.

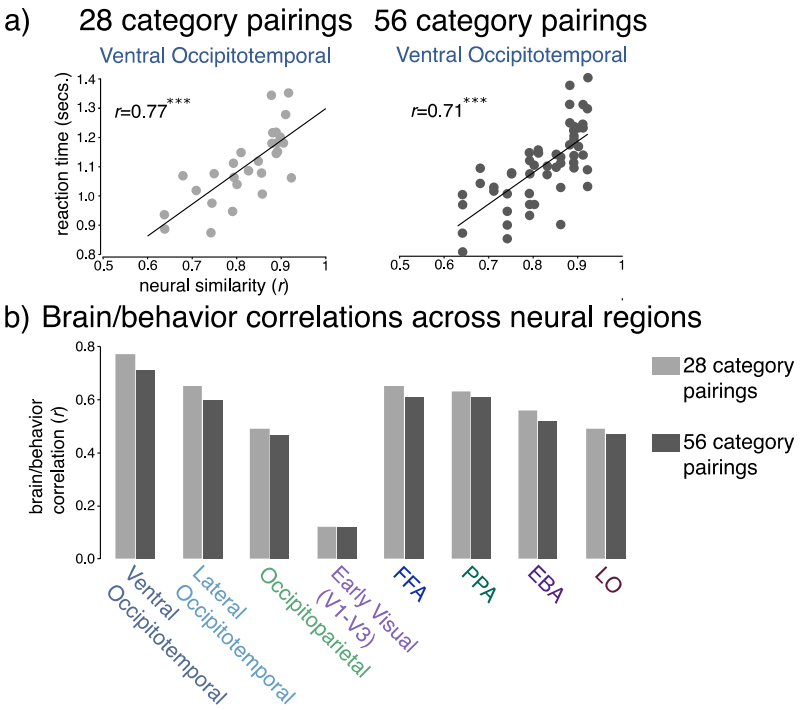


Figure 7. a) Visualization of brain/behavior correlations in ventral occipitotemporal cortex collapsed across target-category (left panel, 28 category pairings) and broken down by target-category (right panel, 56 category pairings). b) Group level brain/behavior correlations in every neural region when examining both 28 and 56 category pairings.

Brain/behavior correlations between individual participants

The figures above show group-level correlations (average behavioral reaction time vs. average neural similarity), with statistics calculated by group-level permutation analyses. We additionally performed linear-mixed effects (LME) analyses, which enable us to determine whether brain/behavior correlations generalize across both behavioral and neural participants. These LME analyses operate over data at the level of individual participants, considering each individual behavioral participant correlated with each individual neural participant (see parameter estimates above).

To visualize the full set of data at the level of individual participants, **Figure 8a** depicts all possible individual-by-individual brain/brain, behavior/behavior, and brain/behavior correlations for the large-scale sectors, category-selective ROIs, and search task. **Figure 8b** visualizes the group averages of each brain/brain and brain/behavior correlation (e.g., the average individual-by-individual correlation when correlating FFA with PPA). This figure highlights that any two individuals are relatively similar both in terms of their behavioral similarities, their neural similarities, and critically in how one individual's behavioral similarity predicts another individual's neural similarity structure. Thus, even when considering data at the individual level, there are robust brain/behavior correlations evident in all high-level visual regions beyond early visual cortex. These results support the conclusion that there is a stable, task-independent

architecture of the visual system that is common across individuals and that is strongly related to visual search performance.

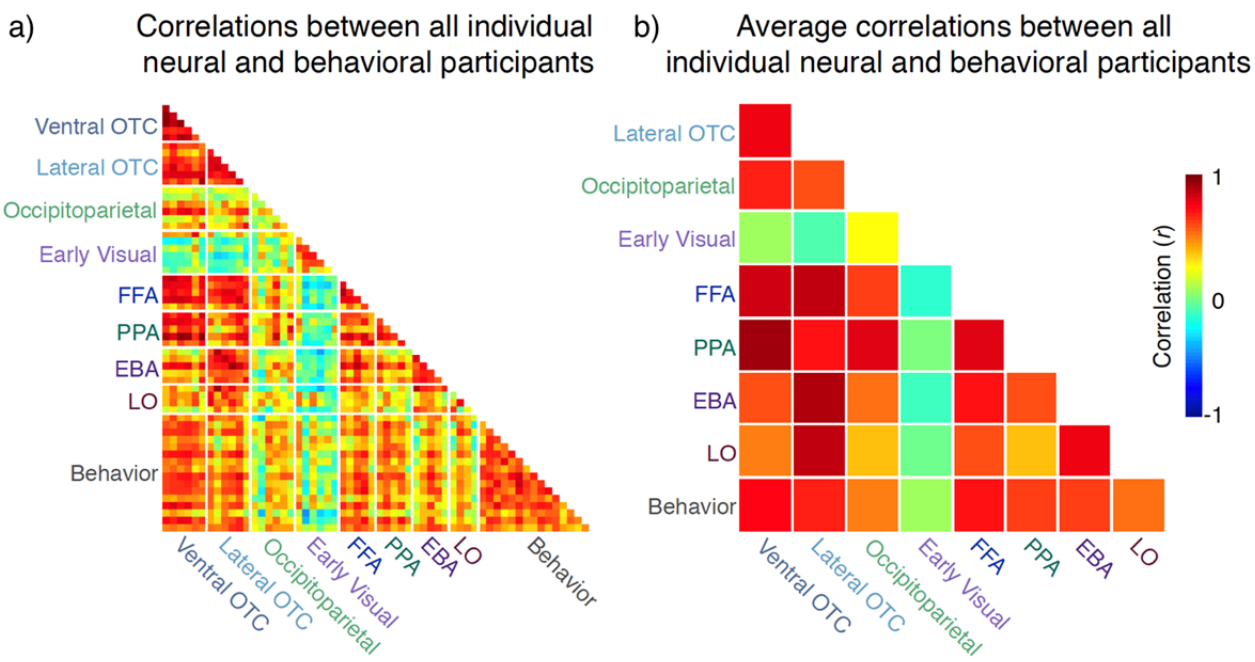


Figure 8. The left panel shows the individual-by-individual correlations between every neural (N=6) and behavioral (N=16) participant. Each cell corresponds to the correlation between two different participants. Instances in which there are only 5 rows or columns for a given region reflect instan3ces when that region was not localizable in one of the participants. The right panel shows the average correlation across all pairs of participants for a particular comparison. For example, the cell corresponding to ventral occipitotemporal and lateral occipitotemporal cortex is the average of all possible individual-by-individual correlations between those two regions. The cell corresponding to ventral occipitotemporal cortex and behavior is the average of 96 individual-by-individual correlations.

One-Item Search Task

It is possible that representational similarity is a factor in visual search because the search displays have both object categories present and there is competition between the simultaneously presented categories (Cohen et al., 2014). However, visual search also requires comparing a target template held in memory to incoming perceptual information. To the extent that the target template matches the evoked responses when seeing pictures (Harrison and Tong, 2009; Serences et al., 2009), then the neural representational similarity will also limit this template-matching process.

To test this possibility, we conducted a second behavioral study where only one item was present in the display and the task was to rapidly categorize the item (e.g., “Is this a face?”), with a yes/no response (**Figure 9a**). Here, the critical trials were when the target category was absent and we measured the time it takes to reject a distractor category (e.g., a body as not-a-face, or a hammer as not-a-house) for all 28 possible pairs of categories.

As with the visual search task, the categorization task yielded significant differences in the time to reject a distractor as a function of what target category they were looking for ($F(1,27)=6.42$, $p<0.001$; **Figure 9b**). This pattern of response times was highly correlated with the visual search task with eight items ($r=0.69$; adjusted for reliability $r=0.84$), and with the neural responses of the higher-level visual system as well (ventral occipitotemporal: $p<0.001$; Mixed-effects model parameter estimate=0.45, $t=6.77$, $p<0.001$; lateral occipitotemporal $p<0.01$; parameter estimate=0.32, $t=4.67$, $p<0.001$; occipitoparietal cortex: $p<0.001$; parameter estimate=0.25, $t=5.95$, $p<0.001$) (**Figure 10**). These results demonstrate that representational similarity is a factor in search, even when isolating the template-matching process.

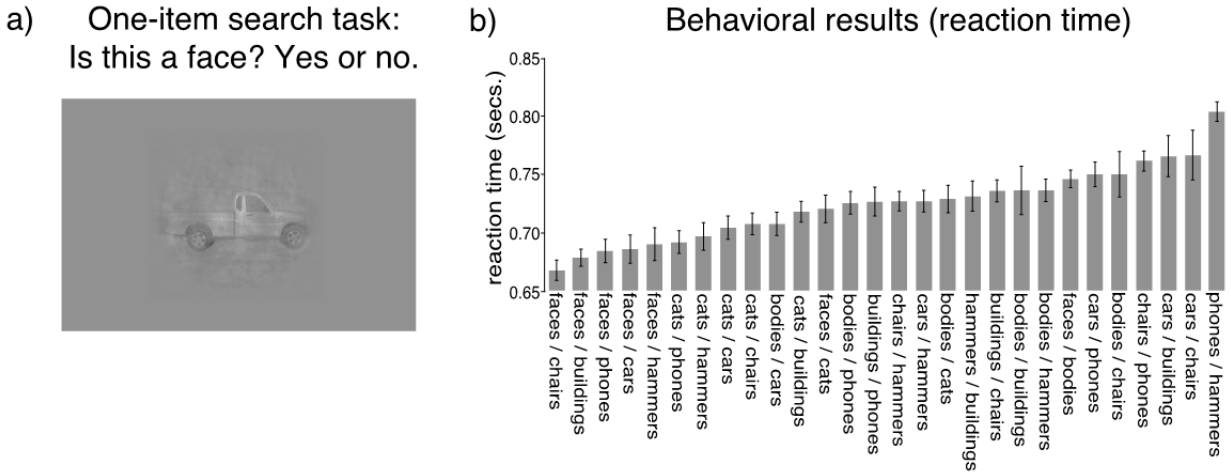


Figure 9. a) Sample display of a target absent trial from the categorization task in which a participant is looking for a face. b) Reaction time results for the one-item search task from correct target absent responses for all possible category pairings. Reaction times are plotted on the y-axis, with each category pairing plotted on the x-axis. Error bars reflect within-subject standard error of the mean.

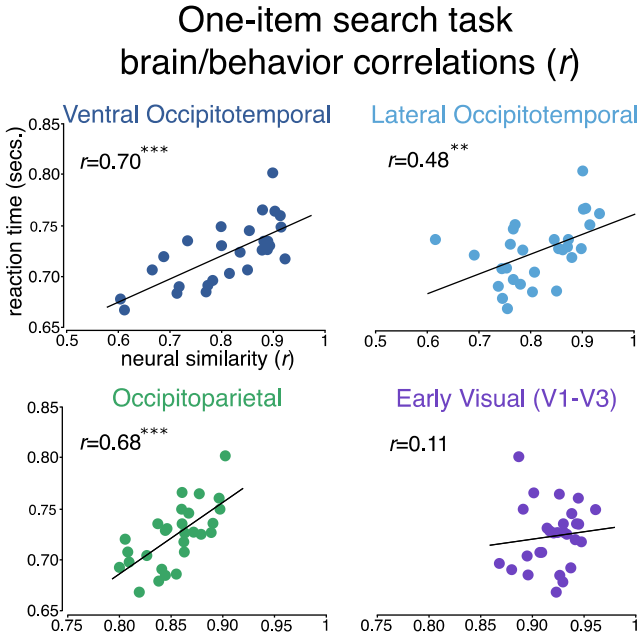


Figure 10. Group level brain/behavior correlations in each macro-scale sector. Each dot corresponds to a particular category pairing (e.g., faces and hammers). Reaction times for each category pairing are plotted on the y-axis, while the similarity of the neural responses for those category pairings are plotted on the x-axis. Note the change in scales of the x-axes between the occipitotemporal sectors and the occipitoparietal/early visual sectors. ** $p<0.01$, *** $p<0.001$.

Brain/behavior correlations and explicit similarity ratings

All of our analyses thus far have focused on visual search behavior (i.e., participants simply looked for a target category on a display and responded as fast as they could). In this case, the similarity relationship between pairs of categories is implicit in performance: although participants were never asked to judge similarity, we assume that search was slower for more similar target-distractor pairs. In contrast, many prior studies have used explicit measures of similarity in which participants directly rate the similarity of different categories, focusing on perceptual similarity, semantic similarity, or just an overall sense of similarity without more specific instructions (Edelman et al., 1998; Op de Beeck et al., 2008; Carlson et al., 2013; Mur et al., 2013; Bracci et al., 2015; Jozwik et al., 2016; Bracci and Op de Beeck, 2016). This previous work has shown that neural similarity within the ventral stream relates to explicit similarity judgments. Thus, to the extent that visual search and explicit ratings tap into the same similarity structure, the relationship between search and neural similarity across the ventral stream is expected from previous work on explicit similarity ratings.

Is the similarity implicitly measured by visual search and the explicit similarity found in a rating task the same, and do these two behavioral similarities have the same relationship to neural representation? To investigate these questions directly, we had a new group of participants explicitly rate the similarity of our stimulus categories and then examined how those ratings relate to visual search performance and the neural responses.

Overall, we found that certain category pairings were rated as being more similar than others ($F(1,27)=9.66$, $p<0.001$; **Figure 11a**). The explicit ratings were significantly correlated with the implicit similarity of the visual search task with eight items ($r=0.44$, $p<0.01$) as well as the three higher-level neural sectors (ventral occipitotemporal: $r=0.62$, $p<0.001$; lateral occipitotemporal: $r=0.58$, $p<0.001$; occipitoparietal: $r=0.44$, $p<0.01$), but not early visual cortex ($r=0.06$, $p=0.38$). Thus, the explicit ratings task produced a relationship with the neural responses that was qualitatively similar to the visual search task (**Figure 11b**).

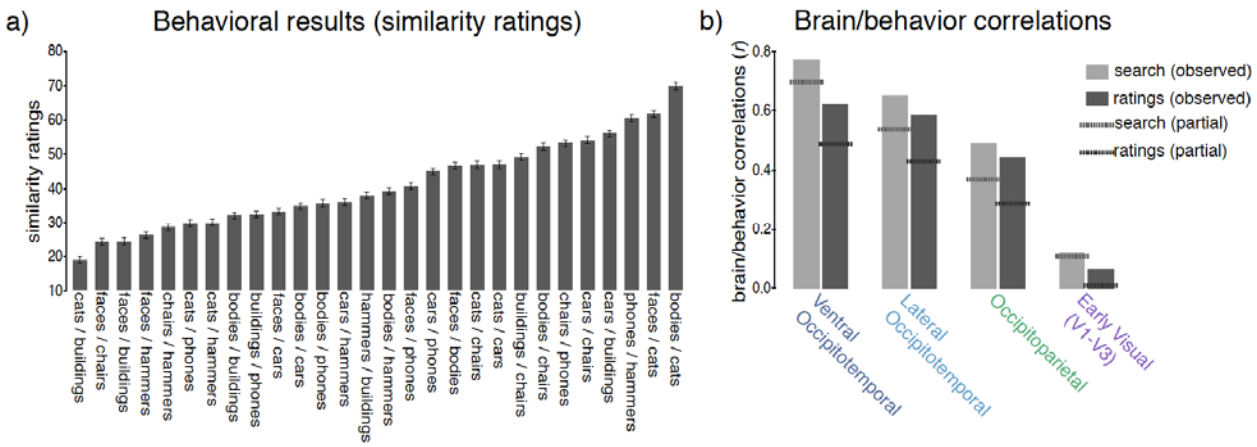


Figure 11. a) Explicit similarity ratings for all category pairings. Average similarity rating is plotted on the y-axis as a function of category pairing along on the x-axis. Error bars reflect within-subject standard error of the mean. b) Group level brain/behavior correlations in each macro-scale sector for both the search task and the rating task. The solid bars indicate the observed correlations, while the dashed lines indicate the partial correlations for each behavioral task after factoring out the other (e.g., "search (partial)" indicates the correlations between the search task and the different neural regions after factoring out the rating task).

Do explicit similarity judgments fully explain the link between visual search behavior and neural similarity? The search and rating tasks were correlated with each other, but this correlation was only moderate (r=0.44) and thus it is possible that these two tasks correlate with the neural results in unique ways and account for different portions of the variance. To directly examine this possibility, we computed a series of partial correlations in which one behavioral task was factored out from the other behavioral task and then correlated with the neural sectors, and visa versa. If the two tasks account for the same neural variance, the brain/behavior correlations should drop to near zero when one of the tasks is factored out. Contrary to this prediction, however, we found that the partial correlations with both tasks remained high in occipitotemporal cortex even after factoring out the other task (search factoring out ratings: ventral occipitotemporal partial correlation=0.70, p<0.001; lateral occipitotemporal partial correlation=0.54 p<0.01; ratings factoring out search: ventral occipitotemporal partial correlation=0.49, p<0.05; lateral occipitotemporal partial correlation=0.43 p<0.05; **Figure 11b** dashed lines). Meanwhile, the partial correlations in both occipitoparietal and early visual cortex were not significant with either task (p>0.05 in all cases). Taken together, these results indicate that the visual search behavior predicts some unique variance in the neural geometry and is not fully explained by an explicit similarity task.

Brain/behavior correlations and the animate/inanimate distinction

We next examined the extent to which the brain/behavior correlations we observed can be explained by the dimension of animacy. The distinction between animate and inanimate objects is one of the most replicated and well-documented distinctions in both the behavioral and cognitive neuroscience literature (Spelke et al., 1995; Caramazza and Shelton, 1998; Kuhlmeier et al., 2004; Martin, 2007; Mahon and Caramazza, 2009), and is currently the biggest

explanatory factor in the representational geometry of high-level categories (e.g. Kriegeskorte et al., 2008; Huth et al, 2012; Konkle & Caramazza et al., 2013). Are the present results solely driven by this distinction?

To examine this issue, we first asked how well a category-model based on animacy could explain the behavioral and neural data. Comparing this model to the behavioral data, we found strong correlations between the animacy model and visual search performance (ρ=0.67, p<0.001) and the one-item search task (ρ=0.68, p<0.001). Comparing this model to the neural data, we also observed a systematic relationship with the neural responses that was similar to those in the visual search task with eight items (ventral occipitotemporal: ρ=0.76, p<0.001; lateral occipitotemporal: ρ=0.74, p<0.001; occipitoparietal: ρ=0.51, p<0.01; early visual: ρ=-0.02, p=0.4778). Together, these results suggest that a significant portion of the results reported thus far is due to the animacy distinction.

Critically, if animacy were the only factor driving these brain/behavior correlations, then we would not expect to see brain/behavior correlations when only examining the within-animate or within-inanimate category pairings. If such correlations are observed, it would suggest that the animacy factor is not solely driving our findings, and that more fine-grained differences in neural similarity are mirrored by fine-grained differences in search performance. To test this possibility, we computed the brain/behavior correlations when only examining the inanimate categories (10 pairwise comparisons). We did not compute correlations with the animate categories because there were only 3 animate categories (yielding only 3 pairwise comparisons), preventing us from having enough power to compute reliable correlations. Within the inanimate categories, we found strong correlations in both ventral and lateral occipitotemporal cortex (ventral occipitotemporal: r=0.72, p<0.001; lateral occipitotemporal: r=0.66, p<0.001) but not in early visual cortex (r=0.05, p=0.46). Interestingly, the correlation in occipitoparietal cortex dropped substantially (from r=0.49; p<0.01 to r=0.19, p<0.3031), which suggests that object representations in the dorsal stream are not as fined grained as in the ventral stream and are driven more by the animacy distinction. More broadly, the fact that we observe these brain/behavior correlations for inanimate categories in the ventral stream suggests that the animacy distinction is not the only organizing principle driving search performance or the correlation between search and neural responses.

Discussion

Visual search takes time, and that time depends on what you are searching for and what you are looking at. Here, we examined whether visual search for real-world object categories can be predicted by the stable representational architecture of the visual system, looking extensively across the major divisions of the visual system at both macro- and meso-scales (e.g., category-selective regions). Initially, we hypothesized that the representational structure within the ventral stream, but not within early visual areas or the dorsal stream, would correlate with visual search behavior. Contrary to this prediction, however, we found strong correlations between visual search speeds and neural similarity throughout all of high-level visual cortex. We also initially predicted that this relationship with behavior would only be observed when pooling

over large-scale sectors of the ventral stream, but found that it was present within all meso-scale regions that showed reliable visual responses in high-level visual cortex.

We conducted a number of further analyses that revealed both the robustness of this relationship between visual search and neural geometry and the novel contributions of these findings. First, these brain/behavior correlations held in category-selective ROIs even when the preferred category was excluded from the analysis. Second, the correlations were evident even when examining individual participants, and are not just aggregate group-level relationships. Third, the brain/behavior correlations cannot be fully explained by either explicit similarity ratings or by the well-known animacy distinction, which also helps situate this work with respect to previous findings. Finally, a simplified visual search task with only one item on the screen also showed a strong brain/behavior relationship, suggesting that search for real-world objects is highly constrained by competition during the template-matching stage of the task. Taken together, these results suggest that there is a stable, widespread representational architecture that predicts and likely constrains visual search behavior for real-world object categories.

Scope of the Present Results

We made two key design choices that impact the scope of our conclusions: (1) focusing on representations at the category level, and (2) using a relatively small number of categories. First, we targeted basic-level categories because of their particular behavioral relevance (Rosch et al., 1976; Mervis and Rosch, 1981) and because objects from these categories are known to elicit reliably different patterns across visual cortex. Consequently, we cannot make any claims about item-level behavior/brain relationships. Second, we selected a small number of categories to ensure that we obtained highly reliable behavioral data (split-half reliability: r=0.88) and neural data (split-half reliability range: r=0.56 to r=0.95 across sectors). It therefore remains an empirical question whether this relationship holds for a wider set of categories, and addressing this question will require large-scale behavioral studies to measure the relationship between more categories. However, we note that a brain/behavior relationship was observed in category-selective regions even when comparing the relationship amongst non-preferred categories (e.g., excluding faces and cats from FFA). This result indicates that this relationship is not driven by specialized processing of specific object categories, suggesting that it will potentially hold for a broader set of categories.

Implications for the Role of Attention

Given what is known about how attention dynamically alters representations in high-level cortex (David et al., 2008; Çukur et al, 2013), one might not have expected to find the high brain/behavior correlations we observed here. For example, previous work has demonstrated that when attention is focused on a particular target object, target processing is enhanced and distractor processing is suppressed (Seidl et al., 2012; Desimone and Duncan, 1998). In general, some evidence suggests that the representational space across the entire brain is warped around the attended object category (Çukur et al., 2013). If these attentional operations dramatically distorted the representational geometry, they would likely change the relationship

between target and distractor neural representations during the search task and a measurement of the stable architecture would not predict visual search reaction times. Contrary to this idea, our data show a clear relationship between neural architecture and visual search reaction times: the neural responses of one group of participants viewing isolated objects without performing a search task predicted most of the variance in search reaction times in a separate group of participants. Taken together, these observations imply that attention alters neural responses in a way that preserves the stable architecture without dramatically changing its geometry. However, direct measures of neural representational geometries under conditions of category-based attention must be obtained and related to behavior to investigate this directly.

Another possibility, not mutually exclusive, is that our task design and stimulus control happened to minimize the role of attentional guidance. This is supported by two main observations. First, the overall reaction times in the main search task were quite slow, indicating a relatively inefficient guidance of attentional allocation to potential targets. This was true even for the fastest category combinations (e.g., when faces were presented among non-face objects or visa versa, **Figure 2**). Previous work has shown that faces can "pop-out" amongst non-face objects (Herschler & Hochstein, 2005), and this pop-out can be removed by matching stimuli on a variety of low-level features (e.g., luminance, spectral energy, etc; VanRullen, 2006), as we did in our stimulus set. If attentional guidance to likely target locations mostly operates at these more basic levels of representation (Wolfe & Horowitz, 2004), then matching stimuli on a variety of these low-level features would minimize the role of this process. Second, we found that the eight-item search task was highly correlated with the one-item search task, which suggests that our behavioral data primarily or exclusively reflects the template matching process. Given these observations, it is possible that only the template-matching process is constrained by the representational architecture of the visual system.

One integrative possibility is that search behavior is limited by multiple bottlenecks that potentially vary as a function of the stimuli and task demands. On such an account, search difficulty hinges on a) how well attention is deployed to likely targets in a display and b) how well a template-matching process is carried out as a function of the stable architecture of the visual system (i.e., what feature representations are distinguished naturally and which ones are not). In cases when lower level features (e.g., color, orientation, etc.) distinguish the target from the distractors, attentional allocation likely plays a larger role than template-matching. Conversely, when simple visual features do not easily distinguish items from one another, as is likely the case with the stimuli used in this study, the template-matching process plays a larger role.

Implications for Representational Similarity

Duncan and Humphreys (1989) proposed what has become a prominent cognitive model of visual search that emphasizes representational factors over attentional factors. Simply put, this model states that search becomes less efficient as target/distractor similarity increases and more efficient as target/distractor similarity decreases. Of course, one of the deeper challenges hidden in this proposal is what is meant by the word "similarity." In their initial studies, Duncan and Humphreys used carefully controlled stimuli so that they could manipulate

target/distractor similarity themselves. With more naturalistic stimuli, such as those used in the present experiments, it is unclear how the similarity between different categories would be determined. Here, we add to their broad framework by demonstrating that target-distractor similarity relationships for high-level visual stimuli can be estimated from the representational architecture of the high-level visual system. Mechanistically, we suggest that it is this very architecture that places constraints on the speed of visual processing.

Of course, the fact that neural measures can be used to predict search behavior also does not fully answer the question about what we mean by "similarity." For example, does the similarity structure driving these results reflect perceptual properties, more semantic properties, or some combination of both? A more complete picture would relate visual search behavior with a computational model that explain these visual search speeds and make accurate predictions for new combinations of categories (e.g. see approaches by Yu et al., 2016). While we do not have that explanatory model, we took some steps to explore this avenue by relating visual search behavior to the dimension of animacy and to explicit similarity ratings.

These analyses yielded one expected and one unexpected result. As expected, the animate/inanimate distinction was a major predictive factor for both neural patterns and behavioral search speeds. However, brain/behavior relationships were found even when only considering pairs of inanimate categories, which do not span the animate/inanimate boundary. This analysis indicates that animacy is not the only representational factor driving the present results. Unexpectedly, we found that explicit similarity ratings were only moderately correlated with visual search speeds, and the two measures largely accounted for different variance in the neural data. Why might this be the case? Explicit similarity judgments are susceptible to the effects of context, frequency, task instructions, and other factors (Tversky, 1977). Thus, one possibility is that these judgments rely on a more complex set of representational similarity relationships than those indexed implicitly by visual search behavior. In this case, it is possible that when doing a visual search task, or when trying to probe ones intuitions about the similarity of two categories, both of these tasks draw on the representational architecture of the high-level visual system, but not in exactly the same way.

Implications for the Dorsal Stream

Another unexpected result was the relationship between visual search and neural responses along the dorsal stream. While this brain/behavior correlation in the parietal lobe is surprising from the perspective of the classic ventral/dorsal, what/where distinction, the current results add to mounting evidence that object category information is also found in the dorsal stream (Konen et al., 2008; Romero et al., 2014; Cohen et al., 2015). There are at least two possible accounts for this neural structure.

On one account, the parietal lobe may be doing the same kind of representational work as the ventral stream. For example, the searchlight analysis reveals that posterior occipitoparietal cortex has the strongest relationship with search. Given our increasing understanding of nearby regions like the occipital place area (Dilks et al., 2013), it could be that

this posterior occipitoparietal cortex has a role that is more closely related to the ventral stream (Bettencourt and Xu, 2013). However, the dorsal stream representations were nearly fully explained by the animacy distinction unlike the ventral stream sectors, pointing to a potential dissociation in their roles.

An alternative possibility draws on recent empirical work demonstrating that certain dorsal stream areas have very specific high-level object information that is largely based on task-relevance (Jeong and Xu, 2016). In the present experiment, participants were only doing a red-frame detection task while in the scanner, but given the simplicity of this task, the dorsal stream responses might naturally have a passive reflection of the object category information in ventral stream. While our data do not distinguish between these alternatives, they clearly show that the posterior aspects of the parietal lobe contain object category information that predicts visual search behavior.

Conclusion

Overall, these results suggest that the stable representational architecture of object categories in the high-level visual system is closely linked to performance on a visual search task. More broadly, the present results fit with previous work showing similar architectural constraints on other behavioral tasks such as visual working memory and visual awareness (Cohen et al., 2014; Cohen et al., 2015). Taken together, this body of research suggests that the responses across higher-level visual cortex reflect a stable architecture of object representation that is a primary bottleneck for many visual behaviors (e.g., categorization, visual search, working memory, visual awareness, etc.).

Appendix

Reaction time results for the eight-item search task are reported in the table. The first column corresponds to the reaction time when the first category listed is the target (i.e., a body target amongst building distractors), while the second column corresponds to the reaction time when the second category is the target (i.e., a building target amongst body distractors). The third column corresponds to the reaction time when averaging across all trials for each category pairing regardless of which category is the target. Note that the numbers in column three are not the average of the numbers in columns one and two since there are sometimes different numbers of trials in the different conditions because of differences in accuracies or reaction time filtering.

Category Pairings	1st Category as Target	2nd Category as Target	Averaged Across Both Targets
Bodies & Buildings	1.079	1.074	1.076
Bodies & Cars	1.028	1.132	1.078
Bodies & Cats	1.223	1.137	1.182
Bodies & Chairs	1.146	1.154	1.149
Bodies & Faces	1.111	0.900	1.007
Bodies & Hammers	1.192	1.112	1.152
Bodies & Phones	1.141	1.097	1.119
Buildings & Cars	1.138	1.232	1.181
Buildings & Cats	1.027	1.015	1.019
Buildings & Chairs	1.237	1.191	1.218
Buildings & Faces	1.004	0.870	0.936
Buildings & Hammers	1.104	0.969	1.040
Buildings & Phones	1.378	1.312	1.344
Cars & Cats	1.100	1.071	1.086
Cars & Chairs	1.187	1.173	1.180
Cars & Faces	0.968	0.932	0.947
Cars & Hammers	1.204	1.094	1.145
Cars & Phones	1.405	1.297	1.352
Cats & Chairs	1.094	1.041	1.069
Cats & Faces	1.089	1.031	1.062
Cats & Hammers	1.120	1.007	1.063
Cats & Phones	1.146	1.074	1.112
Chairs & Faces	0.968	0.806	0.887
Chairs & Hammers	1.249	1.183	1.217
Chairs & Phones	1.312	1.245	1.278
Faces & Hammers	0.851	0.899	0.874
Faces & Phones	0.945	1.006	0.975
Hammers & Phones	1.169	1.231	1.202

Acknowledgements

Thanks to Nancy Kanwisher and Nikolaus Kriegeskote for extensive discussions. Thanks to Alfonso Caramazza, James DiCarlo Rebecca Saxe, Daniel Schacter, Maryam Vaziri-Pashkam, and Daniel Yamins for helpful discussions. This work was supported by an NSF-GRFP and NIH-NRSA (F32EY024483, MAC), NSF CAREER (BCS-0953730, GAA), NIH NEI RO1 (EY01362, KN), and NIH-NRSA (F32EY022863, TK).

References

Baayen, RH. languageR: Data sets and function with “Analyzing linguistic Data: A practical introduction to statistics.” R package version 0.955, 2009.

Barr, DJ, Levy R. Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Mem Lang* 68: 255-278, 2013.

Bates DM, Maechler M. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32, 2009.

Bettencourt K, Xu Y. The role of transverse occipital sulcus in scene perception and its relationship to object individuation in inferior intra-parietal sulcus. *J Cogn Neurosci* 25: 1711-1722, 2013.

Bracci S, Caramazza A, Peelen MV. Representational similarity of body parts in human occipototemporal cortex. *J Neurosci*, 35: 12977-12985, 2015.

Bracci S, Op de Beeck H. Dissociations and associations between shape and category representations in the two visual pathways. *J Neurosci* 36: 432-444, 2016.

Brainard, DH. The psychophysics toolbox. *Spat Vis* 10: 433-436, 1997.

Brown W. Some experimental results in the correlation of mental abilities. *Brit J Psychol* 3: 296-322, 1910.

Buschman TJ, Miller EK. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315: 1860-1862. 2007.

Caramazza A, Shelton JR. Domain specific knowledge systems in the brain: The animate-inanimate distinction. *J Cogn Neurosci* 10: 1-34, 1998.

Carlson TA, Ritchie JB, Kriegeskorte N, Durvasula S, Ma J. Reaction time for object categorization is predicted by representational distance. *J Cogn Neurosci* 26: 132-142, 2014.

Chelazzi L, Miller EK, Duncan J, Desimone R. A neural basis for visual search in inferior temporal cortex. *Nature* 363: 345-347, 1993.

Chun, MM, Wolfe JM. Just say no: How are visual search trials terminated when there is no target present? *Cogn Psych* 30: 39-78, 1996.

Cohen M, Konkle T, Rhee J, Nakayama K, Alvarez GA. Processing multiple visual objects is limited by overlap in neural channels. *Proc Natl Acad Sci USA* 111: 8955-8960, 2014.

Cohen M, Nakayama K, Konkle T, Stantic M, Alvarez GA. Visual awareness is limited by the representational architecture of the visual system. *J Cogn Neurosci* 27: 2240-2252, 2015

Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu Y-C, Abdi H, Haxby JV. The representation of biological classes in the human brain. *J Neurosci* 32: 2608-2618, 2012.

Crouzet SM, Kirchner H, Thorpe SJ. Fast saccades toward faces: face detection in just 100ms. *J Vis* 10: 1-17, 2010.

Çukur T, Nishimoto S, Huth AG, Gallant JL. Attention during natural vision warps semantic representation across the human brain. *Nat Neurosci* 16: 763-770, 2013.

David SV, Hayden BY, Mazer JA, Gallant JL. Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron* 59: 509-521, 2008.

Desimon R, Duncan J. Neural mechanism of selective visual attention. *Ann Rev Neurosci* 18: 193-222, 1998.

DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? *Neuron* 73: 415-434, 2012.

Dilks DD, Julian JB, Paunov AB, Kanwisher N. The occipital place area is causally and selectively involved in scene perception. *J Neurosci* 33: 1331-1336, 2013.

Duncan J, Humphreys GW. Visual search and stimulus similarity. *Psychol Rev* 96: 433-458, 1989.

Edelman S, Grill-Spector K, Kushnir T, Malach R. Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology* 26: 309-321, 1998.

Eimer M. The neural basis of attentional control in visual search. *Trends Cogn Sci* 18: 526-535. 2014.

Goodale MA, Milner AD. Separate visual pathways for perception and action. *Trends Neurosci* 15: 20-25, 1992.

Harrison S, Tong F. Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458: 632-635, 2009.

Haushofer J, Livingstone MS, Kanwisher N. Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PloS Biol* 6: e187, 2008.

Huth AG, Nishimoto S, Vu AT, Gallant JL. A continuous semantic space describes the representational of thousands of object and action categories across the human brain. *Neuron* 76: 1210-1224, 2012.

Jeong SK, Xu Y. Behaviorally relevant abstract object identity representation in the human parietal cortex. *J Neurosci* 36: 1607-1619, 2016.

Jozwik KM, Kriegeskorte N, Mur M. Visual features as stepping stones toward semantics: Explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia* 83: 201-226, 2016.

Itti L, Koch C. Computational modeling of visual attention. *Nat Rev Neurosci* 2: 194-203, 2001.

Kanwisher, N. Functional specificity in the human brain: a window in- to the functional architecture of the mind. *Proc Natl Acad Sci USA* 107: 11163-11170. 2010.

Kastner S, Ungerleider LG. Mechanisms of visual attention in the human cortex. *Annu Rev Neurosci* 23: 315-341, 2000.

Khaligh-Razavi S-M, Henriksson L, Kay K, Kriegeskorte N. Explaining the hierarchy of visual representational geometries by remixing of features from many computational vision models. *bioRxiv*: 009936, 2014.

Konen CS, Behrmann M, Nishimura M, Kastner S. The functional neuroanatomy of visual agnosia: A case study. *Neuron* 71: 49-60, 2008.

Konkle T, Caramazza A. Tripartite organization of the ventral stream by animacy and object size. *J Neurosci* 33: 10235-10242. 2013.

Kuhlmeier VA, Bloom P, Wynn K. Do 5-month-old infants see humans as material objects? *Cognition* 94: 109-112, 2005.

Kriegeskorte N, Goebel R, Bandettini P. Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103: 3863-3868, 2006.

Kriegeskorte N, Mur M, Bandettini P. Representational similarity analysis – connecting the branches of systems neuroscience. *Front Syst Neurosci* 2: 1-28, 2008a.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60: 1126-1141, 2008b.

Kriegeskorte, N, Kievit, R.A. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci* 17: 401-412, 2013

Luck SJ, Chelazzi L, Hillyard SA, Desimone R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol* 77: 24-42, 1997.

Mahon B, Caramazza A. Concepts and categories: A cognitive neuropsyhological perspective. *Annu Rev Psychol* 60: 27-51, 2009.

Martin A. The representation of object concepts in the brain. *Annu Rev Psychol* 58: 25-45, 2007.

Mishkin M, Ungerleider LG. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav Brain Res* 6: 57-77, 1982.

Mervis CB, Rosch E. Categorization of natural objects. *Ann Rev Psychol* 32: 89-115, 1981.

Mur M, Meys M, Bodurka J, Goebel R, Bandettini PA, Kriegeskorte N Human object-similarity judgments reflect and transcend the primate-IT object representation. Front. Psychol. 4:128, 2013

Nakayama K, Martini P. Situating visual search. *Vis Res* 51: 1526-1537.

Nunnally JC, Bernstein JH. Psychometric Theory (3rd ed.) New York: McGraw-Hill. 1994.

Op de Beeck H, Torfs K, Wagemans J. Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *J Neurosci* 28: 10111-10123, 2008.

Peelen MV, Downing PE. Within-subject reproducibility of category-specific visual activation with functional MRI. *Hum Brain Mapp* 25: 402-408, 2005.

Peelen MV, Kastner S. A neural basis for real-world visual search in human occipitotemporal cortex. *Proc Natl Acad Sci USA* 108: 12125-12130, 2011.

Peelen MV, Caramazza A. Conceptual object representations in human anterior temporal cortex. *J Neurosci* 32:15728-15736, 2012.

Pelli, DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10: 437-442, 1997.

Proklova D, Kaiser D, Peelen MV. Disentangling representations of object shape and object category in human visual cortex: the animate-inanimate distinction. *J Cogn Neurosci* 28: 680-692, 2016.

Reddy L, Kanwisher N. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Curr Biol* 17: 3067-2072, 2007.

Romero MC, Pani P, Janssen P. Coding of shape features in the macaque anterior intraparietal area. *J Neurosci* 34: 4006-4021, 2014.

Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-Braem P. Basic objects in natural category. *Cogn Psych* 8: 382-439, 1976.

Scalf PE, Torralbo A. Tapia, E., & Beck, D.M. Competition explains limited attention and perceptual resources: implications for perceptual load and dilution theories. *Front Psychol* 4: 243. 2013.

Seidl KN, Peelen MV, Kastner S. Neural evidence for distracter suppression during visual search in real-world scenes. *J Neurosci* 32: 11812-11819, 2012.

Serences JT, Ester EF, Vogel EK, Awh E. Stimulus-specific delay activity in human primary visual cortex. *Psychol Sci* 20: 207-214.

Sereno MI, Dale AM, Reppas JB, Dwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell, RB. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268: 889-893, 1995.

Spearman CC. Correlation calculated from faulty data. *Brit J Psychol* 3: 271-295, 1910.

Spelke E, Phillips S, Woodward A. Infants' knowledge of object motion and human action. In Sperber D, Premack AJ, Premack D. (eds.) *Causal Cognition: A multidisciplinary debate*, 44-77, 1995.

Taylor J, Downing P. Division of labor between lateral and ventral extrastriate representations of faces, bodies, and objects. *J Cogn Neurosci* 23: 4122-4137, 2011.

Treisman A, Gelade G. A feature-integration theory of attention. *Cogn Psychol* 12: 97-136, 1980.

Treue S. Visual attention: the where, what, how and why of saliency. *Curr Opin Neurobiol* 13: 428-432, 2003.

Tversky A. Features of similarity. *Psych Rev* 84: 327-352, 1977.

VanRullen R. On second glance: Still no high-level pop-out effect for faces. *Vis Res* 18: 3017-3027, 2006.

Wandell BA. Computational neuroimaging of human visual cortex. *Ann Rev Neurosci* 22: 145-173, 1999.

Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW. Controlling low-level image properties: The SHINE toolbox. *Behav Res Meth* 42: 671–684, 2010.

Williams EJ. Experimental designs balanced for the estimation of residual effects of treatment. *Aust J Sci Res* 2: 149-168, 1949.

Winter, B. Linear models and linear mixed effects models in R with linguistic applications. arXiv:1305.5499, 2013.

Wolfe JM, Horowitz TS. What attributes guide the deployment of visual attention and how do they do it? *Nat Rev Neurosci* 5: 495-501, 2004.

Yu CP, Maxfield J, Zelinsky G. Searching for category-consistent features: A computational approach to understanding visual categoroy representation. *Psych Sci* 27: 870-884, 2016.

Figure captions

Figure 1. a) Examples of stimuli from each of the 8 categories. b) Sample display of a target present trial from the visual search task. On this trial, one target item (e.g., a face) was shown amongst seven distracting items from another category (e.g., cars).

Figure 2. Reaction time results from correct target present responses for all possible category pairings. Reaction times are plotted on the y-axis, with each category pairing plotted on the x-axis. Error bars reflect within-subject standard error of the mean. The actual reaction time values are reported in the Appendix.

Figure 3: a) Visualization of four macro-scale sectors from a representative participant. b) Group level brain/behavior correlations in each sector. Each dot corresponds to a particular category pairing (e.g., faces and hammers). Times for each category pairing are plotted on the y-axis, while the similarity of the neural responses for those category pairings are plotted on the x-axis. Note the change in scales of the x-axes between the occipitotemporal sectors and the occipitoparietal/early visual sectors. ** $p<0.01$, *** $p<0.001$.

Figure 4. a) Group-level split-half reliability of the representational structures of the macro-scale sectors. The average reliability estimates (y-axis) for the group-level fMRI data is plotted for each of eight neural regions (x-axis). b) Group-level brain/behavior correlations within the macro-scale sectors. The blue bars represent the observed brain/behavior correlations in these regions. Purple bars represent the adjusted group-level correlations after correcting for attenuation using the split-half reliability of the neural regions.

Figure 5. Searchlight analysis across the entire cortex with brain/behavior correlations computed at every location. The strength of the correlation is indicated at each location along the cortical surface.

Figure 6. a) Visualization of four category selective regions from a representative participant. b) Group level brain/behavior correlations in each region. Each dot corresponds to a particular category pairing (e.g., faces and hammers). Reaction times for each pairing are plotted on the y-axis, while the similarity of the neural responses for those category pairings are plotted on the x-axis, separately for FFA (top), PPA

(middle), and EBA (lower). The left panels show the brain/behavior correlation with all category pairs included. The right panels show the brain/behavior correlation with selected pairings removed. ***p<0.001

Figure 7. a) Visualization of brain/behavior correlations in ventral occipitotemporal cortex collapsed across target-category (left panel, 28 category pairings) and broken down by target-category (right panel, 56 category pairings). b) Group level brain/behavior correlations in every neural region when examining both 28 and 56 category pairings.

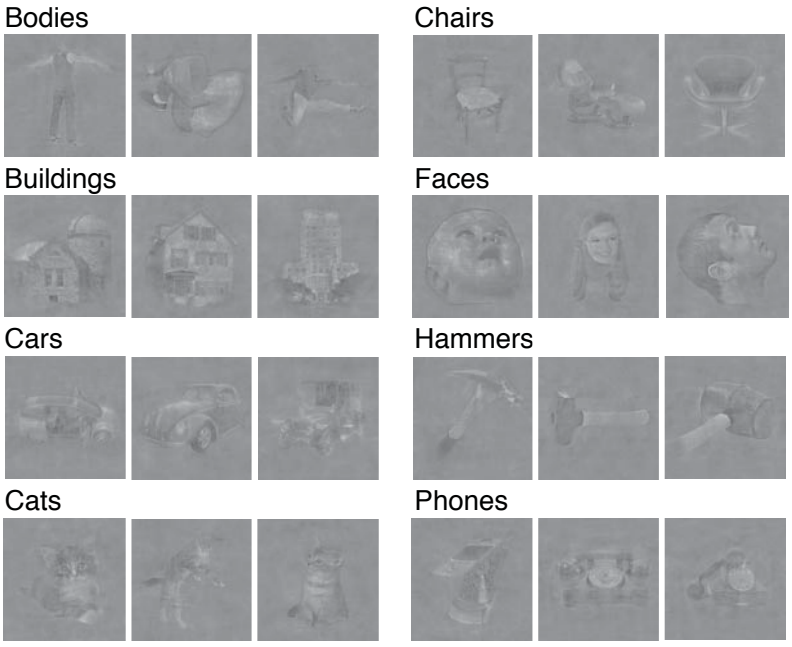
Figure 8. The left panel shows the individual-by-individual correlations between every neural (N=6) and behavioral (N=16) participant. Each cell corresponds to the correlation between two different participants. Instances in which there are only 5 rows or columns for a given region reflect instan3ces when that region was not localizable in one of the participants. The right panel shows the average correlation across all pairs of participants for a particular comparison. For example, the cell corresponding to ventral occipitotemporal and lateral occipitotemporal cortex is the average of all possible individual-by-individual correlations between those two regions. The cell corresponding to ventral occipitotemporal cortex and behavior is the average of 96 individual-by-individual correlations.

Figure 9. a) Sample display of a target absent trial from the categorization task in which a participant is looking for a face. b) Reaction time results for the one-item search task from correct target absent responses for all possible category pairings. Reaction times are plotted on the y-axis, with each category pairing plotted on the x-axis. Error bars reflect within-subject standard error of the mean.

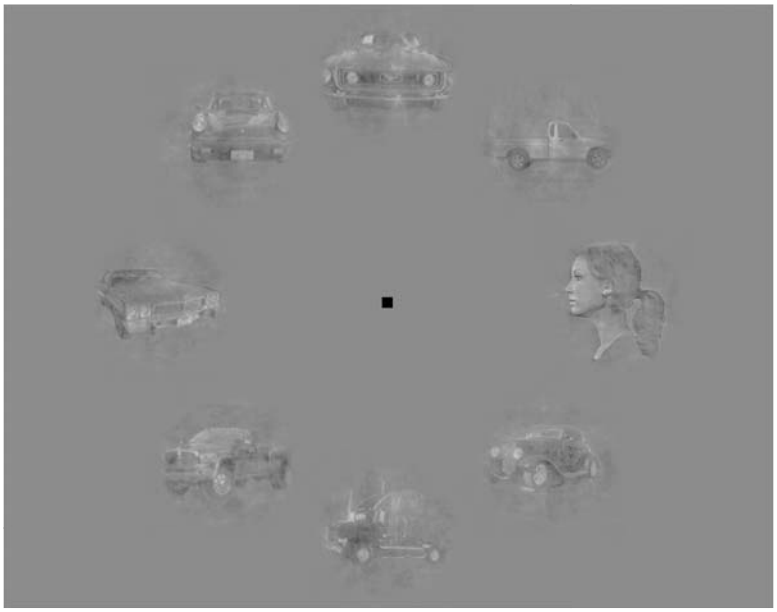
Figure 10. Group level brain/behavior correlations in each macro-scale sector. Each dot corresponds to a particular category pairing (e.g., faces and hammers). Reaction times for each category pairing are plotted on the y-axis, while the similarity of the neural responses for those category pairings are plotted on the x-axis. Note the change in scales of the x-axes between the occipitotemporal sectors and the occipitoparietal/early visual sectors. **p<0.01, ***p<0.001.

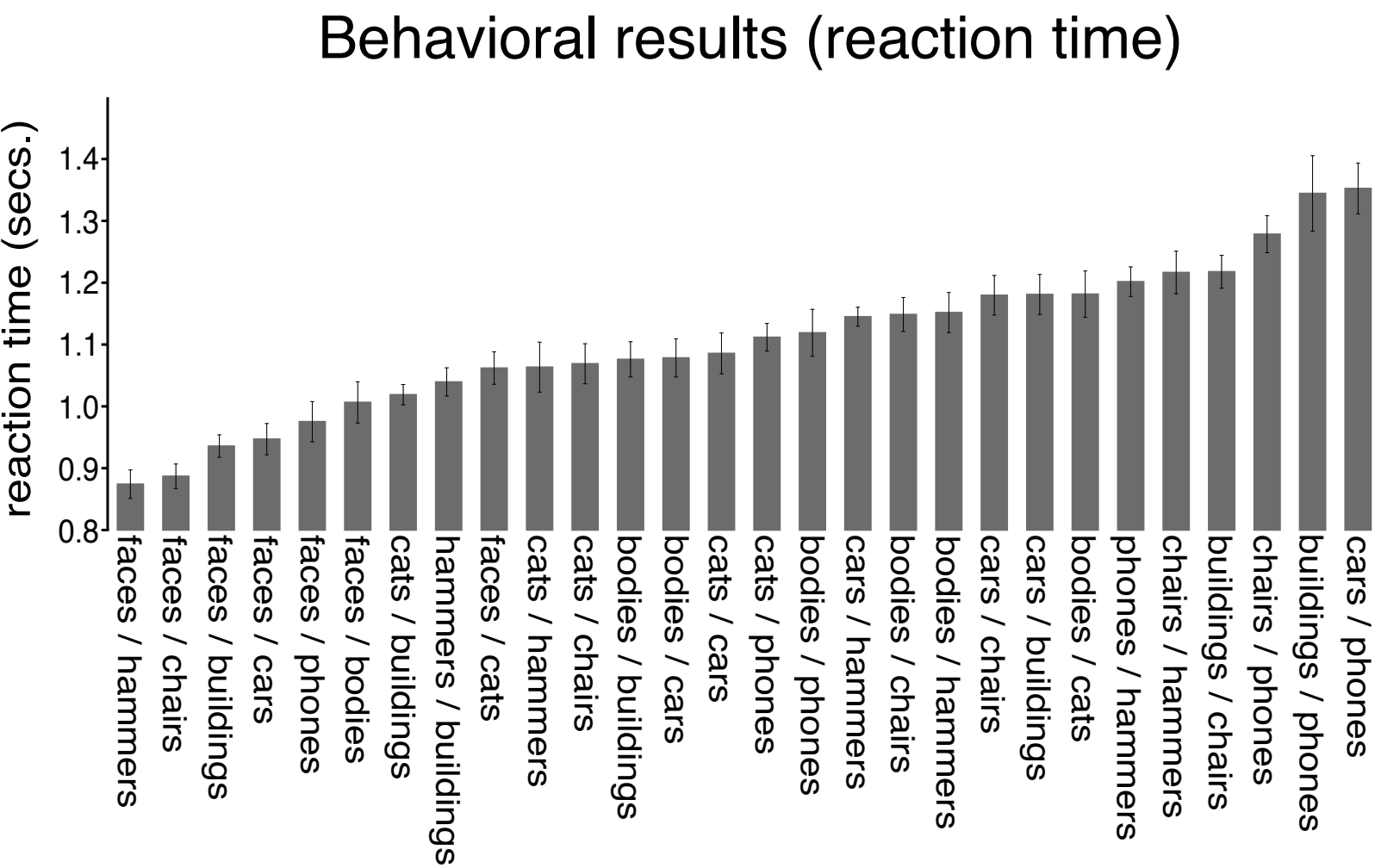
Figure 11. a) Explicit similarity ratings for all category pairings. Average similarity rating is plotted on the y-axis as a function of category pairing along on the x-axis. Error bars reflect within-subject standard error of the mean. b) Group level brain/behavior correlations in each macro-scale sector for both the search task and the rating task. The solid bars indicate the observed correlations, while the dashed lines indicate the partial correlations for each behavioral task after factoring out the other (e.g., “search (partial)” indicates the correlations between the search task and the different neural regions after factoring out the rating task).

a) Stimuli

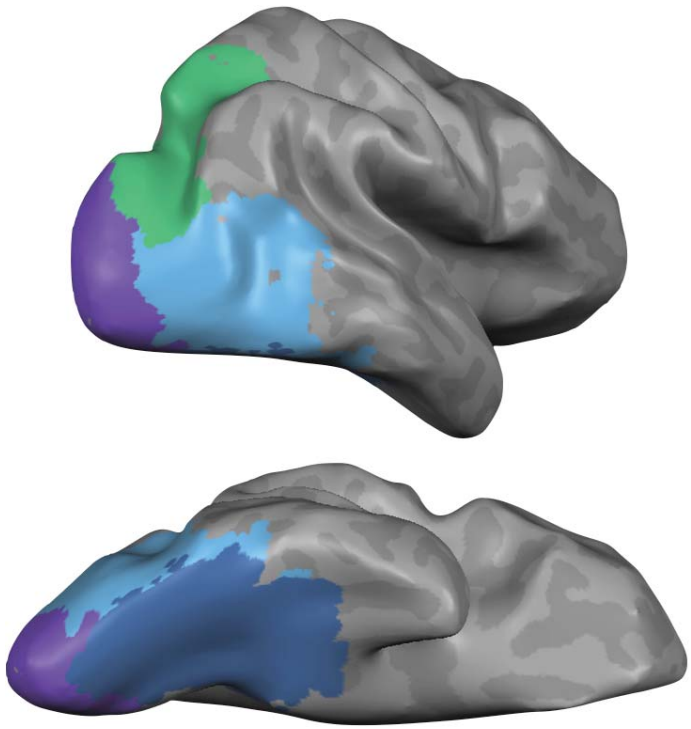


b) Sample Display

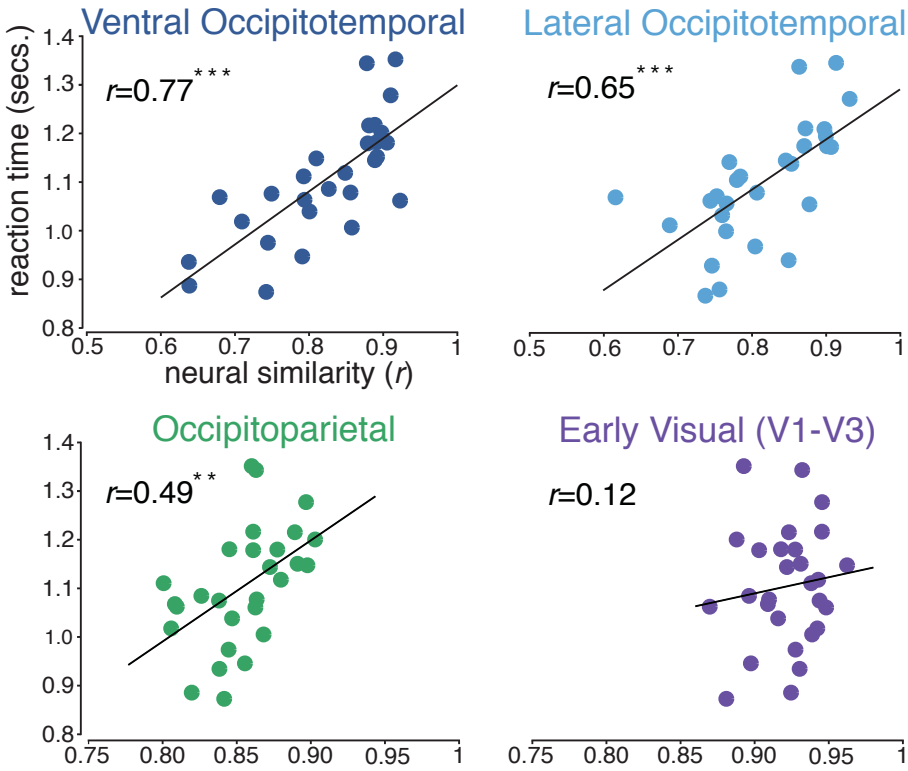



Behavioral results (reaction time)
reaction time (secs.)
0.8
0.9
1.0
1.1
1.2
1.3
1.4
faces / hammers
faces / chairs
faces / buildings
faces / cars
faces / phones
faces / bodies
cats / buildings
hammers / buildings
faces / cats
cats / hammers
cats / chairs
bodies / buildings
bodies / cars
cats / cars
cats / phones
bodies / phones
cars / hammers
bodies / chairs
bodies / hammers
cars / chairs
cars / buildings
bodies / cats
phones / hammers
chairs / hammers
buildings / chairs
chairs / phones
buildings / phones
cars / phones

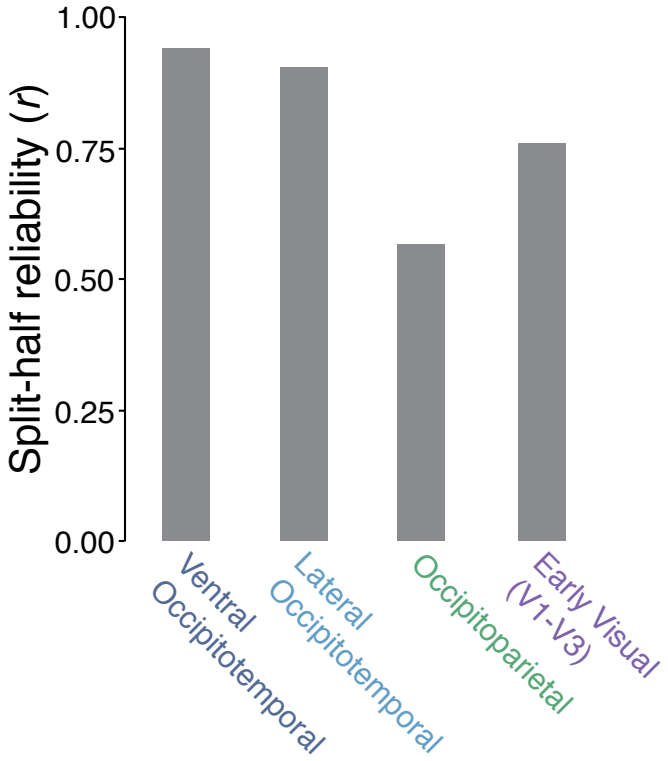
a) Neural Sectors



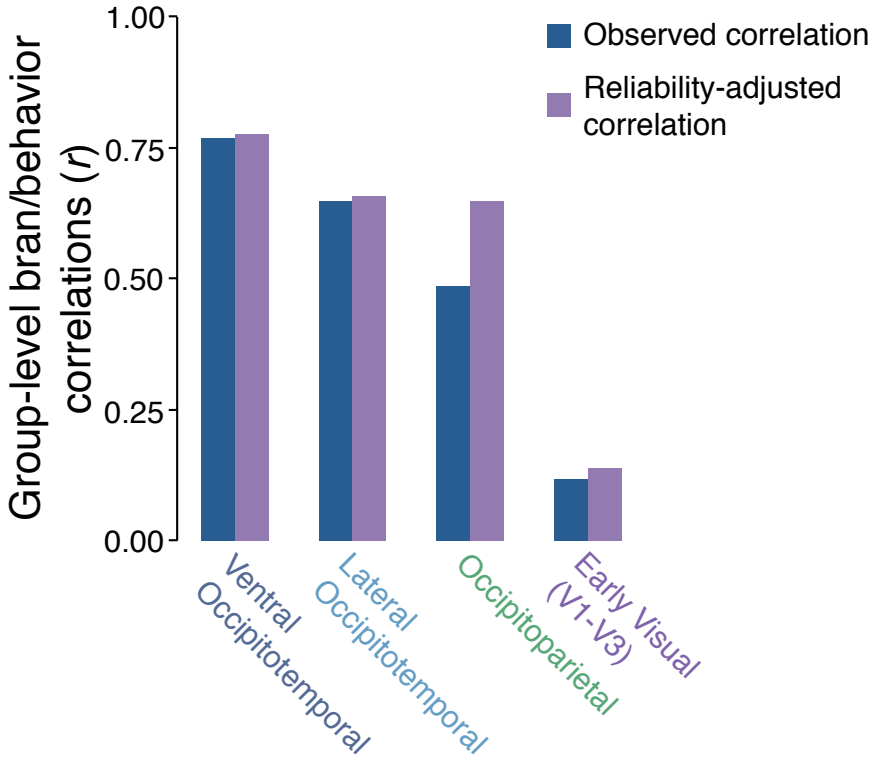
b) Brain/behavior correlations (*r*)



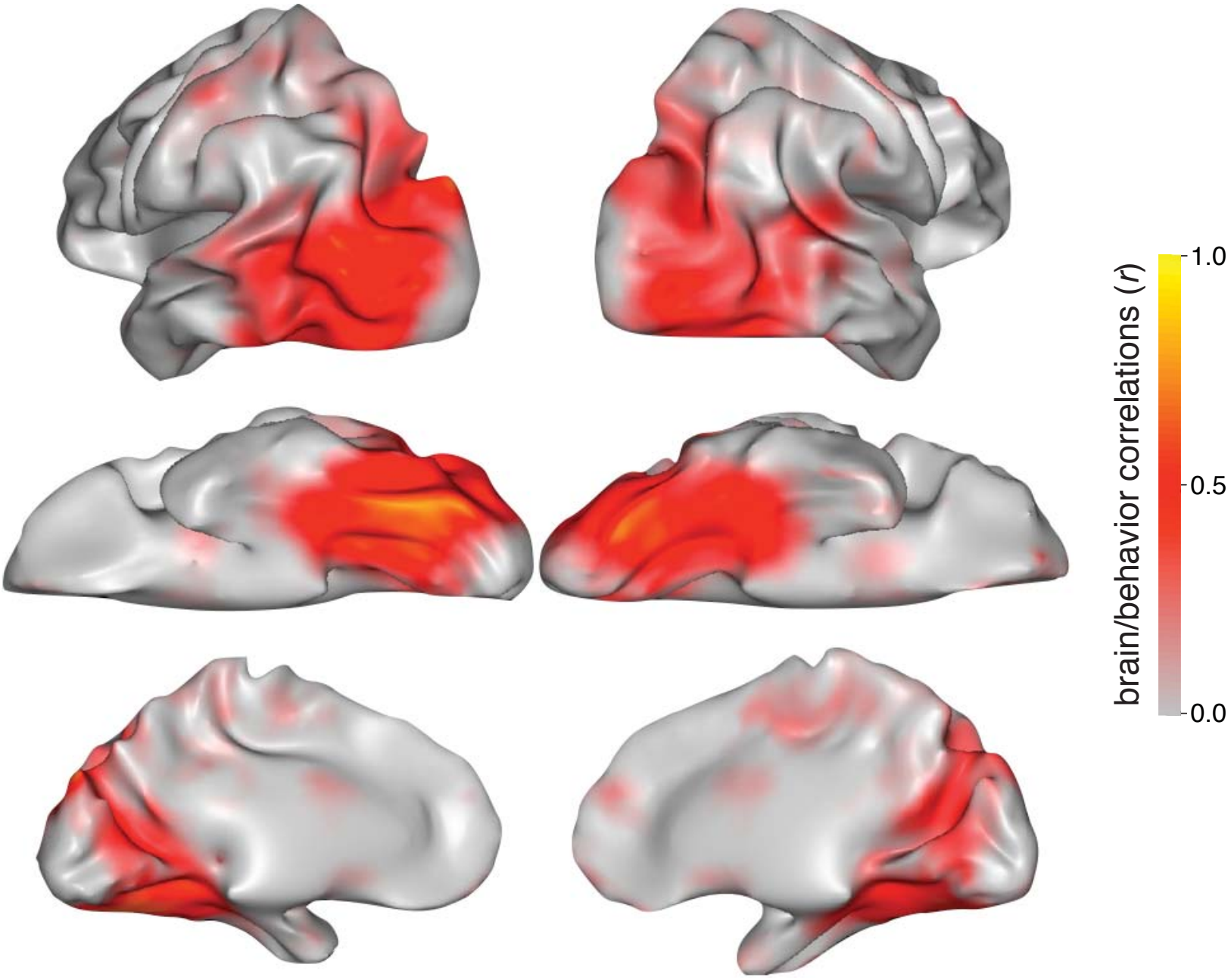
a) Reliability of sectors



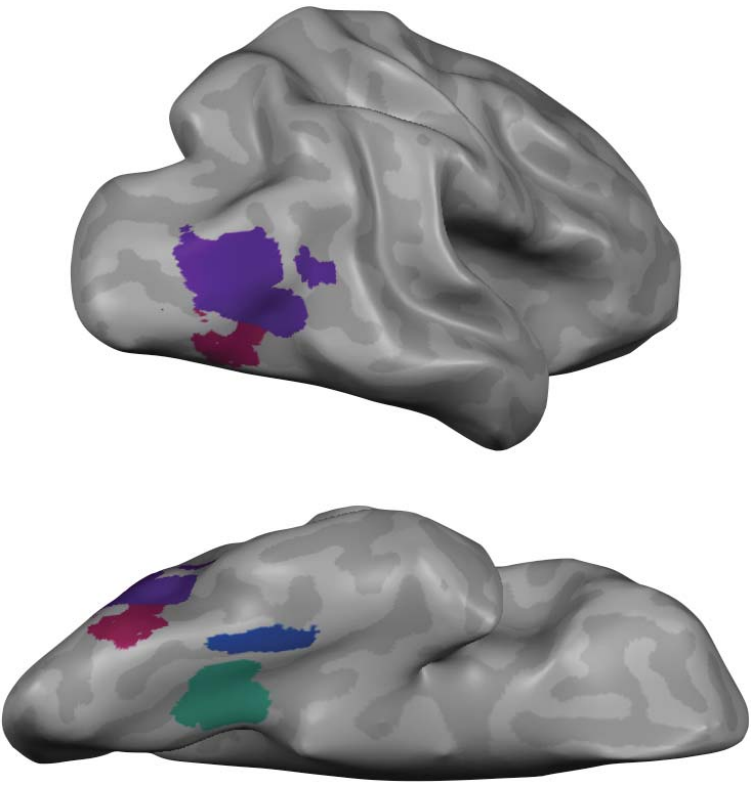
b) Observed vs. adjusted correlations



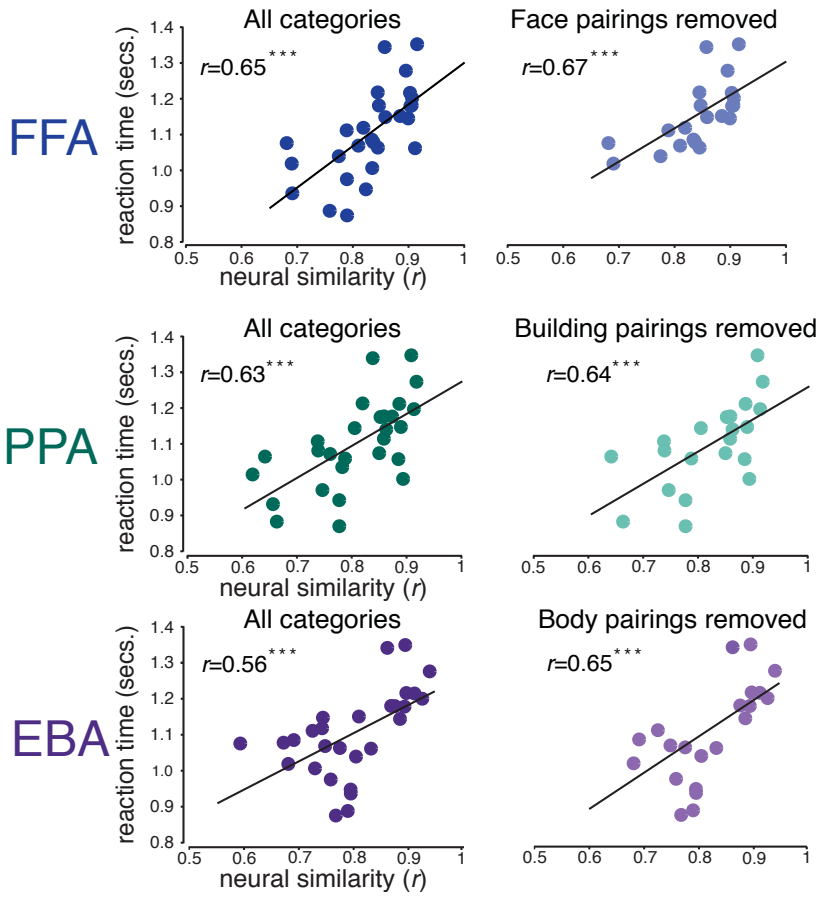
Searchlight analysis across entire cortex

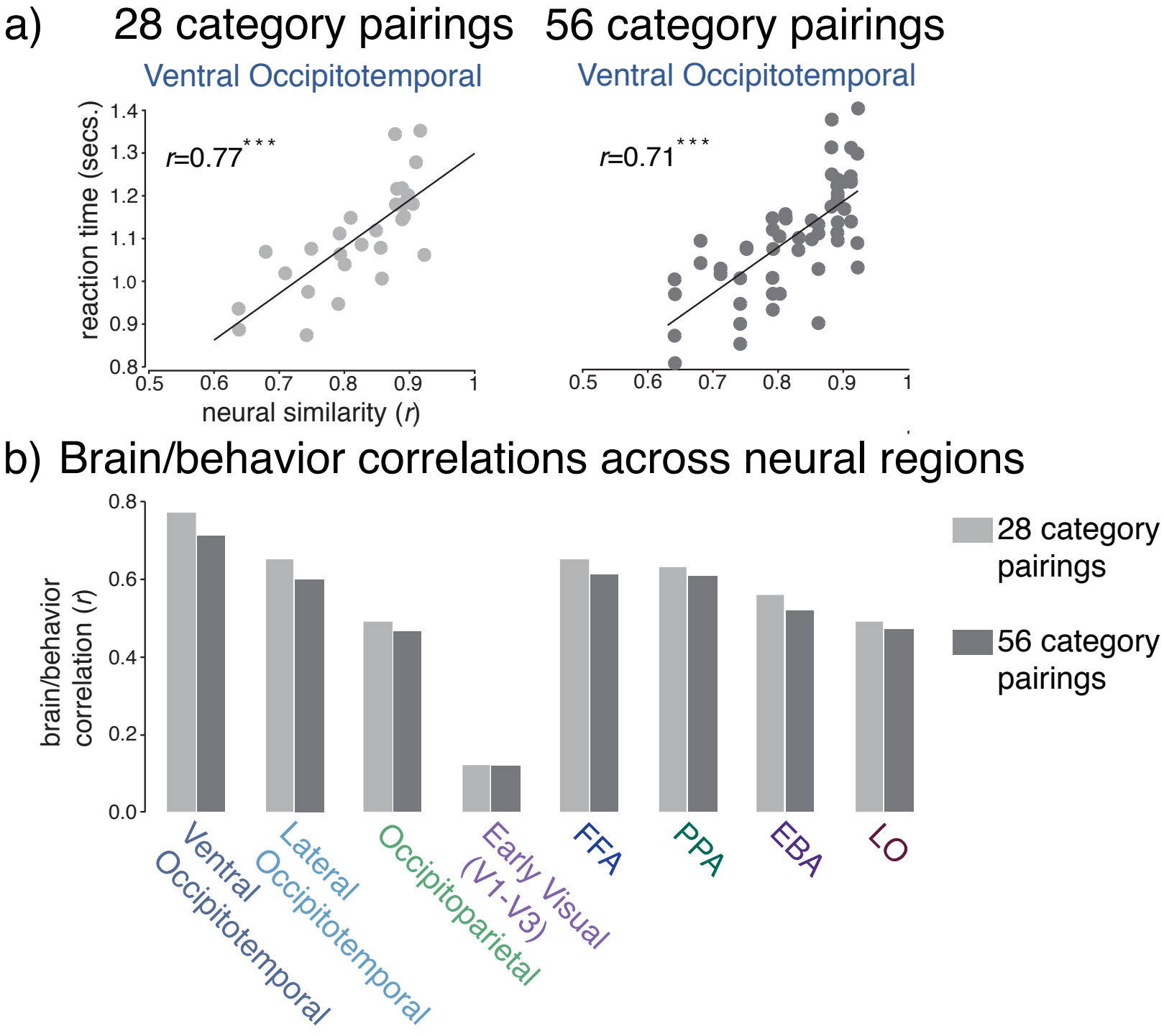


a) Category Selective ROIs



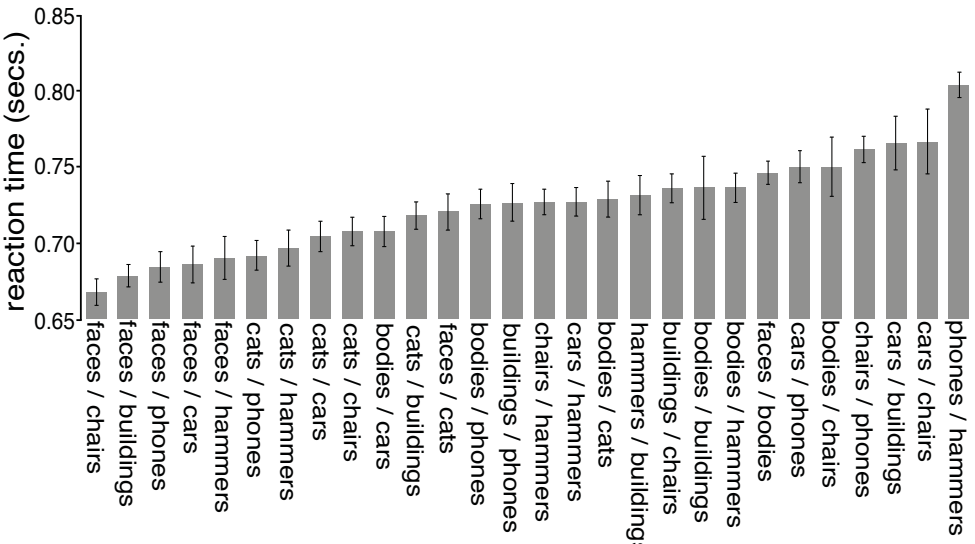
b) Brain/behavior correlations (*r*)



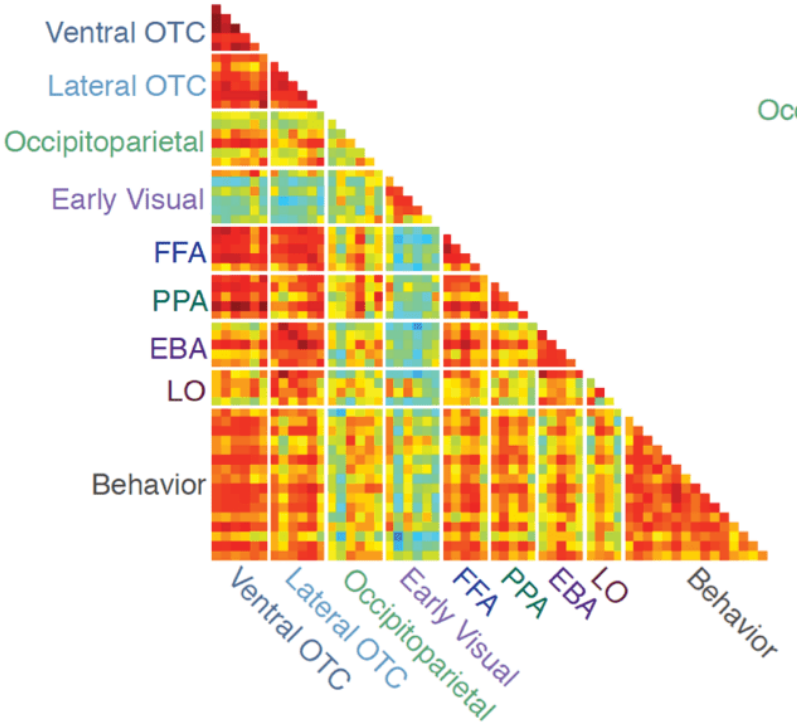

a)
28 category pairings
56 category pairings
Ventral Occipitotemporal
Ventral Occipitotemporal
r=0.77***
r=0.71***
reaction time (secs.)
neural similarity (r)
b) Brain/behavior correlations across neural regions
brain/behavior correlation (r)
28 category pairings
56 category pairings
Ventral Occipitotemporal
Lateral Occipitotemporal
Occipitoparietal
Early Visual (V1-V3)
FFA
PPA
EBA
LO

a) One-item search task:
Is this a face? Yes or no.

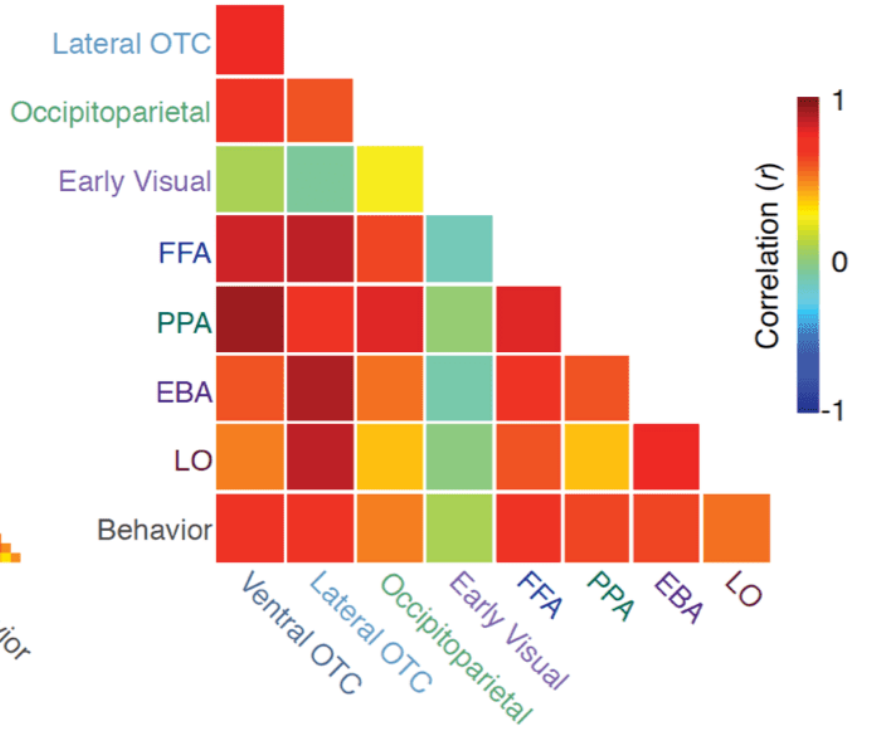
b) Behavioral results (reaction time)



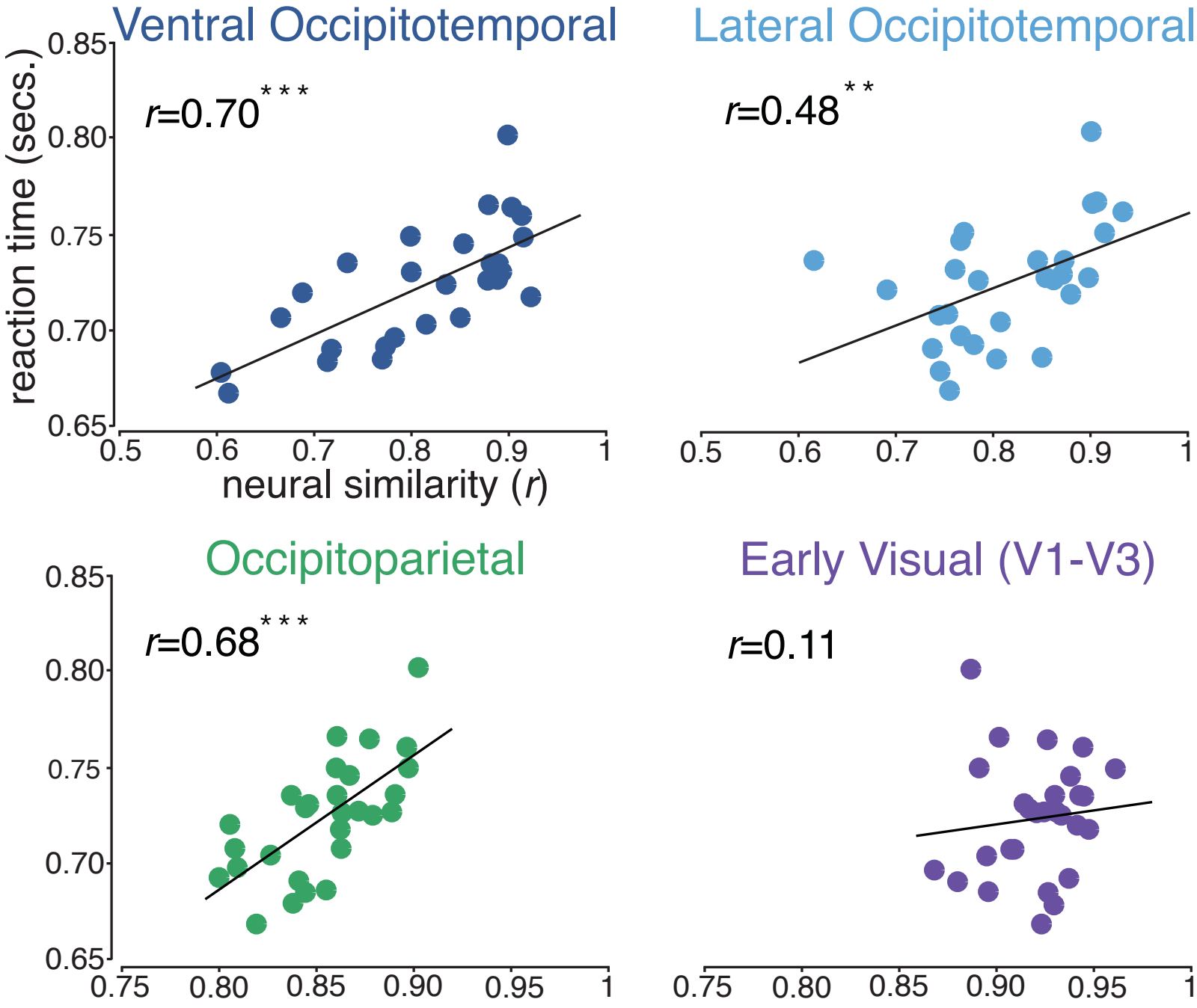
a) Correlations between all individual neural and behavioral participants

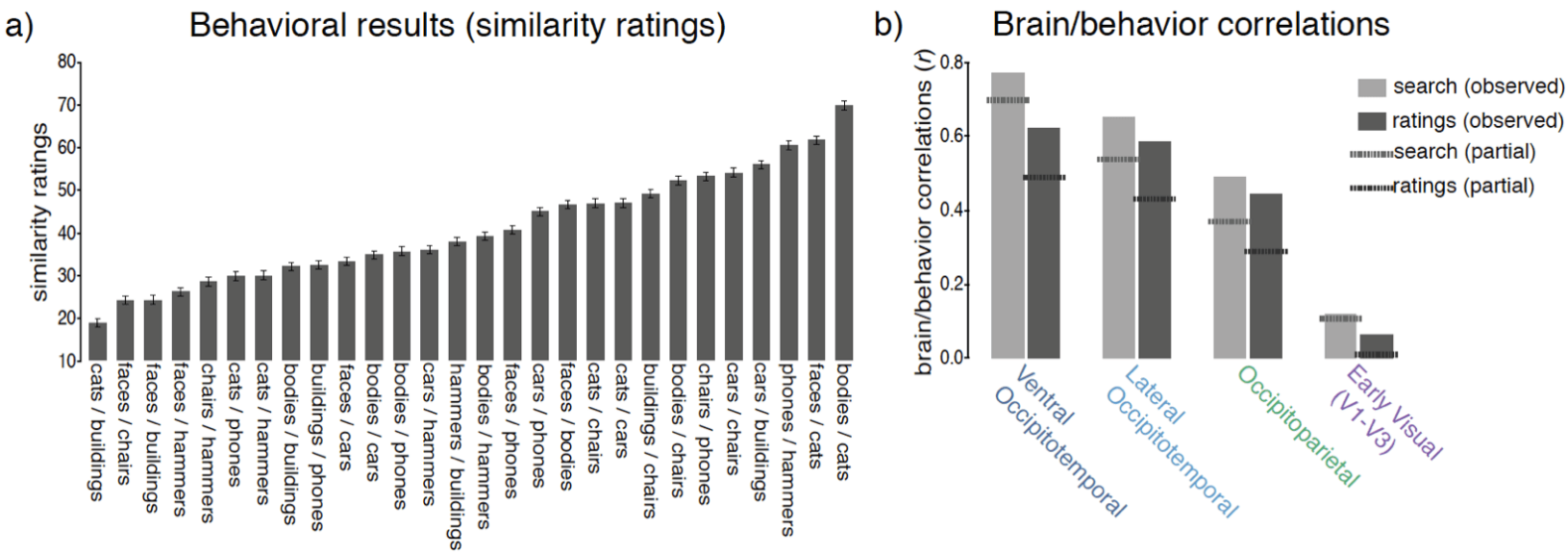


b) Average correlations between all individual neural and behavioral participants



One-item search task brain/behavior correlations (*r*)




a)
Behavioral results (similarity ratings)
similarity ratings
80
70
60
50
40
30
20
10
cats / buildings
faces / chairs
faces / buildings
faces / hammers
chairs / hammers
cats / phones
cats / hammers
bodies / buildings
buildings / phones
faces / cars
bodies / cars
bodies / phones
cars / hammers
hammers / buildings
bodies / hammers
faces / phones
cars / phones
faces / bodies
cats / chairs
cats / cars
buildings / chairs
bodies / chairs
chairs / phones
cars / chairs
cars / buildings
phones / hammers
faces / cats
bodies / cats
b)
Brain/behavior correlations
brain/behavior correlations (r)
0.8
0.6
0.4
0.2
0.0
search (observed)
ratings (observed)
search (partial)
ratings (partial)
Ventral Occipitotemporal
Lateral Occipitotemporal
Occipitoparietal
Early Visual (V1-V3)