

Ultra-high-resolution fMRI of Human Ventral Temporal Cortex Reveals Differential Representation of Categories and Domains

 Eshed Margalit,^{1,7} Keith W. Jamison,^{2,3}  Kevin S. Weiner,^{4,5} Luca Vizioli,²  Ru-Yuan Zhang (张洳源),²
 Kendrick N. Kay,^{2*} and  Kalanit Grill-Spector^{1,6,7*}

¹Neurosciences Graduate Program, Stanford University, Stanford, California 94305, ²Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, Minnesota 55455, ³Department of Radiology, Weill Cornell Medical College, New York, New York 10065, ⁴Department of Psychology, University of California Berkeley, Berkeley, California 94720, ⁵Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, California 94720, ⁶Department of Psychology, Stanford University, Stanford, California 94305, and ⁷Wu Tsai Neurosciences Institute, Stanford, California 94305

Human ventral temporal cortex (VTC) is critical for visual recognition. It is thought that this ability is supported by large-scale patterns of activity across VTC that contain information about visual categories. However, it is unknown how category representations in VTC are organized at the submillimeter scale and across cortical depths. To fill this gap in knowledge, we measured BOLD responses in medial and lateral VTC to images spanning 10 categories from five domains (written characters, bodies, faces, places, and objects) at an ultra-high spatial resolution of 0.8 mm using 7 Tesla fMRI in both male and female participants. Representations in lateral VTC were organized most strongly at the general level of domains (e.g., places), whereas medial VTC was also organized at the level of specific categories (e.g., corridors and houses within the domain of places). In both lateral and medial VTC, domain-level and category-level structure decreased with cortical depth, and downsampling our data to standard resolution (2.4 mm) did not reverse differences in representations between lateral and medial VTC. The functional diversity of representations across VTC partitions may allow downstream regions to read out information in a flexible manner according to task demands. These results bridge an important gap between electrophysiological recordings in single neurons at the micron scale in nonhuman primates and standard-resolution fMRI in humans by elucidating distributed responses at the submillimeter scale with ultra-high-resolution fMRI in humans.

Key words: cortical depth; high-level visual cortex; ventral temporal cortex; visual categories; visual representations

Significance Statement

Visual recognition is a fundamental ability supported by human ventral temporal cortex (VTC). However, the nature of fine-scale, submillimeter distributed representations in VTC is unknown. Using ultra-high-resolution fMRI of human VTC, we found differential distributed visual representations across lateral and medial VTC. Domain representations (e.g., faces, bodies, places, characters) were most salient in lateral VTC, whereas category representations (e.g., corridors/houses within the domain of places) were equally salient in medial VTC. These results bridge an important gap between electrophysiological recordings in single neurons at a micron scale and fMRI measurements at a millimeter scale.

Introduction

Human ventral temporal cortex (VTC) plays a central role in visual recognition (Tong et al., 1998; Grill-Spector et al., 2000,

2004; Moutoussis and Zeki, 2002; Gaillard et al., 2006; Parvizi et al., 2012; Rangarajan et al., 2014). Studies of VTC have shown that ecological domains of objects, faces, body parts, characters, and places elicit clustered responses in predictable anatomical locations within VTC (Malach et al., 1995; Kanwisher et al., 1997; Aguirre et al., 1998; Epstein and Kanwisher, 1998; Grill-Spector

Received Aug. 28, 2019; revised Feb. 10, 2020; accepted Feb. 12, 2020.

Author contributions: E.M., K.N.K., and K.G.-S. designed research; E.M., K.W.J., L.V., R.-Y.Z., and K.N.K. performed research; E.M., K.W.J., and K.N.K. contributed unpublished reagents/analytic tools; E.M., K.S.W., L.V., R.-Y.Z., K.N.K., and K.G.-S. analyzed data; E.M. wrote the first draft of the paper; E.M., K.S.W., L.V., R.-Y.Z., K.N.K., and K.G.-S. edited the paper; E.M. wrote the paper.

This work was supported by National Science Foundation GRFP to E.M.; National Eye Institute 1R01EY023915 to K.G.-S.; National Institutes of Health Grants P41 EB015894, P30 NS076408, S10 RR026783, and S10 OD017974-01 to K.N.K.; and W. M. Keck Foundation to K.N.K.

The authors declare no competing financial interests.

*K.N.K. and K.G.-S. contributed equally to this work as co-senior authors.

Correspondence should be addressed to Eshed Margalit at eshed.margalit@gmail.com.

<https://doi.org/10.1523/JNEUROSCI.2106-19.2020>

Copyright © 2020 the authors

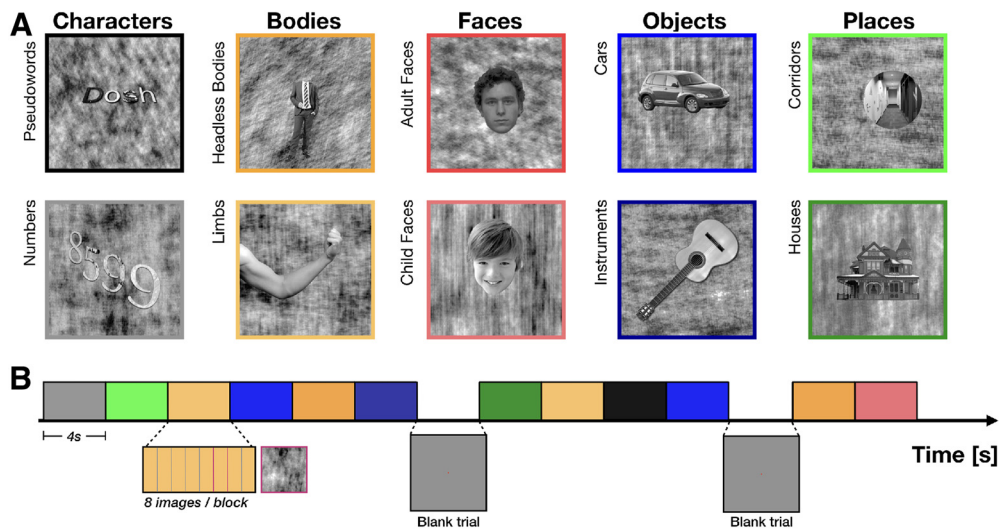


Figure 1. Experimental design. **A**, Stimuli. Images were drawn from five ecological domains: characters, body parts, faces, objects, and places (columns). Each domain contained two visual categories (rows). Exemplars of each category had the same parts and configurations (see sample stimuli). **B**, Experimental design schematic. Each 4 s trial consisted of 8 different images drawn from the same visual category, presented for 500 ms each. Images of phase-scrambled backgrounds with no foreground item (“oddball” images) could appear within a trial, as schematized with magenta boundaries. Color represents the image category (same as border of the sample stimuli in **A**).

et al., 1999; Cohen et al., 2000; Peelen and Downing, 2005; Weiner et al., 2014). In parallel with the ecological domain framework, studies have demonstrated that different object categories elicit distinct and reliable patterns of activity across VTC (Haxby et al., 2001; Cox and Savoy, 2003; Kriegeskorte et al., 2008b; Huth et al., 2012). These distributed representations have a hierarchical organization whereby more abstract categorical distinctions are represented at a larger scale than more specific distinctions. For example, the representation of animate stimuli spans multiple centimeters across lateral VTC, and nested within it are clustered responses to faces and body parts (~1 cm in scale). Similarly, the representation of inanimate stimuli spans medial VTC, and within it are clustered responses to places (Grill-Spector and Weiner, 2014).

Despite the evidence for both theoretical frameworks, the relationship between representations of categories and ecological domains in VTC remains elusive for several reasons. First, ecological domains differ in the variability of visual features of their exemplars; for example, faces are similar in their features and configurations, but body parts, such as limbs and torsos, are visually dissimilar. Second, domains vary with respect to their classically defined level of abstraction (superordinate, basic, and subordinate) (Rosch et al., 1976); for example, places are a superordinate category, but faces are a basic-level category. Third, it is unknown whether the proposed spatial hierarchy of representations extends from the level of domains to categories within a domain (e.g., limbs and headless bodies within the domain of body parts).

Prior studies have evaluated representations across voxels each spanning between 1 mm (McGugin et al., 2012, 2014) and several millimeters on a side, leaving unknown the nature of distributed representations in VTC when measured with submillimeter voxels. With higher spatial resolution, functionally heterogeneous neural populations that would be grouped within a standard fMRI voxel are divided into separate voxels, giving us an opportunity to assess more detailed representations. Recent advances in ultra-high-field (7T) fMRI have enabled such submillimeter measurements (Koopmans et al., 2010; De Martino et al., 2018; Dumoulin et al., 2018). 7T fMRI studies have generally

focused on early visual cortex and have revealed fine-scale structure: ocular dominance (Cheng et al., 2001) and orientation columns in V1 (Yacoub et al., 2008), disparity- and color-selective stripes in V2 and V3 (Nasr et al., 2016), and motion-direction columns in hMT+ (Schneider et al., 2019). However, ultra-high-resolution investigations of object category representations in VTC have not been systematically conducted. Further, as VTC in adults is 1.5- to 3-mm-thick (Sowell et al., 2003; Natu et al., 2019), submillimeter fMRI offers the potential to investigate the variability of object representations across cortical depths (Fujita et al., 1992; Tanaka, 1996; Wang et al., 1996; Tsunoda et al., 2001; Maass et al., 2014).

To fill these gaps in knowledge, we conducted an ultra-high-resolution (0.8 mm) fMRI experiment at 7T in which 7 participants viewed images from five ecological domains with two visual categories per domain (Fig. 1A). Critically, we designed stimuli such that each visual category contained exemplars with similar parts and configuration (Grill-Spector et al., 2004; Grill-Spector and Kanwisher, 2005; Weiner and Grill-Spector, 2010; Stigliani et al., 2015). We measured distributed responses to domains and categories in each participant and tested how this information is represented in lateral and medial VTC as well as across cortical depths. Finally, we compared ultra-high-resolution data with simulated standard-resolution data to determine the necessity of ultra-high-resolution fMRI for revealing the nature of VTC representations.

Materials and Methods

Participants

Seven participants (3 males; a participant, S1, was an author [K.N.K.]) participated in this study. All participants had normal or corrected-to-normal visual acuity. The experimental protocol was approved by the University of Minnesota Institutional Review Board. Informed written consent was obtained from all participants.

Stimulus presentation

Stimuli were presented using a Cambridge Research Systems BOLD-screen 32 LCD monitor positioned at the head of the 7T scanner bed (resolution 1920 × 1080 at 120 Hz; viewing distance 189.5 cm). Participants viewed the monitor via a mirror mounted on the RF coil. A Mac

Pro computer was used to control stimulus presentation using code based on Psychophysics Toolbox. Behavioral responses were recorded using a button box.

Experimental design and task

Participants viewed grayscale images drawn from five domains, with two visual categories per domain: characters (pseudowords and numbers), bodies (headless bodies and limbs), faces (adult and child), places (corridors and houses), and objects (cars and string instruments) (Fig. 1A). Following conventions established in prior work on visual representations in humans, we define a visual category as a set of exemplars sharing the same parts and configuration (Grill-Spector and Kanwisher, 2005; Weiner and Grill-Spector, 2010; Weiner et al., 2010; Stigliani et al., 2015). For example, limb exemplars all have five digits protruding from a cylindrical object. A domain is defined according to Kanwisher (2000, 2010) as a grouping of items from one or more categories by ecological validity (regardless of whether they share visual features), which are hypothesized to require specialized cortical processing (Kanwisher, 2000, 2010). For example, limbs and headless bodies are visually dissimilar categories within the domain of body parts. Each exemplar was presented on a phase-scrambled background, randomly generated from one of the images in the database. Items from each category were randomly jittered in their size, position, and orientation. Stimuli of each visual category were controlled for low-level features, including luminance, contrast, and pixelwise similarity as well as cognitive factors, such as familiarity (Stigliani et al., 2015; Nordt et al., 2019). We note that ecological domains differ in their visual variability; for example, images of children and adult faces from the face domain are visually more similar to each other than images of houses and corridors from the place domain.

Each run contained 48 randomly selected exemplars of a possible 144 unique exemplars from each of the 10 categories. The stimuli (with background) occupied a square region with dimensions $10^\circ \times 10^\circ$. In each run, stimuli from each visual category were presented in 4 s trials, each consisting of 8 images from a given category presented for 500 ms each (Fig. 1B). On a randomly selected one-third of trials, 1 of the 8 images consisted of a background phase-scrambled image with no foreground exemplar. Each run contained 6 trials of each category with different stimuli, as well as blank trials in which a gray screen was presented, for a total of 312 s per run. Images were not repeated within a run. Participants were instructed to fixate on a centrally presented red dot and to report via button press when a background phase-scrambled image (the “oddball”) appeared without a superimposed, intact exemplar. Each participant completed 12 runs in a single scanning session. We excluded one run for Participant S2 and two runs for Participant S6 due to poor visually evoked responses. The order of blocks was pseudorandomized for each run.

MRI data acquisition

Data acquisition procedures are identical to those described in a previous publication (Kay et al., 2019). A summary of these procedures is provided below, and we refer the reader to the previous publication for further details.

MRI data were collected at the Center for Magnetic Resonance Research at the University of Minnesota. Anatomical data were collected using a 3T Siemens scanner at 0.8 mm isotropic resolution. To ensure high signal-to-noise ratio (SNR), we acquired several repetitions of each type of anatomical volume. For each participant, we typically collected 8 scans of a whole-brain T1-weighted MPRAGE sequence (TR 2400 ms, TE 2.22 ms, TI 1000 ms, flip angle 8°) and 2 scans of a whole-brain T2-weighted SPACE sequence (TR 3200 ms, TE 563 ms). We used the prescan-normalized version of the T1 and T2 volumes, which corrects for receive-coil inhomogeneities. Functional data were collected at 7T using a Siemens Magnetom scanner and a custom 4-channel-transmit, 32-channel-receive RF head coil. We used gradient-echo EPI at 0.8 mm isotropic resolution with partial-brain coverage: 84 oblique slices covering occipitotemporal cortex, slice thickness: 0.8 mm, slice gap: 0 mm, FOV: 160 mm (FE) \times 129.6 mm (PE), phase-encode direction: inferior-superior, matrix size: 200×162 , TR: 2.2 s, TE: 22.4 ms, flip angle: 80° , partial Fourier: 6/8, in-plane acceleration factor: 3, and a slice accelera-

tion factor: 2. Gradient-echo fieldmaps were also acquired for *post hoc* correction of EPI spatial distortion (same prescription as the EPI data, resolution: $2 \text{ mm} \times 2 \text{ mm} \times 2.4 \text{ mm}$, TE1 = 4.59 ms, TE2 = 5.61 ms). Fieldmaps were periodically acquired over the course of each scan session to track changes in the magnetic field (before and after the functional runs as well as approximately every 20 min interspersed between the runs).

Data analysis

Data analysis was performed using a combination of custom MATLAB, Python, and R code and includes tools from FreeSurfer, SPM, and FSL (specific instances are noted below). Routines developed for preprocessing and visualization of data are available online (<http://github.com/kendrickkay/>).

Anatomical preprocessing

Preparation of anatomical volumes. T1- and T2-weighted anatomical volumes were corrected for gradient nonlinearities based on scanner calibration measurements. T1 volumes were coregistered (rigid-body transformation estimated using a 3D ellipse that focuses the cost metric on cortical voxels; cubic interpolation) and averaged to improve SNR, and the same was done to the T2 volumes. Each volume was inspected for image artifacts and rejected from the averaging procedure if deemed of poor quality. The FSL tool FLIRT was then used to coregister the averaged T2 volume to the averaged T1 volume (rigid-body transformation; sinc interpolation). We henceforth refer to the averaged and coregistered T1 and T2 volumes as simply the T1 and T2 volumes.

Generation of cortical surface representations. The T1 volume (0.8 mm resolution) was processed using FreeSurfer version 6 beta (build-stamp 20161007) with the *-hires* option. Manual edits of tissue segmentation were performed to maximize accuracy of the resulting cortical surface representations. Several additional processing steps were performed. Using *mrisc_expand*, we generated cortical surfaces positioned at different depths of the gray matter. Specifically, we constructed six surfaces spaced equally between 10% and 90% of the distance between the pial surface and the boundary between gray and white matter. We also increased the density of surface vertices using *mrisc_mesh_subdivide*. This bisected each edge and resulted in a doubling of the number of vertices. For all analyses in this paper, we averaged GLM estimates between depths 1 and 2, 3, and 4, as well as 5 and 6 to effectively create three cortical surfaces: a superficial, middle, and deep surface. Finally, to reduce computational burden, we truncated the surfaces to include posterior portions of the occipital, temporal, and parietal lobes (since this is where functional measurements were made).

Functional preprocessing

Preparation of fieldmaps. Fieldmaps acquired in each session were phase-unwrapped using the FSL utility *prelude*. We then regularized the fieldmaps by performing 3D local linear regression using an Epanechnikov kernel (Epanechnikov, 2005) with radius 5 mm; we used values in the magnitude component of the fieldmap as weights in the regression to improve robustness of the field estimates. This regularization procedure removes noise from the fieldmaps and induces some degree of spatial smoothness. Finally, we linearly interpolated the fieldmaps over time, producing an estimate of the field strength for each functional volume acquired.

Volume-based preprocessing. We performed both temporal and spatial preprocessing of the fMRI data. First, cubic interpolation was performed on each voxel's time-series data to correct for differences in slice acquisition times and to obtain an integer sampling rate (2.0 s). This can be viewed as a temporal correction step. Next, the regularized time-interpolated fieldmap estimates were used to correct EPI spatial distortion (Jezzard and Balaban, 1995). Rigid-body motion parameters were then estimated from the undistorted EPI volumes using the SPM5 utility *spm_realign*. Finally, cubic interpolation was performed on each slice-time-corrected volume to compensate for the combined effects of EPI spatial distortion and motion. This can be viewed as a spatial correction step.

Coregistration to anatomy. We coregistered the average of the preprocessed functional volumes obtained in a scan session to the T2 volume (affine transformation estimated using a 3D ellipsoid that focuses the

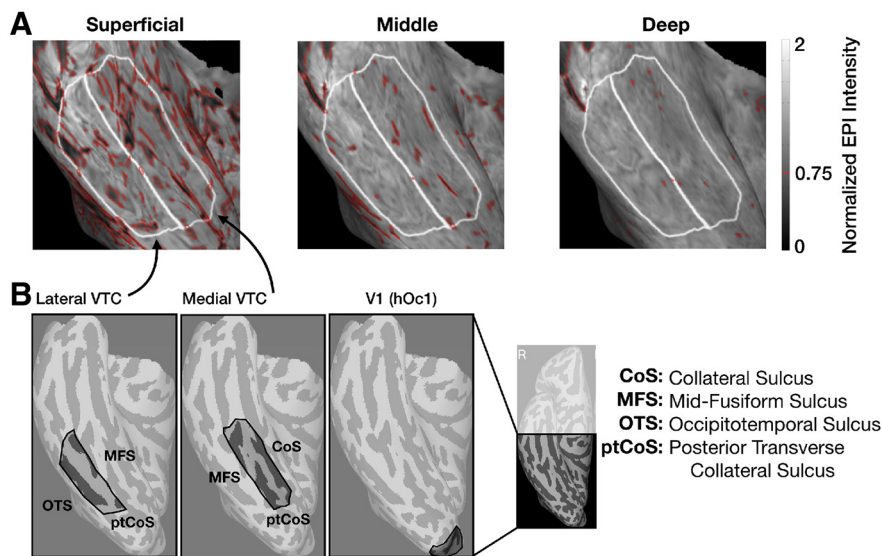


Figure 2. Visualization of likely veins and ROI definitions on the cortical surface of a representative participant's right-hemisphere VTC. **A**, Time-averaged normalized EPI intensities in a zoomed-in view of an example participant's right-hemisphere ventral temporal cortex (VTC) shown at each cortical depth. White lines indicate the boundaries of the lateral VTC and medial VTC ROIs for this participant. Red outlines indicate values < 0.75 , designated as likely veins. The data at the corresponding surface locations were removed at all depths from subsequent analyses. **B**, Example ROI definitions in the right hemisphere of a representative participant for lateral VTC (left), medial VTC (middle), and V1 (right). Far right inset, The posterior region of the right hemisphere to which data acquisition was restricted.

cost metric on cortical ROIs). This resulted in a transformation that maps the EPI data to the participant-native brain anatomy volume.

Surface-based preprocessing. With the anatomical coregistration complete, the functional data were reanalyzed using surface-based preprocessing. The exact same procedures associated with volume-based preprocessing were performed, except that the final spatial interpolation was performed at the locations of the vertices of the six depth-dependent surfaces. Thus, the only difference between volume- and surface-based preprocessing is that the data are prepared either on a regular 3D grid (volume) or an irregular manifold of densely spaced vertices (surface). The use of simple interpolation to map volumetric data onto surface representations helps maximize spatial resolution and avoids making strong assumptions about cortical topology.

Anatomical definition of VTC partitions and early visual cortex

In each participant's native anatomical space, we defined three nonoverlapping ROIs according to anatomical landmarks that have been used in prior work (Weiner and Grill-Spector, 2010; Weiner et al., 2010, 2014; Bugatus et al., 2017; Nordt et al., 2019) (Fig. 2B): lateral VTC, medial VTC, and human occipital cytoarchitectonic area 1 (hOc1), which closely matches functionally defined primary visual area V1 (Hinds et al., 2008; Rosenke et al., 2018). Lateral and medial VTC are separated by the mid-fusiform sulcus (MFS), which identifies transitions in both anatomical and functional properties (Grill-Spector and Weiner, 2014; Weiner et al., 2014; Weiner, 2019). The posterior end of both VTC ROIs was the posterior transverse collateral sulcus; the anterior end of these ROIs was the anterior tip of the MFS. The posterior transverse collateral sulcus is also the separating anatomical landmark between the occipital and temporal lobes, and additionally serves as a boundary marking the transition between hV4 and VO1 (Witthoft et al., 2014). The lateral VTC was laterally bounded by the occipitotemporal sulcus and medially by the MFS. Medial VTC was laterally bounded by the MFS and medially by the collateral sulcus. Whenever possible, the full extent of these bounding sulci was included; for example, the medial VTC includes all of the occipital portion of the collateral sulcus within the anterior and posterior bounds. The hOc1 ROI is defined on the FreeSurfer (<https://surfer.nmr.mgh.harvard.edu>) average surface (Dale et al., 1999; Fischl et al., 1999) and was projected onto each participant's cortical surface using cortex based alignment. All ROIs were defined bilaterally and data from

both hemispheres were concatenated before analysis. In the remainder of the paper, we refer to the hOc1 ROI as "V1" for simplicity.

Masking of potential venous artifacts

The effects of draining veins in superficial cortical layers colocalize with increased signal change, signal variability estimates, and t values (Vu and Gallant, 2015; Kay et al., 2019). While the influence of large veins on the measured BOLD response cannot be completely mitigated with presently available techniques, we implemented an approach to limit their contribution to the effects described here. One of the advantages of ultra-high-resolution fMRI is that it enables identifying likely veins by their low mean EPI intensity (Koopmans et al., 2010; Kay et al., 2019). While EPI intensities can be used as an anatomical marker of the location of venous effects, they are influenced by inhomogeneities in the transmit and receive profiles of the RF coil. To reduce these biases, we estimate the measurement bias map and then correct the measured raw values by the bias map. Bias correction of EPI intensities is used only to help identify venous effects; the additional analyses described below are performed on the preprocessed functional data without bias correction.

Vertices exhibiting low bias-corrected EPI intensity ($< 75\%$ of the median EPI intensity)

at any cortical depth were removed from all cortical depths at that surface location for further analyses (e.g., Fig. 2A, red outlines). This procedure led to the removal of $27 \pm 1\%$ of likely vein vertices across participants. To ensure that our results do not depend on this vein-removal procedure, all analyses were repeated without this preprocessing step.

Surface visualization. To visualize surface-based data, we used a tool (*cvnlookupimages*: <https://github.com/kendrickkay/cvnlookupimages>) that orthographically projects the vertices of a surface onto the image plane and then uses nearest-neighbor interpolation to assign values from vertices to pixels. To summarize the various transformations from fMRI data to surface visualizations: (1) we acquire fMRI data with 0.8 mm voxels, (2) these voxels are resampled via cubic interpolation onto densely packed surface vertices with 0.4 mm resolution, and (3) these surface vertices are visualized using a nearest-neighbor mapping to image pixels. For the figures generated here, we restrict visualized vertices to those included in an anatomically defined VTC mask.

Computation of temporal SNR (tSNR)

Before GLM analysis of the functional data, tSNR was computed on the raw BOLD time-series for each vertex in each participant as the ratio of the time-series mean to its standard deviation (SD). We averaged over all vertices in a given partition and depth for each participant, yielding nine values per participant ($3 \text{ depths} \times 3 \text{ ROIs}$) for comparison.

General Linear Model (GLM)

Preprocessed functional data were analyzed using GLMdenoise software (Kay et al., 2013). However, we chose to disable the denoising option to provide a closer comparison to conventional data analysis methods. Thus, the GLM consisted of experimental regressors (constructed by convolving the design matrix with a canonical hemodynamic response function) and polynomial regressors that characterize the baseline signal level in each run. GLMs were run separately for two splits of the data: odd runs and even runs. The standard error (SE) of the GLM β estimates was computed by bootstrapping 100 times over the runs present for the given data split. On each bootstrap iteration, N runs were sampled with replacement and the β estimate was computed for those runs, where N is the number of runs present for the data split.

The GLM was fit to the time-series data in each vertex of the six depth surfaces generated, then averaged across each pair of adjacent depths to

Table 1. Metric names and MATLAB code for each of the six metrics computed

Metric name	MATLAB code
z norm	$(\text{response}_i - \text{mean_across_conditions}) / \text{se_across_conditions}$
t contrast	$(\text{response}_i - \text{response}_j) / \sqrt{(\text{se}_i^2 / n_i) + (\text{se}_j^2 / n_j)}$
z score	$(\text{response}_i - \text{mean_across_conditions}) / \text{std_across_conditions}$
Mean-subtracted β	$\text{response}_i - \text{mean_across_conditions}$
β norm	$\text{response}_i / \text{se_across_conditions}$
Raw β	response_i

form the three surfaces we analyzed (superficial, middle, and deep). While prior descriptions of these data describe all six surfaces (Kay et al., 2019), we chose to present data across three depths for two reasons. First, given a voxel size of 0.8 mm and an estimated cortical thickness ranging between 1.5 and 3 mm (Sowell et al., 2003; Natu et al., 2019), the functional data projected onto six equidistant surfaces would be similar between adjacent depths, whereas dividing cortex into three depths reduces (but does not necessarily eliminate) this dependence. Second, we found that describing differences between depths in terms of a superficial, middle, and deep depth lent itself to conceptually simpler interpretation. Beta weights from the GLM (reflecting BOLD response amplitudes to the experimental conditions) were converted from raw image units to units of percent BOLD signal change by dividing the signal in each vertex by its mean signal and multiplying by 100. We also computed the variance explained by the GLM (R^2) for each vertex.

Metrics to evaluate distributed responses patterns. For each participant, we calculated six metrics of distributed responses for each split of the data (odd runs and even runs), described in detail in Table 1. Of these, the primary metric for which we report results is the z-norm metric. To remove the effects of mean signal differences between vertices due to incidental factors that are not of interest (e.g., proximity to a draining vein) and to down-weight vertices with poor GLM fits, we computed the z-norm metric of each vertex's responses to each of the 10 categories by (1) subtracting the mean response for that vertex and (2) dividing by the mean SE of β estimates across the 10 categories. For the MATLAB code provided in Table 1, we define the following commonly used variables, where i and j index groups of one or more categories: response_i : mean response to group i ; response_j : mean response to group j ; $\text{mean_across_conditions}$: mean response across all 10 categories; $\text{se_across_conditions}$: mean SE of response estimate across all 10 categories; $\text{std_across_conditions}$: SD of responses across all 10 categories; se_i : mean SE of response estimate for group i ; se_j : mean SE of response estimate for group j ; n_i : number of categories in group i ; n_j : number of categories in group j .

Representational dissimilarity matrices (RDMs). RDMs (Kriegeskorte et al., 2008a) were computed to illustrate the pairwise dissimilarity between distributed responses to different categories at different cortical depths across odd and even splits of the data. z-norm response maps from each combination of categories and cortical depths from one-half of the data were correlated with maps from each combination in the other half of the data (i.e., data from odd runs were compared with data from even runs). Each cell in the RDM reflects the dissimilarity (one minus correlation) between the distributed responses to pairs of categories, across odd and even splits of the data.

Given 10 stimulus types and three cortical depths, each RDM is a 30×30 matrix. RDMs were computed separately for each participant and each anatomical partition (lateral VTC, medial VTC, and V1). The RDM is arranged by stimulus type (domain, then category) and depth (superficial, then middle, then deep); thus, starting from the top left (see, e.g., Fig. 3), each 3×3 square indicates the dissimilarity between the same pairs of categories across the three depths, and each 6×6 square indicates the dissimilarities between the same pairs of domains across categories and depths. We also generated 10×10 RDMs for each cortical depth (superficial, middle, and deep; see Fig. 9) by selecting the corresponding cells from the full 30×30 RDM. These 10×10 RDMs contain the pairwise dissimilarities between distributed responses to different categories within a given cortical depth. To enable RDMs to be used as distance matrices for the construction of multidimensional scaling (MDS) embeddings, we symmetrized the matrices by averaging the up-

per and lower triangles and duplicating the result on either side of the diagonal, which remained unchanged. In the main analyses, RDMs are shown for data excluding likely veins (see Masking of potential venous artifacts). Control analyses include all vertices.

Modeling RDMs. To evaluate the amount of domain-level, category-level, and depth-level structure in the measured RDMs for each partition and each participant, we fit each RDM as a weighted sum of idealized model RDMs (Khaligh-Razavi and Kriegeskorte, 2014) for the three levels of interest (domain, category, and depth) and an intercept term. Because the RDMs and regressors are symmetric, we simplified the model by retaining only the lower triangle (and diagonal) of each matrix used in the regression analysis. After vectorizing the target RDM and predictors, we used linear least squares regression to determine the contribution (β weights) of each regressor (derived from the 3 model RDMs) so that the total contribution minimized the sum squared differences between the measured and predicted RDMs. When modeling depth-specific RDMs, the same procedure was used but without depth as a regressor.

MDS. To better visualize the representational structure from high-dimensional RDMs, we performed multidimensional scaling (MDS) (Kruskal, 1964) of the RDM. MDS embeds the representation of each combination of stimulus type and cortical depth into lower dimensional space (here, two dimensions). The 2D embeddings afford an intuitive understanding of the representational structure; for example, if the point representing cars at the superficial depth is close to the point representing cars at the middle depth, we can infer that the response maps for cars from these two depths are positively correlated. However, because information is necessarily lost in the dimensionality reduction, the MDS embeddings are used only for visualization and statistical tests are based on the full-dimensionality RDMs. Finally, to facilitate comparison of embeddings across partitions, participants, and depths, all embeddings were aligned via a Procrustes transformation (Gower, 1975) to a common space (the embedding from Participant S1's lateral VTC).

Controlling for anatomical differences in GLM variance explained. To control for potential differences in the quality of the GLM fit between lateral and medial VTC partitions, we created a variant of the dataset in which groups of vertices that matched in number and mean variance explained (R^2) were subsampled from the entire population of vertices in the anatomically defined VTC ROIs. We note that it is effectively impossible to select a subset of the data that is matched in R^2 across depths, given that superficial depths have higher BOLD signal magnitudes that translate to higher R^2 values (Kay et al., 2019). Thus, this control dataset is used strictly to evaluate the dependence of partition differences on R^2 .

For the deepest cortical depth, we first sorted all vertices by their R^2 value, then identified the partition with the fewest total vertices, and selected the 70% of vertices from that partition with the highest R^2 . From the larger partition, we again sorted vertices by their R^2 value, then computed all possible windows of sorted vertices containing the same number of vertices as selected from the smaller partition. The subset of vertices whose mean R^2 matched the target value from the smaller partition was used, such that we selected a group of vertices of exactly the same number and mean R^2 value in both partitions. These vertices, selected to match R^2 between lateral and medial VTC in the deepest depth, were then fixed to be in the same positions at the other two depths. This procedure was performed separately for each participant. Across participants, the mean difference in percentage of variance explained between lateral and medial VTC was $<0.01\%$ in the deepest depth and $<0.5\%$ in the middle ($0.14 \pm 0.08\%$) and superficial ($0.29 \pm 0.17\%$) depths. These differences were statistically indistinguishable from 0 for all depths (all t values < 1.8 , all p values > 0.13), indicating that, separately at each depth, this procedure was successful in matching the variance explained between partitions.

Simulation of standard-resolution fMRI data. One comparison of interest is between ultra-high-resolution (submillimeter) data and standard resolution (2.4 mm) data. This comparison was done by simulating the acquisition of standard-resolution data by spatially smoothing the pre-processed functional volumes using an ideal Fourier filter (10th-order low-pass Butterworth filter). Similar approaches have been implemented by other groups (Zaretskaya et al., 2018). In both the ultra-high-resolution and

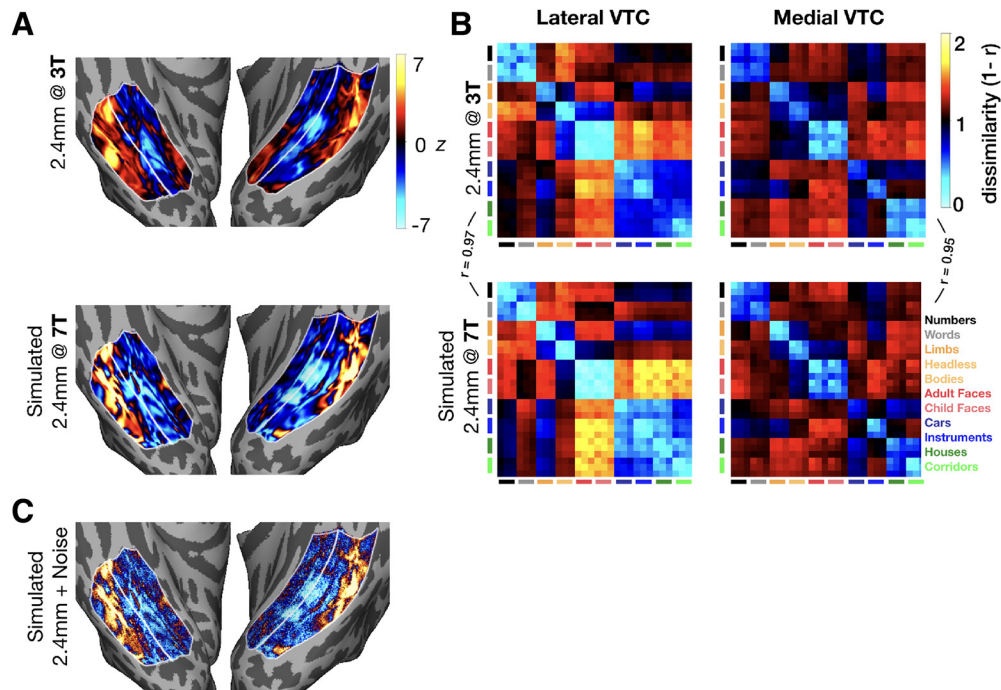


Figure 3. Validation of the Fourier-based smoothing method for simulating 2.4 mm acquisitions. **A**, z-normed response maps for an example category (child faces) are shown for Participant S1. The 2.4 mm data acquired at 3T (top) resemble the simulated 2.4 mm data acquired at 7T (bottom). **B**, RDMs in lateral VTC (left column) and medial VTC (right column) for the 2.4 mm (top) and the simulated 2.4 mm data (bottom). Bottom right, Legend indicates mapping between categories and the colors of the bars indexing rows and columns of RDMs. **C**, z-normed responses to child faces for simulated 2.4 mm data with Gaussian noise added.

simulated standard-resolution cases, the data were subsequently analyzed using the same methods as the original ultra-high-resolution data (surface-based preprocessing, GLM analysis, etc.). Simulating the 2.4 mm data from the 0.8 mm data enabled us to control for factors that may change across sessions, including the scanner and coil, respiration, fatigue, attention, and performance. For consistency with other analyses, the same vertices designated as likely veins (and thus, removed from main analyses) were removed from both the ultra-high-resolution data and the simulated standard-resolution data.

To validate the simulated 2.4 mm data, we obtained functional data for Participant S1 at 2.4 mm in a separate scan session on a 3T scanner. We then compared the simulated standard-resolution data with the true standard-resolution data by computing RDMs as described above and computing the correlation coefficient between lower triangles of the RDMs. A sample category preference map (Fig. 3A) and RDMs (Fig. 3B) are shown for both the measured and simulated 2.4 mm data to demonstrate the similarity between simulated and actual 2.4 mm data. Due to higher SNR at lower resolution, RDM magnitudes (defined as the mean dissimilarity across all cells) were lower for the simulated 2.4 mm than 0.8 mm data. To control for this difference, we added noise to the simulated 2.4 mm data to make the RDM magnitudes more comparable with those in the 0.8 mm data. To generate this dataset, referred to as “simulated 2.4 mm + noise,” we added white Gaussian noise to the category preference maps. We used different magnitudes of noise and report results with noise normally distributed around 0 with a SD of 2.5. The SD of the Gaussian was selected such that the resultant data produced RDM magnitudes that were statistically indistinguishable from those in the original data.

Statistical analyses

When data from two conditions (e.g., lateral VTC vs medial VTC) were directly compared, we used paired *t* tests, in which each pair consists of the two values obtained for each participant. When data consisted of multiple factors, we used repeated-measures ANOVAs, with participant number as a random effect and other factors as fixed effects. Where cortical depth is included as an independent variable, it is coded as a continuous regressor, as opposed to a multilevel factor, thus making the

analysis an ANCOVA. Effect sizes are reported as η^2 , the proportion of variance in the dependent variable that is explained by the term or interaction of terms specified. *Post hoc* tests were done by computing Tukey’s Honest Significant Difference where the familywise confidence level is set at 0.95. *p* values reported from these *post hoc* tests are thus corrected for multiple comparisons.

Statistics were computed using the *scipy* package for Python (<https://www.scipy.org/>) and the *nlme* and *sjstats* package for scripts in R (<https://cran.r-project.org/web/packages/nlme/nlme.pdf>; <https://cran.r-project.org/web/packages/sjstats/index.html>).

Results

During an ultra-high-resolution fMRI experiment, participants performed an oddball detection task while viewing images of exemplars from five domains (characters, bodies, faces, objects, and places). Each domain was composed of two categories, where exemplars of each category had similar parts and configuration (Fig. 1). We estimated the response magnitude to each category using a GLM (see Materials and Methods) at each vertex in lateral and medial VTC, the two anatomically defined partitions examined in this study. As VTC anatomy (Glasser and Van Essen, 2011; Saygin et al., 2011, 2016; Caspers et al., 2013; Weiner et al., 2014; Lorenz et al., 2017) and function (Martin et al., 1996; Levy et al., 2001; Hasson et al., 2002; Nasr et al., 2011; Connolly et al., 2012; Weiner et al., 2014) vary across a lateral-medial axis, we examined visual representations in these two partitions separately (see Materials and Methods). To examine distributed responses across each anatomical partition, we computed z-normed (see Materials and Methods) response maps within each partition and visualized the distributed responses to each category across VTC (Fig. 4).

Visualizations of response maps on the cortical surface show fine-scale structure in VTC that is largely consistent with prior characterizations of its functional architecture (Fig. 4A) (Haxby

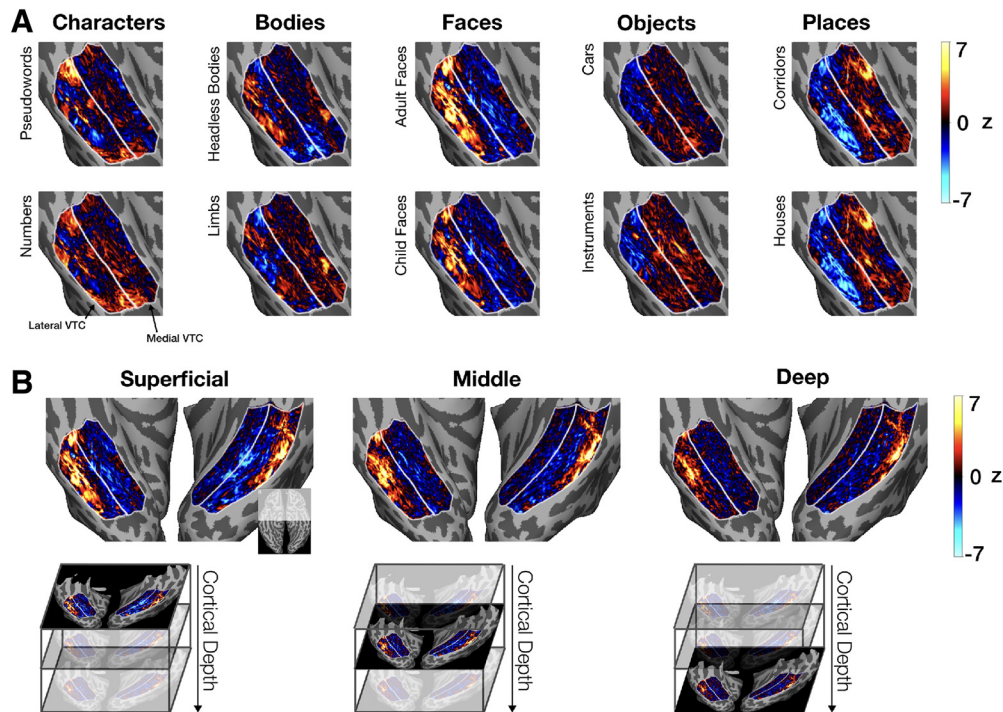


Figure 4. Response maps in VTC shown on the inflated cortical surface of Participant S1. **A**, The z-normed response map for each category is shown for the right hemisphere VTC of a representative participant (S1) at the superficial cortical depth. White lines indicate the lateral and medial VTC partitions. **B**, Maps of z-normed responses for adult faces in both hemispheres of Participant S1 shown across cortical depths: superficial, middle, and deep. Top left inset, The cortical region being visualized with respect to the entire ventral cortical surface.

et al., 2001; Grill-Spector and Weiner, 2014). Ultra-high-resolution acquisition also allowed us to not only investigate differences in responses across lateral and medial VTC, but also across different cortical depths (Fig. 4B). Response maps to each of the categories from three cortical depths (superficial, middle, and deep) suggest both consistency between the spatial profile of category responses across cortical depth (e.g., face-responsive vertices are found in the lateral fusiform gyrus at all depths) as well as differences across depths (e.g., the number of face-responsive vertices decreases between the superficial and deep depths; Fig. 4B).

An advantage of ultra-high-resolution fMRI is that it allows identification of likely veins as vertices with low mean signal, which are more prevalent in superficial cortical depths (Fig. 2A). Thus, in subsequent analyses, we examined response maps, excluding vertices that were identified as likely veins (see Masking of potential venous artifacts).

How reliable are distributed responses measured at ultra-high-resolution?

Given the novelty of ultra-high-resolution measurements throughout human VTC, we sought to establish the split-half reliability of distributed response maps used in our analyses. Here, reliability refers to the correlation of distributed responses from independent halves of the data in which participants viewed randomly selected images from each category. Thus, this measure can be thought of as an estimate of the generalizability of distributed responses across different images from these categories. We computed reliability separately for each category at each depth and then averaged across categories and depths in each participant. The reliability of distributed responses is significantly >0 in both lateral VTC ($r = 0.38 \pm 0.04$, mean \pm SEM, across participants; $t_{(6)} = 11.4$, $p = 2.7 \times 10^{-5}$) and medial VTC ($r = 0.18 \pm$

0.02 , $t_{(6)} = 9.2$, $p = 9.5 \times 10^{-5}$), indicating that distributed responses in both partitions are reliable and generalize across exemplars of a category.

What is the nature of representations in lateral and medial VTC?

To examine distributed VTC representations across partitions and depths, we computed RDMs (Kriegeskorte et al., 2008a). RDMs quantify the dissimilarity between distributed response maps computed at different depths and to different stimuli. Each cell in the RDM indicates the dissimilarity ($1 - \text{correlation}$) between a given pair of response maps across independent halves of the data. Because 10 categories and three depths are compared from each half of the data, each RDM consists of 900 cells.

We considered several possible representational structures in VTC. Schematized RDMs reflecting hypotheses consistent with domain-level, category-level, or depth-level similarity are illustrated in Figure 5A, B, and C, respectively. Other possible representational structures can be described by a combination of these hypothetical RDMs. We tested which of these hypothesized representational structures (Fig. 5) might be consistent with the measured ultra-high-resolution fMRI data. To more intuitively visualize these 900-cell RDMs, we also computed 2D embeddings of the same hypothetical representational structures using MDS (Fig. 5D–F; see Materials and Methods).

We found that RDMs are strikingly similar across participants: the mean correlation between the lower triangle of RDMs (excluding the diagonal) from different participants is 0.88 ± 0.01 in lateral VTC and 0.78 ± 0.02 in medial VTC (mean \pm SEM). By removing the diagonal, the common difference between low on-diagonal and high off-diagonal dissimilarities no longer dominates the correlation coefficient; rather, it is the similarity among pairs of patterns across participants that

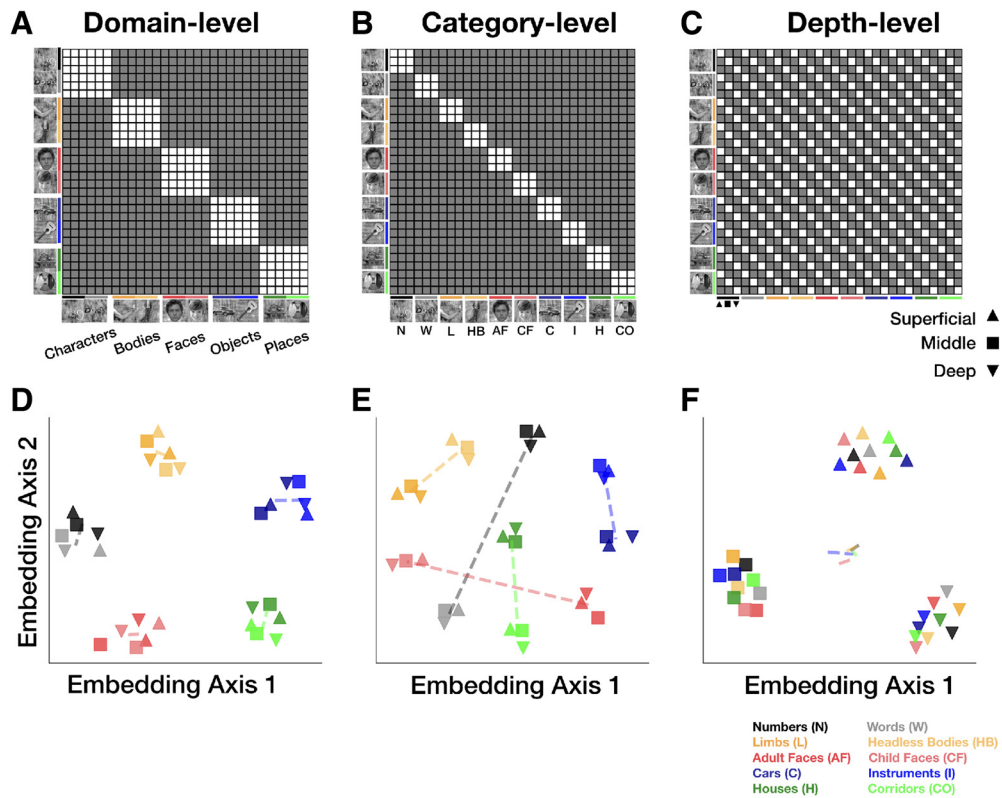


Figure 5. Hypothetical representational structures. Expected representational dissimilarity matrix (RDM) structure for (A) domain-level, (B) category-level, and (C) depth-level distributed representations. Filled (gray) cells represent higher dissimilarities than empty (white) cells. Each RDM is arranged by stimulus type (domain, then category) and depth (from superficial to deep). Thus, starting from the top left, each 3×3 square represents the similarity among items from the same category across the 3 depths, and each 6×6 square represents the similarity among items from the same domain across categories and depths. D–F, 2D MDS embeddings of the RDMs (in A–C) after adding white noise to the RDMs to prevent points from being perfectly overlapping. Colors represent the categories (same as in A–C). Shapes represent depths as follows: upward triangle represents superficial; square represents middle; downward triangle represents deep. Dashed lines connect the centroids of each pair of categories from a domain.

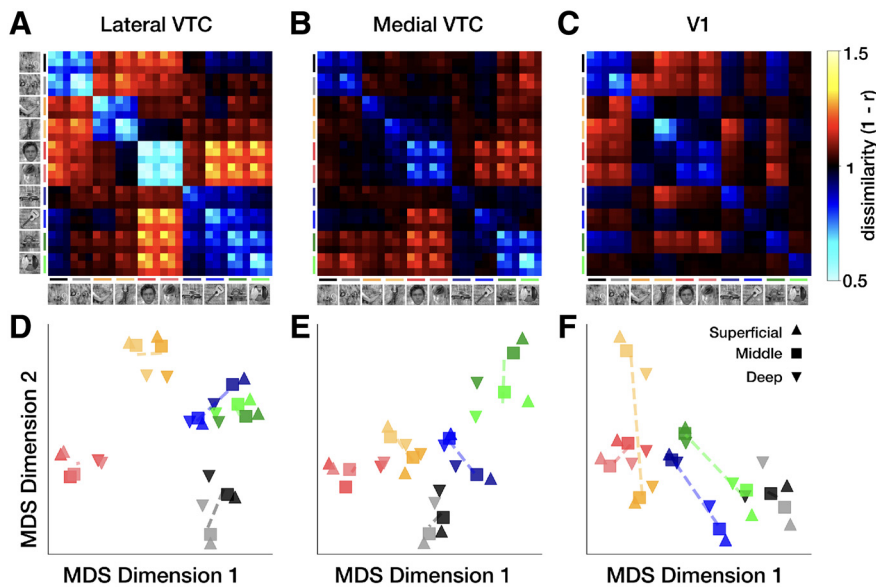


Figure 6. Representational structure in lateral and medial VTC. Group-averaged RDMs in (A) lateral VTC, (B) medial VTC, and (C) V1. Color of each cell represents pairwise dissimilarity between distributed representations across independent halves of the data. D, 2D MDS embeddings of the RDMs for lateral VTC, (E) medial VTC, and (F) V1. Colors represent categories as follows: black represents numbers; gray represents words; dark yellow represents limbs; yellow represents bodies; dark red represents adult faces; red represents child faces; dark blue represents cars; blue represents string instruments; dark green represents houses; green represents corridors. Shapes represent depths as follows: upward triangle represents superficial; square represents middle; downward triangle represents deep. Dashed lines connect the centroids of each pair of categories from a domain.

is being compared (Ritchie et al., 2017). Given the similarity of RDMs from different participants, data in subsequent Figures 6 and 9 show group-averaged RDMs. The group-averaged RDMs are only used for visualization, and all statistical analyses are based on individual participant data.

The group-averaged RDM from lateral VTC shown in Figure 6A appears largely consistent with the domain-level representation illustrated in Figure 5A. This is particularly evident for faces, bodies, characters, and places, but less so for objects, which also have low representational dissimilarity to places. In contrast, category-level representations appear stronger in medial VTC as shown in Figure 6B, which is more consistent with the predictions illustrated in Figure 5B. Neither VTC partition demonstrates representations that are modulated only by depth, as schematized in Figure 5C. Instead, we observe similarity among distributed representations for items of the same category and the same domain (especially in lateral VTC) across cortical depths.

Two-dimensional MDS embeddings of the group-averaged RDMs clarify the mixture of domain-level and category-level representations in each VTC partition. In lateral VTC, we observe a hierarchical structure to the embedding: with the exception of objects (cars and instruments), stimuli from a common domain are clustered, with some additional separation of categories within the domain and across depths (Fig. 6D–F). For example, the domain of characters forms a cluster; and its two constituent categories, pseudowords (black) and numbers (gray), are additionally separable within the cluster. Additionally, the deep layer (downward triangles) seems further separated from the others. In medial VTC, on the other hand, domain-level clusters are less clear, as the separation between categories and domains is similar (Fig. 6E). For example, the separation between bodies and faces is similar to the separation between bodies and limbs.

An important question is whether the representational structure evident in VTC is simply inherited from early retinotopic areas. To examine this question, we defined a V1 ROI on the basis of cytoarchitectonic boundaries (human occipital cytoarchitectonic area 1) (Hinds et al., 2008; Rosenke et al., 2018). We reasoned that finding similar RDMs across V1 and VTC partitions would indicate that the representational structure is inherited from earlier visual stages. While we observe low representational dissimilarity between images of the same category (3×3 on-diagonal blocks), even in V1 (Fig. 6C), the RDM structure is different between V1 and VTC partitions. The highest dissimilarity in V1 is between categories whose typical visual field extent differed; E.g., representations of bodies (which tend to extend vertically), are most dissimilar from representations of cars, houses, and characters (which tend to extend horizontally). In contrast, in both medial and lateral VTC, the largest dissimilarities are between representations of faces and of the inanimate categories cars, houses, and corridors. To more thoroughly quantify the similarity of RDMs from different regions, we computed correlations between the lower triangles (excluding the diagonal) between the V1, lateral VTC, and medial VTC RDMs in each participant. We found that the correlation between the V1 RDM and the RDMs of either lateral (average across participants = 0.56 ± 0.02) or medial VTC ($r = 0.44 \pm 0.05$) is lower than the correlation between the lateral and medial VTC RDMs ($r = 0.70 \pm 0.03$; both two-sample t tests, t values > 4.5 , both p values < 0.01). Together, these analyses demonstrate that domain and category structure in VTC does not simply mirror differences among stimuli that are already present in V1.

We next quantified the amount of domain-level, category-level, and depth-level structure in the RDMs by using a regression analysis in which each participant's RDM is predicted from a linear combination of idealized model RDMs (Fig. 7A). This approach allows us to jointly estimate the degree to which each model RDM contributes to the structure in the measured RDMs. Our goal was not to perform an exhaustive search over candidate models to identify which best fits the data reported here; instead, we sought to test the specific hypotheses that relate to the hierarchical nature of the stimulus set used in this experiment.

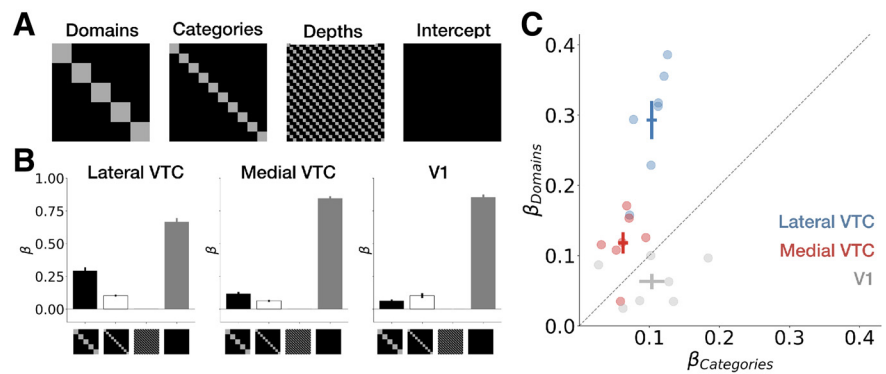


Figure 7. Regression analysis of RDM structure. **A**, Diagram indicating the model RDMs that were linearly combined to fit measured RDMs. Black represents 1; gray represents 0. **B**, Average weights \pm SEM across participants in each ROI. **C**, Individual participants' weights for the domain and category regressors. Each dot represents a single participant; cross represents mean across participants ± 1 SEM along each axis; dashed line represents unity line; color represents ROI.

The contributions of domain, category, and depth to RDM structure are summarized in Figure 7B. In general, across all participants and ROIs, the weights estimated for the depth regressor are close to 0 (lateral VTC: 0.005 ± 0.0005 , mean \pm SEM across participants; medial VTC: 0.003 ± 0.0003 ; V1: 0.001 ± 0.002), suggesting that the within-depth similarity hypothesis (Fig. 5C) does not describe our data well. Weights fit to the domain, and category regressors are substantially larger and significantly differ across ROIs: a two-way ANOVA on weights with factors of ROI (lateral VTC/medial VTC/V1) and regressor (domain/category/depth) reveals a main effect of partition ($F_{(2,54)} = 29.1$, $p = 2.7 \times 10^{-9}$, $\eta^2 = 0.15$), a main effect of ROI ($F_{(2,54)} = 92.6$, $p = 3.6 \times 10^{-18}$, $\eta^2 = 0.47$), and a significant ROI \times regressor interaction ($F_{(4,54)} = 23.8$, $p = 2.3 \times 10^{-11}$, $\eta^2 = 0.24$). This interaction remains significant when disregarding the low-magnitude depth weights; that is, including only weights fit to the domain and category regressors ($F_{(2,36)} = 22.8$, $p = 4.1 \times 10^{-7}$, $\eta^2 = 0.28$), and when removing the control ROI (V1) from the analyses ($F_{(1,24)} = 14.5$, $p = 0.001$, $\eta^2 = 0.12$). These analyses provide further support for differential representational structure across VTC partitions.

The regression analysis shows that both category and domain information contribute to the RDM structure in the VTC ROIs. Notably, domain-level structure is more pronounced than category-level structure in lateral VTC: a *post hoc* test (Tukey's honest significant difference) on the two-way ANOVA described above shows that domain weights are significantly larger than category weights in lateral VTC ($p = 4.2 \times 10^{-7}$). In medial VTC, however, domain-level structure and category-level structure are statistically indistinguishable in magnitude ($p = 0.14$). Indeed, a scatter plot of the contribution of domain and category structure to the RDMs in each participant (Fig. 7C) demonstrates that the contribution of domain was larger than category in both partitions, but more so in lateral VTC than medial VTC for each of the individual participants.

These results held (all significant effects remained significant and all nonsignificant effects remained nonsignificant) for data that were not subject to the vein-removal procedure. In particular, the regressor \times partition interaction remains significant ($F_{(1,24)} = 14.9$, $p = 0.001$, $\eta^2 = 0.12$).

To facilitate comparison with prior research and verify the robustness of our results, we also recomputed these metrics after reconstructing the RDMs from five other commonly used met-

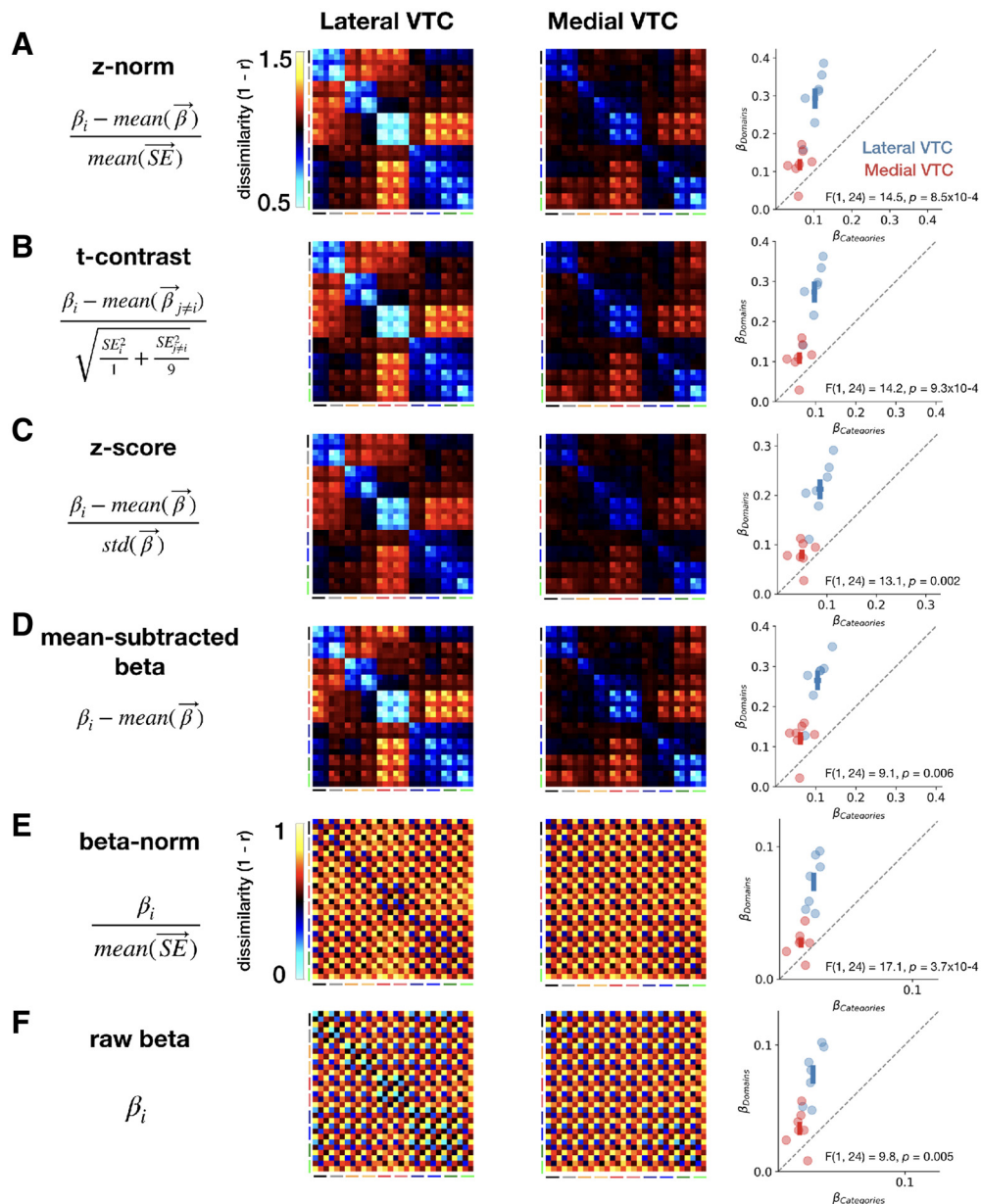


Figure 8. RDMs and regression analysis fits of the RDM for multiple metrics. In each row: leftmost column, metric; middle columns, RDMs for lateral and medial VTC, respectively; rightmost column, β weights of category and domain contributions to the RDM (see regressors in Fig. 7A). Colors represent ROI, crosses represent mean across participants \pm 1 SEM along each axis. **A**, z-norm metric. **B**, t-contrast metric. **C**, z-score metric. **D**, Mean-subtracted β metric. **E**, β -norm metric. **F**, Raw β metric.

rics, including raw betas and category-preference maps (Table 1). Representational structure for the other metrics is shown in Figure 8. For metrics that do not subtract the responses of other categories (i.e., β -norm and raw β), the weight fit to the depth regressor ranged between 0.23 and 0.32. For the remaining four metrics, the range of the depth weight was 0.002–0.007, which includes the z-norm metric we primarily use. Critically, we found that, for all six metrics, there is a significant interaction between VTC partition and regressor: all F values $>$ 9.0, all p values $<$ 0.01, all η^2 values $>$ 0.08 (Fig. 8), indicating that there is differential domain and category information across lateral and medial VTC.

Does representational structure in VTC vary with cortical depth?

While the depth-specific model RDM does not fit our measured RDMs well, the relative contributions of domains and categories

might still vary across cortical depth. Recent fMRI studies of early visual cortex have identified heterogeneity in function across cortical depths (Muckli et al., 2015; Fracasso et al., 2016; Klein et al., 2018), prompting us to test whether the differences we found between lateral and medial VTC might also depend on cortical depth. Thus, we examined visual representations separately for each cortical depth in both VTC partitions.

We split each RDM into three depth-specific RDMs: one for each of the superficial, middle, and deep depths (Fig. 9A–C, lateral VTC; and Fig. 9D–F, medial VTC). Examination of the RDMs in Figure 9 reveals that category and domain distinctions are more prominent in superficial than deep layers: within-domain and within-category dissimilarities are lower (cyan) in the superficial than deep layers, whereas between-category and between-domain dissimilarities are higher (yellow) in the superficial than deep layers.

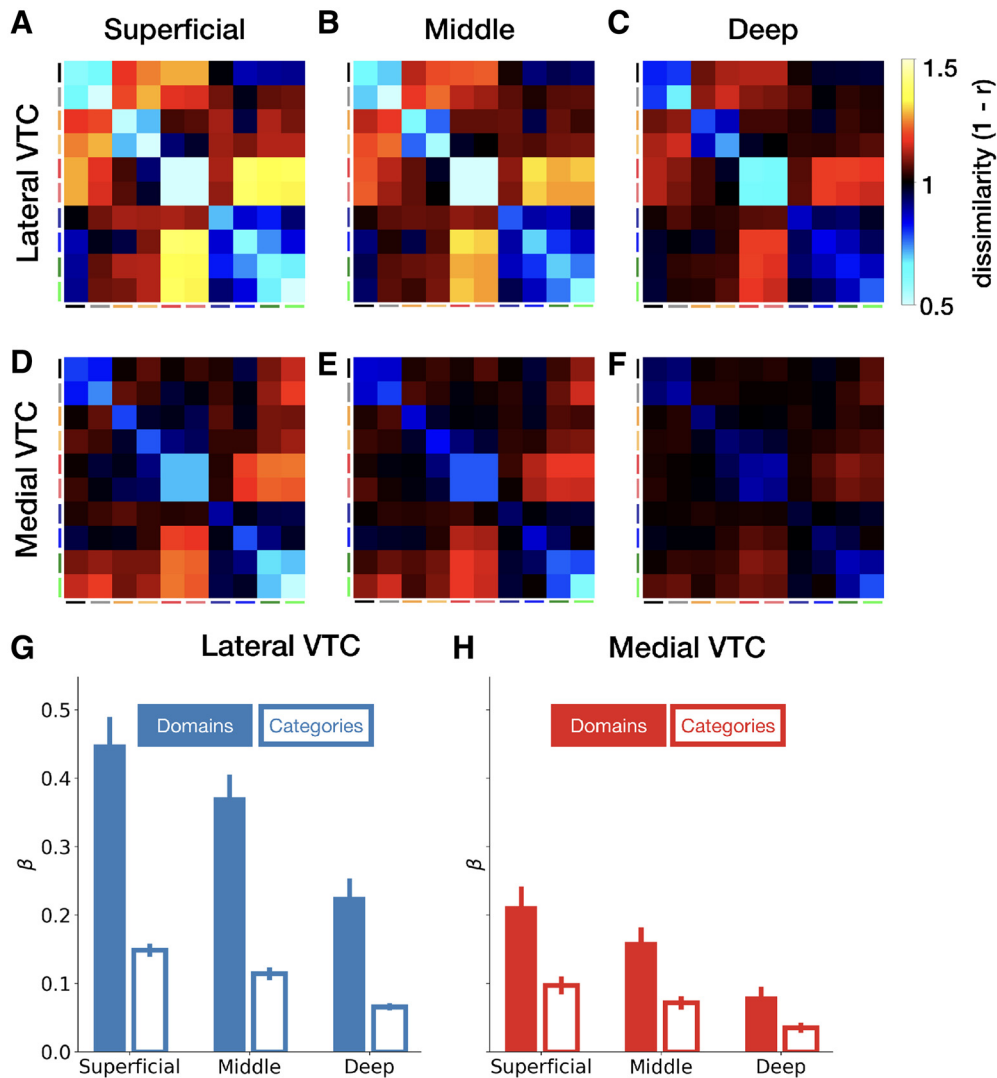


Figure 9. Representational structure in VTC partitions across cortical depths. Group-averaged RDM for lateral VTC (A–C) and medial VTC (D–F) at each cortical depth. Average domain and category weights for lateral VTC (G) and medial VTC (H) across depths. Error bars indicate SEM across participants. Category order matches Figure 6.

We quantified the contribution of domain and category structure for each depth and partition using a regression analysis in which each RDM is predicted as a weighted sum of idealized domain-level and category-level RDMs. In both lateral (Fig. 9G) and medial VTC (Fig. 9H), weights for the domain-level predictor are larger than category-level weights, and both domain and category weights decrease with depth. Notably, the difference between domain-level and category-level structure is largest in superficial lateral VTC and smallest in deep medial VTC. In contrast, deep lateral VTC and superficial medial VTC are well matched in domain-level and category-level structure (compare Fig. 9G, deep, with Fig. 9H, superficial). However, this structure is driven by different representational distinctions. For example, in deep lateral VTC (Fig. 9C), words and numbers are dissimilar from faces and bodies; but in superficial medial VTC, words and numbers are most dissimilar from places (Fig. 9D).

To quantify the effect of depth on representational structure in each VTC partition, we conducted a three-way ANCOVA on regression weights with partition (lateral VTC/medial VTC) and regressor (domain/category) as factors and depth as a continuous covariate. We found a main effect of depth ($F_{(1,76)} = 60.8, p = 2.7 \times 10^{-11}, \eta^2 = 0.14$), whereby fit β weights decrease by 0.04

per unit depth (e.g., from superficial to middle). Consistent with the previous analysis that aggregated across depths, we also found a main effect of partition ($F_{(1,76)} = 83.8, p = 6.8 \times 10^{-14}, \eta^2 = 0.2$), a main effect of regressor ($F_{(1,76)} = 152.9, p = 7.0 \times 10^{-20}, \eta^2 = 0.36$), and a partition \times regressor interaction ($F_{(1,76)} = 35.7, p = 6.9 \times 10^{-8}, \eta^2 = 0.08$). Finally, we found a regressor \times depth interaction ($F_{(1,76)} = 10.8, p = 0.002, \eta^2 = 0.03$), suggesting that domain and category weights decrease at different rates. We did not find a partition \times regressor \times depth interaction ($F_{(1,76)} = 1.2, p = 0.28$), indicating that this decrease is similar in both lateral and medial VTC.

Does representational structure in VTC vary across domains?

The results presented thus far summarize representational structure across all 10 categories and 5 domains. However, as there are differences across domains in visual similarity (e.g., child and adult faces are visually similar, but cars and guitars are visually dissimilar) and abstraction level (e.g., faces can be considered a basic-level category, and places a superordinate one), we further quantified representational structure separately by domain. For each domain, we extracted the dissimilarities between representations of exemplars within the same category and domain (e.g.,

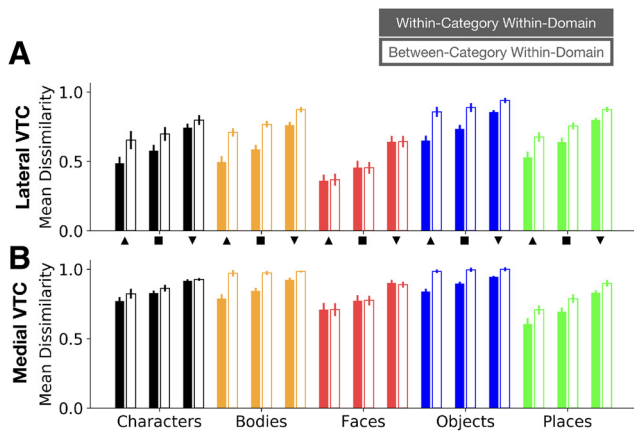


Figure 10. Within-category/within-domain and between-category/within-domain representational dissimilarities in VTC. Mean representational dissimilarity across participants in lateral VTC (**A**) and medial VTC (**B**) for each domain and each depth. Error bars indicate SEM across participants. Filled bars represent mean within-category/within-domain dissimilarity; hollow bars represent mean between-category/within-domain dissimilarity. Dissimilarities are plotted for each depth. Upward triangle represents superficial depth; square represents middle depth; downward triangle represents deep depth. Colors and *x*-axis labels represent domains.

different houses) as well as the dissimilarities between representations of exemplars from different categories within the same domain (e.g., houses vs corridors). The dissimilarities were calculated across independent split halves of the data, separately for each domain and depth (Fig. 10).

We observe three main effects: (1) across domains and partitions, both within-category/within-domain (Fig. 10, filled bars) and between-category/within-domain (Fig. 10, empty bars) representational dissimilarities are lowest in superficial VTC and highest in deep VTC; (2) with the exception of faces (Fig. 10, red bars), between-category/within-domain dissimilarities are larger than within-category/within-domain dissimilarities across depths; and (3) dissimilarities vary across domains and partitions. We quantified these observations using a four-way ANCOVA with factors of partition (lateral/medial VTC), domain (characters/bodies/faces/objects/places), and dissimilarity type (within-category/between-category), with depth as a continuous covariate. In support of our observations, we found: (1) a main effect of depth ($F_{(1,380)} = 227.7, p = 1.2 \times 10^{-40}, \eta^2 = 0.13$), whereby dissimilarity increases by 0.05 per unit depth, (2) a significant interaction between domain and within- versus between-category dissimilarity ($F_{(4,380)} = 12.0, p = 3.2 \times 10^{-9}, \eta^2 = 0.03$), and (3) both a main effect of domain ($F_{(4,380)} = 81.5, p = 6.9 \times 10^{-50}, \eta^2 = 0.18$) and a significant domain \times partition interaction ($F_{(4,380)} = 35.4, p = 4.1 \times 10^{-25}, \eta^2 = 0.08$), indicating that the representational dissimilarities for these five domains differ between partitions. Together, these results provide evidence that depth effects are consistent across domains, categories, and partitions, but representational structure varies across domains and VTC partitions.

Can differences in measurement quality explain differences between partitions and depths?

As veins are more prevalent in superficial than deeper cortical depths (Fig. 2A) (Duvernoy et al., 1981; Lauwers et al., 2008; Polimeni et al., 2010; Siero et al., 2011; Goense et al., 2016; Kay et al., 2019), it is important to consider how our results may depend on cortical vasculature. In this section, we consider whether measurement quality, which is affected by draining veins, impacts the

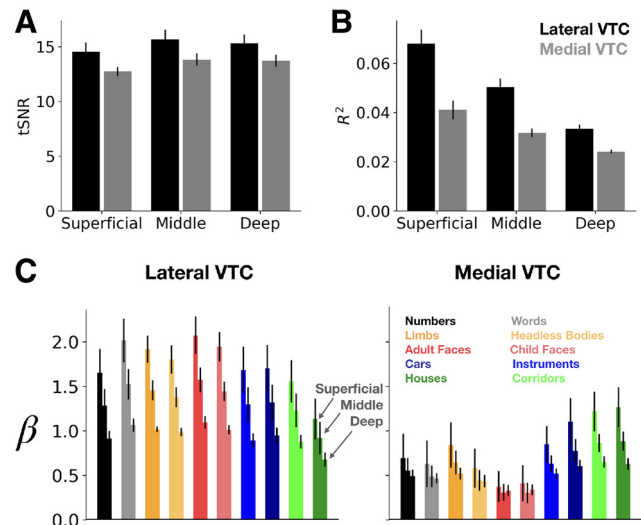


Figure 11. tSNR, variance explained, and response amplitude across cortical depth. Each metric was calculated at each vertex and then averaged across vertices of the ROI. tSNR (**A**) and variance explained by the GLM (**B**) in lateral and medial VTC at the three cortical depths examined in this study. **C**, Mean response amplitudes to the 10 stimulus categories in lateral and medial VTC. Normed amplitudes were averaged first across vertices in an ROI and then across participants. Error bars indicate SEM across participants. Bar colors represent stimulus category (see legend at right). Each triplet of bars sharing a color proceed through depth (superficial, middle, and deep) from left to right.

results. To determine whether measurement differences across depths and partitions can explain the data, we compared two additional metrics in each VTC partition and depth: (1) the mean time-series tSNR, which is insensitive to specification of the GLM used to estimate responses, and (2) the mean variance explained (R^2) by the GLM. Further, we also computed the mean response amplitude to each stimulus type to characterize potential differences in mean response magnitude across each ROI. We reasoned that, if any of these measurements could be explained by an interaction between (1) partition and depth or (2) partition and category, we should further test whether these factors account for our results.

Across partitions and cortical depths, the mean tSNR across participants ranges between 12.7 and 15.7 (Fig. 11A), consistent with prior reports at similar voxel volumes (Triantafyllou et al., 2005). A two-way ANCOVA with partition (lateral/medial VTC) and depth as a continuous covariate indicated no significant effect of depth on tSNR ($F_{(1,38)} = 1.46, p = 0.23$) and no partition \times depth interaction ($F_{(1,38)} = 0.016, p = 0.90$). This analysis suggests that tSNR cannot explain differences in representations across cortical depths or partitions. A similar analysis (two-way ANCOVA with partition and depth as a continuous covariate) on variance explained by the GLM (R^2 ; Fig. 11B) revealed a main effect of partition ($F_{(1,38)} = 42.5, p = 1.1 \times 10^{-7}, \eta^2 = 0.30$), a main effect of depth ($F_{(1,38)} = 56.7, p = 4.8 \times 10^{-9}, \eta^2 = 0.39$), and a significant interaction between partition and depth ($F_{(1,38)} = 6.6, p = 0.01, \eta^2 = 0.05$). As we found differences in representational structure across lateral and medial VTC, we tested whether our results hold after reexamining domain-level and category-level structure in a subset of our data that was closely matched in mean R^2 across partitions (see Material and Methods). Recomputing domain and category weights for RDMs generated from these R^2 -matched data replicate our results: we observe a significant regressor (domain/category) \times partition (lateral/medial VTC) interaction ($F_{(1,24)} = 15.1, p = 7.1 \times 10^{-4}$,

$\eta^2 = 0.12$), whereby domain-level structure exceeds category level structure in lateral, but not medial, VTC. Thus, differences between representations across partitions are likely not driven by differences in R^2 .

Inspection of GLM betas (% BOLD signal change) by category, depth, and partition (Fig. 11C) reveals that, as previously reported (Polimeni et al., 2010; Kay et al., 2019), response amplitudes decrease across cortical depth. A three-way ANCOVA on the average % BOLD signal change (averaged across all vertices in the partition of interest) with category and partition (lateral VTC/medial VTC) as factors and depth as a continuous covariate revealed a main effect of depth ($F_{(1,380)} = 106.7, p = 3.2 \times 10^{-22}, \eta^2 = 0.12$), and of partition (lower amplitudes in medial than lateral VTC, $F_{(1,380)} = 273.8, p = 1.1 \times 10^{-46}, \eta^2 = 0.31$). There is no main effect of category ($F_{(9,380)} = 0.93, p = 0.50$), and no category \times depth interaction ($F_{(9,380)} = 0.12, p > 0.99$), indicating that response amplitude was well matched across conditions and decreased with depth evenly across all conditions. We also found a significant category \times partition interaction ($F_{(9,380)} = 7.795, p = 1.47 \times 10^{-10}, \eta^2 = 0.08$), which is driven by higher responses to inanimate (cars, guitars, houses, and corridor) than animate stimuli in medial, but not lateral, VTC (Fig. 11C). However, this overall higher mean response to inanimate versus inanimate categories in medial VTC does not explain the categorical structure in medial VTC, as distributed responses to houses, guitars, and cars are dissimilar from one another (Fig. 6B), even as their mean responses are similar. Together, these analyses suggest that differential category and domain structure across medial and lateral VTC cannot be explained by tSNR, variance explained by the GLM, or mean response amplitudes to certain categories. However, depth effects are coupled with lower variance explained and lower signals in deeper than superficial layers.

Is ultra-high-resolution necessary to reveal these representations?

The differences between anatomical partitions and cortical depths described thus far have been derived from submillimeter resolution data. We tested whether these findings of differential representations across depths and partitions require ultra-high-resolution acquisition, or whether the same findings would be accessible to more standard fMRI measurements at 2.4 mm resolution. Some features of these data, such as depth-dependent effects, will clearly depend on ultra-high-resolution acquisition; if the voxel size matches the thickness of the cortical sheet (between 2 and 3 mm) (Fischl and Dale, 2000), functional differences at different cortical depths would be challenging, if not impossible, to identify with 2.4 mm resolution. The differences between lateral and medial VTC in the level of abstraction of their representations, however, may be apparent at lower spatial resolutions as well.

To investigate differences between visual representations at ultra-high-resolution (0.8 mm voxels) and standard-resolution (2.4 mm voxels), we downsampled each participant's functional data to 2.4 mm using a Fourier-space smoothing method (see Materials and Methods). Downsampling a fixed set of data, as opposed to collecting a separate dataset, has the advantage of controlling for session-to-session variability in attention, fatigue, and performance (Peelen and Downing, 2005; Weiner et al., 2016; Zaretskaya et al., 2018). The simulated 2.4 mm data were projected onto the same FreeSurfer surfaces as the high-resolution data to facilitate comparison. Additionally, the entire experiment was repeated at 3T to validate this downsampling method in Participant S1 (Fig. 3).

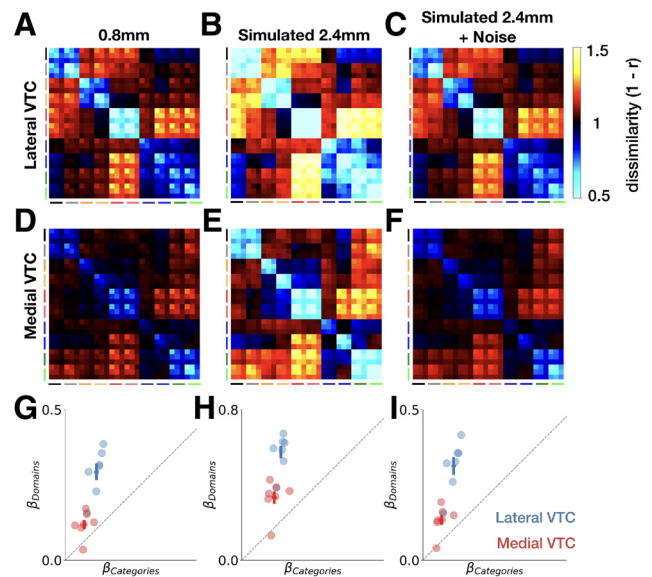


Figure 12. Effects of spatial resolution on representational structure. Group-averaged RDM for lateral VTC at (A) 0.8 mm, (B) simulated 2.4 mm, and (C) simulated 2.4 mm with added Gaussian noise. D–F, Same as in A–C, but for medial VTC. Model RDM fits for domains and categories for lateral VTC (blue) and medial VTC (red) at (G) 0.8 mm, (H) simulated 2.4 mm, and (I) simulated 2.4 mm with added Gaussian noise. Crosses represent ± 1 SEM across participants.

How closely do the simulated 2.4 mm data match data acquired natively at 2.4 mm? The simulated and measured 2.4 mm maps have similar spatial structure, mean absolute RDM magnitude (the average dissimilarity in the RDM), and RDM structure (Fig. 3). Indeed, the correlation between the lower triangles (excluding the diagonal) of RDMs computed from Participant S1's measured 3T data and downsampled 7T data is $r = 0.97$ in lateral VTC and $r = 0.95$ in medial VTC, indicating high consistency in the representational structure across resolutions. This comparison for Participant S1 suggests that the downsampled data in the remaining 6 participants are an appropriate approximation of the data one may observe in an experiment done with 2.4 mm data acquired at 3T.

Comparing the ultra-high-resolution 0.8 mm and simulated standard-resolution 2.4 mm data revealed two key results. First, the representational dissimilarities are lower overall in the simulated 2.4 mm data than in the 0.8 mm data (Fig. 12A,B and Fig. 12D,E). A two-way ANOVA on the RDM magnitudes with anatomical partition (lateral/medial) and data resolution (0.8 mm/2.4 mm) as factors revealed a main effect of resolution ($F_{(1,24)} = 54.7, p = 1.2 \times 10^{-7}, \eta^2 = 0.48$), a main effect of partition ($F_{(1,24)} = 32.2, p = 7.7 \times 10^{-6}, \eta^2 = 0.28$), but no interaction between data resolution and partition ($F_{(1,24)} = 2.4, p = 0.13$). To account for the difference in overall magnitude of representational dissimilarities between the data at each resolution, we simulated another variant of the 2.4 mm data in which we added Gaussian noise to the simulated 2.4 mm preference maps (see Materials and Methods; Fig. 12C,F). Comparison of the mean RDM magnitude for the 0.8 mm data with the 2.4 mm + noise data preserved the main effect of anatomical partition ($F_{(1,24)} = 37.6, p = 2.5 \times 10^{-6}, \eta^2 = 0.61$); critically, however, there was no effect of dataset (0.8 mm/2.4 mm + noise; $F_{(1,24)} = 0.04, p = 0.85$). Thus, in subsequent comparisons in which we also used the simulated 2.4 mm + noise data, RDM magnitude is appropriately controlled across resolutions.

Second, we found that, while RDM magnitude differs between spatial resolutions, the representational structure was largely similar. Regression analyses using the domain, category, and depth predictors (as in Fig. 7A) revealed that, in all three data preparations (0.8 mm, simulated 2.4 mm, and simulated 2.4 mm with added noise), domain weights exceed category weights more strongly in lateral than medial VTC (Fig. 12G–I). As in the 0.8 mm data, a two-way ANOVA on the regression weights with regressor (domain/category) and partition (lateral/medial VTC) as factors indicated a significant interaction when using the simulated 2.4 mm data ($F_{(1,24)} = 14.9, p = 7.4 \times 10^{-4}, \eta^2 = 0.09$). This result held with the simulated 2.4 mm + noise data ($F_{(1,24)} = 13.4, p = 0.001, \eta^2 = 0.11$).

To further investigate the effect of spatial resolution on representational structure, we directly compared weights for the domain-level and category-level regressors for each data preparation. A three-way ANOVA with partition (lateral VTC/medial VTC), data resolution (0.8 mm/2.4 mm), and regressor (domain/category) indicated main effects of all three factors, as expected from prior analyses (all p values $< 1 \times 10^{-9}$). Critically, we found no evidence for a three-way interaction ($F_{(1,48)} = 1.3, p = 0.26$), which would have suggested that the regressor \times partition interaction reported here varies with data resolution. Adding noise to the simulated 2.4 mm data yields the same results: the three-way interaction remained nonsignificant ($F_{(1,48)} = 0.01, p = 0.92$). Together, these analyses demonstrate that representational differences of categories and domains across lateral and medial VTC are robust across resolutions from 0.8 to 2.4 mm.

Discussion

Using ultra-high-resolution fMRI, we measured responses in VTC to a hierarchically structured stimulus set of five domains, each consisting of two categories. These measurements allowed us to compare submillimeter representations both parallel to the cortical surface and through cortical depth, revealing three novel results. First, we find that anatomically distinct lateral and medial partitions of VTC represent the visual input at multiple levels of abstraction: there was both domain structure and more granular category structure in each VTC partition. However, domain-level representations were stronger in lateral VTC than in medial VTC. Second, both domain-level and category-level structure decreased with cortical depth. Third, downsampling the data to 2.4 mm mirrored the differential representational structure across VTC partitions that we observed with ultra-high-resolution. In the sections below, we discuss the utility of ultra-high-resolution fMRI for measuring category representations, relate our findings to prior research, and discuss the theoretical implications of our findings.

The strength, but not nature, of VTC representations varies with cortical depth

Submillimeter resolution enabled us to test the structure of visual representations across cortical depths in human VTC for the first time. We found a quantitative, but not qualitative, change in VTC representations across depths: while domain- and category-level structure both decrease with cortical depth, their relative contributions are similar across depths. Future studies will be needed to examine whether other factors, such as cortical thickness and curvature, might also affect representations across depths.

The stability of representations across cortical depth is consistent with findings by Tanaka, Tanifuji, and colleagues (Fujita et al., 1992; Tanaka, 1996; Wang et al., 1996; Tsunoda et al., 2001) who reported a columnar structure in macaque IT (which is

thought to be homologous to human VTC). Using optical imaging and electrophysiology, they found that neurons with similar functional properties were clustered across cortical depths. More recent ultra-high-resolution fMRI studies of early visual cortex have reported that attentional modulation of population receptive fields in V1 varies by cortical depth and is highest in deeper layers (Klein et al., 2018). Such findings suggest the possibility that tasks with higher top-down load than the present oddball task may differentially modulate VTC representations across cortical depths.

The quantitative effect of depth, which is evinced by stronger domain and category structure in superficial than deep cortical depths, is coupled with stronger response amplitudes and higher variance explained by the GLM in the superficial cortical depth. This raises the possibility that differences across depths may be explained by the underlying vasculature; that is, stronger responses in the superficial layers may be a consequence of draining veins in superficial cortex (Koopmans et al., 2010; Vu and Gallant, 2015; Kay et al., 2019). While we took care to exclude cortical vertices that are likely veins in our data, we may not have excluded all venous effects. Nonetheless, representations measured from deep cortical depths, which are predominantly free of the influence of large draining veins, resemble those from the superficial and middle depths. Thus, one important insight from our data is that the domain-level and category-level structure of VTC representations is not driven by draining veins but is a functional property of the underlying representations.

Ultra-high-resolution fMRI reveals consistency of category and domain representations across spatial resolutions

We found that the differences in representational structure between lateral and medial VTC are apparent even when downsampling our submillimeter data to a standard fMRI resolution of 2.4 mm. This direct comparison of 0.8-mm-resolution data to 2.4-mm-resolution data suggests that distributed representations in VTC studied with standard resolution fMRI are consistent with representations probed with measurements at a finer scale. This finding is hardly trivial. It suggests that the vast literature on domain and category information in human VTC (Malach et al., 1995; Kanwisher et al., 1997; Aguirre et al., 1998; Epstein and Kanwisher, 1998; Cohen et al., 2000; Haxby et al., 2001; Cox and Savoy, 2003; Peelen and Downing, 2005; Kriegeskorte et al., 2008b; Huth et al., 2012; Weiner et al., 2014) is unlikely to be a strict artifact of spatial smoothing by voxels that are several millimeters on a side. Instead, our 0.8 mm measurements validate that domain and category representations are an inherent functional property of VTC representations.

We note, however, that other aspects of VTC representations, including semantic information and visual features (Bracci and Op de Beeck, 2016; Jozwik et al., 2016; Proklova et al., 2016; Bracci et al., 2019), environmental regularities (Peelen and Downing, 2017; Kaiser et al., 2019), typical size (Konkle and Oliva, 2012), and eccentricity (Levy et al., 2001; Hasson et al., 2002; Gomez et al., 2019) might differ when examined with ultra-high-resolution fMRI. For example, we have not tested finer categorical distinctions at the subordinate scale (e.g., different types of string instruments) or exemplar-level representations (e.g., face identity across views). It is also possible that representations of object parts and features, which have been found in macaque IT with electrophysiology (Fujita et al., 1992; Tsunoda et al., 2001; Issa et al., 2013) may only be apparent with ultra-high-resolution fMRI. These hypotheses can be tested in future research.

Different visual representations across VTC partitions

Our finding that distributed representations in medial VTC differ from those in lateral VTC adds to a large body of work on anatomical and functional differences between these partitions. Anatomically, medial and lateral VTC differ in their cytoarchitectonic properties (Weiner et al., 2014, 2017; Gomez et al., 2017; Lorenz et al., 2017) and connectivity fingerprint (Saygin et al., 2011, 2016; Osher et al., 2016). Functionally, they differ in (1) eccentricity preference (Levy et al., 2001; Hasson et al., 2002; Weiner et al., 2014), (2) real-world object size preference (Konkle and Oliva, 2012), (3) stimulus animacy preference (Martin et al., 1996; Connolly et al., 2012), and (4) the domains they represent: clustered representations for faces (Kanwisher et al., 1997), words (Cohen et al., 2000; Nordt et al., 2019), and bodies (Downing et al., 2001; Weiner and Grill-Spector, 2010) are in lateral VTC, whereas clustered representations for places are in medial VTC (Aguirre et al., 1998; Epstein and Kanwisher, 1998).

Our use of a hierarchical stimulus set consisting of five domains and two categories per domain allowed us to investigate whether ecological domain (Kanwisher, 2000, 2010) or category (Haxby et al., 2001) is the dominant organizing principle of VTC. We found that both domain and category structure contribute to VTC representations and that these representations are not merely inherited from early visual areas. Critically, however, we discovered that these representations are not uniform across VTC: domain representations are more prominent than category representations in lateral VTC, but domain and category structure similarly contribute to medial VTC representations.

Differential representations across VTC partitions and depths form a functional-architectural foundation for flexible readout

What might be the utility for this difference in the granularity of representations across VTC partitions? Central to our understanding of the functional architecture of VTC are theories that link the neuroanatomical and functional implementational features of the visual system to information processing (Zeki and Shipp, 1988; Van Essen et al., 1992; Malach, 1994; Grill-Spector and Weiner, 2014). According to these frameworks, the functional divergence of representations across lateral and medial VTC may produce a diversity of representations that can be used by downstream regions for different tasks in a flexible manner. One inference from our findings is that representations in lateral VTC are better suited toward readouts that favor more abstract domain distinctions, whereas representations in medial VTC are better suited for reading out categorical distinctions. For example, finding a face in a hallway may rely on a representation that discriminates domains (faces vs places), which is prominent in lateral VTC. However, deciding whether that place is a hallway at work or a corridor in your house may depend on a representation that discriminates among place categories, such as those in medial VTC.

We note that, because we used categories and domains that have been typically used in prior research and are associated with ecological constructs, the domains are not controlled for visual similarity or level of abstraction. For example, the face domain contains child and adult faces, which, being important social categories, are both the most visually homogeneous and most specific in terms of level of abstraction compared with the other four domains. Future experiments using carefully controlled stimuli (Jozwik et al., 2016; Proklova et al., 2016; see, e.g., recent elegant work by Bracci and Op de Beeck, 2016; Bracci et al., 2019) can

further examine the role of visual similarity and conceptual knowledge on both domain and category representations.

In conclusion, ultra-high-resolution measurements allowed us to compare representational structure across cortical depths and partitions of human high-level visual cortex. We show that anatomically distinct partitions of VTC exhibit different representational structure and that the strength of both domain and category-level representations varies with cortical depth.

References

- Aguirre GK, Zarahn E, D'Esposito M (1998) An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron* 21:373–383.
- Bracci S, Op de Beeck H (2016) Dissociations and associations between shape and category representations in the two visual pathways. *J Neurosci* 36:432–444.
- Bracci S, Ritchie JB, Kalfas I, Op de Beeck HP (2019) The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. *J Neurosci* 39:6513–6525.
- Bugatus L, Weiner KS, Grill-Spector K (2017) Task alters category representations in prefrontal but not high-level visual cortex. *Neuroimage* 155:437–449.
- Caspers J, Zilles K, Eickhoff SB, Schleicher A, Mohlberg H, Amunts K (2013) Cytoarchitectonical analysis and probabilistic mapping of two extrastriate areas of the human posterior fusiform gyrus. *Brain Struct Funct* 218:511–526.
- Cheng K, Waggoner RA, Tanaka K (2001) Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron* 32:359–374.
- Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff MA, Michel F (2000) The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123:291–307.
- Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu YC, Abdi H, Haxby JV (2012) The representation of biological classes in the human brain. *J Neurosci* 32:2608–2618.
- Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* 19:261–270.
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* 9:179–194.
- De Martino F, Yacoub E, Kemper V, Moerel M, Uludag K, De Weerd P, Ugurbil K, Goebel R, Formisano E (2018) The impact of ultra-high field MRI on cognitive and computational neuroimaging. *Neuroimage* 168:366–382.
- Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. *Science* 293:2470–2473.
- Dumoulin SO, Fracasso A, van der Zwaag W, Siero JC, Petridou N (2018) Ultra-high field MRI: advancing systems neuroscience towards mesoscopic human brain function. *Neuroimage* 168:345–357.
- Duvernoy HM, Delon S, Vannson JL (1981) Cortical blood vessels of the human brain. *Brain Res Bull* 7:519–579.
- Epanechnikov VA (2005) Non-parametric estimation of a multivariate probability density. *Theory Prob Appl* 14:153–158.
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601.
- Fischl B, Dale AM (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci* 97:11050–11055.
- Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9:195–207.
- Fracasso A, Petridou N, Dumoulin SO (2016) Systematic variation of population receptive field properties across cortical depth in human visual cortex. *Neuroimage* 139:427–438.
- Fujita I, Tanaka K, Ito M, Cheng K (1992) Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360:343–346.
- Gaillard R, Naccache L, Pinel P, Clémenceau S, Volle E, Hasboun D, Dupont S, Baulac M, Dehaene S, Adam C, Cohen L (2006) Direct intracranial, fMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron* 50:191–204.

- Glasser MF, Van Essen DC (2011) Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J Neurosci* 31:11597–11616.
- Goense J, Bohraus Y, Logothetis NK (2016) fMRI at high spatial resolution implications for BOLD-models. *Front Comput Neurosci* 10:66.
- Gomez J, Barnett MA, Natu V, Mezer A, Palomero-Gallagher N, Weiner KS, Amunts K, Zilles K, Grill-Spector K (2017) Microstructural proliferation in human cortex is coupled with the development of face processing. *Science* 355:68–71.
- Gomez J, Barnett M, Grill-Spector K (2019) Extensive childhood experience with Pokémon suggests eccentricity drives organization of visual cortex. *Nat Hum Behav* 3:611–624.
- Gower J (1975) Generalized procrustes analysis. *Psychometrika* 40:33–51.
- Grill-Spector K, Kanwisher N (2005) Visual recognition: as soon as you know it is there, you know what it is. *Psychol Sci* 16:152–160.
- Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci* 15:536–548.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203.
- Grill-Spector K, Kushnir T, Hendler T, Malach R (2000) The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 3:837–843.
- Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform face area subserves face perception, not generic within-category identification. *Nat Neurosci* 7:555–562.
- Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an organizing principle for human high-order object areas. *Neuron* 34:479–490.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.
- Hinds OP, Rajendran N, Polimeni JR, Augustinack JC, Wiggins G, Wald LL, Diana Rosas H, Potthast A, Schwartz EL, Fischl B (2008) Accurate prediction of V1 location from cortical folds in a surface coordinate system. *Neuroimage* 39:1585–1599.
- Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76:1210–1224.
- Issa EB, Papanastassiou AM, DiCarlo JJ (2013) Large-scale, high-resolution neurophysiological maps underlying fMRI of macaque temporal lobe. *J Neurosci* 33:15207–15219.
- Jezzard P, Balaban RS (1995) Correction for geometric distortion in echo planar images from B0 field variations. *Magn Reson Med* 34:65–73.
- Jozwik KM, Kriegeskorte N, Mur M (2016) Visual features as stepping stones toward semantics: explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia* 83:201–226.
- Kaiser D, Quek GL, Cichy RM, Peelen MV (2019) Object vision in a structured world. *Trends Cogn Sci* 23:672–685.
- Kanwisher N (2000) Domain specificity in face perception. *Nat Neurosci* 3:759–763.
- Kanwisher N (2010) Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc Natl Acad Sci U S A* 107:11163–11170.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Kay KN, Rokem A, Winawer J, Dougherty RF, Wandell BA (2013) GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Front Neurosci* 7:247.
- Kay K, Jamison KW, Vizioli L, Zhang R, Margalit E, Ugurbil K (2019) A critical assessment of data quality and venous effects in sub-millimeter fMRI. *Neuroimage* 189:847–869.
- Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput Biol* 10:e1003915.
- Klein BP, Fracasso A, van Dijk JA, Paffen CL, Te Pas SF, Dumoulin SO (2018) Cortical depth dependent population receptive field attraction by spatial attention in human V1. *Neuroimage* 176:301–312.
- Konkle T, Oliva A (2012) A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74:1114–1124.
- Koopmans PJ, Barth M, Norris DG (2010) Layer-specific BOLD activation in human V1. *Hum Brain Mapp* 31:1297–1304.
- Kriegeskorte N, Mur M, Bandettini P (2008a) Representational similarity analysis: connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008b) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60:1126–1141.
- Kruskal JB (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27.
- Lauwers F, Cassot F, Lauwers-Cances V, Puwanarajah P, Duvernoy H (2008) Morphometry of the human cerebral cortex microcirculation: general characteristics and space-related profiles. *Neuroimage* 39:936–948.
- Levy I, Hasson U, Avidan G, Hendler T, Malach R (2001) Center-periphery organization of human object areas. *Nat Neurosci* 4:533–539.
- Lorenz S, Caspers J, Mohlberg H, Schleicher A, Bludau S, Eickhoff SB, Zilles K, Amunts K, Weiner KS, Grill-Spector K (2017) Two new cytoarchitectonic areas on the human mid-fusiform gyrus. *Cereb Cortex* 27:373–385.
- Maass A, Schütze H, Speck O, Yonelinas A, Tempelmann C, Heinze HJ, Berron D, Cardenas-Blanco A, Brodersen KH, Stephan KE, Düzel E (2014) Laminar activity in the hippocampus and entorhinal cortex related to novelty and episodic encoding. *Nat Commun* 5:5547.
- Malach R (1994) Cortical columns as devices for maximizing neuronal diversity. *Trends Neurosci* 17:101–104.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135–8139.
- Martin A, Wiggs CL, Ungerleider LG, Haxby JV (1996) Neural correlates of category-specific knowledge. *Nature* 379:649–652.
- McGugin RW, Gatenby JC, Gore JC, Gauthier I (2012) High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proc Natl Acad Sci U S A* 109:17063–17068.
- McGugin RW, Newton AT, Gore JC, Gauthier I (2014) Robust expertise effects in right FFA. *Neuropsychologia* 63:135–144.
- Moutoussis K, Zeki S (2002) The relationship between cortical activation and perception investigated with invisible stimuli. *Proc Natl Acad Sci U S A* 99:9527–9532.
- Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E (2015) Contextual feedback to superficial layers of V1. *Curr Biol* 25:2690–2695.
- Nasr S, Liu N, Devaney KJ, Yue X, Rajimehr R, Ungerleider LG, Tootell RB (2011) Scene-selective cortical regions in human and nonhuman primates. *J Neurosci* 31:13771–13785.
- Nasr S, Polimeni JR, Tootell RB (2016) Interdigitated color- and disparity-selective columns within human visual cortical areas V2 and V3. *J Neurosci* 36:1841–1857.
- Natu VS, Gomez J, Barnett M, Jeska B, Kirilina E, Jaeger C, Zhen Z, Cox S, Weiner KS, Weiskopf N, Grill-Spector K (2019) Apparent thinning of human visual cortex during childhood is associated with myelination. *Proc Natl Acad Sci U S A* 116:20750–20759.
- Nordt M, Gomez J, Natu V, Jeska B, Barnett M, Grill-Spector K (2019) Learning to read increases the informativeness of distributed ventral temporal responses. *Cereb Cortex* 29:3124–3139.
- Osher DE, Saxe RR, Koldewyn K, Gabrieli JD, Kanwisher N, Saygin ZM (2016) Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex. *Cereb Cortex* 26:1668–1683.
- Parvizi J, Jacques C, Foster BL, Witthoft N, Rangarajan V, Weiner KS, Grill-Spector K (2012) Electrical stimulation of human fusiform face-selective regions distorts face perception. *J Neurosci* 32:14915–14920.
- Peelen MV, Downing PE (2005) Within-subject reproducibility of category-specific visual activation with functional MRI. *Hum Brain Mapp* 25:402–408.
- Peelen MV, Downing PE (2017) Category selectivity in human visual cortex: beyond visual object recognition. *Neuropsychologia* 105:177–183.
- Polimeni JR, Fischl B, Greve DN, Wald LL (2010) Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *Neuroimage* 52:1334–1346.
- Proklova D, Kaiser D, Peelen MV (2016) Disentangling representations of object shape and object category in human visual cortex: the animate-inanimate distinction. *J Cogn Neurosci* 28:680–692.

- Rangarajan V, Hermes D, Foster BL, Weiner KS, Jacques C, Grill-Spector K, Parvizi J (2014) Electrical stimulation of the left and right human fusiform gyrus causes different effects in conscious face perception. *J Neurosci* 34:12828–12836.
- Ritchie BJ, Bracci S, Op de Beeck H (2017) Avoiding illusory effects in representational similarity analysis: what (not) to do with the diagonal. *Neuroimage* 148:197–200.
- Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-Braem P (1976) Basic objects in natural categories. *Cogn Psychol* 8:382–439.
- Rosenke M, Weiner KS, Barnett MA, Zilles K, Amunts K, Goebel R, Grill-Spector K (2018) A cross-validated cytoarchitectonic atlas of the human ventral visual stream. *Neuroimage* 170:257–270.
- Saygin ZM, Osher DE, Koldeewyn K, Reynolds G, Gabrieli JD, Saxe RR (2011) Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nat Neurosci* 15:321–327.
- Saygin ZM, Osher DE, Norton ES, Youssoufian DA, Beach SD, Feather J, Gaab N, Gabrieli JD, Kanwisher N (2016) Connectivity precedes function in the development of the visual word form area. *Nat Neurosci* 19:1250–1255.
- Schneider M, Kemper VG, Emmerling TC, De Martino F, Goebel R (2019) Columnar clusters in the human motion complex reflect consciously perceived motion axis. *Proc Natl Acad Sci U S A* 116:5096–5101.
- Siero JC, Petridou N, Hoogduin H, Luijten PR, Ramsey NF (2011) Cortical depth-dependent temporal dynamics of the BOLD response in the human brain. *J Cereb Blood Flow Metab* 31:1999–2008.
- Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW (2003) Mapping cortical change across the human life span. *Nat Neurosci* 6:309–315.
- Stigliani A, Weiner KS, Grill-Spector K (2015) Temporal processing capacity in high-level visual cortex is domain specific. *J Neurosci* 35:12412–12424.
- Tanaka K (1996) Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19:109–139.
- Tong F, Nakayama K, Vaughan JT, Kanwisher N (1998) Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21:753–759.
- Triantafyllou C, Hoge RD, Krueger G, Wiggins CJ, Potthast A, Wiggins GC, Wald LL (2005) Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *Neuroimage* 26:243–250.
- Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M (2001) Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci* 4:832–838.
- Van Essen DC, Anderson CH, Felleman DJ (1992) Information processing in the primate visual system: an integrated systems perspective. *Science* 255:419–423.
- Vu AT, Gallant JL (2015) Using a novel source-localized phase regressor technique for evaluation of the vascular contribution to semantic category area localization in BOLD fMRI. *Front Neurosci* 9:411.
- Wang G, Tanaka K, Tanifuji M (1996) Optical imaging of functional organization in the monkey inferotemporal cortex. *Science* 272:1665–1668.
- Weiner KS (2019) The Mid-Fusiform Sulcus (sulcus sagittalis gyri fusiformis). *Anat Rec (Hoboken)* 302:1491–1503.
- Weiner KS, Grill-Spector K (2010) Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* 52:1559–1573.
- Weiner KS, Sayres R, Vinberg J, Grill-Spector K (2010) fMRI-adaptation and category selectivity in human ventral temporal cortex: regional differences across time scales. *J Neurophysiol* 103:3349–3365.
- Weiner KS, Golarai G, Caspers J, Chuapoco MR, Mohlberg H, Zilles K, Amunts K, Grill-Spector K (2014) The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage* 84:453–465.
- Weiner KS, Jonas J, Gomez J, Maillard L, Brissart H, Hossu G, Jacques C, Loftus D, Colnat-Coulbois S, Stigliani A, Barnett MA, Grill-Spector K, Rossion B (2016) The face-processing network is resilient to focal resection of human visual cortex. *J Neurosci* 36:8425–8440.
- Weiner KS, Barnett MA, Lorenz S, Caspers J, Stigliani A, Amunts K, Zilles K, Fischl B, Grill-Spector K (2017) The cytoarchitecture of domain-specific regions in human high-level visual cortex. *Cereb Cortex* 27:146–161.
- Witthoft N, Nguyen ML, Golarai G, LaRocque KF, Liberman A, Smith ME, Grill-Spector K (2014) Where is human V4? Predicting the location of hV4 and VO1 from cortical folding. *Cereb Cortex* 24:2401–2408.
- Yacoub E, Harel N, Ugurbil K (2008) High-field fMRI unveils orientation columns in humans. *Proc Natl Acad Sci U S A* 105:10607–10612.
- Zaretskaya N, Fischl B, Reuter M, Renvall V, Polimeni JR (2018) Advantages of cortical surface reconstruction using submillimeter 7 T MEMPRAGE. *Neuroimage* 165:11–26.
- Zeki S, Shipp S (1988) The functional logic of cortical connections. *Nature* 335:311–317.