Low-level tuning biases in higher visual cortex reflect the semantic informativeness of visual features

Margaret M. Henderson
Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA
Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA
Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

Michael J. Tarr
Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA
Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA
Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

Leila Wehbe
Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA
Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA
Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

Representations of visual and semantic information can overlap in human visual cortex, with the same neural populations exhibiting sensitivity to low-level features (orientation, spatial frequency, retinotopic position) and high-level semantic categories (faces, scenes). It has been hypothesized that this relationship between low-level visual and high-level category neural selectivity reflects natural scene statistics, such that neurons in a given category-selective region are tuned for low-level features or spatial positions that are diagnostic of the region's preferred category. To address the generality of this "natural scene statistics" hypothesis, as well as how well it can account for responses to complex naturalistic images across visual cortex, we performed two complementary analyses. First, across a large set of rich natural scene images, we demonstrated reliable associations between low-level (Gabor) features and high-level semantic categories (faces, buildings, animate/inanimate objects, small/large objects, indoor/outdoor scenes), with these relationships varying spatially across the visual field. Second, we used a large-scale functional MRI dataset (the Natural Scenes Dataset) and a voxelwise forward encoding model to estimate the feature and spatial selectivity of neural populations throughout visual cortex. We found that voxels in category-selective visual regions exhibit systematic biases in their feature and spatial selectivity, which are consistent with their hypothesized roles in category processing. We further showed that these low-level tuning biases are not driven by selectivity for categories themselves. Together, our results are consistent with a framework in which low-level feature selectivity contributes to the computation of high-level semantic category information in the brain.

Introduction

Cortical responses to visual inputs demonstrate organization according to both high-level and low-level stimulus properties. High-level information about images, such as their membership in semantic categories, is reflected in the activation of spatially localized areas of the ventral visual cortex selective for categories such as faces, body parts, places, food, and words (Downing et al., 2006; Epstein & Kanwisher, 1998;

Citation: Henderson, M. M., Tarr, M. J., & Wehbe, L. (2023). Low-level tuning biases in higher visual cortex reflect the semantic informativeness of visual features. *Journal of Vision*, *23*(4):8, 1–32, https://doi.org/10.1167/jov.23.4.8.

https://doi.org/10.1167/jov.23.4.8 Received September 1, 2022; published April 27, 2023 ISSN 1534-7362

Jain et al., 2023; Kanwisher et al., 1997; Khosla et al., 2022; McCandliss et al., 2003; Pennock et al., 2023; Sergent et al., 1992). At the same time, low- and mid-level visual features also elicit topographically regular patterns of activation in visual cortex, such as retinotopic maps of spatial position (Arcaro et al., 2009; Sereno et al., 1995; Swisher et al., 2007) and large-scale maps of selectivity for orientation (Freeman et al., 2011; Issa et al., 2000; Sasaki et al., 2006), spatial frequency (Aghajari et al., 2020; Bonhoeffer & Grinvald, 1991), color (Conway & Tsao, 2009; Zeki, 1973), and curvature (Yue et al., 2014; Yue et al., 2020). Given the hierarchical nature of processing in the visual system, understanding the relationship between selectivity for features at these different levels of complexity is critical for explicating the neural mechanisms by which high-level category information is computed in the brain.

Past work has suggested that low-level and high-level selectivity may be intertwined in their neural organization, in that category-selective visual regions exhibit systematic biases toward particular low-level properties of the visual environment. Not surprisingly, the low-level visual features most strongly represented in a given category-selective region tend to reflect the image statistics of the category in question. For example, scene-selective cortical regions, such as the parahippocampal place area (PPA), have been shown to be more responsive to cardinal (vertical and horizontal) orientations and rectilinear contour features than diagonal orientations and curved contours (Nasr et al., 2014; Nasr & Tootell, 2012; Li & Bonner, 2022). PPA has also been shown to be biased toward high spatial frequencies over low (Rajimehr et al., 2011). In contrast, areas of the visual system selectively responsive to faces have been shown to overlap with selectivity for curved features (Srihasam et al., 2014; Yue et al., 2014; Yue et al., 2020), and these face-selective areas may be more responsive to low spatial frequencies (Rajimehr et al., 2011). In addition to feature selectivity, spatial selectivity also appears to covary with category responsiveness, with face-selective and word-selective cortical regions tending to have biases toward the central visual field and scene-selective cortical regions tending to have more peripheral biases (Hasson et al., 2002; Levy et al., 2001).

A unifying explanation for these findings is based on the observation that, as mentioned, for many categories of natural images, the statistics of low-level visual features differ depending on the semantic content of images (Oliva & Torralba, 2001; Torralba & Oliva, 2003). For example, large-scale outdoor scene images are dominated by horizontal orientations and high spatial frequencies, while close-up images of objects may have a more isotropic distribution of orientations and more energy at low spatial frequencies (Torralba & Oliva, 2003). Beyond spectral features, mid-level features like the overall curvature in an image may covary with object-level distinctions such as real-world size and animacy (Long et al., 2016; Long et al., 2017). Such statistical associations may be reflected in the organization of visual cortex due to learning—either over the course of evolution or during an individual's lifetime. These observations suggest that category-selective visual regions follow a principle whereby they are biased in favor of low-level image properties that are informative for their "preferred" category, and these biases may play a functional role in both learning and categorization (Op de Beeck et al., 2008, Bracci et al., 2017).

Supporting the hypothesis that image statistics constrain the low-level biases found in category-selective visual cortex, the selectivity for cardinal orientations and rectilinear contours in scene-selective visual areas has been related to the fact that scene stimuli contain more cardinal orientations and rectilinear angles than nonscene stimuli (Nasr et al., 2014). A similar idea may hold for color selectivity, based on the finding that color tuning of neurons in ventral object-selective cortex is biased in favor of colors that are associated with objects (Rosenthal et al., 2018). Biases in spatial coverage of the visual field may also be understood in this framework; for example, the central (foveal) eccentricity biases found in face- and word-selective cortical regions may be related to the use of high spatial frequency information in identifying these classes of stimuli and the fact that they tend to be foveated, leading to an association with the central visual field (Hasson et al., 2002; Levy et al., 2001).

While these results provide some insight into the origins of low-level biases in visual cortex, much of the supporting work has focused on a small range of visual stimulus classes, often using controlled synthetic stimuli or objects on isolated backgrounds rather than natural scenes (Hasson et al., 2002; Levy et al., 2001; Nasr et al., 2014). In addition, there has been a tendency to focus on only one brain region or a small group of regions at a time (Hasson et al., 2002; Levy et al., 2001; Li & Bonner, 2022; Nasr et al., 2014; Nasr & Tootell, 2012). As a result, the generality of this hypothesis is somewhat uncertain, particularly with respect to how well it can account for findings across a range of visual areas during naturalistic image viewing. As an alternative, it has been suggested that semantic category selectivity reflects representations at a high level of abstraction only (Mahon & Caramazza, 2011). That is, the critical dimension in the organization of high-level visual cortex is category membership in and of itself rather than selectivity for continuously varying low-level features that may be associated with each given category. On this view, low-level feature biases do not play a functional role in computing category membership and, instead, may reflect more general structural constraints, such as inheritance of selectivity from an earlier area in the processing stream.

The implications of this alternative are that low-level feature biases measured in category-selective visual areas may *not* correlate with the low-level features and/or positions that are most diagnostic of each area's preferred category.

To distinguish between these alternatives, we provide two complementary sets of analyses. First, we take advantage of a large, richly annotated natural scene image database (Common Objects in Context [COCO]; Lin et al., 2014) to demonstrate consistent associations between low-level features (orientation, spatial frequency) and high-level semantic category labels. Next, we use a recently released, large-scale open functional MRI (fMRI) dataset collected by Allen et al. (2021; the "Natural Scenes Dataset" or NSD) to show that these same associations are reflected in the patterns of feature and spatial selectivity estimated from neural population responses measured while human participants view natural scene images. Importantly, our use of naturalistic images allows us to demonstrate a link between neural response properties and natural scene statistics within an ecologically relevant setting, in contrast to past work with simpler stimuli. Our results also provide a proof of concept that low-level feature and spatial biases can be reliably measured in the human brain using complex, naturalistic stimuli. Furthermore, our use of a high-resolution, whole-brain fMRI dataset provides an opportunity to assess our hypothesis across the brain rather than in a limited region of cortex. Based on the convergence among our analyses, we argue that low-level feature and spatial biases may play a key role in the processing of semantic categories within high-level visual areas.

Methods

Human participants and acquisition of fMRI data

We used a large-scale publicly available dataset, the NSD, for all analyses. A detailed description of the data is provided in Allen et al. (2021). Briefly, the NSD includes measurements of whole-brain blood-oxygen-level-dependent (BOLD) fMRI from eight participants who each viewed between 9,000 and 10,000 colored natural scenes over the course of 30 to 40 scan sessions. All functional scans were conducted at 7T using whole-brain gradient-echo echo-planar imaging (EPI) at a 1.8-mm resolution and a 1.6-s repetition time. Images were taken from the Microsoft COCO database (Lin et al., 2014) and were square-cropped and presented in color, at a size of 8.4° × 8.4° (° = degrees of visual angle). Of the approximately 10,000 images viewed by each participant, ∼9,000 images were seen only by that participant, and ∼1,000 were overlapping across participants. Each image was viewed for a duration of 3 s, with 1 s between trials. Over the course of the experiment, each image was viewed approximately three times, for a total of roughly 30,000 trials per participant. Participants were required to fixate centrally on a small fixation dot superimposed on each image while performing a task in which they reported whether or not each image had been presented before in any session.

Preprocessing of fMRI data

fMRI data were preprocessed by performing one temporal interpolation to correct for slice time differences and one spatial interpolation to correct for head motion within and across scan sessions (Allen et al., 2021), resulting in volumetric fMRI time-series data at a 1.8-mm resolution in participant native space. Beta weights for each voxel and each trial were then estimated using a general linear model. To improve the signal-to-noise ratio in estimating beta weights, a three-stage procedure was used that consisted of selecting a hemodynamic response function (HRF) from a library of candidate HRFs, denoising the data using a set of noise regressors estimated from voxels not related to the experimental paradigm, and using fractional ridge regression to regularize the beta weight estimation on a single voxel basis (Prince et al., 2022). Finally, beta weights for each voxel were averaged over trials on which the same image was shown (typically three trials per image), and the averaged beta weights for each image were used for all further analyses.

Defining regions of interest

Each NSD participant also performed several runs of a category functional localizer task (Stigliani et al., 2015) and a population receptive field (pRF) mapping task (Benson et al., 2018). The pRF mapping task was used to define early retinotopic visual regions of interest (ROIs) V1, V2, V3, and hV4 (Allen et al., 2021). The category localizer task was used to define face-selective ROIs (occipital face area [OFA] and fusiform face area [FFA]; for our analyses, we combined FFA-1 and FFA-2 into a single FFA region), scene-selective regions (parahippocampal place area [PPA], occipital place area [OPA], and retrosplenial cortex [RSC]), and a body-selective region (extrastriate body area [EBA]). In addition to these functional ROIs, we also utilized a probabilistic atlas (Wang et al., 2015) to define areas V3ab and the intraparietal sulcus (IPS; all six subregions of IPS0–5 were combined into a single region). Since the probabilistic atlas also included definitions for early visual areas V1, V2, V3, and hV4, which were already

defined using the pRF mapping task, we combined both sets of definitions for these regions, and where they disagreed, we deferred to the pRF-based definitions. For example, if a voxel was labeled V1 in the atlas and was not included in the pRF-based labels, it was added to V1, but if it was labeled V1 in the atlas and V2 in the pRF-based labels, it was kept as V2. As a result, the retinotopic ROI labels were nonoverlapping with one another. To prevent the category-selective ROIs from overlapping with the retinotopic ROIs, we removed any voxels that overlapped from the retinotopic ROIs and included them only in the category-selective ROI to which they corresponded; the area most affected by this was V3ab, which was originally overlapping with OPA. To prevent category-selective ROIs from overlapping with one another (which can happen for regions defined based on different contrasts), we removed the overlap by prioritizing face-selective definitions over body- and scene-selective definitions and prioritizing scene-selective definitions over body-selective definitions. Thus, the final set of 12 ROIs was entirely nonoverlapping. Finally, we applied an additional threshold to the voxels in each ROI based on their noise ceiling (i.e., the theoretical proportion of variance in the data that can be predicted; for details on noise ceiling calculation, see Allen et al., 2021), using a threshold of 0.01.

Encoding model fitting

Overview

We fit an encoding model for each individual fMRI voxel that captured its spatial selectivity as well as its feature selectivity (Figure 1B; see also St-Yves & Naselaris, 2018). Fitting was done for two different visual feature spaces and a semantic category feature space (see Methods: Feature spaces). The basic procedure was to first loop over a grid of candidate pRFs (see Methods: Population receptive fields) for each voxel and use regularized regression to fit the weights of a linear model that describes the voxel response as a weighted sum of the feature activations corresponding to that pRF. The best-fitting candidate pRF and the feature weights for that pRF made up the final encoding model. We then evaluated that model's ability to predict voxel responses on a held-out partition of data. The fitted encoding models were also used to estimate voxel feature selectivity. Below, we outline each of these steps in detail.

Population receptive fields

Following from past work (Dumoulin & Wandell, 2008; Kay et al., 2013; St-Yves & Naselaris, 2018; Vo et al., 2017), we modeled the spatial selectivity of each fMRI voxel using a two-dimensional Gaussian response profile over the spatial extent of the viewed image. This approach is similar to classic approaches of fitting spatial receptive fields for single neurons in visual cortex and is often termed a *population receptive field* (pRF) to denote summation over the population of neurons in a voxel. The pRF can be described by three parameters, x_0, y_0, and σ, where $[x_0,y_0]$ and σ, respectively, indicate the center and standard deviation of the two-dimensional Gaussian response profile:

$$g_{x_0,y_0,\sigma}(x,y) = \exp\left(-\left(\frac{(x-x_0)^2}{2\sigma^2}+\frac{(y-y_0)^2}{2\sigma^2}\right)\right) \quad (1)$$

To select each voxel's optimal pRF, we constructed a grid over candidate pRF parameters. Our grid had even spacing between adjacent pRF centers in terms of their polar angle position (θ) and nonlinear spacing in eccentricity (r), where candidate centers were spaced closer to the center of the visual field. The purpose of the nonlinear eccentricity spacing was to account for the cortical magnification factor in human visual cortex, where the neuronal sampling of visual space is more dense close to the fovea (Duncan & Boynton, 2003). More concretely, our 16 candidate polar angle positions were linearly spaced, ranging from 0° to 337.5° in steps of 22.5°, and our 10 candidate eccentricities were logarithmically spaced, ranging from 0° to 7°. Our 10 σ values were also spaced logarithmically, with σ ranging from 0.17° to 8.4°. To generate the complete grid, we first computed every possible combination of r, θ, and σ, which resulted in 1,600 pRFs. We then converted the centers from polar angle coordinates $[r,\theta]$ into Euclidean coordinates $[x_0,y_0]$. Finally, we eliminated any pRFs that landed completely outside the image region (an 8.4° × 8.4° square), by the criterion that their rough spatial extent (center ± σ) was nonoverlapping with the image region. The result of this was that for the most peripheral pRFs, only the larger size values were included in the grid. The final grid included 1,456 pRFs in total.

Encoding model design

Our encoding model framework assumed that each voxel response could be modeled as a linear weighted sum of the activations in a set of underlying feature channels. In mathematical notation, this can be formulated as

$$y = X\beta \quad (2)$$

where y is a column vector of length n containing the voxel responses across n images, and β is a column vector of length w containing w weights, corresponding to each of $(w-1)$ feature channels plus an intercept. X is the design matrix for the feature space, describing the

activation in each feature channel for each image plus a column of ones, size $[n \times w]$.

The features in each of our feature spaces were computed in a spatially specific manner (St-Yves & Naselaris, 2018), such that the design matrix depended on the pRF parameters x_0, y_0, and σ. Referring back to the pRF definition given in Equation 1, the design matrix can be expressed as

$$X = f\left(g_{x_0,y_0,\sigma}\right) \quad (3)$$

where f is a function that depends on the feature space under consideration. In most cases, f simply refers to taking the dot product of the pRF with a spatial feature map describing the activation of some feature channel at each position in the image. In those cases, for the spatial activation map $a_{t,c}$ (size $p \times p$) corresponding to image t and feature channel c, and pRF profile g having the same resolution as a, the corresponding element of X can be computed by

$$X_{t,c} = \sum_{x=1}^{p}\sum_{y=1}^{p} a_{t,c}(x,y) \cdot g_{x_0,y_0,\sigma}(x,y) \quad (4)$$

But this form differs when we use a semantic category feature space (see Methods: Semantic features for details).

Note that Equation 4 requires generating each pRF with a flexible resolution to fit the resolution of the feature map output from a given model. To achieve this, we scaled the x_0, y_0, and σ parameters along with the feature map resolution, so that they always corresponded to the same positions in image coordinates.

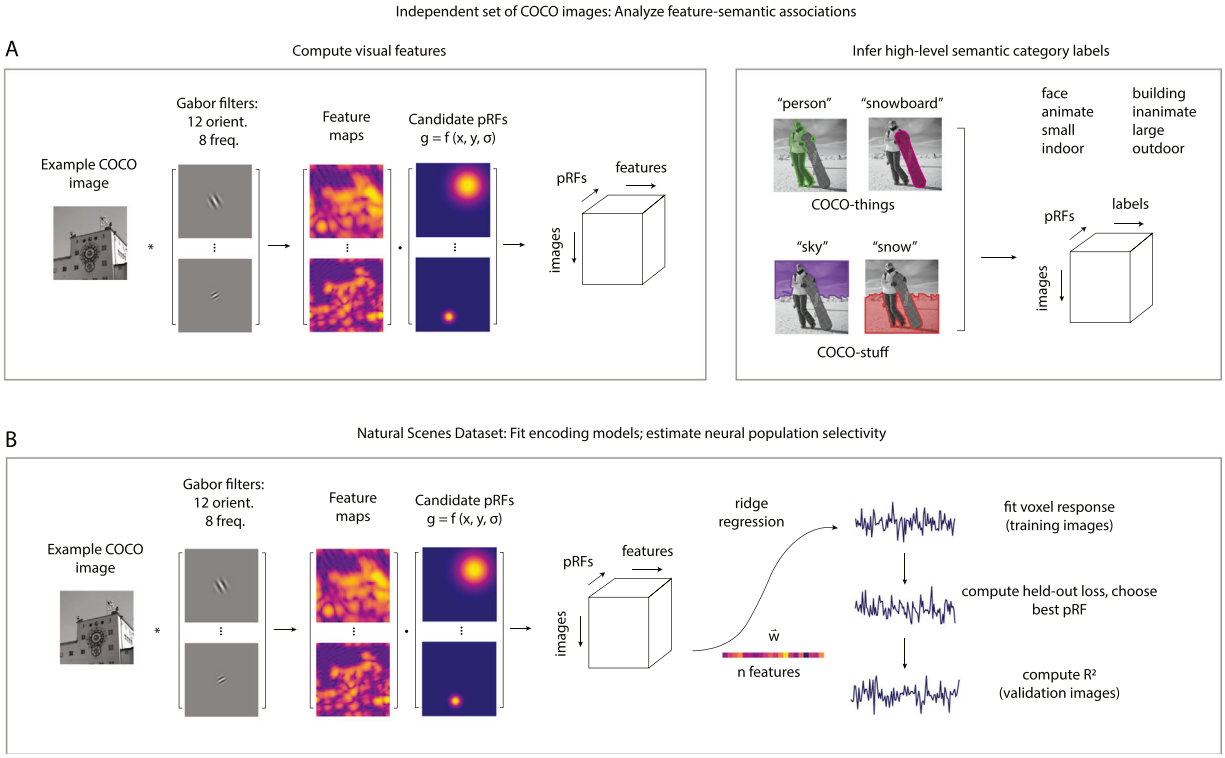

Figure 1. Overview of our two analysis procedures. (A) In the first set of analyses, we analyzed the associations between Gabor features and high-level semantic labels across a set of 50,000 COCO images (note that no fMRI data are involved in this step). The left box depicts a schematic of the feature extraction procedure: For each natural scene image, we used a Gabor feature bank to compute a stack of feature maps and then applied various spatial weighting matrices (i.e., candidate pRFs; Dumoulin & Wandell, 2008) to compute the estimated activation in each feature map for a range of spatial positions and pooling field sizes (see Methods for details). The right box depicts examples of the semantic category labels included in the COCO dataset (Lin et al., 2014); each label is accompanied by a spatial mask indicating the corresponding object's position in the image. We used these labels to infer high-level semantic category content for each image, labeling either the entire image (indoor, outdoor) or each pRF individually (face, building, animate, inanimate, small, large); see Methods for details. (B) In the second set of analyses, we used the NSD (Allen et al., 2021) to learn an encoding model predicting each fMRI voxel's response as a weighted sum of image features in the voxel's pRF. The feature extraction procedure was identical to (A) but performed on a different set of images (the images viewed by each fMRI participant). To fit the model for each voxel, we first learned a set of weights for each candidate pRF (using ridge regression), then chose the pRF that resulted in the lowest loss on held-out data. Note that while this figure illustrates use of a Gabor feature bank, we used the same approach for other feature spaces; see text for details.

Also note that when f takes the form described in Equation 4, the σ parameter of the pRF does not provide a complete description of the spatial extent of the image that contributes to computing X. This is because each pixel in the activation map a has its own pooling region, determined by the operations used to compute that activation (e.g., the kernel size of a convolution). Thus, σ should be interpreted as a lower bound on the pRF size rather than an exact estimate of its size (St-Yves & Naselaris, 2018).

Model-fitting procedure

Following previous work (Güçlü & van Gerven, 2014; Huth et al., 2016; Wehbe et al., 2014), we solved for the weights for each voxel using ridge regression (L2-regularization). The ridge regression estimator of β is given by

$$\widehat{\beta} = (X^T X + \lambda I)^{-1} X^T y \quad (5)$$

where I is a $[w \times w]$ identity matrix, λ is a regularization parameter, and y is a vector containing the voxel response for n images. Once $\hat{\beta}$ is computed, the voxel response can be predicted from the design matrix associated with any arbitrary stimulus input by

$$\hat{y} = X\hat{\beta} \quad (6)$$

The regularization parameter (λ) was selected using cross-validation on a per-voxel basis, from a set of 10 candidate λ values logarithmically spaced between 0 and 10^5. The full cross-validation procedure is detailed below (see also Figure 1B).

First, we held out ~1,000 images from the ~10,000 total images for each participant, to serve as a validation set. The validation set images always consisted of the set of "shared images" that were seen by all participants (see Methods: Human participants and acquisition of fMRI data). The remaining ~9,000 images made up the training data. Before fitting, we z-scored the values in each column of the design matrix, separately for the training images and validation images. To select the ridge parameter and the best pRF parameters for each voxel, we held out a random 10% of the training data as a nested validation set. We then used the remaining 90% of training images to compute $\hat{\beta}$ for each of our candidate λ values and candidate pRF models (recall that the pRF parameters determine the design matrix X used; see Equation 3). For each of the candidate pRFs and λ values, we computed a prediction of the nested validation data $\hat{y}$ based on the estimate of $\hat{\beta}$, and computed the loss of that estimate, $\sum (y - \hat{y})^2$. The pRF parameters and λ value that resulted in the lowest loss were selected as the best pRF parameters and λ for that voxel. Correspondingly, the $\hat{\beta}$ associated with that pRF and λ were selected as the best weights for that voxel. Finally, these best-fit parameters were used to predict each voxel's response on the held-out validation data, and we computed the coefficient of determination (R^2) between y and $\hat{y}$ as a measure of overall model accuracy.

The above procedure describes our method for fitting spatial and feature selectivity simultaneously, as done in St-Yves and Naselaris (2018). However, once a stable estimate of the pRF for each voxel has been obtained using some feature space, we can adapt this method to fit just the feature selectivity for each voxel (for any arbitrary feature space), assuming that its spatial selectivity remains fixed. To do this, we use each voxel's precomputed pRF estimate to select the correct design matrix X and fit its $\hat{\beta}$ for that design matrix only, for each candidate λ value. We follow the same procedure as outlined above with respect to training/validation image splits and λ selection.

For all analyses presented here, our approach was to first fit both the feature and spatial selectivity of each voxel using the AlexNet concatenated feature space (see Methods: AlexNet features). Once the pRF estimates were obtained based on AlexNet, we assumed that these pRFs remained fixed for all other feature spaces and used them to fit each voxel's weights for all other feature spaces of interest (i.e., Gabor and COCO-all; see Methods: Feature spaces). We chose the AlexNet feature space (Krizhevsky, 2014) for the first pRF fitting step because AlexNet has often been used to model voxel responses in human and nonhuman primate visual cortex and gives good predictive performance across a range of cortical areas (Cadieu et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014; Schrimpf et al., 2020, Li & Bonner, 2022). In our data, AlexNet generally yielded higher predictive accuracy than the Gabor model in higher-level visual areas. Similar results for both spatial and feature selectivity were obtained when we fit the entire model, including pRFs, from scratch on the Gabor feature space (Supplementary Figure S8).

To determine whether encoding model accuracy was significantly better than chance, we used a permutation test. This consisted of shuffling the image label sequence randomly 1,000 times, refitting the model weights, and computing R^2. Shuffling was done separately within the training images, validation images, and nested validation images. When performing the permutation test, we used the "real" estimate of each voxel's pRF as a starting point and only refit the feature weights on these data, rather than fitting the pRF on the shuffled data. Once the shuffled R^2 values for each voxel were obtained, we computed one-tailed p-values by computing the proportion of iterations where the shuffled R^2 was greater than or equal to the real R^2. We then performed false discovery rate (FDR) correction on the p-values across voxels for each participant, using a threshold of $\alpha = 0.01$ (Benjamini & Hochberg, 1995). An identical procedure was used to test significance of

fits on the original data and to test significance of fits on the residuals of the semantic model (see Methods: Semantic category encoding model).

Feature spaces

Overview

For each stimulus image viewed by fMRI participants, we extracted several different sets of features that were intended to capture different aspects of the image's visual and semantic content. These features were used to construct voxelwise encoding models, as well as to estimate the statistical associations between lower-level features and semantic categories (see Methods: Measuring feature-semantic associations). Each set of features was extracted in a spatially specific manner, such that the features associated with each pRF grid position described the visual or semantic content within a specified region of the image only. Unless otherwise specified, all feature extraction was performed on grayscale images at a resolution of 240 × 240 pixels. Below, each feature space is described in detail.

Gabor features

Our first set of features was based on Gabor filters that extract the energy at specified orientations and spatial frequencies (see Figure 1A). Similar models have previously been used to model the responses of early visual cortex to natural images (Kay et al., 2008; Lescroart & Gallant, 2019; St-Yves & Naselaris, 2018). Our Gabor filter bank included filters at 12 unique orientations, linearly spaced between 0° and 165° in increments of 15°. Each filter consists of a two-dimensional complex-valued sinusoid with a specified frequency and orientation, multiplied by a two-dimensional Gaussian envelope (standard deviation of the Gaussian was 2.9 pixels for a filter 12 × 12 pixels in size). The real and imaginary components of the sinusoid are at 0° and 90° phase, respectively, to form a quadrature pair. The final activation of each filter was obtained by convolving both the real and imaginary filters with the input image (using circular padding for the stimulus edges), squaring the output of both the real and imaginary parts, summing the real and imaginary parts, and taking the square root. We then applied a nonlinearity to the resulting activation values, $f(x) = log_e(1 + \sqrt{x})$.

We applied this bank of filters at eight spatial frequencies that were logarithmically spaced between 0.35 and 8.56 cycles per degree of visual angle (cyc/°). To achieve filtering at each frequency, we first resized the input images to an appropriate size (i.e., smaller for lower frequencies) using bilinear resampling. We then applied the stack of filters, which were always a fixed size and frequency in pixels (12 × 12 pixels and 4.13 pix/cycle), to the resized images. The end result was a set of eight stacks of feature maps, one for each spatial frequency, where the height and width dimensions of each stack depended on its corresponding spatial frequency. Each stack contained 12 feature maps for the 12 orientation channels.

Finally, to extract the feature activations within each pRF in our grid, we took the dot product of the pRF with each feature map to obtain a single value for the activation in each feature channel (Equation (4)). This resulted in a 96-dimensional feature space, computed separately at each pRF grid position.

AlexNet features

We extracted visual features from a convolutional neural network model referred to as "AlexNet," which is trained on a 1,000-way image classification task (for more details on the model's construction and training, see Krizhevsky, 2014). We extracted activations from the first five convolutional layers of a pretrained AlexNet model. Activations were extracted after the rectifying nonlinearity (ReLU) function that follows each convolutional operation. To extract the features for each pRF grid position, we took the dot product of each feature map with the pRF of interest (Equation 4), which resulted in a single value for each feature channel. The dimensions of the resulting feature sets corresponding to each AlexNet layer are [64, 192, 384, 256, 256].

Before using these features in our encoding models, we reduced the dimensionality of features from each AlexNet layer using principal components analysis (PCA). PCA was always performed using the training data only to solve for the principal components and then using those components to transform all data into the same subspace. We retained a sufficient number of components to explain at least 95% of the variance in the training data. We performed PCA on the features from one pRF at a time. As a result, the dimensionality of features from different pRFs was allowed to differ following PCA, even though the dimensionality of the features from each pRF was the same before PCA. After performing PCA on the features from each layer, we concatenated the features across all layers.

Semantic features

In addition to modeling the visual features present in our image set, we created a feature set that explicitly modeled the semantic categories present at each spatial position (i.e., each pRF) in each image. To achieve this, we took advantage of the object segmentations associated with each image in the Microsoft COCO database (Lin et al., 2014), as well as an additional

set of semantic segmentations that label the "stuff" (meaningful but amorphous regions such as sky, walls, etc.) in each image (COCO-stuff; Caesar et al., 2018). Each object or stuff instance in each of these labeling schemes is accompanied by a polygon that defines the spatial extent of the instance. To determine whether each instance was overlapping with a given pRF in a given image, we first generated a binary mask for the label polygon, at a resolution of 425 × 425 pixels. We then created a second binary mask of the same size, which captured a circular region ± 2σ from the pRF center (see Equation 1). The pRF was considered to be overlapping with the label if the two masks overlapped by at least 10 pixels. This very lenient overlap threshold was meant to account for the possibility of noise in our pRF parameter estimates, as well as the fact that receptive fields in category-selective regions of visual cortex tend to be large. In initial tests, using a more stringent overlap threshold led to poorer fits of the semantic model.

Using this method, we extracted several sets of semantic features for each image and each pRF. The first set of semantic features, which we termed "COCO-all," included 80 basic-level object categories, 12 superordinate object categories, 92 basic-level stuff categories, and 16 superordinate stuff categories, for a total of 200 features (Supplementary Table S1). Each feature is a binary label that denotes whether a category is present within the pRF of interest. When building encoding models from this feature set (COCO-all model), we directly used the binary features for each pRF as our design matrix (see Equation 2). These encoding models were used in order to regress out the contributions of semantic selectivity from voxel responses (see Methods: Semantic category encoding model).

Next, we created eight additional semantic category labels that captured coarse-level semantic information about the image contents: faces, buildings, animate objects, inanimate objects, small objects (in terms of real-world size), large objects, indoor scenes, and outdoor scenes. Most of these higher-level semantic labels were defined on the basis of the "things" and "stuff" category labels. For example, if a pRF contained any animate object (i.e., a person or any animal), it was labeled as having the "animate" label, and if it contained any inanimate object, it was labeled with the "inanimate" label. This allows for the possibility that pRFs could be labeled with neither the animate nor the inanimate label, or they could be labeled with both. Similarly, the small and large object labels were assigned to pRFs based on whether they contained items we defined as being "small" (e.g., a banana) or "large" (e.g., an elephant). See Supplementary Table S2 for a list of the object categories we defined as small and large. For the "building" category label, we assigned the label of building to any pRF that had a label from the COCO-stuff superordinate category named "building."

To label the "face" category, we used the RetinaFace network (Deng et al., 2020) to detect faces in each image. We used a pretrained version of the model implemented in Tensorflow 2.0, using a ResNet-50 backbone, and ran it on images at a resolution of 240 × 240 pixels. The model labels each instance of a face in each image with a rectangular bounding box. We then used the same approach described above to determine whether each face bounding box overlapped with each pRF (again using a threshold of 10 pixels for the overlap). To verify the general accuracy of these labels, we visually checked the bounding boxes and also compared the labels against the presence of the "person" label in COCO. This comparison revealed good agreement between the COCO labels and the RetinaFace labels: approximately 30% of images were labeled as both person and face, another 20% were labeled as person but not face (these are the images that contain a human body part but not a face), and less than 1% were labeled with "face" only (these indicate false positives of the RetinaFace model, sometimes indicating dog or cat faces).

For the indoor–outdoor distinction, since the indoor or outdoor category is most naturally understood as a property of entire scenes, rather than something that can vary across the spatial extent of a single image, we created labels for the entire image rather than for each pRF separately. To define whether an image was indoor or outdoor, we looked for the presence of particular diagnostic objects or stuff classes (i.e., "car" and "grass" are diagnostic of outdoor images, "bed" and "carpet" are diagnostic of indoor images). See Supplementary Table S3 for a list of the categories we used to make these distinctions. When images included both indoor and outdoor classes, we resolved the tie by counting the number of total indoor categories (objects and stuff) and outdoor categories included in the image and taking the maximum. This left only around 7% of images that could not be unambiguously labeled by our method.

Measuring feature-semantic associations

To investigate the association between each Gabor feature channel and each of our eight semantic categories (see Methods: Semantic features), we used two approaches. First, we computed a partial correlation coefficient between each of the 96 Gabor features and each category label of interest (represented as a binary value for the presence of the category), controlling for the contributions of the other seven category labels. Partial correlations were computed using the linear regression method, which consisted of learning a multiple linear regression that maps from the

"other" category labels to the category label of current interest, and another regression that maps from the "other" category labels to the Gabor feature of interest. The partial correlation was then obtained by computing the correlation coefficient of the residuals of these two regression fits. This was done for each feature within each pRF individually.

Second, we used a linear decoder to measure how much object category information can be extracted from the combined pattern of activation across all 96 Gabor features. We performed decoding for one category at a time, using labels for whether that category was present or absent from each image. Decoding was performed using a linear discriminant analysis classifier, implemented in scikit-learn, and a 10-fold cross-validation procedure. To measure decoding performance, we computed d' from signal detection theory, based on the formula $d' = \text{Z(hit rate)} - \text{Z(false-positive rate)}$. The hit rate is defined as the proportion of test samples in Condition X accurately classified as belonging to Condition X, and the false-positive rate is the proportion of test samples in Condition Y inaccurately classified as belonging to Condition X, and Z is the inverse of the cumulative distribution of the Gaussian distribution.

To ensure that comparisons of decoding performance across pRFs were fair, we always used the same number of images to train and test the classifier in each pRF (since semantic labels were created for each pRF individually, the number of images that were labeled with each category varied across pRFs). To achieve this, we first identified the minimum number of images (n) in any category for any pRF. When performing decoding within each pRF and each category, we created a randomly downsampled set of images that consisted of n images each with and without that label. This also ensured that the classifier was trained on a perfectly balanced dataset. Before performing this subsampling, we first removed the most peripheral pRFs (those with eccentricity $> 4°$), as well as the smallest pRFs (those with size $< 1°$) from the grid, using only the remaining 640 pRFs. The purpose of this was to ensure that there were sufficient images available to support robust decoding, since category labels were more rare in the smallest and most peripheral pRFs (see Supplementary Figure S1).

To test whether decoding performance varied significantly based on pRF parameters, we estimated the slope of a linear regression fit where the y-coordinate is d' and the x-coordinate is each pRF parameter of interest (pRF σ, eccentricity, horizontal position, and vertical position). When performing this test for eccentricity, horizontal position, and vertical position, we used only the pRFs corresponding to the smallest size used for decoding (after excluding the smallest sizes as described above, this was 1.48°) because this provides the maximum resolution to detect differences across the visual field. To determine whether the relationship with each parameter was significantly different from zero, we used a permutation test where we randomly shuffled the x-coordinate data 10,000 times and refit the line. We then computed a two-tailed p-value by computing the number of iterations on which the real slope exceeded the shuffled slope and the number of iterations on which the shuffled slope exceeded the real slope, taking the minimum and multiplying by 2. The resulting p-values were FDR corrected (Benjamini & Hochberg, 1995) across all pRF parameters and semantic categories (24 total values) at $\alpha = 0.01$.

Importantly, both of these analyses were performed on a completely independent, randomly selected set of 50,000 COCO images that were *not* seen by any participant in the NSD. This was done to ensure that the comparison between our image statistics analyses and our voxel selectivity analyses was noncircular and not driven by the possibility of overlap in the images. We preprocessed these independent COCO images in a similar manner to the preprocessing used for the NSD images, but instead of computing the "semantic loss" described in Allen et al. (2021) to determine cropping boxes, we simply cropped the long edge of each rectangular image in a symmetric way to generate square images.

Estimating voxel semantic selectivity

We estimated the selectivity of voxel activations for each of our eight high-level semantic categories (Methods: Semantic features) by using a partial correlation coefficient that controls for the contributions of the other seven semantic categories, analogous to the method described in the previous section (Methods: Measuring feature-semantic associations). We used these partial correlation coefficient values to define the top 500 category-selective voxels for each category in each participant. To choose the top voxels, we first thresholded voxels based on significant accuracy of the Gabor encoding model (see Methods: Model-fitting procedure) and also removed any voxels in early retinotopic areas V1, V2, V3, and hV4. This ensured that the final set of voxels we selected were within higher visual cortex and were well fit by the Gabor model.

Estimating voxel visual feature selectivity

After fitting Gabor encoding models to each voxel, we used the fitted models to estimate voxel selectivity for individual Gabor feature channels. These analyses were only performed on well-fit voxels, defined as those whose validation set accuracy was above chance

using a permutation test (see Methods: Model-fitting procedure; see Supplementary Table S4 for the number of voxels passing this threshold). To estimate feature selectivity, we first computed the sensitivity of the voxel response (as predicted by the encoding model) to changes in the activation within each feature channel. In mathematical notation, for a given feature channel c and a voxel v with a fitted encoding model for the feature space of interest, our measure of feature sensitivity $\rho_{v,c}$ is

$$\rho_{v,c} = Corr(\widehat{y_v}, x_c) \quad (7)$$

where $\widehat{y_v}$ denotes the model-predicted response of the voxel to the validation set images, and x_c denotes the activation in channel c for the same validation set images.

This sensitivity measure is meant to capture approximately how strongly each feature channel is "weighted" in the overall encoding model prediction for each voxel. Importantly, however, it is not the same as simply using the raw $\widehat{\beta}$ weights as a measure of feature sensitivity. The raw weights of our encoding models are not likely to provide a stable estimate of voxel feature tuning due in part to the high degree of collinearity between feature channels and in part to the fact that we fit the models with ridge regression, which can lead to biased weight estimates (Farrar & Glauber, 1967; Hoerl & Kennard, 1970). In contrast, because $\rho_{v,c}$ is measured using actual validation set data, it provides a measure of the functional alignment between the feature channel and the response of the encoding model, within the context of the real covariance structure of the data. At the same time, we note that $\rho_{v,c}$ is not intended to dissociate the contributions to the encoding model prediction made by feature channels that are highly correlated.

Once the feature sensitivity (ρ) values are computed for each Gabor feature (12 orientations × 8 spatial frequencies) and each voxel, it is straightforward to create a sensitivity profile that captures each voxel's sensitivity to changes in feature intensity at different positions along the orientation or spatial frequency axis. Orientation sensitivity profiles are obtained by plotting the sensitivity values as a function of feature orientation, and spatial frequency sensitivity profiles are obtained by plotting the sensitivity values as a function of feature spatial frequency. Averaging these profiles across values of the other dimension (e.g., averaging orientation sensitivity profiles across all spatial frequency levels) yields an average feature sensitivity profile. These average feature sensitivity profiles were used to compute each voxel's "preferred" orientation and spatial frequency (i.e., values plotted in Supplementary Figure S2) through an argmax operation.

Counting peaks in feature sensitivity profiles

For each voxel's average orientation and spatial frequency sensitivity profiles (see previous section), we identified the approximate number of peaks in the curve as a supplementary analysis to ensure that our results did not depend on assuming a single peak (Supplementary Figure S4). To identify peaks, we first identified all the local maxima in the curve by comparing each sensitivity value against the values to its left and right. For orientation, we accounted for the circularity of the feature space by wrapping the curve circularly when computing local maxima, while for spatial frequency, we treated the endpoints as local maxima if they exceeded the points to their left or right. Next, we removed any peaks that had a negative sensitivity value, since those peaks do not indicate a positive feature preference. Then, we computed the height of each peak by subtracting the minimum value across the sensitivity profile and divided the height of each peak by the height of the largest peak. Based on the resulting ratio values, we retained only those peaks whose value exceeded 0.50. This ensured that when multiple peaks were counted for a given voxel, each of these peaks was comparable in height.

Based on the finding that many voxels had two peaks in their orientation sensitivity profiles, we analyzed the orientation preferences of these voxels by sorting them into groups. For each of the two-peaked (bimodal) voxels, we identified the two orientations at which the two peaks occurred, without regard to the relative height of the two peaks. We then grouped together voxels that were selective for the same pair of orientations. To create the plots in Supplementary Figure S4D, we selected the three most common pairs of orientations across all voxels and combined all other pairs of orientations into a separate group. We did not perform this procedure for spatial frequency, since most voxels had only one true peak in their frequency sensitivity profiles.

pRF coverage analysis

To quantify the visual field biases in each ROI, we computed an aggregated estimate of visual field coverage across all pRFs (similar to Silson et al., 2015). Specifically, we combined all pRF estimates across all voxels in a given brain region (after first thresholding the voxels based on performance of the AlexNet encoding model at an R^2 value of 0.01). We combined voxels using an averaging operation across all pRFs, as this takes into account differences in the density of pRF coverage across the visual field, but similar results were obtained when aggregating using a

maximum operation. This led to one aggregated pRF coverage map for each participant and each ROI. We then took the average of these maps in each visual field quadrant to yield an estimate of the coverage for each quadrant.

Next, we analyzed the differences in coverage among ROIs and quadrants by performing a three-way repeated-measures analysis of variance (ANOVA), with ROI, vertical position, and horizontal position as factors (implemented using the Python package *statsmodels*). Because this test revealed a significant interaction between ROI and vertical position, we performed post hoc pairwise tests to identify ROIs where there was a difference in coverage between the upper and lower visual fields. Pairwise tests were done using a nonparametric, two-tailed paired *t*-test. To achieve this, we first computed a *t*-statistic for the actual coverage values in each ROI and each half of the visual field (upper vs. lower), then randomly shuffled the vertical position labels for the coverage values across 10,000 iterations and computed a *t*-statistic for each shuffling iteration. Then, we computed a two-tailed *p*-value by calculating the number of iterations on which the real *t*-statistic exceeded the shuffled statistic and the number of iterations on which the shuffled *t*-statistic exceeded the real *t*-statistic, took the minimum of these values, and then multiplied by 2. Finally, we performed FDR correction on the *p*-values across all ROIs with an α value of 0.01 (Benjamini & Hochberg, 1995).

Semantic category encoding model

To determine whether visual feature selectivity was driven in part by category selectivity, we constructed an encoding model (COCO-all model) that predicts each voxel's response based on the COCO categories present at each location in the image (see Methods: Semantic features). As described previously (Methods: Model-fitting procedure), when fitting this model, we utilized the pRFs for each voxel that were already estimated based on the AlexNet encoding model, so fitting the category model for each voxel required fitting only a single set of weights across the 200 semantic category features. Once the model was fit, we used it to generate predicted voxel responses for all images in the dataset (both training and validation) and then subtracted the actual voxel activations from these predicted responses to yield a residual for each voxel's response to each image. These residuals represent the portion of voxel responses that cannot be modeled as a linear combination of category features. We then refit our Gabor encoding model with the residuals in place of raw voxel activation data, using the same approach described earlier (outlined in Methods: Encoding model fitting). We also used an analogous approach to fit the COCO-all model to the residuals of the Gabor encoding model. Importantly, the exact same data splits were used when fitting both the COCO-all model and the Gabor model, such that the validation set data did not contribute to training of either model.

When analyzing feature selectivity results from the Gabor encoding model fit to the COCO-all residuals and comparing them to the results of the Gabor model fit to the raw data, we always thresholded voxels according to their R^2 for both the raw data fit and the residuals fit, based on the results of a permutation test for significance (see Methods: Model-fitting procedure). This ensured that the same set of voxels was being compared between the raw and residual fit models (see Supplementary Table S4 for the number of voxels in each ROI meeting this threshold).

Fitting models for single categories

As an additional test of whether feature selectivity was influenced by differential processing of categories, we performed Gabor model fitting using images from only one high-level category at a time. For this analysis, we focused on only one semantic dimension at a time: For instance, when separating images into indoor and outdoor (i.e., the indoor–outdoor dimension), we ignored the face, building, animacy, and real-world size labels. This was because balancing across multiple dimensions resulted in too few images to robustly fit the model. We also excluded the "face" and "building" categories from this analysis because these categories were relatively rare, especially for small pRFs (Supplementary Figure S1). When splitting the images based on each dimension, we always split images based on labels on a per-pRF basis, meaning that the images assigned to a given category were different depending on which pRF was currently of interest. For example, when fitting the set of voxels whose best pRF was pRF *n*, we used the category labels for pRF *n* to split the images into category groupings and performed fitting for these voxels using the data in each split separately. For a set of voxels with a different pRF estimate, the set of images that went into each split could be different. The exception to this was the indoor versus outdoor distinction, where entire images had the same label, and thus the same data split was used for all pRFs (and thus all voxels).

Given that the amount of data used to fit the encoding model can influence its overall predictive power, we always used the same number of images in the splits corresponding to a given semantic dimension and pRF. For example, if there were 5,000 animate labels for a given pRF and only 4,000 inanimate labels, we randomly selected 4,000 images from the set with the animate label. To provide an additional comparison, we also generated a subsampled set of

images for each semantic dimension that was balanced with respect to category (i.e., 50% each label) and had the same total number of images as the smaller category (in this example, it would include 2,000 animate-labeled images and 2,000 inanimate-labeled images). Thus, for a given voxel and a given semantic dimension, the single-category and balanced category models were always fit with the same number of images, facilitating a balanced comparison between the three sets of results (i.e., indoor only, outdoor only, balanced indoor–outdoor). However, the number of images was not necessarily matched for voxels with different pRFs or for splits over different semantic dimensions.

Fitting models with simulated data

To determine whether the nonuniform image statistics of our dataset led to any overall bias in the model's feature sensitivity estimates, we ran a test using simulated voxel data. The rationale for this analysis is that if we simulate the responses of voxels having known orientation and frequency tuning, and we can accurately recover the tuning properties of these simulated voxels, then this suggests our modeling procedure is not biased by the image statistics of the dataset—and that the feature sensitivity values we have measured on the real dataset are due to actual differences in response properties. To construct the simulated data, we started with a grid of 160 pRFs (combinations of eight angular positions evenly spaced between 0° and 315°, four eccentricities log-spaced between 0.24 and 3.93 degrees of visual angle [dva], and five sizes log-spaced between 0.17 and 5.46 dva). For each pRF, we simulated 96 possible "ground-truth" preferred feature values, representing every possible combination of the 8 spatial frequencies and 12 orientations in our original Gabor model. The final simulated population of voxels consisted of every possible combination of preferred feature value and pRF, for a total of 15,360 simulated voxels. Each simulated voxel's response was generated by taking the matrix of features extracted from the images shown to S1 within the voxel's pRF and generating a response that was perfectly correlated with the feature channel corresponding to the voxel's preferred feature value. We then added random Gaussian noise to each response, drawing from a distribution with $\mu = 0$ and $\sigma = 0.10$. We also tested other values of σ for the noise and obtained similar results. Using the simulated dataset, we then performed our model-fitting procedure. Because we were primarily interested in how accurately we could estimate feature selectivity, we used each voxel's ground-truth pRF as a precomputed pRF estimate for the voxel (as described in Methods: Model-fitting procedure) and used the fitting procedure to estimate its feature weights only. We computed the preferred feature values based on these fits as described in the preceding sections.

Data and code availability

The NSD dataset is openly available at http://naturalscenesdataset.org/. All code needed to run our analyses can be accessed at https://github.com/mmhenderson/image_stats_gabor.

Results

Our overall experimental goal was to measure the selectivity of cortical populations for visual features and spatial positions and determine if the properties of selectivity in each visual area can be understood in terms of natural image statistics. We used a two-pronged approach to critically assess different hypotheses. First, we analyzed the relationship between low-level visual (Gabor) features and high-level semantic categories across a large set of natural scene images (Figure 1A). Second, we performed a comprehensive analysis of the orientation, spatial frequency, and spatial selectivity of neural populations throughout visual cortex. The measured neural activity in these populations was taken from the NSD (Allen et al., 2021) in which fMRI scanning was performed while human observers viewed COCO images (an independent set from those used in our first analysis). We constructed voxelwise predictive encoding models based on image-computable features and used these models to estimate voxel selectivity for low-level features as well as spatial position (Figure 1B). To examine the relationship between feature selectivity and category selectivity, we compared our results across a range of category-selective cortical regions and determined whether the specific biases detected in each area corresponded with that area's presumed role in category processing.

Comparing feature statistics across high-level semantic categories

Across a set of 50,000 natural scene images sampled from COCO, we observed several key relationships between low-level Gabor features and high-level semantic categories (Figure 2A). Here we focus on eight commonly studied high-level semantic categories: faces, buildings, animate objects, inanimate objects, small objects (in terms of real-world size), large objects, indoor scenes, and outdoor scenes; see Methods for details on how high-level category labels were

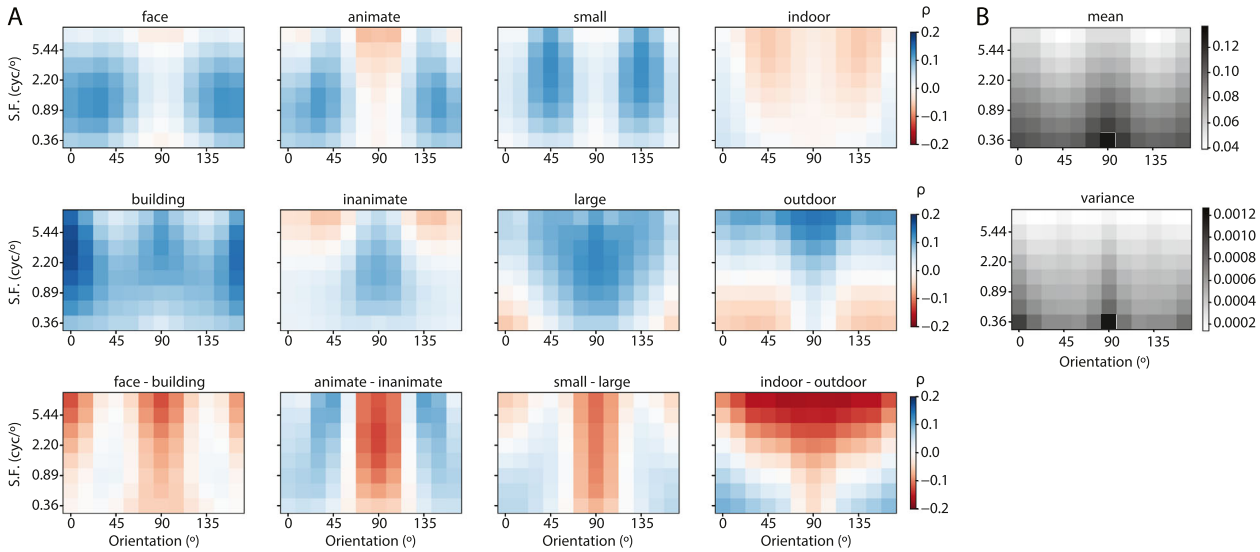

Figure 2. High-level semantic categories are each associated with distinct patterns of low-level visual features. (A) The partial correlation coefficient (ρ) for each Gabor feature channel with each semantic category, with channels organized by spatial frequency on the y-axis and orientation on the x-axis. Our orientation axis is defined such that 0° = vertical and 90° = horizontal. (B) The mean (top) and variance (bottom) across all images for each Gabor feature channel. All analyses in this figure were performed using a set of 50,000 COCO images independent from the ones used in the NSD dataset; see Methods for details.

determined. In the orientation domain, we found that images labeled as outdoor scenes, as well as image patches containing large, inanimate objects, were positively associated with horizontal (90°) orientations (Figure 2A). Consistent with prior results, buildings were also associated with horizontal (90°) orientations, as well as being more strongly associated with vertical (0°) orientations. In contrast, image patches containing faces, animate objects, and/or objects with a small real-world size were positively associated with diagonal orientations (45°/135°). Focusing on the spatial frequency dimension, we observed that outdoor scenes were positively associated with high spatial frequencies, while faces and other animate objects were associated with lower spatial frequencies, particularly around 30° and 150° in orientation. In addition to these semantic associations, the overall mean and variance of feature channels varied across orientation and spatial frequency (Figure 2B), with mean and variance tending to be higher for cardinal orientations (0/90°) as compared to diagonal orientations, and higher for low spatial frequency than high spatial frequency; these findings are likewise consistent with past reports (Girshick et al., 2011; Henderson & Serences, 2021; van der Schaaf & van Hateren, 1996).

Statistical associations between Gabor features and high-level semantic categories indicate that low-level visual features may serve as informative cues for the detection of such categories. To further investigate this association, we used a linear decoding analysis to examine how the category-diagnostic information contained across all Gabor feature channels varied as a function of spatial position in the visual field. Since both our Gabor features and semantic labels were computed in a spatially specific manner across a grid of candidate model pRFs, we performed decoding for each pRF separately and then examined how decoding performance (d'; see *Methods* for details) varied as a function of pRF size and position (Figure 3). First, there was an effect of pRF size that differed depending on the category considered: The status of scenes as indoor or outdoor, as well as the presence of small, inanimate objects, was more easily decodable from large pRFs than small, but the presence of faces and animate objects was more easily decodable from small pRFs than large. Second, the effect of eccentricity also differed across categories, with peripheral pRFs containing more information about the indoor and outdoor scene categories but central pRFs containing more information about faces, animate objects, small objects, and large objects. Third, comparing the upper and lower visual fields, we found that pRFs in the lower visual field were more informative for decoding faces, buildings, inanimate objects, small objects, and outdoor scenes than the upper visual field. No differences were detected between the left and right visual fields.

Note that these category decoding analyses were performed using a fixed number of images for every pRF, so any differences in category diagnosticity across the visual field are necessarily due to differences in the informativeness of the Gabor features themselves, not differences in the frequency of each object category

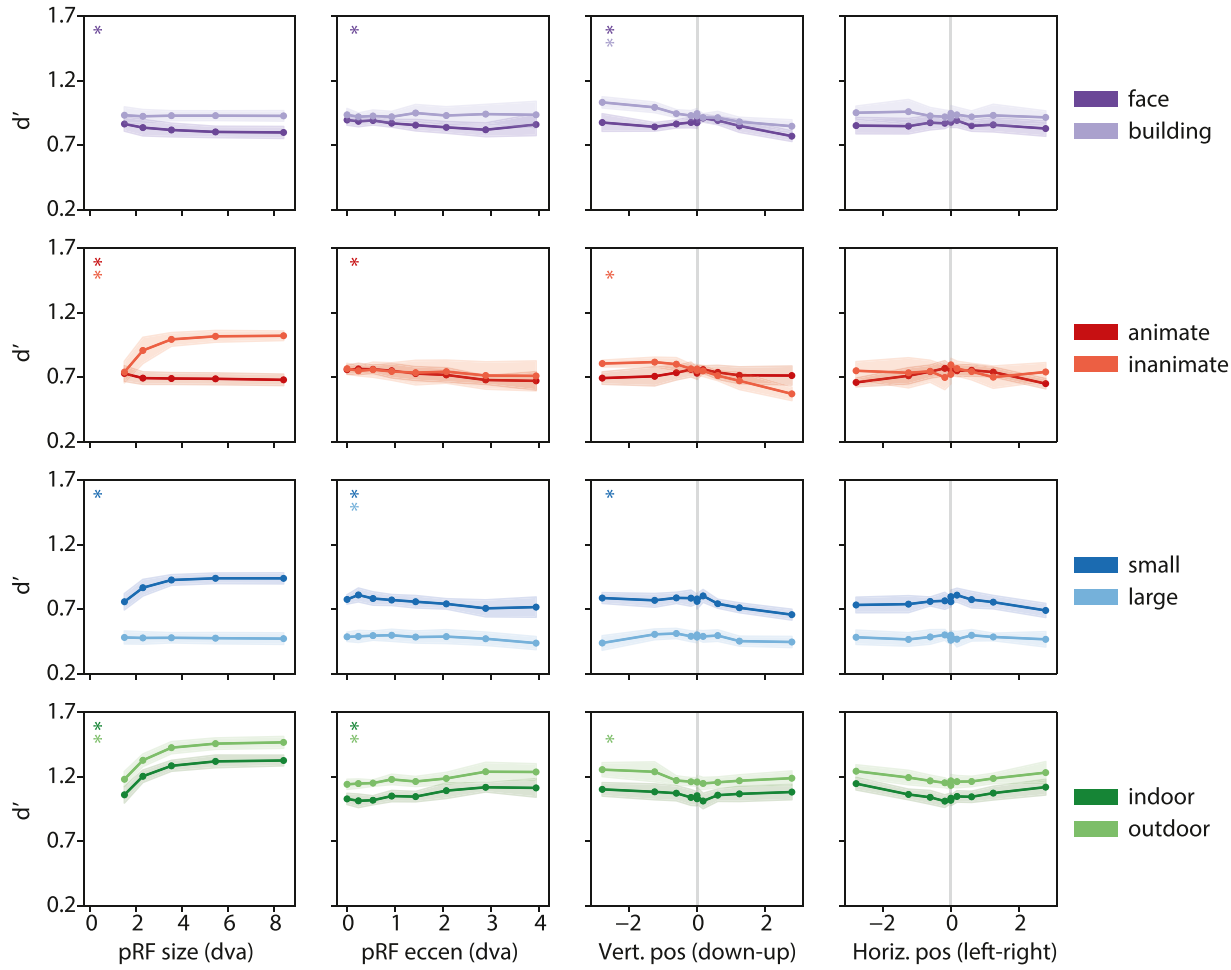

Figure 3. Category diagnosticity of Gabor features differs depending on spatial position. Each plot shows the cross-validated performance (d') of a linear decoder at detecting the specified category based on patterns of activation across all 96 Gabor feature channels (see Methods). Rows represent different semantic axes; each column represents values binned according to one pRF parameter. Error bars indicate $\pm$ 1 *SD* across pRFs within each parameter bin, and * indicates a significant linear relationship evaluated using a permutation test, FDR-corrected $\alpha = 0.01$. All analyses used 50,000 COCO images independent from those used in the NSD dataset; see Methods for details.

across the visual field. However, differences in the frequency of object categories were also observed across pRFs with different parameters (Supplementary Figure S1). All six of the object categories considered in our analysis were more frequent for larger pRFs and for more central pRFs, while the categories of animate, inanimate, small, and large objects were all more frequent in the lower visual field. Interestingly, faces and buildings were both more frequent in the upper visual field, which is in contrast to the finding that these categories were more easily decodable from Gabor features in the lower visual field. Thus, the informativeness of Gabor features for category detection appears to vary with spatial position in a way that is partly dissociable from overall label frequency.

Feature selectivity in visual cortex

Given this evidence for associations between image features and semantic categories, we next evaluated whether neural populations in human cortex exhibit biased tuning that reflects these associations. We focused on voxels within ROIs in visual cortex: early retinotopic ROIs (V1, V2, V3, and hV4), dorsal visual ROIs (V3ab, IPS), scene-selective ROIs (OPA, PPA, RSC), face-selective ROIs (OFA, FFA), and a body-selective ROI (EBA). All ROIs were defined using independent functional mapping data; see Methods for details. The rationale for selecting these regions is that the category-selective ROIs will allow us to test preexisting hypotheses regarding the functional roles of these regions in category processing, while including

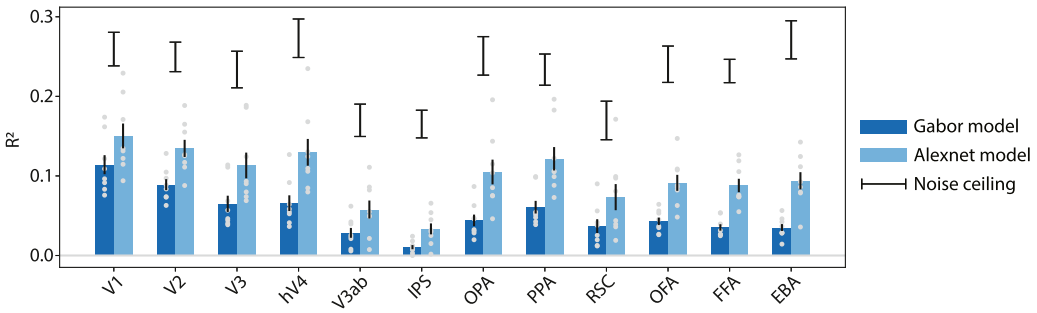

Figure 4. Average cross-validated accuracy (R^2) of the encoding models. The Gabor model was used to estimate voxel selectivity for low-level features, while the AlexNet model was used to fit pRFs; see Methods for details. Each bar represents R^2 averaged over voxels within one ROI; gray points indicate single participants; error bars represent ± 1 *SEM* across 8 participants. Black brackets above each bar indicate the noise ceiling (mean ± 1 *SEM*) for each ROI.

early visual and dorsal ROIs will allow us to explore the generalizability of our findings beyond the typically explored category-selective regions of ventral visual cortex.

To measure selectivity for low-level visual features and spatial positions, we constructed a forward encoding model for each voxel that modeled its response to each image as a linear combination of a set of image-computable features. Our framework also incorporated a model of each voxel's spatial selectivity or pRF (see Figure 1B and Methods for details). We used two different visual feature spaces to construct encoding models. The first feature space was a set of Gabor features with channels corresponding to different combinations of orientation and spatial frequency; this model allowed us to assess voxel selectivity for easily interpretable low-level visual features. The Gabor encoding model was able to predict voxel responses to held-out images with good accuracy (R^2) across a range of visual ROIs, with highest average R^2 in V1, and performance generally declining in more anterior ROIs (Figure 4). The second feature space was a set of features from a deep neural network (AlexNet; Krizhevsky, 2014); this model was used primarily to fit the pRF parameters for each voxel and was chosen because it provided higher predictive performance than the Gabor model, particularly in higher levels of visual cortex (Figure 4).

Once each voxelwise encoding model was fit, we used the models to estimate voxel sensitivity for individual model features. Our measure of feature sensitivity was computed by generating encoding model predicted responses to images in the validation image set (i.e., a set of images not used during model training) and then correlating the predicted responses with the continuous activation values in each feature channel (see Methods for details). Since the orientation and spatial frequency of each Gabor feature channel is known, we then plotted the sensitivity values as a function of orientation and spatial frequency to yield feature sensitivity profiles (Figure 5). Plotting the feature sensitivity profiles, averaged across all voxels in each ROI, revealed several key differences among ROIs. First, early visual ROIs, while having positive sensitivity on average for all Gabor feature channels, had the highest sensitivity for oblique (45°/135°) orientations. This effect, which diverges from the common finding of a cardinal orientation bias in early visual cortex, may be related to the broad spatial frequency content of our natural image stimuli; we return to this issue in the Discussion. In contrast, scene-selective regions tended to have largest average sensitivity for vertical (0°) and horizontal (90°) orientations. Face- and body-selective regions displayed a different pattern, with the highest sensitivity values for oblique orientations 30° and 150°. Along the spatial frequency axis, early visual areas tended to have the highest average sensitivity for spatial frequencies between 2 and 5 cycles/°, while spatial frequency sensitivity profiles in face-selective regions peaked at a lower spatial frequency of around 1 to 2 cycles/°. Scene-selective region RSC showed maximal sensitivity for high spatial frequencies.

Consistent with the ROI-averaged sensitivity profiles, differences were also evident in the distribution of peak sensitivity values (i.e., "preferred" feature values) across individual voxels within each ROI (Supplementary Figure S2). With regard to orientation sensitivity, early visual as well as face-selective and body-selective ROIs had a majority of voxels that preferred oblique orientations (45°/135°), although as in the previous analysis, the distribution in face-selective ROIs was shifted slightly toward vertical relative to the distribution in early visual cortex. In contrast, both IPS and the scene-selective ROIs included a mixture of voxels that preferred vertical (0°) and horizontal (90°) orientations. In these areas, there was a relationship between preferred orientation and preferred spatial frequency, with many of the horizontal-preferring voxels tending to be tuned for high spatial frequencies

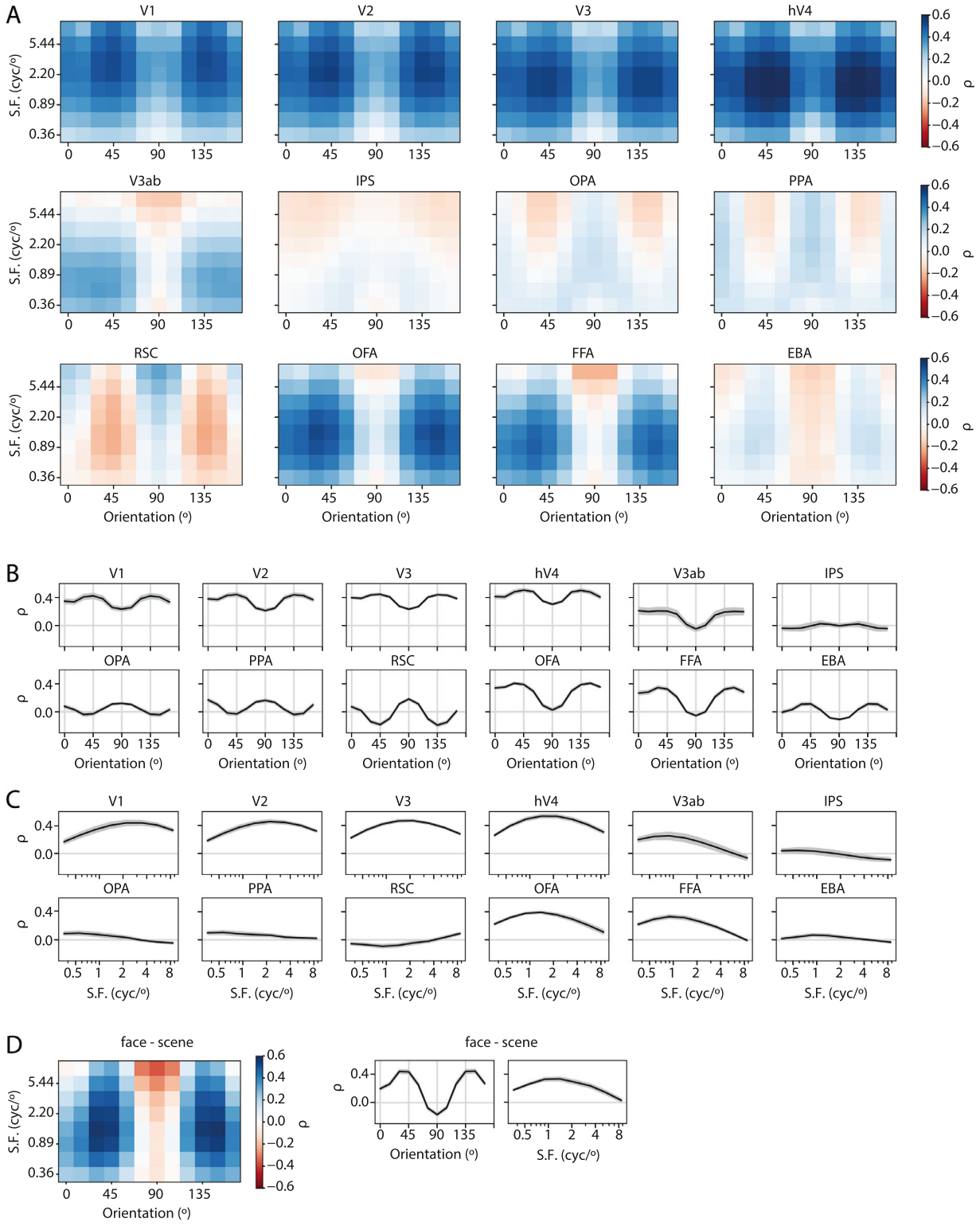

Figure 5. Feature selectivity of voxels differs across early visual and category-selective ROIs. (A) The feature sensitivity profile of each ROI is plotted in a two-dimensional representation, where the x-axis indicates orientation and the y-axis indicates spatial frequency. Feature sensitivity (ρ) was estimated by computing the correlation between the predicted encoding model response for validation set images and the activation in each feature channel for the same images (see Methods for details). (B) The average orientation sensitivity profile (collapsed across spatial frequency) for each ROI. (C) The average spatial frequency sensitivity profile (collapsed across orientation) for each ROI. In (B) and (C), black lines indicate the participant average, and gray shaded error bars indicate $\pm$ 1 *SEM* across participants. See Supplementary Figure S2 and Supplementary Figure S3 for the distribution of peak feature sensitivity values across voxels within ROIs and on a flattened cortical surface. (D) The difference between the average feature sensitivity profiles for face-selective areas (OFA, FFA) and the profiles from scene-selective areas (OPA, PPA, RSC).

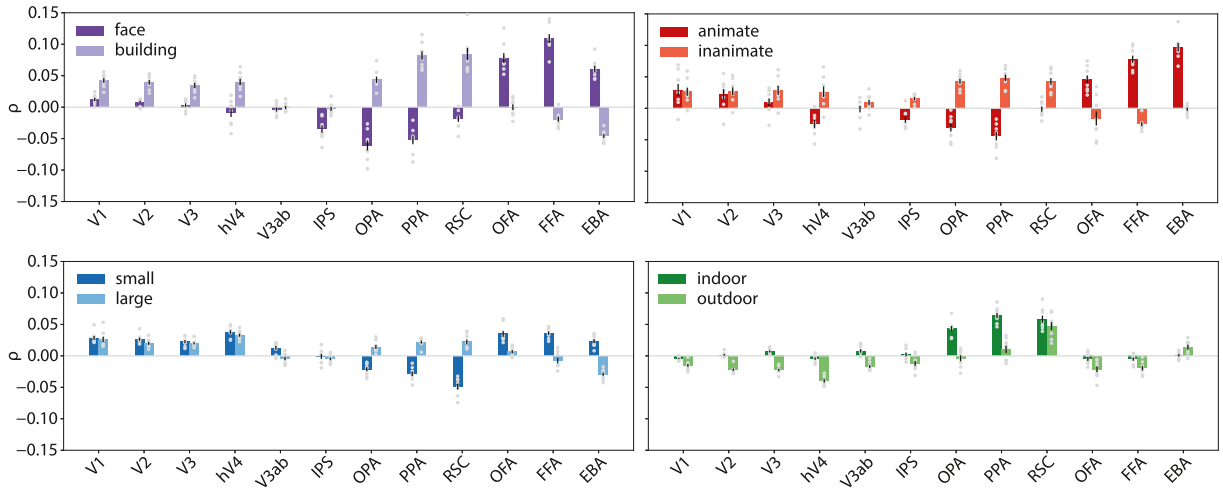

Figure 6. Semantic category selectivity for each ROI. The partial correlation (ρ) between voxel responses in each ROI and each of eight high-level semantic categories (see Methods for details), averaged over all voxels within each ROI and participant. Gray points indicate single participants, and error bars represent $\pm$ 1 *SEM* across eight participants. See Supplementary Figure S5 for the voxelwise selectivity values plotted on a flattened cortical surface.

and many of the vertical-preferring voxels tending to be tuned for lower spatial frequencies. EBA also included a group of voxels that preferred horizontal orientations and high spatial frequencies.

Of note, these results were not dependent on assuming any particular shape for orientation and frequency sensitivity profiles. A supplementary analysis (Supplementary Figure S4) revealed that while the spatial frequency sensitivity profiles for most voxels tended to have only one peak, a significant proportion of voxels across all areas had orientation sensitivity profiles that were bimodal, having two peaks (see Methods for details). When these bimodal voxels were analyzed based on the orientations at which their two peaks occurred, they largely fell into groups that matched the results of our previous analysis (i.e., bimodal voxels in face-selective areas tended to have two peaks at 30°/150°, and bimodal voxels in scene-selective areas tended to have two peaks at 0°/90°). This suggests that the orientation biases in these areas were widespread across voxels, regardless of the exact shape of their sensitivity profiles.

To further aid interpretation of these results and link them to the image statistics analyses (Figure 2), we also analyzed the degree to which average activation in each ROI was correlated with each high-level semantic category (Figure 6). Unsurprisingly, this revealed a distinction between the face- and body-selective areas versus the scene-selective areas, in that activation of voxels in OFA, FFA, and EBA was more correlated with faces and animate objects than buildings and inanimate objects, while activation in RSC, PPA, and OPA was more correlated with buildings and inanimate objects. RSC additionally showed selectivity for large over small objects, which was also evident more weakly in OPA and PPA, while OFA, FFA, and EBA each exhibited selectivity for small over large objects. Both OPA and PPA showed selectivity for indoor scenes over outdoor scenes, while RSC was positively correlated with both indoor and outdoor scenes.

These results, along with the feature sensitivity analyses, provide support for our hypothesis that low-level feature biases in higher visual areas are aligned with the roles of these areas in category processing. On the one hand, scene-selective ROIs have feature sensitivity profiles that match the most diagnostic features for buildings and inanimate objects (Figure 2A), such as vertical and horizontal orientations (see Supplementary Table S5 for a quantification of this alignment). Consistent with RSC's selectivity for large objects, the most preferred features for RSC are similar to those that distinguish large objects from small (Figure 2, bottom row). On the other hand, in face- and body-selective areas, feature sensitivity profiles are aligned with the most diagnostic features for faces, animate objects, and small objects, consistent with the measured category selectivity of these areas. Finally, early visual areas, for example, V1 and V2, which exhibited little category selectivity, were aligned with the features that had the lowest overall mean and variance across all images (Figure 2B), in addition to being aligned with features that were diagnostic of small objects. This latter finding appears consistent with the importance of generic (i.e., category-independent) image statistics as a characteristic of tuning properties in early visual areas; we expand on this point in the Discussion.

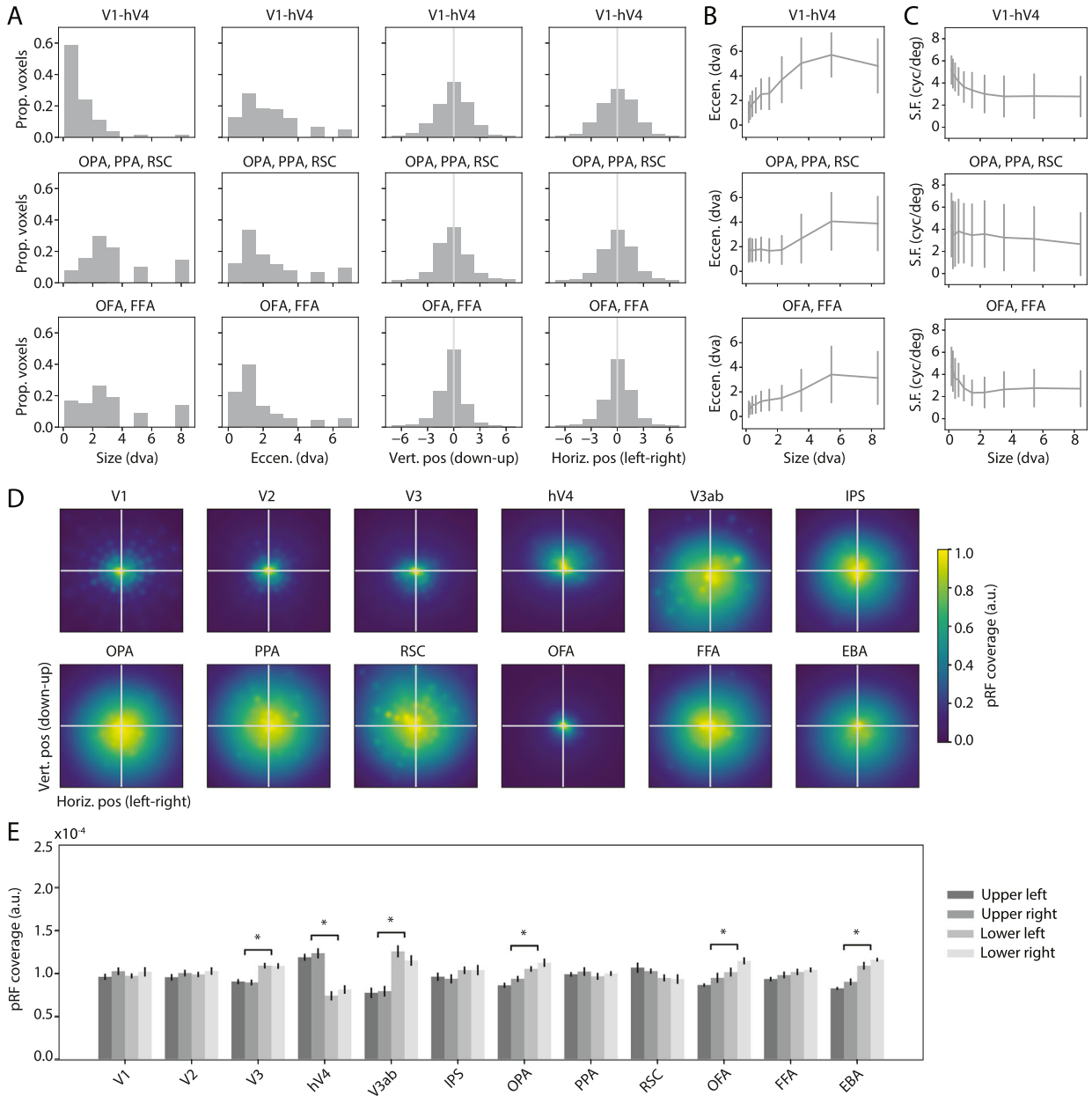

Figure 7. Spatial selectivity of voxels differs across early visual and category-selective ROIs. (A) Distribution of the best pRF parameters across all voxels in early visual (top), scene-selective (middle), and face-selective (bottom) ROIs; see Supplementary Figure S6 for all individual ROIs. (B) Relationship between pRF eccentricity and pRF size. (C) Relationship between the preferred spatial frequency of each voxel (computed as in Figure 5) and the pRF size. In (B) and (C), voxels are binned according to pRF size, and error bars reflect mean $\pm$ 1 *SD* across voxels within each bin. (D) Visualization of the coverage of the visual field by pRFs in each ROI, obtained by averaging the pRFs over individual voxels in each ROI. (E) The mean of the aggregated pRF values in each visual field quadrant (i.e., mean of quadrants in each image of panel (D)). * indicates significance of paired *t*-test for the upper versus lower visual field difference, FDR-corrected $\alpha = 0.01$. Error bars indicate $\pm$ 1 *SEM* across participants. Note that values in panel (D) have been normalized to have a maximum of 1 for visualization purposes, and values in (E) are not normalized.

Spatial selectivity in visual cortex

Analyzing the spatial selectivity of voxels estimated by our encoding model (pRFs; see Methods for details) revealed several trends among brain regions. We found that the median size (σ) of pRFs was smallest in V1 and tended to increase progressively along the anterior axis of the brain, with more large pRFs observed in higher areas like PPA, RSC, FFA, and EBA (Figure 7A and Supplementary Figure S6). Within each

ROI, pRF eccentricity tended to scale positively with pRF size (Figure 7B). These results are consistent with past work (Dumoulin & Wandell, 2008; Klink et al., 2021; Vo et al., 2017) and thus provide validation of our model-fitting procedure. The preferred spatial frequency of voxels also exhibited a relationship with pRF size, with smaller pRFs tending to be associated with higher spatial frequencies, particularly in early areas (Figure 7C). Also consistent with past work, we found that in early visual areas, a correlation was evident between preferred orientation and preferred angular position (Supplementary Figure S7). This result aligns with previous findings of radial bias in early visual cortex (Freeman et al., 2011; Sasaki et al., 2006) and provides additional support for the validity of both our spatial selectivity and feature selectivity estimates.

In addition to these general trends, further examination of the pRF parameters in each ROI indicated biases toward particular portions of the visual field. Focusing first on the eccentricity of pRF centers, we found that face-selective ROIs, especially OFA, had a relatively high proportion of voxels with pRFs close to the central visual field, while scene-selective regions had relatively more voxels with pRFs in the periphery (Figure 7A and Supplementary Figure S6). This result is consistent with previous findings of a central bias for face representations and a peripheral bias for scene representations (Hasson et al., 2002; Levy et al., 2001). In addition to eccentricity, different ROIs also exhibited differences in the distribution of pRF vertical positions (Figure 7A). Both hV4 and RSC tended to have more pRFs with centers in the upper visual field, while other areas, including OPA and EBA, tended to have more centers concentrated in the lower visual field. To quantify this effect, we computed a measure of pRF coverage across the entire visual field by aggregating all pRFs across all voxels for each participant and then taking the mean (Figure 7D; see Methods). Taking the average of these coverage plots within each visual field quadrant provided a measure of pRF coverage for each quadrant (Figure 7E; Silson et al., 2015). Note that this measure is different from counting the number of pRF centers in each quadrant because it takes into account pRF size as well as center. Across participants, a significant interaction between pRF coverage of the upper versus lower visual field and ROI was found (three-way repeated-measures ANOVA with ROI, vertical position, and horizontal position as factors: main effect of ROI, $F_{(11,\ 77)} = 0.00$, $p = 1.0000$; main effect of vertical position, $F_{(1,\ 7)} = 3.29$, $p = 0.1124$; main effect of horizontal position, $F_{(1,\ 7)} = 3.89$, $p = 0.0893$; interaction between ROI and vertical position, $F_{(11,\ 77)} = 10.68$, $p < 10^{-4}$; see Supplementary Table S6 for all interactions). Post hoc tests revealed that pRF coverage was higher for the upper versus the lower visual field in hV4, while average pRF coverage was higher for the lower versus the upper visual field in V3, V3ab, OPA, OFA, and EBA (paired t-test with permutation; FDR-corrected $\alpha = 0.01$). No differences in coverage of the left versus the right visual field were observed.

Comparing these spatial biases to the distribution of high-level semantic category information across the visual field (Figure 3) reveals several relationships. The lower half of the visual field, which contains more information about several object categories—building, face, inanimate, and small—is overrepresented within the category-selective ROIs OPA, OFA, and EBA. This suggests a bias of these areas toward a portion of the visual field useful for object detection. The central visual field, which was found to contain more information about faces, animate objects, large objects, and small objects, was represented most strongly in face-selective areas, particularly OFA. In contrast, scene-selective areas like PPA and RSC had a greater frequency of voxels with peripheral pRFs. This finding is consistent with the observation that peripheral pRFs were more informative for detecting indoor and outdoor scenes. Finally, our decoding analysis (Figure 3) revealed that while large pRFs were more informative for detecting indoor and outdoor scenes as well as inanimate and small objects, small pRFs were more informative for detecting faces and animate objects. Thus, the finding that face-selective areas, particularly OFA, had relatively more voxels with the smallest pRF sizes is consistent with the task-relevant processing demands for detecting faces and other animate objects.

Disentangling feature selectivity and category selectivity

Given that our image set consists of natural scene images conveying semantic category information in addition to visual features, an important question is the extent to which our measurements of visual feature selectivity are influenced by responses to the semantic information itself. To address this possibility, we constructed a second encoding model whose features described the presence of 200 different object categories in the image ("COCO-all" model; see Methods for details). This model is intended to capture the contribution of explicit selectivity for semantic categories to voxel responses. We evaluated how much variance in the voxel responses was explained by the COCO-all feature space by fitting the COCO-all model first on the raw data and then on the residuals of the Gabor model fit (Figure 8A). This analysis revealed that the COCO-all model explained a substantial portion of variance in all visual areas, with highest R^2 in higher visual areas such as PPA. At the same time, a portion of the variance explained by the COCO-all model was shared with the Gabor model, as indicated by a drop in R^2 when fitting on the Gabor model residuals (light blue

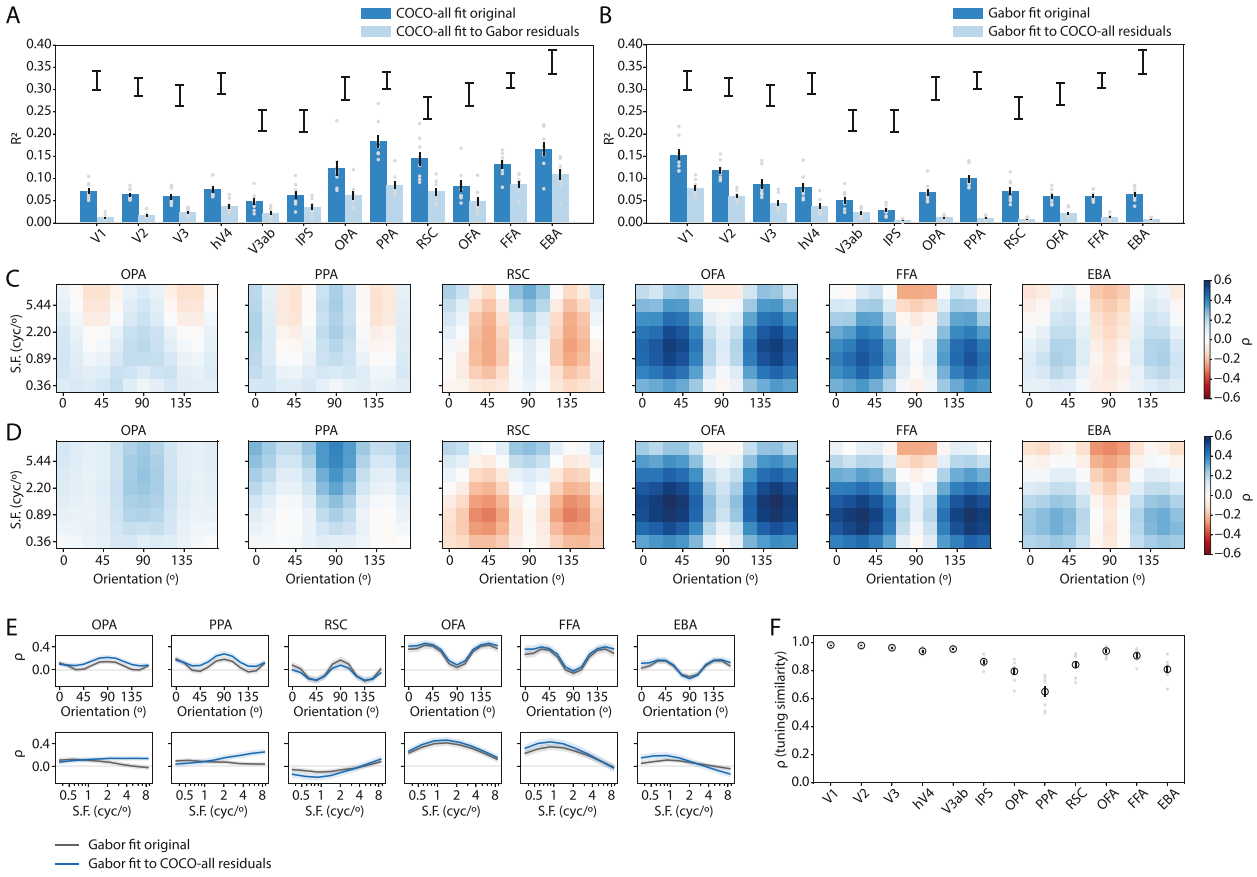

Figure 8. Visual feature selectivity persists after regressing out the contributions of explicit category selectivity from the voxel responses. (A) R^2 for the COCO-all semantic encoding model (a model whose features describe the presence of diverse object and stuff categories in the image; see Methods for details), obtained by fitting the raw data or using the residuals of the Gabor encoding model. (B) R^2 for the Gabor model, obtained using the raw data or the residuals of the COCO-all semantic encoding model. Error bars indicate ± 1 *SEM* across eight participants, and gray dots indicate individual participants. Black brackets above each bar indicate the noise ceiling (mean ± 1 *SEM*) for each ROI. (C, D) The average feature sensitivity profile for each category-selective ROI (x-axis represents orientation, y-axis represents spatial frequency), computed using the model fit on the raw data (C) or fit on the residuals of the COCO-all model (D). Note that the panels in (C) are similar to Figure 5, except that the voxels have been additionally thresholded based on R^2 for the Gabor model fit to the COCO-all residuals (see Supplementary Table S4 for voxel counts after thresholding). (E) Feature sensitivity profiles marginalized over one dimension at a time (top = orientation sensitivity profiles, bottom = spatial frequency sensitivity profiles). Dark gray indicates the model fit on the raw data, and blue indicates the model fit on the COCO-all residuals. Shaded error bars indicate ± 1 *SEM* across participants. (F) The average correlation between individual voxel feature sensitivity profiles (i.e., sensitivity for each of the 96 Gabor features) that were estimated using the raw data or the residuals of the COCO-all semantic model, averaged across voxels in each ROI. High values indicate that feature selectivity was similar whether fitting on the raw data or the residuals. Open circles and error bars indicate mean ± 1 *SEM* across eight participants, and gray dots indicate individual participants.

bars vs. dark blue bars in Figure 8A). This drop was proportionally largest in early visual areas and smaller in higher visual areas, consistent with semantic category features making a larger unique contribution to the responses of higher visual areas. Next, we evaluated how much variance was uniquely contributed by the Gabor model features by performing the analysis in the opposite direction: fitting the Gabor model on the residuals of the COCO-all model (Figure 8B). This revealed a substantial drop in R^2 relative to fitting the Gabor model on the raw data, suggesting that the unique variance contributed by the Gabor features was modest, particularly in higher visual areas. However, a permutation test revealed that many voxels in every ROI were still predicted with above-chance accuracy by the Gabor model fit to the COCO-all model residuals. This result indicates that the Gabor model explained a significant amount of unique variance in these voxels (all p-values significant following FDR correction with $\alpha = 0.01$, see Methods; see Supplementary Table S4 for list of voxel counts in each area after thresholding).

Given that many voxels had significant unique variance explained by the Gabor model relative to the semantic category model, we were then able to ask whether there was any change in the estimated feature sensitivity of these voxels before and after regressing out the contributions of the COCO-all feature space. The rationale for this is that any feature selectivity that is consistent across these methods likely reflects true sensitivity to low-level features, exceeding what is explainable based on category selectivity alone. Critically, this analysis revealed that feature selectivity remained largely consistent after removing the contributions of explicit semantic category selectivity from the voxel responses (Figures 8C–F). This can be seen by comparing the overall feature sensitivity profiles between the raw fits (Figure 8C) and the residual fits (Figure 8D) and was quantified by computing a correlation coefficient between the feature sensitivity profiles across the two methods (Figure 8F). The largest changes in feature sensitivity when regressing out category selectivity were observed in PPA, followed by OPA, RSC, and EBA, although the correlation between feature sensitivity from the raw data and from the residuals was still positive on average in all areas. The changes in both OPA and PPA appeared to be related in part to the spatial frequency sensitivity of voxels. In the raw data, OPA and PPA were, on average, negatively correlated with feature channels corresponding to high spatial frequencies and diagonal orientations, while after regressing out the residuals of the COCO-all model, the average feature sensitivity for these channels was close to zero, and the overall sensitivity to high spatial frequencies increased (Figure 8E). One possible explanation is that the diagonal high spatial frequency channels were also inversely correlated with the "indoor" category as well as the "inanimate" category (Figure 2), both categories with which both OPA and PPA were positively correlated (Figure 6). Thus, the shift in their feature tuning when semantic category selectivity was regressed out may indicate that in our original analysis, some aspects of the feature sensitivity in these areas were actually reflective of sensitivity for indoor scene images and/or inanimate objects. In contrast, feature tuning in all early visual areas as well as in face-selective areas OFA and FFA was similar whether estimated from the raw data or from the residuals (Figure 8F). This latter result suggests that the measured feature selectivity in most visual areas was not driven by signals related to category selectivity.

As an additional way to disentangle the effects of category selectivity and feature selectivity, we next refit the encoding models using images of only one category label at a time (i.e., only indoor scenes, only images with an animate object; see Methods). We compared the tuning properties for models fit using only one category of images at a time, as well as for models fit using balanced sets of images that represented multiple categories equally (see Methods). In agreement with the results of the previous analysis, we found that visual feature selectivity was largely stable across categories (Figure 9). Across all areas, feature sensitivity profiles for single categories were positively correlated with those for balanced sets of images (Figure 9B), with the highest correlation (tuning similarity) values obtained in early visual cortex and OFA and lowest values in scene-selective ROIs and EBA. In scene-selective ROIs OPA, PPA, and RSC, low tuning similarity values were obtained for the "outdoor" and "animate" categories, which may reflect that these areas are all positively correlated with the "indoor" and "inanimate" categories and thus may have a lower signal-to-noise ratio when those preferred image categories are absent (Figure 6). As suggested in the previous analysis, this result may also signify that sensitivity of these ROIs for indoor scenes and/or inanimate objects can explain a portion of their selectivity for low-level features.

In addition to these effects in scene-selective areas, EBA showed a change in its feature sensitivity profile when fit on small objects only as opposed to a balanced set of images, showing mostly negative sensitivity to all feature values. This result was unexpected given that EBA is on average more positively correlated with small objects than large objects (Figure 6); therefore, it is unlikely that these effects are entirely due to the removal of category-selective signals. One possibility is that there is a difference in EBA's low-level feature sensitivity as a function of real-world size, which may reflect a meaningful difference in sensitivity profiles or, alternatively, a difference in the signal-to-noise ratios between responses to large and small objects. A second possibility is that these effects are due to differences in the properties of the model training dataset across categories; for example, images with small objects may contain especially low variance for some feature channels or other properties that could bias our estimates of feature sensitivity toward negative values. To address this latter possibility, we generated a set of simulated voxel responses to the images in our dataset and measured how well the true orientation and spatial frequency selectivity of these simulated voxels could be recovered using our fitting procedure (Supplementary Figure S9). This analysis revealed that our procedure generated similarly accurate estimates of feature selectivity whether fitting was performed using images of just one category (such as "small") or a set that included multiple categories. This argues against the idea that the sampling of the data itself plays a role in the tuning differences seen in Figure 9, instead suggesting that these patterns may be due to an interaction between neural coding of semantic category information and low-level feature information.

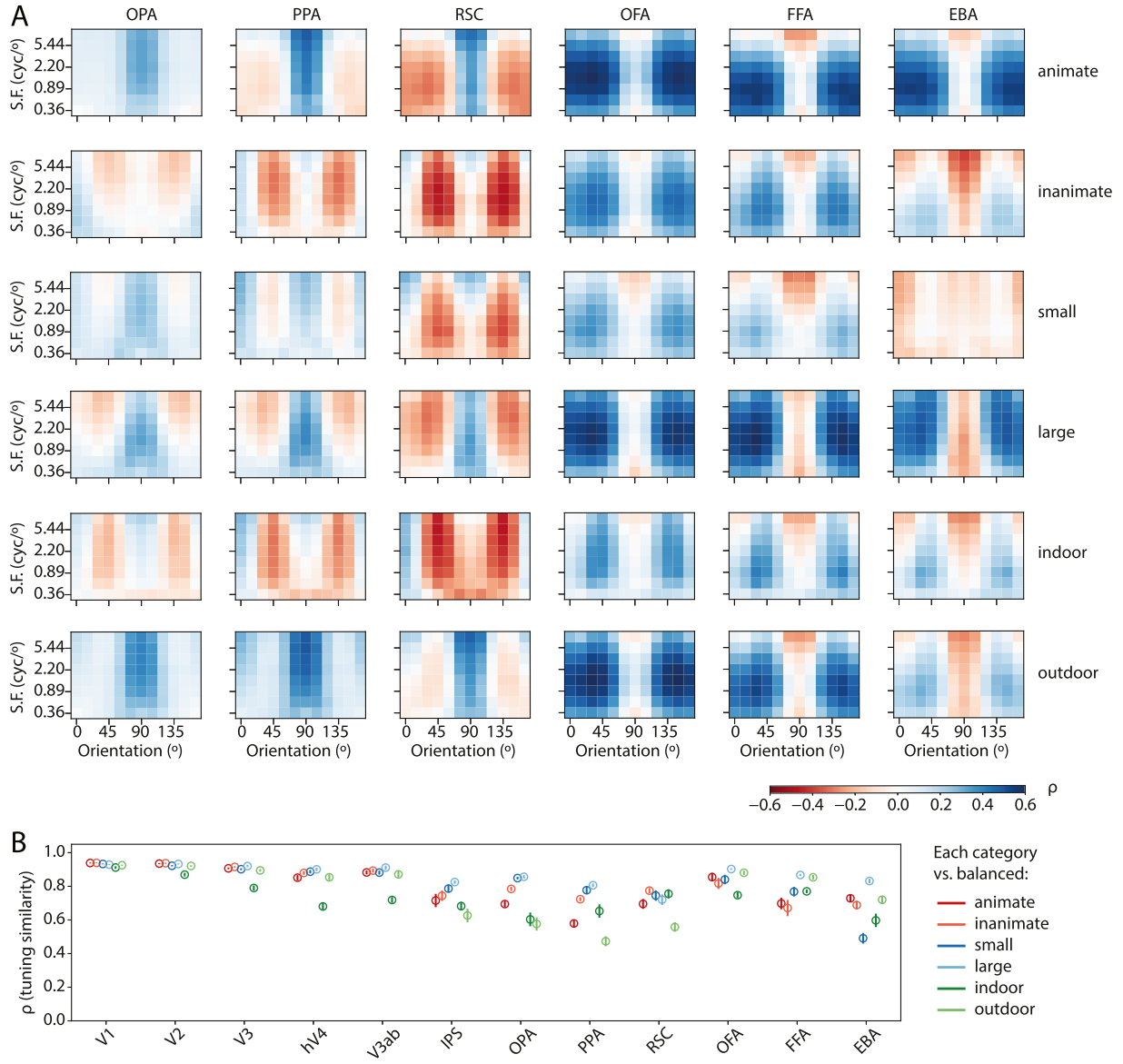

Figure 9. Visual feature selectivity is similar across semantic categories, with interactions evident in some areas. (A) Feature sensitivity profiles computed using images from a single semantic category at a time; each row represents one semantic category, and each column represents one higher visual cortex ROI. (B) The correlation between feature sensitivity profiles computed using images from a single semantic category at a time versus those computed using a balanced set of images from two categories (i.e., for "animate," the balanced set contains equal numbers of animate and inanimate images; see Methods for details). Correlations were computed for individual voxels and then averaged within ROIs and across participants; error bars indicate mean ± 1 *SEM* across eight participants.

Feature selectivity for the top category-selective voxels

The above results provide support for our overall hypothesis with respect to several functionally defined ROIs. However, these ROIs do not provide a perfect parcellation of visual cortex based on category selectivity, as there are also category-selective populations that fall outside the bounds of our ROIs and there is heterogeneity in category selectivity within individual ROIs (Supplementary Figure S5). To address whether our results are generalizable beyond the set of ROIs we have analyzed, we next identified groups of voxels across all of visual cortex that had the highest selectivity for each of our eight high-level categories (excluding retinotopically defined early visual areas; see Methods) and examined their feature selectivity. Before fitting the Gabor encoding model, we used the COCO-all model described previously to regress out the contributions of explicit category selectivity from the voxel responses (as in Figure 8). This ensured that the results obtained were not a trivial consequence

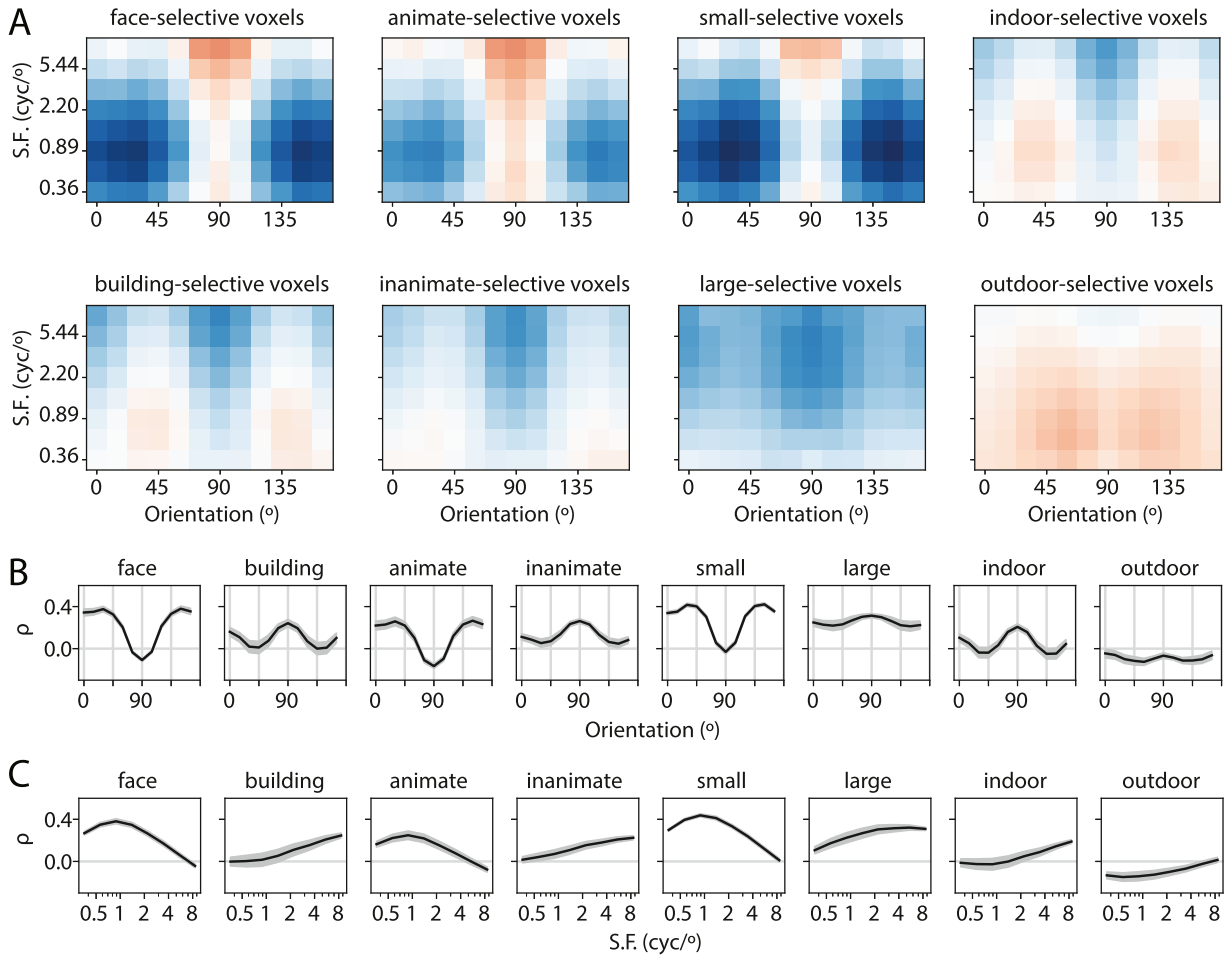

Figure 10. Orientation and spatial frequency selectivity of the most highly category-selective voxels. Selectivity was assessed based on a partial correlation coefficient, and the top 500 voxels for each category were selected (see Methods). Feature sensitivity for each group was computed using the Gabor model fit to the residuals of the COCO-all model (as in Figure 8). (A) The average feature sensitivity profile for each group of voxels is plotted in a two-dimensional representation, where the x-axis indicates orientation and the y-axis indicates spatial frequency. (B) The average orientation sensitivity profile (collapsed across spatial frequency) for each voxel group. (C) The average spatial frequency sensitivity profile (collapsed across orientation) for each voxel group. In (B) and (C), black lines indicate participant average, and gray shaded error bars indicate ± 1 *SEM* across participants.

of the fact that the top category-selective voxels by definition had strong signals related to semantic category membership. Note that in this analysis, the top category-selective voxels are selected using the same data used to quantify the voxels' feature selectivity, which does create a potential issue with circularity. However, with this caveat in mind, the analysis provides a test of the broad generalizability of our results and complements our ROI-based analyses.

As shown in Figure 10, this analysis revealed several patterns. First, the top face-selective, animate-selective, and small-selective voxel groups had very similar average profiles of feature sensitivity; this is likely because these populations partially overlapped with one another. The feature sensitivity of these voxels resembled that of FFA, with highest sensitivity for orientations around 30° and 150°, especially at lower spatial frequencies around 1 to 2 cyc/°. The bias toward these features is consistent with the positive association between these same features and the face, animate, and small labels (Figure 2A). In contrast, the building-selective, inanimate-selective, large-selective, indoor-selective, and outdoor-selective voxel groups were all biased toward cardinal orientations (0°/90°) over obliques (Figure 10B), a finding broadly consistent with the greater diagnosticity of cardinal orientations for each of these high-level categories. Each of these voxel groups was also more sensitive to high spatial frequencies than low (Figure 10C). This spatial frequency bias is consistent with the image statistics of buildings, large objects, and outdoor scenes, which are each positively associated with high spatial frequencies, but is not entirely

consistent with the image statistics of inanimate objects and indoor scenes, which are both associated more strongly with low spatial frequencies than high. Thus, the inanimate-selective and indoor-selective populations may represent a case where the spatial frequency biases of neural populations are not entirely consistent with their role in semantic processing. However, these results overall suggest a broad pattern of consistency between feature sensitivity of neural populations and image statistics, demonstrating that this effect can be identified beyond the boundaries of our original set of ROIs.

Discussion

To extract meaning from the world, the visual system leverages regularities of natural images such as the co-occurrence statistics between sensory features and semantic categories. Correspondingly, such co-occurrence statistics may be reflected within the high-level organization of human visual cortex. We investigated this hypothesis by examining the relationship between low-level visual and high-level category selectivity in visual cortex. First, we demonstrated that images of different semantic categories exhibit distinct patterns of low-level visual features and that information about these distinctions is distributed nonuniformly throughout the visual field. Second, we compared the selectivity for these same features across brain areas thought to play different roles in high-level visual processing and found several correspondences. Areas selective for faces and bodies were biased toward diagonal orientations, aligning with the finding that diagonal orientations were associated with faces and other animate, small objects, while on the other hand, areas selective for scenes were biased toward vertical and horizontal orientations, consistent with the diagnosticity of these orientations for buildings and other large, inanimate objects. In terms of spatial selectivity, face-selective and scene-selective regions were biased toward the central and peripheral visual field, respectively, in agreement with our observation that the central visual field was more informative for detecting object categories, while the peripheral visual field provided more information about scene categories. Several ROIs were also biased toward the lower visual field, which contained diagnostic features for multiple high-level categories. Together, these results suggest that cortical regions tend to overrepresent components of the visual world that are informative for detecting their preferred category. This supports the idea that low-level feature biases observed throughout visual cortex reflect the structure of the visual environment and thus may provide organizational scaffolding for high-level category representations.

Biases in orientation and spatial frequency selectivity

With respect to the dimension of orientation, our results revealed that scene-selective visual ROIs and IPS were preferentially selective for the cardinal orientations (0°/90°), while early visual, face- and body-selective ROIs exhibited stronger selectivity for diagonal orientations. The finding of cardinal bias in scene-selective areas is consistent with past reports (Nasr & Tootell, 2012), but the lack of such a bias in early visual areas is surprising in light of past work demonstrating that at the single-unit level, more neurons in V1 and other early areas tend to be tuned for vertical and horizontal orientations than diagonals (Mansfield, 1974; Li et al., 2003; Shen et al., 2014). A possible explanation for this discrepancy is the fact that we measured neural responses to natural image stimuli, whereas most past studies have used simple, synthetic stimuli such as oriented gratings. Indeed, at the behavioral level, it has been suggested that the commonly observed "oblique effect," in which observers tend to show better performance close to the cardinal orientations and worse at the obliques (Appelle, 1972), may reverse when measured with more naturalistic, broadband spatial frequency images (Essock et al., 2003). This "horizontal effect" has been explained in terms of a divisive normalization mechanism in which the normalization pool acting to suppress neuronal activity might be larger for cardinal orientations than oblique orientations (Essock et al., 2003). Given that horizontal and vertical orientations are overall more common in natural images (see Figure 2B and Coppola et al., 1998; Girshick et al., 2011; Henderson & Serences, 2021), this mechanism might serve to suppress the most common orientations in natural images while enhancing processing of uncommon (i.e., unexpected) information (Essock et al., 2003). Such an explanation would be consistent with the idea that orientation anisotropies in early visual cortex reflect efficient coding of natural image statistics and suppression of redundant information (Barlow, 1961; Coen-Cagli et al., 2015; Klímová et al., 2021). At the same time, another factor that may contribute to the discrepancy between our results and past electrophysiology work is the difference in recording method. Since the fMRI signal reflects pooled synaptic activity over many neurons, it may give less precise measures of orientation selectivity than single-neuron recordings (O'Herron et al., 2016) and could also be differentially sensitive to the divisive normalization mechanism described above. Supporting this, some past fMRI studies, in agreement with our results, have found greater activation of early visual cortex for oblique than cardinal orientations (Mannion et al., 2010; Swisher et al., 2010). On the other hand, one fMRI study found that the relative strength of activation for oblique orientations versus cardinals

is dependent on stimulus properties such as contrast (Maloney & Clifford, 2015), which lends support to the idea that orientation selectivity may exhibit a different functional signature when measured using natural images versus gratings.

Considering these factors in toto, one interpretation of the difference between early visual cortex and scene-selective cortex in our results is that feature selectivity in early areas may be more strongly constrained by generic efficient coding of images, without regard to semantic content, while higher visual areas may reflect stronger semantic constraints, including the association of cardinal orientations with scene-diagnostic content such as large, inanimate objects. An increase in the magnitude of cardinal orientation bias from early to late stages of visual processing is also consistent with past reports of stronger cardinal biases at deeper layers of a convolutional neural network (Henderson & Serences, 2021). This pattern is also compatible with our findings in face- and body-selective ROIs, in which orientation biases were similar to those in early areas, but with a slight shift in the distribution of preferred orientations toward vertical (shifting from peaks at 45°/135° to peaks at 30°/150°; Figure 5). Given that 30° and 150° were the orientations most strongly associated with faces and other animate objects (Figure 2), this shift in orientation biases may reflect an increase in the influence of semantic constraints on feature selectivity when moving from early visual to higher visual regions.

With respect to a second low-level dimension, spatial frequency, we again found evidence for differences in the spatial frequency selectivity of voxels across different ROIs. One notable finding was that RSC and PPA each exhibited maximum sensitivity for high spatial frequencies, although in PPA, this was only the case after regressing out the contributions of category selectivity from the voxels' responses (Figure 8). The finding of high spatial frequency selectivity in PPA has been reported previously (Kauffmann et al., 2015; Rajimehr et al., 2011), although others have reported relatively less sensitivity to high spatial frequency in RSC (Kauffmann et al., 2015). Further supporting the relevance of high spatial frequencies for processing in scene regions, a recent study also showed that decoding of scene categories in PPA and RSC was driven predominantly by high spatial frequency information (Berman et al., 2017). Our results are consistent with this finding, and we further show that high spatial frequency features carry information about semantic categories such as buildings and outdoor scenes (Figure 2). Thus, the high spatial frequency biases in scene-selective ROIs appear to be consistent with their proposed roles in scene processing. In contrast to scene-selective regions, face-selective areas OFA and FFA were observed to be more sensitive to low spatial frequencies, with maximum sensitivity at around 1 to 2 cyc/°. This is consistent with both the observation that these lower frequencies were associated with faces and other animate objects and past work implicating low spatial frequencies in face processing, particularly for global configural processing (Goffaux et al., 2005; Goffaux & Rossion, 2006).

Biases in spatial selectivity

With respect to spatial selectivity, we found evidence for biased coverage of the visual field in several ROIs, as well as corresponding biases in the distribution of category-diagnostic feature information around the visual field. In particular, category-selective areas OPA, OFA, and EBA all showed bias in favor of the lower half of the visual field (Figure 7), and correspondingly, the decodability of several high-level semantic categories was better in the lower visual field (Figure 3). These findings are consistent with past reports of high-level visual areas on the lateral surface of the brain tending to be biased in favor of the lower hemifield (Sayres & Grill-Spector, 2008; Schwarzlose et al., 2008; Silson et al., 2016; Silson et al., 2015), as well as with past findings of scene-diagnostic objects tending to be more concentrated in the lower hemifield (Greene et al., 2013). We build on this work by demonstrating that high-level object and scene properties are more easily decodable from features in the lower versus the upper visual hemifield and that this spatial bias in the visual environment is reflected in patterns of selectivity within specific visual areas. Of note, there is some evidence from past studies for an upper visual field bias in visual areas on the ventral surface, including PPA and FFA (Groen et al., 2017; Silson et al., 2015), although at least one study found no evidence for an upper visual field bias in FFA (Finzi et al., 2021). We did not find evidence for a visual field bias in FFA or PPA in our main set of pRF analyses. However, in a supplementary analysis, where we fit pRFs using a different model, some evidence for an upper visual field bias was observed in both PPA and RSC, although not FFA (Supplementary Figure S8). These disparate results suggest that methodological choices may be a contributing factor to the different measurements of visual field bias in ventral visual cortex across studies.

Previous work has also reported a lower visual field bias in early visual cortex, with more cortical surface area allocated to representing the lower versus upper vertical meridian in V1 (Benson et al., 2021). Similarly, there are reports that pRF sizes are smaller and the cortical magnification factor is larger in the lower versus upper visual field in V1, V2, and V3 (Silva et al., 2018). These results are broadly consistent with our finding that the lower visual field was overrepresented within V3, as well as mid-level retinotopic area V3ab, although we did not find evidence for a lower visual field advantage in V1 or V2. These differences are likely due

to differences in methodology, such as the fact that past results were based on dedicated retinotopic mapping experiments, whereas ours were based on model-based pRF estimates obtained from responses to natural scene viewing, as well as the fact that we computed pRF coverage of each hemifield as opposed to cortical surface area. We also did not analyze the dependence of the vertical meridian asymmetry on eccentricity, as has been done in past work. Such considerations are beyond the scope of our current study, but whether converging results can be obtained using our modeling approach remains an interesting question for future research.

In addition to biased coverage of the lower versus upper visual hemifield, we observed biases along the central to peripheral axis. A central bias was measured in face-selective areas, particularly OFA, while scene-selective areas PPA and RSC had relatively more voxels with peripheral pRFs. These findings are broadly consistent with past findings of central and peripheral biases in face- and scene-selective areas, respectively (Finzi et al., 2021; Hasson et al., 2002; Levy et al., 2001). Compared to the large effects measured in these past studies, however, the central-to-peripheral bias we measured was relatively subtle. One possible contributing factor to this difference is that our stimulus was smaller (an 8.4° square) than that used in past work (Hasson et al., 2002, used a circle of max diameter 20°). In addition to these findings regarding neural tuning properties, we also extended this past work by showing that information carried by Gabor features about the presence of faces, animate objects, small objects, and large objects was higher for more central pRF positions, while information about the indoor versus outdoor scene distinction was higher for more peripheral pRF positions. The role of peripheral and/or global image features in supporting scene perception has been suggested by past work (Greene & Oliva, 2009a, 2009b; Groen et al., 2017). The correspondence between these sets of findings is consistent with the idea that face-selective areas are biased in favor of the detailed parts of the image that may be informative for face detection (i.e., the central visual field), while scene-selective areas are biased in favor of the parts of the image more informative for processing large-scale scene distinctions.

Interactions between feature and category coding

Taken together, the above results provide evidence for reliable, selective sensitivity to low-level features throughout the entire visual cortical hierarchy. Moreover, we demonstrate that the feature selectivity seen in higher visual cortex cannot be attributed solely to coding of semantic category information. That is, even after removing the contributions of explicit category selectivity from voxel responses, a Gabor encoding model predicted a significant amount of the variance, and biases in feature selectivity for each brain region were largely stable following this manipulation. Our observation is in keeping with past work showing low-level feature selectivity and biases for particular features in higher visual cortex regions (Bermudez et al., 2009; Nasr & Tootell, 2012; Rajimehr et al., 2011; Vogels & Orban, 1994; Yue et al., 2014) and supports the interpretation that higher visual cortex responses reflect a mixture of visual and categorical information.

At the same time, our results also suggest potential interactions between feature coding and category coding in some higher-level areas, particularly in scene-selective visual regions and EBA. In OPA and PPA, the sensitivity to spatial frequency changed when category selectivity was regressed out of the voxel responses (Figure 8), and in OPA, PPA, and EBA, feature sensitivity profiles changed somewhat when fitting models on images from different high-level categories (Figure 9). Simulations suggested that these effects were not driven by differences in dataset sampling across these categories (Supplementary Figure S9), indicating that feature sensitivity was at least partly dependent on the category of images. These interactions could potentially take the form of a difference in signal-to-noise ratio (SNR) across categories—for example, since OPA and PPA are each on average more responsive to large, inanimate objects and indoor scenes, this could lead to lower SNR for detecting low-level feature sensitivity with images of small, animate objects and/or outdoor scenes. Such interactions could also reflect more complex coding schemes, such as nonlinear mixed selectivity for high-level categories and low-level visual features. Modeling the presence of additional mid-level or high-level visual features may provide further insight into these possibilities; this remains a question for future work.

In past work, several theoretical proposals have been put forth to explain the relationship between visual and category selectivity. One hypothesis is that functional selectivity for a category reflects a combination, either additive or nonlinear, of selectivity for a set of underlying features that comprise the category (Op de Beeck et al., 2008). Under such a coding scheme, neural populations might be selective for sets of features that are diagnostic of a particular category but also exhibit some residual selectivity for the category itself (Bracci & Op de Beeck, 2016; Bracci et al., 2017). Similarly, work using synthetic stimuli has shown that mid-level visual features, in the absence of semantic information, can generate signatures of categorical selectivity in ventral visual cortex but that other (possibly semantic) features may be required to elicit the strongest level of

category selectivity in these regions (Long et al., 2018). Our results are consistent with this model, based on our finding that low-level visual features and semantic category features, beyond the substantial overlap in the variance explained by each factor, each explained significant unique variance within higher visual cortex (Figure 8).

Associations between low-level features and semantic categories

Our measurements of the associations between low-level features and semantic categories (Figure 2) are consistent with several earlier studies. For example, Torralba and Oliva (2003) compared the spectral content for a range of image categories and found differences between scene categories, as well as between images of different object classes. Consistent with our findings, certain classes of outdoor images, such as forests and cities, showed larger power at high spatial frequencies. Similarly, certain classes of outdoor images, such as beaches and highways, had higher power at horizontal orientations. At the object category level, Torralba and Oliva reported spectral differences across images containing animals versus other objects, which is consistent with our observation that certain orientation channels were associated with animate over inanimate objects. However, an important difference between our approach and theirs is that we labeled images according to their animacy or inanimacy in a spatially specific manner—when we computed the features that were correlated with animacy, we only used features extracted from image patches that actually contained an inanimate object, an animal, or a person. Thus, the features correlated with animacy in our analyses are not likely to have been as strongly driven by background content as those used in Torralba and Oliva's work. Despite this, there is still a correspondence between our finding that diagonal orientations were positively associated with the animate–inanimate axis and Torralba and Oliva's finding that images with animals had relatively more power at diagonal orientations (i.e., were less strongly dominated by cardinal orientations) than images with other types of objects. Our finding that vertical and horizontal orientations were associated with buildings and other large objects is also consistent with their observation of strong cardinal biases in scenes including buildings and/or cars. The relationship between cardinal orientations and large objects is also consistent with the observation that objects having a large real-world size are dominated by boxy contour features (Long et al., 2016; Nasr et al., 2014). Importantly, our results build on these past observations by providing detailed comparisons of the correspondence between a fine-grained bank of Gabor features and several high-level semantic categories. These measurements provide a foundation for interpreting the fine-grained biases in feature representations within visual cortex.

Limitations of the natural image statistics hypothesis

Although our results are primarily in agreement with the hypothesis that low-level tuning biases in higher visual cortex reflect natural image statistics, some aspects of our results do not align with this hypothesis. First, we found evidence that in OPA and PPA, feature sensitivity changed when explicit category selectivity was regressed out, with the corrected feature sensitivity of these areas reflecting greater sensitivity to high spatial frequencies. The "true" spatial frequency sensitivity of these areas was thus more closely aligned with the diagnostic features for outdoor scenes than indoor, which is at odds with the observation that OPA and PPA were more correlated with the indoor scene label. Similarly, when we looked at feature sensitivity of the top indoor-selective and the top inanimate-selective voxels, each of these populations showed biases toward high spatial frequencies over low, which is inconsistent with the statistics of indoor scenes and inanimate objects. Thus, the categories of indoor scenes and inanimate objects may represent cases where the statistics of images are not perfectly matched by the spatial frequency sensitivity of category-selective neural populations. This could suggest that detection of indoor scenes and/or inanimate objects is more strongly driven by mid-level or high-level features than Gabor-like features.

The organization of higher visual cortex

We find support for our overall hypothesis both within commonly used category-selective ROIs, as well as within voxel populations beyond these defined ROIs (Figure 10). The consistency of these results is in line with the idea that the large-scale organization of higher visual cortex may be better understood in terms of broad visual feature dimensions that are mapped onto the cortex in a continuous fashion (Bao et al., 2020; Konkle & Caramazza, 2013; Op de Beeck et al., 2008), rather than as a discrete set of regions processing disparate types of information. Within this framework, the biases that we observed in feature tuning within functionally localized regions of visual cortex may be interpreted as reflecting the correspondence between maps encoding features at different levels of complexity. This view is also compatible with our

observation that there was heterogeneity for the feature sensitivity profiles and semantic selectivity across voxels within a given ROI (Supplementary Figure S3 and Supplementary Figure S5). It seems plausible that the heterogeneity in feature tuning of voxels across an ROI is related to heterogeneity in their semantic selectivity; again, future work will be needed to fully explore this possibility.

The structured relationship between category selectivity and visual feature selectivity in the brain may emerge due to multiple factors, including visual experience during development and functional or anatomical neural constraints. One account for the organization of visual cortex suggests that from infancy, the primate brain includes a "proto-organization" that may be a precursor to the large-scale maps of category and feature selectivity observed in adult brains (Arcaro & Livingstone, 2017; Livingstone et al., 2019). These early topographic constraints, including selectivity for low-level features such as spatial frequency and curvature, may interact with visual input to constrain where mature category-selective visual areas will develop (Op de Beeck et al., 2019). For example, retinotopic biases present in the early visual system interact with the portion of the visual field in which certain classes of stimuli tend to fall (i.e., faces and words tend to be foveated, buildings tend to land in the periphery), and this may lead to a correspondence between retinotopy and category selectivity (Hasson et al., 2002; Levy et al., 2001; see Groen et al., 2022, for a review). Weak early biases for features such as curvature may similarly lead cortical populations to develop selectivity for those categories in which such features are prominent (Livingstone et al., 2019). Our results provide some support in favor of this hypothesis, in that we find a correspondence between the features and spatial positions associated with a given category and the low-level tuning of neural populations involved in processing that category. Thus, it is plausible that experience with the statistics of natural images during development, along with some degree of early topographic organization, may be sufficient to predict the organization of feature, spatial, and category selectivity in the adult visual system.

The functional role of tuning biases

Category-selective visual areas appear to contain representations of diagnostic low-level features, which could suggest that these low-level feature representations play a functional role in category perception. Although our study did not directly assess this functional role, past work supports the idea that low-level visual features may influence behavioral judgments of object and/or scene category. For example, rapidly classifying scenes into basic-level categories may be mediated by global properties associated with the spectral content of scenes (Greene & Oliva, 2009a, 2009b; Oliva & Torralba, 2001). Detecting object categories, such as animals, may also be supported by spectral differences across images (Torralba & Oliva, 2003). At the same time, other work has suggested that in the case of animal detection, spectral features may not be sufficient to predict behavior (Wichmann et al., 2010) and that features more complex than spectral content, such as mid-level textural features (Long et al., 2016; Long et al., 2017) or contour junctions (Walther & Shen, 2014), may be more useful in computing semantic properties of objects and scenes. Additional experiments are needed to determine whether the feature biases we measured play a functional role in behavior, but our results do add to a growing body of evidence that low-level features contain potentially informative cues to the semantic meaning of images. We build on this past work by showing that these cues are reflected in the brain across a range of visual areas and semantic category distinctions.

Conclusion

Our results provide evidence supporting the theory that representations of low-level features within category-selective regions of visual cortex are aligned with the high-level computational goals of these regions. Moreover, this principle appears to hold for category-selective neural populations across a wide range of visual areas, even beyond ROI boundaries, and for multiple semantic categories. Such findings suggest that the computation of semantic meaning in the visual system may reflect contributions from features at multiple levels of complexity.

Keywords: category selectivity, orientation tuning, population receptive field, natural image statistics, encoding model

Acknowledgments

Funded by a Distinguished Postdoctoral Fellowship from the Carnegie Mellon Neuroscience Institute to MMH. Collection of the NSD dataset was supported by NSF IIS-1822683 and NSF IIS-1822929.

Commercial relationships: none.
Corresponding author: Margaret M. Henderson.
Email: mmhender@cmu.edu.
Address: Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA.

References

Aghajari, S., Vinke, L. N., & Ling, S. (2020). Population spatial frequency tuning in human early visual cortex. *Journal of Neurophysiology, 123*, 773–785.

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., . . . Kay, K. (2021). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience, 25*, 116–126.

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin, 78*, 266–278.

Arcaro, M. J., & Livingstone, M. S. (2017). A hierarchical, retinotopic proto-organization of the primate visual system at birth. *eLife, 6*, e26196.

Arcaro, M. J., McMains, S., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *Journal of Neuroscience, 29*, 10638–10652.

Bao, P., She, L., McGill, M., & Tsao, D. Y. (2020). A map of object space in primate inferotemporal cortex. *Nature, 583*, 103–108.

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*, 289–300.

Benson, N. C., Jamison, K. W., Arcaro, M. J., Vu, A. T., Glasser, M. F., Coalson, T. S., . . . Kay, K. (2018). The human connectome project 7 tesla retinotopy dataset: Description and population receptive field analysis. *Journal of Vision, 18*, 23, https://doi.org/10.1167/18.13.23.

Benson, N. C., Kupers, E. R., Barbot, A., Carrasco, M., & Winawer, J. (2021). Cortical magnification in human visual cortex parallels task performance around the visual field. *eLife, 10*, e67685.

Berman, D., Golomb, J. D., & Walther, D. B. (2017). Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. *PLoS ONE, 12*, e0189828.

Bermudez, M. A., Vicente, A. F., Romero, M. C., Perez, R., & Gonzalez, F. (2009). Spatial frequency components influence cell activity in the inferotemporal cortex. *Visual Neuroscience, 26*, 421–428.

Bonhoeffer, T., & Grinvald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature, 353*, 429–431.

Bracci, S., & op de Beeck, H. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience, 36*(2), 432–444.

Bracci, S., Ritchie, J., & de Beeck, H. O. (2017). On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia, 105*, 153–164.

Cadieu, C. F., Hong, H., Yamins, D. L., Pinto, N., Ardila, D., Solomon, E. A., . . . DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS Computational Biology, 10*, e1003963.

Caesar, H., Uijlings, J., & Ferrari, V. (2018). Coco-stuff: Thing and stuff classes in context. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1209–1218). Los Alamitos, CA: IEEE Computer Society.

Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nature Neuroscience, 18*, 1648–1655.

Conway, B. R., & Tsao, D. Y. (2009). Color-tuned neurons are spatially clustered according to color preference within alert macaque posterior inferior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 18034–18039.

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 4002–4006.

Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5203–5212). Los Alamitos, CA: IEEE Computer Society.

Downing, P., Chan, A., Peelen, M., Dodds, C., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex (New York, N.Y.: 1991), 16*, 1453–1461.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage, 39*, 647–660.

Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron, 38*, 659–671.

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

Essock, E. A., DeFord, J. K., Hansen, B. C., & Sinai, M. J. (2003). Oblique stimuli are seen best (not worst!) in naturalistic broad-band stimuli: A horizontal effect. *Vision Research, 43*, 1329–1335.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics, 49*, 92.

Finzi, D., Gomez, J., Nordt, M., Rezai, A. A., Poltoratski, S., & Grill-Spector, K. (2021). Differential spatial computations in ventral and lateral face-selective regions are scaffolded by structural connections. *Nature Communications, 12*, 1–14.

Freeman, J., Brouwer, G. J., Heeger, D. J., & Merriam, E. P. (2011). Orientation decoding depends on maps, not columns. *Journal of Neuroscience, 31*, 4792–4804.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience, 14*, 926–932.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception, 34*, 77–86.

Goffaux, V., & Rossion, B. (2006). Faces are "spatial"—holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1023–1039.

Greene, M. R., & Oliva, A. (2009a). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20*, 464–472.

Greene, M. R., & Oliva, A. (2009b). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*, 137–176.

Greene, M. R., Zelinsky, G., Tatler, B., & Rosenholtz, R. (2013). Statistics of high-level scene context. *Frontiers in Psychology, 4*, 777.

Groen, I. I. A., Dekker, T. M., Knapen, T., & Silson, E. H. (2022). Visuospatial coding as ubiquitous scaffolding for human cognition. *Trends in Cognitive Sciences, 26*(1), 81–96.

Groen, I. I. A., Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*, 20160102.

Güçlü, U., & van Gerven, M. A. (2014). Unsupervised feature learning improves prediction of human brain activity in response to natural images. *PLoS Computational Biology, 10*, e1003724.

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.

Henderson, M., & Serences, J. T. (2021). Biased orientation representations can be explained by experience with nonuniform training set statistics. *Journal of Vision, 21*, 10, https://doi.org/10.1167/jov.21.8.10.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*, 55–67.

Huth, A. G., Heer, W. A. D., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*, 453–458.

Issa, N. P., Trepel, C., & Stryker, M. P. (2000). Spatial frequency maps in cat visual cortex. *Journal of Neuroscience, 20*, 8504–8514.

Jain, N., Wang, A., Henderson, M. M., Lin, R., Prince, J. S., Tarr, M. J., . . . Wehbe, L. (2023). Selectivity for food in human ventral visual cortex. *Communications Biology, 6*, 175.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*, 4302–4311.

Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage, 112*, 86–95.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452*, 352–355.

Kay, K. N., Winawer, J., Mezer, A., & Wandell, B. A. (2013). Compressive spatial summation in human visual cortex. *Journal of Neurophysiology, 110*, 481–494.

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS Computational Biology, 10*, e1003915.

Khosla, M., Murty, N. A. R., & Kanwisher, N. (2022). A highly selective response to food in human visual cortex revealed by hypothesis-free voxel decomposition. *Current Biology, 32*, 1–13.

Klímová, M., Bloem, I. M., & Ling, S. (2021). The specificity of orientation-tuned normalization

within human early visual cortex. *Journal of Neurophysiology, 126*, 1536–1546.

Klink, P. C., Chen, X., Vanduffel, V., & Roelfsema, P. (2021). Population receptive fields in non-human primates from whole-brain fMRI and large-scale neurophysiology in visual cortex. *eLife, 10*, e67304.

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience, 33*, 10235–10242.

Krizhevsky, A. (2014). One weird trick for parallelizing convolutional neural networks. *arXiv*, https://doi.org/10.48550/arXiv.1404.5997.

Lescroart, M. D., & Gallant, J. L. (2019). Human scene-selective areas represent 3d configurations of surfaces. *Neuron, 101*, 178–192.e7.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience, 4*, 533–539.

Li, B., Peterson, M. R., & Freeman, R. D. (2003). Oblique effect: A neural basis in the visual cortex. *Journal of Neurophysiology, 90*(1), 204–217.

Li, D. S. P., & Bonner, M. F. (2022). Emergent selectivity for scenes, object properties, and contour statistics in feedforward models of scene-preferring cortex. *bioRxiv*, https://doi.org/10.1101/2021.09.24.461733.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., . . . Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science, 8693 LNCS,* 740–755.

Livingstone, M. S., Arcaro, M. J., & Schade, P. F. (2019). Cortex is cortex: Ubiquitous principles drive face domain development. *Trends in Cognitive Sciences, 23*, 3–4.

Long, B., Konkle, T., Cohen, M. A., & Alvarez, G. A. (2016). Mid-level perceptual features distinguish objects of different real-world sizes. *Journal of Experimental Psychology: General, 145*, 95–109.

Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision, 17*, 20, https://doi.org/10.1167/17.6.20.

Long, B., Yu, C. P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America, 115*, E9015–E9024.

Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences, 15*, 97–103.

Maloney, R. T., & Clifford, C. W. (2015). Orientation anisotropies in human primary visual cortex depend on contrast. *NeuroImage, 119*, 129–145.

Mannion, D. J., McDonald, J. S., & Clifford, C. W. (2010). Orientation anisotropies in human visual cortex. *Journal of Neurophysiology, 103*, 3465–3471.

Mansfield, R. J. (1974). Neural basis of orientation perception in primate vision. *Science, 186*, 1133–1135.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences, 7*(7), 293–299.

Nasr, S., Echavarria, C. E., & Tootell, R. B. (2014). Thinking outside the box: Rectilinear shapes selectively activate scene-selective cortex. *Journal of Neuroscience, 34*, 6721–6735.

Nasr, S., & Tootell, R. B. (2012). A cardinal orientation bias in scene-selective visual cortex. *Journal of Neuroscience, 32*, 14921–14926.

O'Herron, P., Chhatbar, P. Y., Levy, M., Shen, Z., Schramm, A. E., Lu, Z., . . . Kara, P. (2016). Neural correlates of single-vessel haemodynamic responses in vivo. *Nature, 534*, 378–382.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*, 145–175.

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nature Reviews Neuroscience, 9*(2), 123–135.

Op de Beeck, H. P., Pillet, I., & Ritchie, J. B. (2019). Factors determining where category-selective areas emerge in visual cortex. *Trends in Cognitive Sciences, 23*(9), 784–797.

Pennock, I. M. L., Racey, C., Allen, E., Wu, Y., Naselaris, T., Kay, K., . . . Bosten, J. (2023). Color-biased regions in the ventral visual pathway are food-selective. *Current Biology, 33*, 134–146.

Prince, J. S., Charest, I., Kurzawski, J. W., Pyles, J. A., Tarr, M. J., & Kay, K. N. (2022). Improving the accuracy of single-trial fMRI response estimates using GLMsingle. *eLife, 11*, e77599.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biology, 9*, 1000608.

Rosenthal, I., Ratnasingam, S., Haile, T., Eastman, S., Fuller-Deets, J., & Conway, B. R. (2018). Color statistics of objects, and color tuning of object cortex in macaque monkey. *Journal of Vision, 18*, 1, https://doi.org/10.1167/18.11.1.

Sasaki, Y., Rajimehr, R., Kim, B. W., Ekstrom, L. B., Vanduffel, W., & Tootell, R. B. (2006). The

radial bias: A different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron, 51*, 661–670.

Sayres, R., & Grill-Spector, K. (2008). Relating retinotopic and object-selective responses in human lateral occipital cortex. *Journal of Neurophysiology, 100*, 249–267.

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., . . . DiCarlo, J. J. (2020). Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv*. 407007, https://doi.org/10.1101/407007.

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 4447–4452.

Sereno, M., Dale, A., Reppas, J., Kwong, K., Belliveau, J., Brady, T., . . . Tootell, R. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science, 268*, 889–893.

Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain, 115*, 15–36.

Shen, G., Tao, X., Zhang, B., Smith, E. L., Chino, Y. M., & Chino, Y. M. (2014). Oblique effect in visual area 2 of macaque monkeys. *Journal of Vision, 14*, 3, https://doi.org/10.1167/14.2.3.

Silson, E. H., Groen, I. I., Kravitz, D. J., & Baker, C. I. (2016). Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. *Journal of Vision, 16*, 14, https://doi.org/10.1167/16.6.14.

Silson, E. H., Chan, A. W.-Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *Journal of Neuroscience, 35*, 11921–11935.

Silva, M. F., Brascamp, J. W., Ferreira, S., Castelo-Branco, M., Dumoulin, S. O., & Harvey, B. M. (2018). Radial asymmetries in population receptive field size and cortical magnification factor in early visual cortex. *NeuroImage, 167*, 41–52.

Srihasam, K., Vincent, J. L., & Livingstone, M. S. (2014). Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nature Neuroscience, 17*, 1776–1783.

Stigliani, A., Weiner, K. S., & Grill-Spector, K. (2015). Temporal processing capacity in high-level visual cortex is domain specific. *Journal of Neuroscience, 35*, 12412–12424.

St-Yves, G., & Naselaris, T. (2018). The feature-weighted receptive field: An interpretable encoding model for complex feature spaces. *NeuroImage, 180*, 188–202.

Swisher, J. D., Halko, M. A., Merabet, L. B., McMains, S. A., & Somers, D. C. (2007). Visual topography of human intraparietal sulcus. *Journal of Neuroscience, 27*, 5326–5337.

Swisher, J. D., Gatenby, J. C., Gore, J. C., Wolfe, B. A., Moon, C. H., Kim, S. G., . . . Tong, F. (2010). Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *Journal of Neuroscience, 30*, 325–330.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network Computation in Neural Systems, 14*, 391–412.

van der Schaaf, A., & van Hateren, J. (1996). Modelling the power spectra of natural images: Statistics and information. *Vision Research, 36*(17), 2759–2770.

Vo, V. A., Sprague, T. C., & Serences, J. T. (2017). Spatial tuning shifts increase the discriminability and fidelity of population codes in visual cortex. *The Journal of Neuroscience, 37*, 3386–3401.

Vogels, R., & Orban, G. A. (1994). Activity of inferior temporal neurons during orientation discrimination with successively presented gratings. *Journal of Neurophysiology, 71*, 1428–1451.

Walther, D. B., & Shen, D. (2014). Non-accidental properties underlie human categorization of complex natural scenes. *Psychological Science, 25*, 851.

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex, 25*, 3911–3931.

Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., Ramdas, A., & Mitchell, T. (2014). Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. *PLoS ONE, 9*, e112575.

Wichmann, F. A., Drewes, J., Rosas, P., & Gegenfurtner, K. R. (2010). Animal detection in natural scenes: Critical features revisited. *Journal of Vision, 10*, 6, https://doi.org/10.1167/10.4.6.

Yue, X., Pourladian, I. S., Tootell, R. B., & Ungerleider, L. G. (2014). Curvature-processing network in macaque visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 111*, E3467.

Yue, X., Robert, S., & Ungerleider, L. G. (2020). Curvature processing in human visual cortical areas. *NeuroImage, 222*, 117295.

Zeki, S. M. (1973). Colour coding in rhesus monkey prestriate cortex. *Brain Research, 53*, 422–427.