Title: Visual featural topography in the human ventral visual pathway

Authors: Shijia Fan[1,2], Xiaosha Wang[1,2], Xiaoying Wang[1,2], Tao Wei[1,2], Yanchao Bi[1,2*]

Author Affiliations: [1] State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China. [2] Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, 100875, China.

Corresponding author: Yanchao Bi, State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China. Email: ybi@bnu.edu.cn.

Key words: ventral occipital temporal cortex; computational vision model; domain organization; response mapping

Abstract

Visual object recognition in humans and nonhuman primates is achieved by the ventral visual pathway (ventral occipital-temporal cortex: VOTC). A classical debate is whether the seemingly domain-based structure in higher-order VOTC simply reflects distributional patterns of certain visual features. Combining computational vision models, fMRI experiments using a parametric-modulation approach, and natural image statistics of common objects, we depicted the neural distribution of a comprehensive set of visual features in VOTC, identifying voxel sensitivities to specific feature sets across geometry/shape, Fourier power, and color. We found that VOTC's sensitivity pattern to these visual features fully predicts its domain-based organization (adjusted R^2 around .95), and is partly independent of object domain information. The visual feature sensitivity pattern, in turn, is significantly explained by relationships to types of response/action computation (Navigation, Fight-or-Flight, and Manipulation), more so than the "object domain" structure, as revealed by behavioral ratings and natural image statistics. These results offer the first comprehensive visual featural map in VOTC and a plausible theoretical explanation as a mapping onto different types of downstream response systems.

Significance Statement

Human higher-order ventral visual pathway (VOTC) has a well-documented object domain organization (animate vs. inanimate), but the underlying mechanisms remain debated. Combining computational vision, functional neuroimaging and behavioral rating experiments, we depicted a first comprehensive visual featural map in VOTC, which almost perfectly explained classical object-domain organization, and was partly independent of object domain information. We further showed that one factor that may explain why the visual feature sensitivities distribute in VOTC this way is how they associates with computations for salient types of responses/actions (Navigation, Fight-or-Flight, and Manipulation). That is, the mappings onto downstream response systems are a key driving force that shape the neural sensitivity in the visual cortex.

Introduction

Ventral occipital-temporal cortex (VOTC), which underlies visual object recognition in humans and nonhuman primates, has a hierarchical architecture, from a retinotopic organization of simple features in the early visual cortex to a domain-based (e.g., faces, scenes/places/large objects, animate vs. inanimate) organization in higher-order visual cortex (1–4). An ongoing debate concerns whether this latter organization reflects a real, possibly innate, domain-based structure or if it instead reflects distributional patterns of low- or mid-level visual features that systematically correlate with domain (5, 6, see ref. 2 for a review). This debate has led to many studies about the properties of VOTC (7–9, see reviews in ref. 10, ref. 11), which are important in understanding the transition from simple feature processors to higher-order recognition mechanisms.

The distributions of two specific types of visual features in higher-order VOTC have been very recently demonstrated in humans and nonhuman primates. One type is the mid-level shape contrast between rectilinear and curved features. High rectilinearity, especially right angles, is more prevalent in images of scenes and places and activates scene-preferring regions including the parahippocampal place area (PPA) and transverse occipital sulcus more strongly than curved lines in humans (12). Low rectilinearity, or high curvature, tends to be associated with animate items (13, 14), which tend to activate regions close to the face patches in the macaque brain (15). The other visual feature of recent interest in the VOTC is color. Different colors have been shown to associate with objects versus their backgrounds, and with animate versus inanimate objects (16). Three VOTC patches were identified in the macaque monkeys to be sensitive to color and the more anterior medial patch showed both a yellow/red preference and face/body preference (17). These studies focus on individual visual features, and the effects may be driven by other features that correlate with them. More importantly,

if there is a systematic pattern of various visual feature sensitivity across VOTC, what are the organizational principles of this organization? While there are no explicit hypotheses about the organization of visual features in VOTC, predictions might be inferred from broader hypotheses about the driving forces and computational purpose of this cortical territory. One hypothesis states that domain categorization serves visual object recognition (2). Another, non-mutually exclusive, hypothesis states that the organization arises due to mapping onto various action/response systems such as Navigation, Fight-or-Flight and Manipulation (11, 18, 19). Both hypotheses predict that the visual feature distribution is not random, but rather associates with major domain and/or response systems, respectively.

We aimed to depict a comprehensive topographical map of visual features across VOTC, taking into consideration their correlational nature in the context of common objects. We combined computational vision modelling and parametric modulation analysis on fMRI responses. The parametric modulation approach exploits the natural variation in salience of various visual features across object images (obtained from computation vision modelling) and identifies brain regions responsive to each feature or combination of features by computing the degree of association between brain response and image feature weights. The obtained visual feature maps were compared to the domain distributions, and an additional experiment using isolated visual features was conducted to examine the relationship between object domain effects and feature effects. Two hypotheses about the driving forces of the VOTC organization – object domain and response mapping – were tested to explain the visual feature topography of VOTC.

Results

Twenty visual features covering a broad range of shape, spatial frequency, orientation, and color information were tested, and their weights were extracted for each of 95 object images using computational vision models (see Materials and Methods and **Fig. S1** for details) (12, 15, 20–24). fMRI responses for these images were also obtained from 26 participants, and parametric modulation models were used to compute the effects of visual features across VOTC voxels, taking into consideration their inter-correlations (Object Experiment; see **Fig. 1**). Then the relation between domains and features in VOTC were investigated (Object Experiment and Isolated-Features Experiment). Finally, two theoretical models of VOTC computation (by object domain and by response mapping) were examined in terms of their explanatory power for the VOTC visual feature patterns.

Computation of visual feature weights in object images

A set of 95 real object images (28 large nonmanipulable objects, 32 animals, and 35 small manipulable objects) were analyzed using computational vision models to obtain their properties for 20 visual features (see Materials and Methods): in geometry/shape space these features were right angle, curvature, number of pixels and elongation; in Fourier power space high/low spatial frequencies and four orientations (0, 45, 90, 135°); in color space eight hues, luminance, and chroma. The descriptive statistics, including distribution plots for each feature across the whole image set, as well as the mean and standard deviation (SD) by domains, are shown in **Fig. S2**. The correlations (Pearson) among features are shown in **Fig. S3**. As often observed, we found significant differences (FDR corrected $q < .05$) among the three object domains across some visual features: right angle ($F(2,92) = 6.77$, $p = .002$), number of pixels ($F(2,92) = 16.37$, $P = 8.27\times 10^{-7}$) and elongation ($F(2,92) = 15.47$, $p = 1.61 \times 10^{-6}$) in geometry/shape space; low spatial frequency ($F(2,92) = 6.59$, $p = .002$), 0°

orientation (F(2,92) = 6.21, p = .003), 90° orientation (F(2,92) = 5.08, p = .008) and 135° orientation (F(2,92) = 8.06, p = .001) in Fourier power space; orange (F(2,92) = 5.11, p = .008) and yellow (F(2,92) = 5.43, p = .006) in color space. The post-hoc comparisons across domain pairs are shown in **Table S1**. Pairs of highly-correlated visual features (Pearson r > 0.85) were collapsed into one by taking the means (cyan/indigo, r = 0.92, red/purple, r = 0.86). To reduce chances of multicollinearity, low spatial frequency was further excluded from the full parametric modulation model analysis because it had a variance inflation factor (VIF) > 10 (VIF = 48.25; other features' VIFs are within the range of 1.26 – 5.41)(25). Thus, 17 features were retained for the subsequent parametric modulation analysis, with pairwise correlations within the range of -.56 to .64.

Visual feature topography in VOTC: Results of the Object Experiment

For all fMRI analyses below, we adopted a threshold of cluster-level FWE corrected p < .05 within the VOTC mask(26), with voxel-wise p < .001 unless explicitly stated otherwise.

The results of the full model analysis, where the 17 visual feature weights were entered into the parametric modulation model for BOLD activity estimates, are shown in **Fig. 2**. In higher-order VOTC, for geometry/shape-space features, right angle modulated responses in bilateral medial fusiform gyrus (medFG) and left lateral occipital temporal cortex (LOTC); number of pixels modulated responses in left medFG. For Fourier-power-space features, high spatial frequency modulated responses in bilateral medFG; 0° orientation modulated responses in right medFG and bilateral LOTC; oblique orientations (45°, 135°) modulated responses in right lateral fusiform gyrus (latFG) and orientation 135° additionally modulated responses in left latFG. For color-space features, red/purple and green modulated broad regions in bilateral FG; and luminance modulated responses in bilateral

latFG. Red/purple additionally modulated responses in left LOTC.

Independent models, in which each feature was entered into the parametric modulation model separately without considering the correlations among features, were also performed and are shown in **Fig. S4**. Here more commonalities across features can be observed, with most features showing regions largely consistent with those obtained in the full model above with effects covering broader regions in higher-order VOTC. Five features showed differences between the two analyses: The effects of elongation (in left LOTC), 90^{o} orientation (in bilateral medFG), and blue (in bilateral medFG) were significant in the independent model but not in the full model; the effects of 0^{o} orientation (in bilateral LOTC) and luminance (in bilateral latFG) were significant in the full model but not in the independent model. These differences are likely due to their correlations with other features (see **Fig. S3**).

Featural effects in accounting for the domain effects in VOTC

To what extent can the classical object domain distribution in VOTC be explained by visual feature effects? We carried out the following two analyses, one across all VOTC voxels, one in classical domain-preferring clusters.

How much of VOTC voxels' domain-selectivity is explained by the visual feature effects? We constructed a multiple linear regression model using the visual feature sensitivity patterns to predict selectivity strength for a given domain across all VOTC voxels. That is, the 17-feature sensitivity maps in VOTC from the full parametric modulation model were taken as the independent variables. The dependent variable was the VOTC domain-selectivity strength map, which was computed separately for each object domain by contrasting the brain responses (beta values) to images from one domain to those of the other two. (Note that this analysis does not involve double dipping as the visual-

feature maps were generated completely independently of the domain information of each image.) The results (**Fig. 3A)** show almost perfect explanatory powers of the linear regression models: for large nonmanipulable object selectivity, adjusted-R^2 = 0.946; for animal selectivity: adjusted-R^2 = 0.957; for small manipulable object selectivity: adjusted-R^2 = 0.973. That is, how much a voxel prefers each of the three object domains was perfectly predicted by its sensitivity pattern to the 17 visual features.

Do domain-preferring regions overlap with the feature maps? We localized classical domain-preferring regions in VOTC and checked for visual features showing effects in these clusters. In the VOTC mask, we contrasted each object domain to the other two domains, which yielded clusters highly correspondent to those observed in the literature (**Fig. 3B**)(18, 27): large nonmanipulable objects correspond to clusters in bilateral PPA (cluster-level FWE corrected $p < .05$ within the VOTC mask, voxel-wise $p < .0001$); animals in bilateral latFG (cluster-level FWE corrected $p < .05$ within the VOTC mask, voxel-wise $p < .0001$); and small manipulable objects in left LOTC (cluster-level FWE corrected $p < .05$ within the VOTC mask, voxel-wise $p < .01$). The overlapping voxels between the domain preferring regions and the visual feature modulation clusters in the full model analysis were counted, and those greater than 10 voxels were considered. Results (**Fig. 3B)** showed that the PPA-large nonmanipulable object region overlapped with right angle (424 voxels), high spatial frequency (257 voxels), red/purple (103 voxels), green (92 voxels), 45° orientation (33 voxels), number of pixels (31 voxels), luminance (18 voxels) and 0° orientation (10 voxels); the latFG-animal region overlapped with 135° orientation (46 voxels), red/purple (45 voxels), green (40 voxels), luminance (23 voxels) and 45° orientation (18 voxels); and finally the LOTC-small manipulable object region overlapped with red/purple (21 voxels) and right angle (19 voxels).

Featural effects independent of object domain information

The above findings showed a strong association between effects of visual features and object domains in VOTC. In this section we test whether there are feature effects independent of the object domains.

Visual feature effects regressing out the domain model. We first re-tested the visual feature effects in the feature-domain overlap ROIs by running the full parametric modulation model analysis using feature weights, with domain structure regressed out (see Materials and Methods). The ROIs were defined as the overlapping voxels between the domain preferring regions and the significant clusters of each feature modulation. There were 15 ROIs in total (**Fig. 3B**) and all effects remained the same to those before regressing out the domain model (**Fig. S5,** FDR corrected $q < .05$), indicating the featural effects exist independently from object domain.

Visual feature effects in the non-preferring domains. We further examined the visual feature effects using object images belonging to the non-preferring domains to fully exclude any potential domain effects. In each feature-domain overlap ROI, we excluded stimuli from the corresponding object domain and re-ran the full model analysis. (Note that in this analysis one third of the data was not used, lowering the statistical power.) For example, we tested whether the right-angle effect in PPA, the large nonmanipulable object-preferring region, remained when participants were looking at images of animals and small manipulable objects. The results (**Fig. 4A,** FDR corrected $q < .05$) showed that effects in 10 of the 15 ROIs remained stable, again confirming the relative independence of visual featural effects from object domain.

Visual feature effects without object contexts (Isolated-Features Experiment). To test whether the

observed visual feature effects in higher-order VOTC remain without object contexts, we conducted a second fMRI experiment in which the stimuli were arrays of lines and random dot patterns varying in hue (see **Fig. 4B** for sample stimulus shapes). We also included a functional localizer run with images of scenes, animals and small manipulable objects. As shown in **Fig. 4B**, semicircles, compared to horizontal/vertical lines, or to right angles, activated bilateral or right latFG close to animal-preferring clusters. Right angles compared to semicircles activated bilateral medFG corresponding to PPA only at a lenient threshold (voxel-wise $p < .05$, uncorrected). No effects of horizontal or vertical lines greater than semicircles were found in the FG. For dot patterns rendered with different hues (red, orange, green, cyan, and blue) compared to grayscale dot patterns, weak activations (voxel-wise $p < .05$, uncorrected) were observed posterior to the domain-preferring clusters or sandwiched between latFG and PPA. In summary, when presented as isolated single-feature stimuli without object contexts, the effects of right angles in PPA and curvature in right latFG tended to remain; the effects of right angles in LOTC, and of hues and orientations broadly, disappeared.

Across experiments, the effect of right angle in PPA was reliable across all domain-independent analyses – it is present when the domain structure was regressed out, when participants were looking at non-preferring domains (animal and small manipulable objects), and when presented in isolation from object contexts. The effects of right angle in LOTC and of different hues and orientations (at least in part) in VOTC remained whether or not domain information is taken out (by regressing out domain structure or excluding preferring domain items), but disappeared when participants were looking at isolated featural stimuli, indicating that their effects may be linked to other visual features and/or object contexts.

Factors driving the feature distribution patterns in VOTC voxels

We have described the distributional topography of a comprehensive set of visual features in VOTC, identifying voxels by their sensitivities to specific feature sets. We also demonstrated that each voxel's sensitivity pattern to feature sets almost perfectly accounts for its domain-selectivity and that certain features like right angle continue to produce effects in isolation. To understand why visual features distribute in this way, we examined whether the visual feature topography can be explained by two hypotheses about the computational purpose of VOTC: object domain categorization (potentially for more specific downstream object recognition), or mapping with downstream response computations, by examining the correlation between "prototypical" visual feature vectors based on these two hypotheses (**Fig. 5A**) and the observed neural visual feature vector in VOTC.

Relationship between observed VOTC voxel feature vectors and domain-feature vectors. We first generated "prototypical" visual feature vectors associated with the three object domains (i.e., domain-feature vectors) based on natural image statistics to approximate the feature-domain association profile in the natural world. For this purpose, we used a broader image set containing 767 images, which included the 95 images from the current Object Experiment and 672 images (isolated objects with clear domain membership on white background) selected from three previous studies (28–30). Three binary domain vectors were then constructed (e.g., animal vector: animals = 1, others = 0). For each image, weights on visual features were obtained using the identical computational vision models mentioned above. Logistic regression was conducted between each binary domain vector and each of visual feature weights, and the resulting beta values were taken as a prototypical domain-feature vector for each domain. This vector reflects the visual feature patterns that best distinguish each domain from the others in natural object images. The results (FDR corrected $q < .05$)

are shown in **Fig. 5B left**; see **Table S2** for β and *p*-values).

We then tested whether the VOTC voxels' visual feature sensitivity profile reflects these domain-feature vectors. For each VOTC voxel, we correlated its neural-feature-vector (obtained by the independent parametric modulation model; **Fig. S4**) and each of the three prototypical domain-feature-vectors based on natural image statistics, resulting in a correlation map for each domain (**Fig. 5B right**). Note that the independent parametric modulation model results were used (with 18 visual features), because the beta weights were more transparently interpretable and the results were also largely consistent with those of the full model (17 features, with "low spatial frequency" excluded due to high VIF). The large nonmanipulable object domain-feature-vector was significantly positively correlated with the neural-feature vector of VOTC voxels in 3 clusters located in bilateral medFG and left middle occipital cortex. The animal domain-feature-vector was significantly correlated with the neural-feature vector of VOTC voxels in one cluster located in right lateral occipital cortex. The small manipulable object domain-feature-vector was significantly correlated with the neural- feature vector of VOTC voxels in 2 clusters located in LOTC and left lingual gyrus. These results suggest that VOTC voxels' feature-sensitivity patterns are associated with the natural image statistics of three major object domains.

Relationship between observed VOTC voxel feature vectors and response-feature vectors. Here we obtained prototypical visual feature vectors associated with three hypothesized response systems (i.e., response-feature vector): Navigation, Fight-or-Flight and Manipulation (11, 18, 19), by first asking an independent group of participants to rate each object image on its relevance to each response (see Materials and Methods for details). Linear regression was conducted between the rated value for each response system and each of the 18 visual feature weights; the resulting beta

values were used as the prototypical response-feature vector for each response system in natural object images. The results (FDR corrected $q < .05$) are show in **Fig. 5C left** (see **Table S2** for β and *p*-values).

Similar to above, for each VOTC voxel we correlated its neural-feature vector (**Fig. S4**) with each of the three prototypical response-feature-vectors, resulting in 3 VOTC correlation maps (**Fig. 5C right**). The navigation-response-feature vector was significantly correlated with the neural-feature vector in clusters located in bilateral medFG, bilateral middle occipital gyrus and right lingual gyrus. The fight/flight-response-feature vector showed no significant correlations at the standard threshold. When we lowered the threshold (voxel-wise $p < .01$, uncorrected), it correlated with the neural-feature-vector of VOTC voxels in bilateral lateral occipital cortex, right latFG and bilateral occipital pole. The manipulation-response-feature vector was significantly correlated with the neural-feature-vector of VOTC voxels in clusters located in left LOTC and left lingual gyrus.

Comparison between domain-driven and response-driven hypotheses. Because the prototypical domain-feature vectors and prototypical response-feature vectors were highly correlated (Pearson *r*s: large nonmanipulable objects with navigation response = .84; animals with fight-or-flight response = .80; small manipulable objects with manipulation response = .87) and showed similar relationship patterns in VOTC (**Fig. 5B** & **5C**), we directly compared the explanatory power of these two types of feature vectors. To do this, we first generated a domain-driven maximum R map by selecting the highest R value for each voxel out of the three domain-driven R maps in **Fig. 5B,** and generated the response-driven maximum R map in the same way using the three maps in **Fig. 5C**. Then the two max R maps were Fisher-z transformed and compared by paired t-test. Results showed that the "response-driven" map was significantly higher than the "domain-driven" map (global mean Rs ± SD: 0.57 ± 0.27

vs. 0.54 ± 0.28, t (3914) = 12.21, p = 1.07 X 10^{-33}).

Discussion

Combining computational vision models, a parametric modulation analysis of fMRI data, and natural image statistics, we depicted the distributional topography of a comprehensive set of visual features in VOTC, identifying voxels' sensitivities to specific feature sets. We demonstrated that the sensitivity pattern across geometry/shape, Fourier power, and color (visual feature vectors) almost perfectly predicts each voxel's domain-selectivity strength for large nonmanipulable objects, animals, and small manipulable objects. Some features (e.g., right angle in PPA) also produce effects when isolated from other features and/or object contexts. We also demonstrated that two organizational principles – for domain categorization and for efficient response mapping – may drive visual features' distribution, with the response mapping principle having slightly but significantly more explanatory power.

By contrast to recent studies that focused on one or two specific visual features, our approach tested a much more comprehensive set of visual features and correlations among them. Our finding that natural image statistics are associated with object domains is consistent with a set of recent studies, including those showing that large nonmanipulable objects tend to have more right angles (12) and animals tend to have more yellow colors (16). Our results with combinations of multiple features highlight the importance of not only positive associations but also negative ones in informing about domains, e.g., animals *do not* have right angles whereas large nonmanipulable objects are *not* likely to be elongated.

Remarkably, we found an almost-perfect association between the domain selectivity patterns

and visual feature sensitivity patterns in VOTC. Previous studies have shown associations between certain visual feature and domain preferences: A preference for rectilinearity, high spatial frequency, and cardinal orientation features was observed in regions preferring scenes/large objects (12, 22, 31, 32) and a preference for high curvature, low spatial frequency, and red/yellow hues was observed in regions preferring faces (15, 17, 22, 33). However, it has been questioned whether VOTC domain preferences can be fully explained by the effects of these visual features, given that the selectivity strengths for features are lower than for the corresponding domains (34), the anatomical overlap between feature effects and domain effects is far from perfect (15), and the domain preferences are still present when visual shape are controlled (7). Here, by incorporating the combinational effects of multiple visual features together, we showed remarkably high explanatory power of visual features to domain-preference: Voxels' visual-feature-preference vectors accounted for about 95% of VOTC's variance in selectivity for animals, large nonmanipulable objects, and small manipulable objects.

Although strongly associated with the domain-preference effects in VOTC, not all feature effects were dependent on the presence of object domain information. The effect of right angle in bilateral medFG (aligning with the PPA) was present when the features are shown in isolation without object contexts and/or other features, and even during presentation of objects from non-preferring domains (i.e., when objects are small manipulable objects and animals). The effects of other features like hue and orientation were only observed when presented within objects and disappeared when shown in isolation, showing that they are processed in combination with other visual features and/or object contexts in VOTC (35).

Showing the robustness of visual feature combinations in explaining domain preferences does not imply that domains are not a productive or accurate way to conceptualize the organization of

VOTC. The hard question remains: Why does VOTC have this specific type of visual feature topography? Why do specific VOTC regions prefer specific combinations of visual features? We considered two (related) hypotheses of VOTC organization, one classically-embraced stating that object domain categorization is in the serve of more-specific object recognition (2), and one more recently articulated stating that the organization arises due to how visual properties map onto different types of responses systems (11, 18, 19). We tested whether either or both principles explain the VOTC feature sensitivity patterns by correlating each VOTC voxel's visual-feature vector with prototypical visual-feature vectors corresponding to object domains and response systems. We found that indeed both principles significantly explain the VOTC's feature sensitivity patterns: large nonmanipulable object domain-vectors and navigation response-vectors were associated with bilateral medFG; animal domain-vectors and fight-or-flight response-vectors with right lateral occipital cortex; and small manipulable object domain-vectors and manipulation response-vectors with left LOTC. These two principles (domain-categorization and response-mapping) are intrinsically correlated, as the evolutionary saliency of domains has been proposed to reflect the fact that different object domains entail different brain connectivity with downstream action computations (11, 18, 19). In this sense, the salient "domain" structures are vaguely defined and may be understood as proxy terms for different vision-response mappings. Indeed, we showed that the response-relevance-feature vectors have a slight but significant advantage in explaining the VOTC voxels' visual-feature-sensitivity patterns, providing support for response-mapping as the critical driving force for VOTC organization.

Taking all the results together, the following picture of VOTC properties emerges: 1) VOTC's seemingly domain-related organization can be fully explained by voxel sensitivity patterns to a comprehensive set of visual features; 2) The specific VOTC visual feature sensitivity pattern associates

with the object domain structure but is better explained by how features map onto different types of responses; 3) Although the feature sensitivity structure may be intrinsically driven by response mapping (through evolution and/or individual level experiences), the resulting VOTC may behave as general feature-, rather than domain- processors, as at least some features were effective even in the absence of the corresponding object domain information.

There are two caveats to consider. One is that the visual features we tested are based on knowledge and algorithms from computational vision practice. There is always a possibility that other relevant types of visual features were missed, and that the algorithm choice was not optimal. For instance, the current curvature computation considers 5 arbitrarily-selected concavity features, and its effects on VOTC based on this computation were not significant yet were visible when using a direct contrast (top 25% amount of curvature – top 25% amount of right angle, **Fig. S6**), which is more in line with studies using subjective curvature ratings, which may reflect a composite index of various types of curvatures (36). However, our results that the feature combination model explains the domain-preference strength in VOTC voxels almost perfectly indicate the power of the included features. Second, we only examined three major object domains, not testing other classical domains for VOTC: scenes and faces. The current framework makes the same predictions about preferences for these two types of images, which remain to be empirically tested.

In summary, we report the first comprehensive visual feature map in human VOTC. We showed that visual-feature-sensitivity patterns fully explain domain-selectivity patterns in VOTC. Finally, we conclude that the mappings onto downstream response systems are a key driving force for the observed VOTC feature distribution patterns.

Materials and Methods

Participants

Twenty-nine participants (age range 19-25 years; 20 females) participated in the Object Experiment. An independent group of 19 participants (age range 19-29 years; 11 females) participated in the Isolated-Features Experiment. An independent group of 20 participants (age range 18-29 years, 14 females) participated in the rating study. All participants had no history of neurological or psychiatric impairment, had normal or corrected-to-normal vision, were native Chinese speakers, and provided written informed consent. The Object Experiment was approved by the Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University. The Isolated-Features Experiment was approved by the Institutional Review Board of Department of Psychology, Peking University.

fMRI stimuli

Stimuli in the Object Experiment consisted of 95 colorful real-world objects centered on a white background belonging to three common domains: large nonmanipulable objects (28 images), animals (32 images), and small manipulable objects (35 images). Images of large nonmanipulable objects included buildings, furniture, appliances, communal facilities and large transport. Images of animals included mammals, birds, reptiles and insects. Images of small manipulable objects included common household tools, kitchen utensils, stationery and accessories. These images were obtained from the Internet and resized to 400 x 400 pixels (10.5° x 10.5° of visual angle).

Stimuli in the Isolated-Features Experiment consisted of 112 images from 14 conditions (see **Fig. 4B** for sample stimulus shape). Nine conditions were individual visual features: 4 different line-based

features as shape arrays (8 arrays of right angles; 8 arrays of semicircles; 8 arrays of horizontal lines; 8 arrays of vertical lines), and the remaining 5 were hue-based features (40 random dot patterns in each of: red, orange, green, cyan, and blue). Another condition of grayscale random dot patterns was included as controls for hue-based features. The remaining conditions (non-cardinal lines, phase-scrambled images of non-cardinal lines, and two types of combination of multiple visual features) were designed for other research purposes. Only the nine conditions of individual features and the grayscale random dot pattern condition were analyzed here. Each shape array contained 40 elements, and all the shape arrays were located randomly within a virtual circular limit (radius = 200 pixels) (constructed largely following previous studies (12)). Orientations of shape arrays except for horizontal and vertical lines were not cardinal and varied semi-randomly to minimize orientation biases. Images of dot patterns were randomly distributed small dots (< 0.1°) rendered with gray or different color hues. For the colored dot patterns, we sampled 8 equally spaced points in (CIE) L*C*H space (Commission Internationale de l'Eclairage), that started from h = 0, at L* = 69, c* = 36, and for the following 5 hues: red (h = 0°), orange (h = 45°), green (h = 135°), cyan (h = 180°), blue (h = 270°). All stimuli had equal number of pixels and were presented against a Gaussian noise background (using the Matlab function imnoise). The SHINE toolbox (37) was used to match spatial frequency and luminance across all stimuli and to match the Fourier spectrums across dot conditions.

The Isolated-features Experiment included further domain-functional localizer runs that included 90 black-and-white images taken from a previous study (38), including scenes (30 images), animals (30 images) and small manipulable objects (30 images). All images subtended 7.6° x 7.2° of visual angle (400 x 400 pixels).

Computation of visual feature weights in object images

The weights of 20 visual features covering a broad range of shape, spatial frequency, orientation, and color properties were extracted using computational vision models for each of 95 object images (see **Fig. S1** for schematic modelling steps).

Geometry/shape space We examined four geometry/shape features: number of pixels, right angle, curvature, and elongation. For number of pixels, a binary object mask (defined as pixels with grayscale values lower than 240) was created and each pixel in the mask was counted. Overall right angle and curvature information was measured largely following previous approaches with some modification (12, 15, 20). Specifically, for right angle, 64 right-angle Gabor filters (using an absolute function (12)) were constructed using 4 spatial scales (1/5, 1/9, 1/15, and 1/27 cycles per pixel) and 16 orientations (22.5°-360° in 22.5° steps). Images were converted to grayscale and edge maps were constructed using Canny edge detection at a threshold of 0.1 (39). Each edge map was convolved with 64 Gabor filters of different spatial scales and orientations. This produced 64 Gabor coefficient images, which were then normalized by dividing by the mean magnitude of each Gabor filter. For each spatial scale, the largest magnitude across the 16 coefficient images of different orientations was extracted for each pixel to obtain a peak Gabor coefficient image, which was then averaged across all pixels of each image and z-scored across the image set. The resulting Gabor coefficient values for each image were finally averaged across 4 spatial scales and z-scored to provide a single value for each image to represent the amount of right-angle information in that image. For curvature, the same procedure was used using the bank of 320 curved Gabor filters (using a square root function (40), composed of 4 spatial scales, 16 orientations, and 5 levels of curvature (π/256, π/128, π/64, π/32, π/16), to generate a single value for the amount of overall curvature information for each image. Elongation

was measured as the aspect ratio of the rectangle that encloses the object parallel to the object's longest axis.

Fourier power space Images were converted to grayscale and submitted to a 2D fast Fourier transform (built-in MATLAB function fft2). The high/low spatial frequency and 4 orientations (0, 45, 90 and 135°) were measured based on previous approaches (22, 23, 31) to parameterize energy variation in Fourier power space. The overall energy of high and low spatial frequency was calculated by averaging the energy of the high (>5 cycles/degree) and low (<1 cycles/degree) band for each image (22). For orientations, we selected four directions which centered on vertical (0°), left oblique (45°), horizontal (90°), and right oblique (135°) with a bandwidth of 20° (23). For each orientation range, the energy across spatial frequencies was averaged.

Color space Three main perceptual dimensions of color—luminance, chroma, and hue—were quantified using (CIE) L*C*H space following previous studies (16, 24). Pixel colors in each image were converted from RGB space into (CIE) L*C*H space using the MATLAB "colorspace" package (41). The white point for the transformation of the image colors was set to D65 (the standard illumination for noon daylight). The luminance and chroma of each image were calculated by averaging these values across pixels within the object. The hue space was divided into 8 bins with equal width, which started from the 338°-23° bin, in 45° steps, and roughly corresponded to red, orange, yellow, green, cyan, indigo, blue, and purple (24). The number of pixels in each bin was then counted as the hue-specific measures; pixels that did not belong to objects or were ambiguous (defined as luminance or chroma values less than 10) were excluded.

Distribution of visual features across and within domains Distribution plots of z-transformed visual feature weights across 95 images were plotted and the mean and standard deviation (SD) by

domains were calculated (**Fig. S2**). One-way ANOVA was used to test whether there were significant differences among domains for each feature and multiple comparisons were corrected using FDR. For the features showing domain differences, Tukey's HSD post hoc comparison was conducted. The analysis was conducted using SPSS Statistics Software version 26 (IBM). Pairwise Pearson correlations between visual features were also computed across the 95 stimulus images.

fMRI Experiment Design

In the Object Experiment, participants named the 95 object images overtly and as quickly and accurately as possible. There were 6 runs, each lasting for 528 s. Each image (0.5 s fixation followed by 0.8 s image) was presented once per run. Inter-trial intervals ranged from 2.7-14.7 s. The order of stimuli and length of ITI were optimized using optseq2 (http://surfer.nmr.mgh.harvard.edu/optseq/). The order of items was randomized across runs. Each run started with 10 s of blank screen.

In the Isolated-Features Experiment, participants viewed images of isolated shape arrays and dot patterns (described above). There were 7 runs in total, each lasting for 236 s. The stimuli were presented in 8-s blocks, separated by 8 s of fixation. Each block consisted of 8 images, each presented for 0.8 s and separated by 0.2 s of blank screen. Each condition appeared once in each run and the order of conditions was randomized. Each run started and ended with 10 s of fixation. Participants were instructed to pay attention to the stimuli and to press a button when a white dot (< 0.5°) appeared in the center of the screen, which occurred 0, 1 or 2 times per block with equal possibility. Six of the 19 participants also participated in a second scanning session of 9 runs on a separate day; we thus analyzed 16 runs for those 6 participants.

After the isolated-feature runs, 2 domain localizer runs were conducted, in which participants

viewed images from 3 domains (scenes, animals, and small manipulable objects). Images were presented in 24-s blocks, separated by 8 s of fixation. Each block consisted of 30 images, each presented for 0.3 s and separated by 0.5 s blank screen. Each condition was repeated 3 times per run. During the localizer, participants performed a 1-back repetition detection task, pressing a button using their right index finger whenever two consecutive images were the same. There were 0, 1, or 2 times repetitions per block (same frequency for each condition).

MRI acquisition and data preprocessing

The Object Experiment was conducted at the Beijing Normal University Neuroimaging Center using a 3T Siemens Trio Tim scanner. Functional data were collected using an EPI sequence (33 axial slices, TR = 2000 ms, TE = 30 ms, flip angle = 90°, matrix size = 64 x 64, voxel size = 3 x 3 x 3.5 mm with gap of 0.7 mm). T1-weighted anatomical images were acquired using a 3D MPRAGE sequence: 144 slices, TR = 2530 ms, TE = 3.39 ms, flip angle = 7°, matrix size = 256 x 256, voxel size = 1.33 x 1 x 1.33 mm.

The Isolated-Features Experiment was conducted at the Imaging Center for MRI Research, Peking University, using a Siemens Prisma 3-T scanner with a 64-channel phase-array head coil. Functional data were collected with a simultaneous multi-slice (SMS) sequence (62 slices, multi-band factor = 2; TR = 2000 ms; TE = 30 ms; flip angle = 90°; matrix size = 112 x 112; voxel size =2 x 2 x 2 mm with gap of 0.3 mm). T1-weighted anatomical images were acquired using a 3D MPRAGE sequence: 192 sagittal slices; 1 mm thickness; TR = 2530 ms; TE = 2.98 ms; inversion time = 1100 ms; flip angle = 7°; FOV = 256 × 224 mm; voxel size = 0.5 × 0.5 × 1 mm, interpolated; matrix size = 512 × 448.

Functional images were preprocessed and analyzed using Statistical Parametric Mapping (SPM12,

http://www.fil.ion.ucl.ac.uk/spm), Statistical Non-parametric Permutation Testing Mapping (SnPM13, http://warwick.ac.uk/snpm), and DPABI (42). The first 5 volumes in each run of the Object Experiment and Isolated-Features Experiment were discarded. Image preprocessing included slice time correction, head motion correction, and normalization to the Montreal Neurological Institute (MNI) space using unified segmentation (resampling voxel size = 3 x 3 x 3 mm in the Object Experiment; 2 x 2 x 2 mm in the Isolated-Features Experiment), and spatial smoothing with a Gaussian kernel of 6 mm full width at half maximum. Three participants in the Object Experiment were excluded from analyses due to excessive head motion (>3 mm maximum translation or 3° rotation).

Statistical analyses were carried out within a bilateral VOTC mask (containing 3915 voxels for 3-mm voxel size) constructed in our previous study (26), which was defined as brain regions activated by the contrast of all objects versus fixation in an object picture perception task in the ventral occipitotemporal cortex. Activation maps for parametric modulation and contrasts between conditions (see below for details) were first created in individual participants and then submitted to group-level random-effects analyses using SnPM13. No variance smoothing was used and 5,000 permutations were performed. A conventional cluster extent-based inference threshold (voxel level at $p < .001$; cluster-level FWE corrected $p < .05$ within VOTC mask) was adopted unless stated explicitly otherwise.

Visual feature topography in VOTC: Object Experiment

To identify brain regions associated with each feature, parametric modulation was employed to investigate the correlations between activity levels and feature weights across the 95 stimulus images in the Object Experiment. For the full model that considers the correlations among multiple features,

the variance inflation factor (VIF) for each feature was calculated using SPSS Statistics Software version 26 (IBM) and features with VIF above 10 was excluded from analysis to reduce multicollinearity (25). Then the preprocessed functional images of each participant were entered into a General Linear Model (GLM), which included the onsets of items as one regressor, all features' weights for each image in the parametric modulation module, and 6 head motion regressors for each run. A high-pass filter cutoff was set at 128 s. Contrast images for each feature versus baseline were then calculated and submitted for random-effects analyses. Because there is no *a priori* expectation that any brain region should become "less" active as the processing demands for a given feature increase, making the interpretation of negative correlations speculative, only positive modulations were reported. To obtain raw feature maps without considering correlations among features, we also conducted parametric modulation analyses for each feature by including one feature at a time in the GLM.

Feature effects in accounting for the domain effects in VOTC

How much of the VOTC voxels' domain selectivity is explained by visual feature effects? To test how much variance in domain selectivity in VOTC voxels could be explained by visual features, we built a multiple linear regression model across VOTC voxels for each domain, using the data from the Object Experiment. Using the conventional contrast approach, a GLM was built including three regressors, each per domain, as well as 6 head motion regressors for each run. A high-pass filter cutoff was set at 128 s. The beta values contrasting one domain with the other two (e.g., for the animal domain: animal – [large nonmanipulable & small manipulable objects]) were computed in each participant and then averaged across participants to construct the domain selectivity VOTC maps. In

the linear regression model for each domain (in SPSS), the domain selectivity VOTC map was treated as the dependent variable and the full-model visual feature maps were included as independent variables. The adjusted R-squared of each model was reported.

Do domain-preferring regions overlap with feature maps? To investigate the overlap between domain preferring regions and brain regions significantly modulated by each feature, we localized 3 domain-preferring regions by thresholding the group-level activation maps at cluster-level FWE-corrected $p < .05$ within the VOTC mask with voxel-wise $p < .0001$ for large nonmanipulable objects and animals, and voxel-wise $p < .01$ for small manipulable objects. The details of the identified regions were as follows: for large nonmanipulable objects > others, the bilateral parahippocampal place area (PPA), 464 voxels; for animal > others, the bilateral lateral fusiform gyrus (latFG), 51 voxels; and for small manipulable objects > others, the left lateral occipital temporal cortex (LOTC), 93 voxels. These domain-preferring regions were then overlaid on each feature map and overlapping clusters containing more than 10 voxels were reported.

Feature effects independent of object domain information

Visual feature effects regressing out the domain model. To test the independence of visual feature from domain effects in the ROIs showing feature and domain overlap (see above), we regressed out dummy-coded domain vectors (assigning items from the corresponding domain as 1 and other items as 0) from each feature vector and re-ran the full model parametric modulation analysis using feature residuals as the modulating parameters. This procedure statistically eliminates the shared variance between feature weights and domains. The resulting parameter estimates for each feature in each ROI were averaged across voxels and compared with zero across participants

using one-tailed one-sample *t*-tests, followed by FDR correction for multiple comparisons.

Visual feature effects in non-preferring domains. To further exclude any potential domain effects on visual feature effects, the full model parametric modulation analysis was re-run using items from non-preferring domains. At the individual participant level, GLMs were built including two regressors coding stimuli as "preferring domain" items and "non-preferring domain" items, and for the "non-preferring domain" condition, visual feature weights were entered for the full parametric modulation analysis. As above, the resulting parameter estimates for each feature in the ROIs showing feature and domain overlap were averaged across voxels and compared with zero across participants using one-tailed one-sample *t*-tests, followed by FDR correction for multiple comparisons.

Visual feature effects without object contexts (Isolated-Features Experiment) In the analysis of the Isolated-Features Experiment, all 14 stimulus conditions were included in the GLM as predictors, along with head motion regressors. The following contrasts were examined: dot patterns in different hues vs. gray-scale dots, right angles vs. semicircles, horizontal lines vs. semicircles, and vertical lines vs. semicircles. The group-level domain-preferring regions were identified using the two localizer runs. At the conventional threshold, the effects of animals and large nonmanipulable objects in this experiment spread to early visual cortex, so the localization of PPA and animal-latFG were constrained in an anatomical mask including bilateral fusiform gyrus and parahippocampal gyrus in the Harvard-Oxford atlas (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases).

Factors driving the feature distribution patterns in VOTC voxels

To test the organizational principles of the observed voxel feature distribution patterns, prototypical visual-feature vectors for domains/responses based on natural image statistics were

constructed and compared with feature distribution patterns in the VOTC. To gain an unbiased understanding of feature distribution among objects, we built a larger object image dataset containing 767 images from three previous image sets (28–30) and the 95 images from the current Object Experiment. We used these image sets because they had isolated objects presented on a white background (Twenty-six images from Brodeur et al., 2014 were excluded for lack of clear domain membership, e.g., pillow). One object image was the same in our current experiment and in Downing et al. 2006 and thus only one was included. There were 419 animals (e.g., mammals, marine creatures, birds, insects, fish, reptiles), 168 large nonmanipulable objects (e.g., buildings, furniture, appliances, communal facilities, large transportation), and 180 small manipulable objects (e.g., common household tools, kitchen utensils, stationery, accessories). All images were resized to 256 x 256 pixels with 72 DPI using Adobe Photoshop CS6. For each image, the feature weights were measured using computational vision models, as described above for the Object Experiment stimuli.

Prototypical visual-feature vectors for domains and response systems. For domain-driven prototypical visual-feature vectors, we constructed three binary domain vectors (e.g., animal vector: animals = 1, others = 0) and performed pairwise logistic regression analyses between each domain vector and each of visual features, resulting in 3 domain-feature beta vectors. For response-driven prototypical visual-feature vectors, we examined 3 theorized response systems: Navigation, Fight-or-Flight and Manipulation (11, 18, 19). The relevance of the 767 images (set described above) to each response system was rated by an independent group of participants (N = 20, age range 18-29 years, 14 females) on a 1 – 5 scale. For Navigation, the participants were asked to rate “to what extent the object depicted in image could offer spatial information to help you explore the environment.” For Fight-or-Flight, the participants were asked to rate “to what extent the object depicted in image would

make you to show a stress response, e.g., run away, attack, or freeze." For Manipulation, the participants were asked to rate "to what extent the object depicted in image can be grasped easily and used with one hand." The ratings were averaged across participants to get one relevance index for each image to each response. Then pairwise linear regression analyses were conducted between each response relevance type and each visual feature, resulting in 3 response-feature beta vectors.

Comparison between domain-driven and response-driven hypotheses. We examined whether and how the VOTC voxels' visual feature profiles could be explained by domain- or response-feature vectors based on natural image statistics. The VOTC voxels' visual feature profile was generated by extracting group-averaged parameter estimates of each feature in independent models, which resulted in a feature beta-vector for each voxel. The independent model results were used here because the beta values were more transparently interpretable and the results of the independent models are largely consistent with those of the full model. Then the neural visual feature profiles were correlated with each of three domain- or response-driven feature vectors using Pearson correlation, resulting in six R maps. The significance of the R maps above zero (one-tailed) was thresholded at cluster-level FWE-corrected $p < .05$ within the VOTC mask with voxel-wise $p < .001$. To compare the explanatory powers between domain-driven and response-driven hypotheses, the three R maps derived from domain-feature vectors were collapsed by extracting the highest R value in each voxel in VOTC and the three R maps derived from response-feature vectors were also collapsed use the same method. The resulting two maximum R maps were then Fisher-Z transformed and compared using paired t-test across voxels.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31671128 and 31925020 to Y.B., 31700999 to T.W., 31700943 to X.S.W., 31500882 to X.Y.W.), the National Program for Special Support of Top-notch Young Professionals (Y.B.), the Changjiang Scholar Professorship Award (T2016031 to Y.B.), the Fundamental Research Funds for the Central Universities (2017EYT35 to Y.B.), and the 111 Project (BP0719032). We thank Shiguang Shan for discussions about computation vision models, and Joshua B. Julian for generously sharing codes for the right angle and curvature computation.

Conflict of interest: The authors declare no competing financial interests.

References

1. D. J. Felleman, D. C. Van Essen, Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cereb Cortex* **1**, 1–47 (1991).

2. K. Grill-Spector, K. S. Weiner, The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci* **15**, 536–548 (2014).

3. N. Kanwisher, Functional specificity in the human brain: A window into the functional architecture of the mind. *PNAS* **107**, 11163–11170 (2010).

4. T. Konkle, A. Caramazza, Tripartite Organization of the Ventral Stream by Animacy and Object Size. *J. Neurosci.* **33**, 10235–10242 (2013).

5. I. Levy, U. Hasson, G. Avidan, T. Hendler, R. Malach, Center–periphery organization of human object areas. *Nat Neurosci* **4**, 533–539 (2001).

6. U. Hasson, I. Levy, M. Behrmann, T. Hendler, R. Malach, Eccentricity Bias as an Organizing Principle for Human High-Order Object Areas. *Neuron* **34**, 479–490 (2002).

7. D. Proklova, D. Kaiser, M. V. Peelen, Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction. *Journal of Cognitive Neuroscience* **28**, 680–692 (2016).

8. S. Bracci, H. O. de Beeck, Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways. *J. Neurosci.* **36**, 432–444 (2016).

9. D. Kaiser, D. C. Azzalini, M. V. Peelen, Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology* **115**, 2246–2250 (2016).

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

11. M. V. Peelen, P. E. Downing, Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia* (2017) https:/doi.org/10.1016/j.neuropsychologia.2017.03.033 (May 8, 2017).

12. S. Nasr, C. E. Echavarria, R. B. H. Tootell, Thinking Outside the Box: Rectilinear Shapes Selectively Activate Scene-Selective Cortex. *J. Neurosci.* **34**, 6721–6735 (2014).

13. B. Long, V. S. Störmer, G. A. Alvarez, Mid-level perceptual features contain early cues to animacy. *Journal of Vision* **17**, 20–20 (2017).

14. L. U. Perrinet, J. A. Bednar, Edge co-occurrences can account for rapid categorization of natural versus animal images. *Scientific Reports* **5**, 11400 (2015).

15. X. Yue, I. S. Pourladian, R. B. H. Tootell, L. G. Ungerleider, Curvature-processing network in

macaque visual cortex. *PNAS* **111**, E3467–E3475 (2014).

16. I. Rosenthal, *et al.*, Color statistics of objects, and color tuning of object cortex in macaque monkey. *Journal of Vision* **18**, 1–1 (2018).

17. L. Chang, P. Bao, D. Y. Tsao, The representation of colored objects in macaque color patches. *Nature Communications* **8**, 2064 (2017).

18. Y. Bi, X. Wang, A. Caramazza, Object Domain and Modality in the Ventral Visual Pathway. *Trends in Cognitive Sciences* **20**, 282–290 (2016).

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

20. V. Zachariou, A. C. D. Giacco, L. G. Ungerleider, X. Yue, Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization. *Journal of Vision* **18**, 3–3 (2018).

21. V. Troiani, A. Stigliani, M. E. Smith, R. A. Epstein, Multiple Object Properties Drive Scene-Selective Regions. *Cereb Cortex* **24**, 883–897 (2014).

22. R. Rajimehr, K. J. Devaney, N. Y. Bilenko, J. C. Young, R. B. H. Tootell, The "Parahippocampal Place Area" Responds Preferentially to High Spatial Frequencies in Humans and Monkeys. *PLoS Biology* **9**, e1000608 (2011).

23. V. Goffaux, F. Duecker, L. Hausfeld, C. Schiltz, R. Goebel, Horizontal tuning for faces originates in high-level Fusiform Face Area. *Neuropsychologia* **81**, 1–11 (2016).

24. G. J. Brouwer, D. J. Heeger, Decoding and Reconstructing Color from Responses in Human Visual Cortex. *J. Neurosci.* **29**, 13992–14003 (2009).

25. J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, *Multivariate Data Analysis: Pearson New International Edition* (Pearson Education Limited, 2013).

26. X. Wang, *et al.*, How Visual Is the Visual Cortex? Comparing Connectional and Functional Fingerprints between Congenitally Blind and Sighted Individuals. *J. Neurosci.* **35**, 12545–12559 (2015).

27. L. L. Chao, J. V. Haxby, A. Martin, Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience* **2**, 913–919 (1999).

28. M. B. Brodeur, K. Guérard, M. Bouras, Bank of Standardized Stimuli (BOSS) Phase II: 930 New Normative Photos. *PLOS ONE* **9**, e106953 (2014).

29. P. E. Downing, A. W.-Y. Chan, M. V. Peelen, C. M. Dodds, N. Kanwisher, Domain Specificity in Visual Cortex. *Cereb Cortex* **16**, 1453–1461 (2006).

30. F. J. Moreno-Martínez, P. R. Montoro, An Ecological Alternative to Snodgrass & Vanderwart: 360

High Quality Colour Images with Norms for Seven Psycholinguistic Variables. *PLOS ONE* **7**, e37527 (2012).

31. S. Nasr, R. B. H. Tootell, A Cardinal Orientation Bias in Scene-Selective Visual Cortex. *J. Neurosci.* **32**, 14921–14926 (2012).

32. M. D. Lescroart, D. E. Stansbury, J. L. Gallant, Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. *Front. Comput. Neurosci.* **9** (2015).

33. R. Caldara, *et al.*, The fusiform face area is tuned for curvilinear patterns with more high-contrasted elements in the upper part. *NeuroImage* **31**, 313–319 (2006).

34. P. B. Bryan, J. B. Julian, R. A. Epstein, Rectilinear Edge Selectivity Is Insufficient to Explain the Category Selectivity of the Parahippocampal Place Area. *Front. Hum. Neurosci.* **10** (2016).

35. R. Lafer-Sousa, B. R. Conway, N. G. Kanwisher, Color-Biased Regions of the Ventral Visual Pathway Lie between Face- and Place-Selective Regions in Humans, as in Macaques. *J. Neurosci.* **36**, 1682–1697 (2016).

36. B. Long, C.-P. Yu, T. Konkle, Mid-level visual features underlie the high-level categorical organization of the ventral stream. *PNAS*, 201719616 (2018).

37. V. Willenbockel, *et al.*, The SHINE toolbox for controlling low-level image properties. *Journal of Vision* **10**, 653–653 (2010).

38. C. He, *et al.*, Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *NeuroImage* **79**, 1–9 (2013).

39. J. Canny, A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-8**, 679–698 (1986).

40. N. Krüger, G. Peters, C. V. D. Malsburg, Object Recognition with a Sparse and Autonomously Learned Representation Based on Banana Wavelets in *Learned Representation Based on Banana Wavelets, Technical Report Ir-Ini 96-11. Institut Fur Neuroinformatik, Ruhr-Universitat Bochum*, (1996).

41. Pascal Getreuer (2020). Colorspace Transformations (https://www.mathworks.com/matlabcentral/fileexchange/28790-colorspace-transformations), MATLAB Central File Exchange. Retrieved March 20, 2020.

42. C.-G. Yan, X.-D. Wang, X.-N. Zuo, Y.-F. Zang, DPABI: Data Processing & Analysis for (Resting-State) Brain Imaging. *Neuroinform* **14**, 339–351 (2016).

Figures

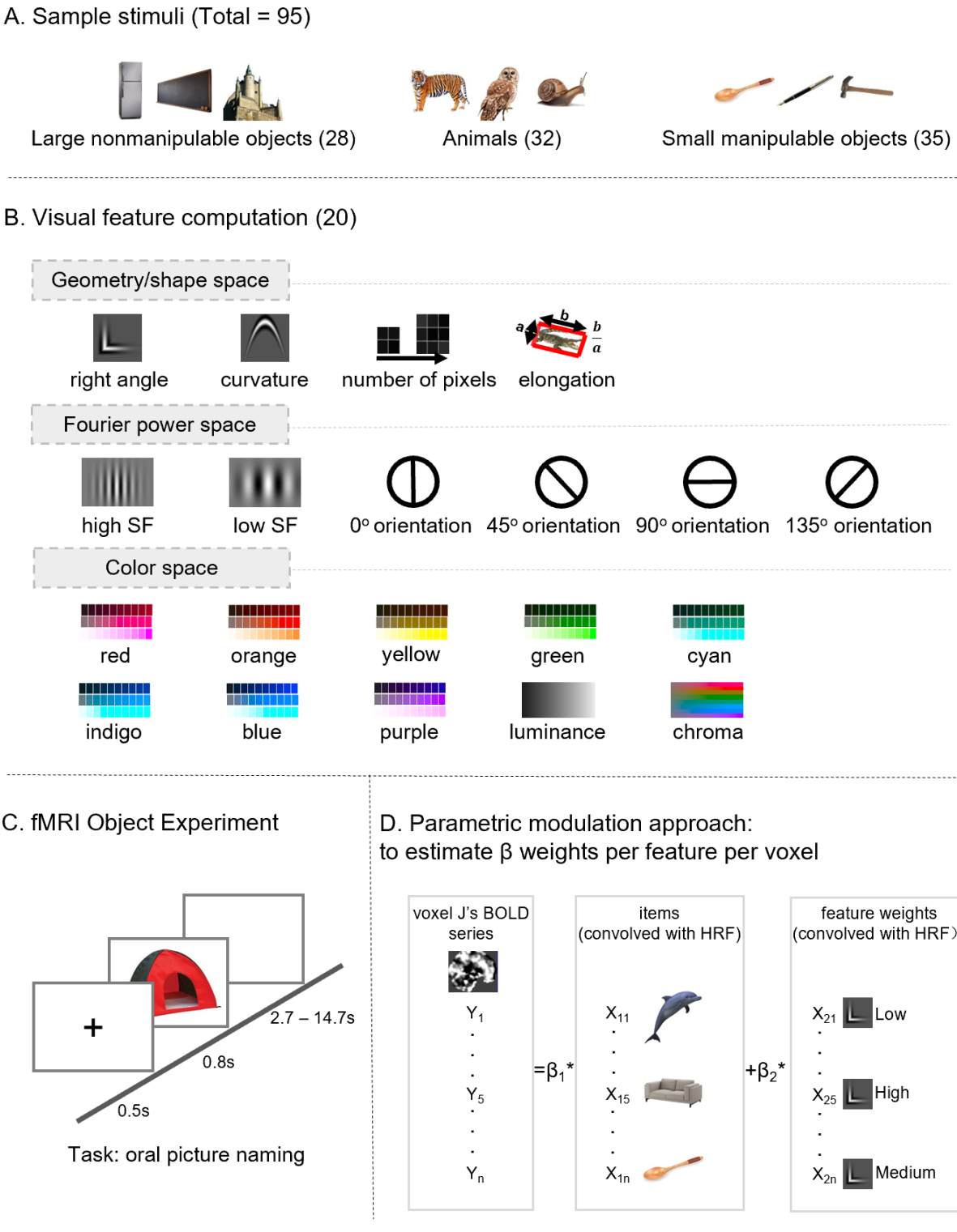

Figure 1. Schematic overview of the methods in Object Experiment. **(A)** Sample stimuli of the Object Experiment. Images of 95 common objects (28 large objects, 32 animals, and 35 small nonmanipulable objects) were used. **(B)** Visual feature construction from computational vision models. For each picture, computational vision models were used to obtain values of 20 visual

features, including geometry/shape (based on modified Gabor filters), Fourier power features (using 2D fast Fourier transform), color (based on Cmmission Internationale de l'Eclairage (CIE) L*C*H space). See **Fig. S1** and Materials and Methods for model construction method details. **(C)** fMRI experiment. In an event-related fMRI experiment, participants viewed and named these objects. **(D)** Parametric modulation analysis. Parametric modulation was used to estimate the degree of association between brain responses and visual feature weights across the whole VOTC.

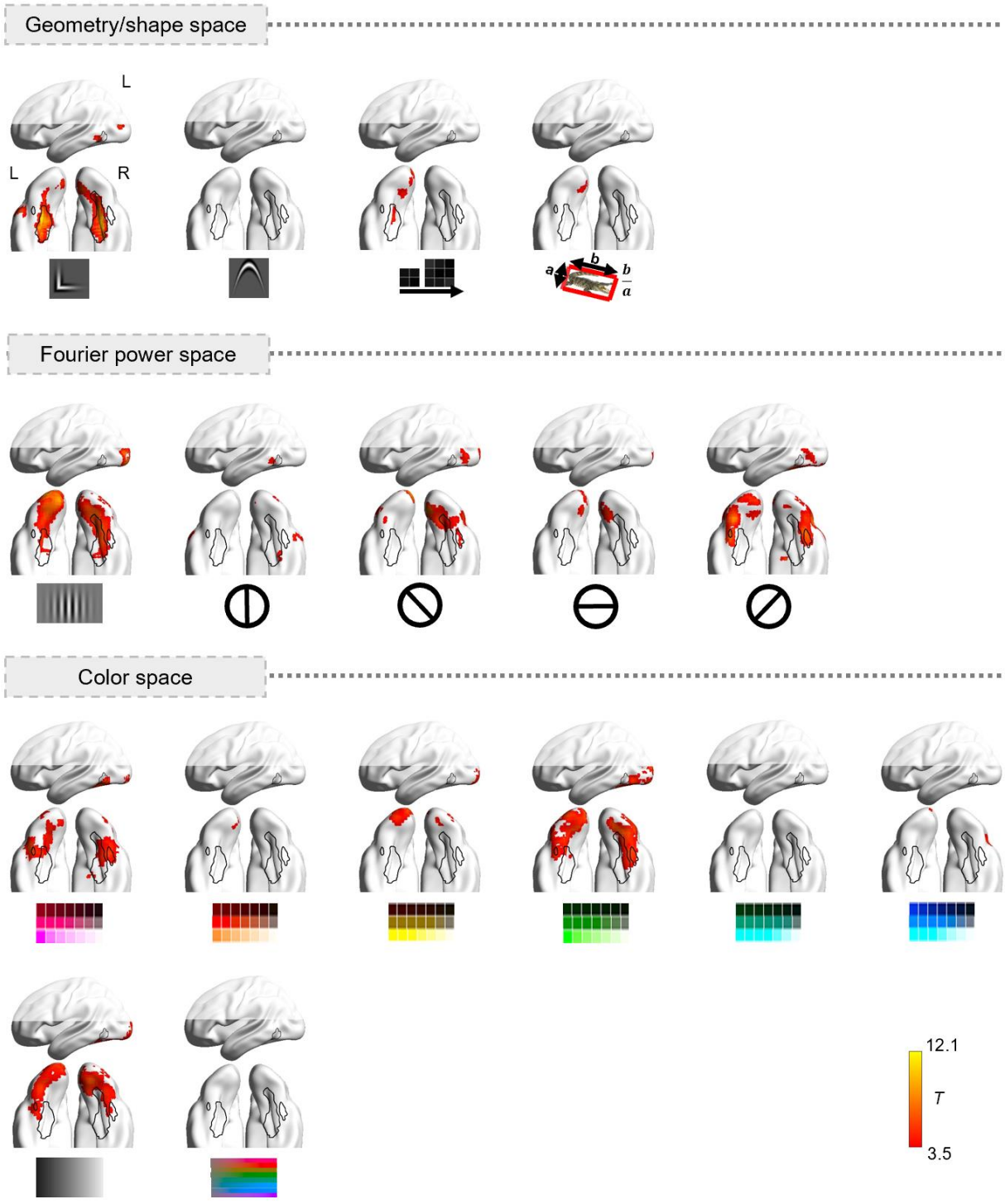

Figure 2. Object visual feature topography in a full-model parametric modulation analysis. All visual feature weights were entered into the parametric modulation model for BOLD activity estimates, yielding an activation map for each visual feature in the VOTC mask. The maps are thresholded at cluster-level FWE corrected $p < .05$ within the VOTC mask, with voxel-wise $p < .001$. The outlines show the domain-preferring clusters for large objects (bilateral PPA), animals (bilateral latFG), and small

manipulable objects (left LOTC) localized by contrasting each object domain with the other two domains in Object Experiment (see Materials and Methods).

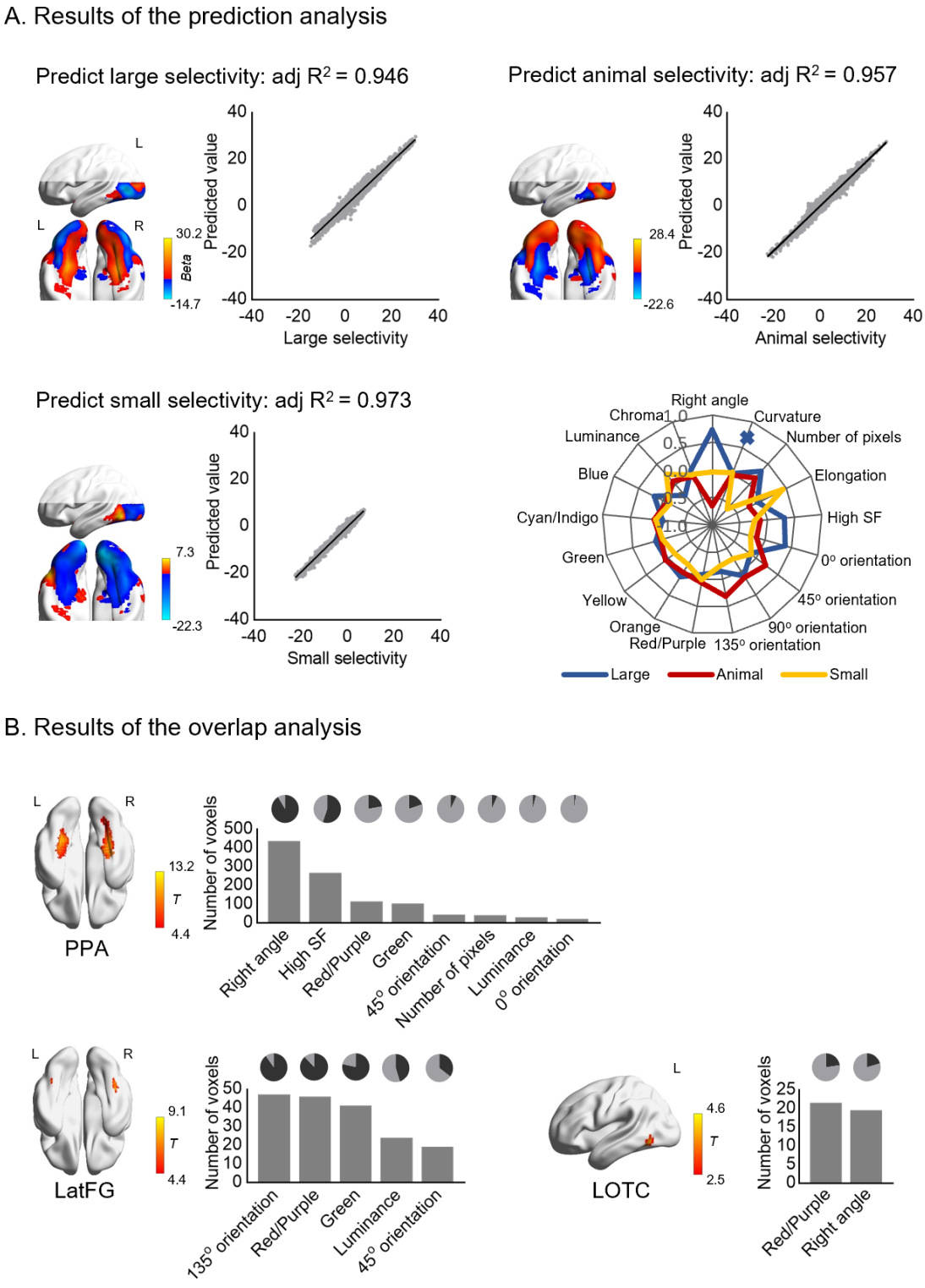

Figure 3. The association between visual feature topography and object domain effects. **(A)** Results of prediction analysis: A multiple linear regression model was constructed to predict domain selectivity strength for each domain, using 17 visual features' beta values as predictors, across all VOTC voxels. The brain maps are the unthresholded activation maps for each domain, showing the group-averaged selectivity strength (beta values of target domain – those of other domains) for all

VOTC voxels. The scatter plots show correlations between predicted domain-selectivity strength using VOTC visual-feature maps and the observed domain-selectivity strength across all VOTC voxels. The radar chart shows the standardized beta weights of each visual feature in the regression model for each domain. All but the one labelled by the cross mark (curvature in predicting large selectivity) were significant at FDR corrected $q < .05$ for 51 comparisons. **(B)** Results of overlapping analysis: The classical object domain preference clusters were localized by contrasting each object domain with the other two domains (same as **Fig. 2**). The obtained domain-preference maps were overlapped with each visual feature map in full model analysis (cluster-level FWE-corrected $p < .05$ within the VOTC mask, voxel-wise $p < .001$; see **Fig. 2**). The overlapping voxels between the domain preferring regions and the visual feature modulation clusters were counted, and those greater than 10 voxels were considered, with the number of overlapping voxels shown in the bar plots. The pie charts show the percentage of overlapping voxels in domain preferring regions.

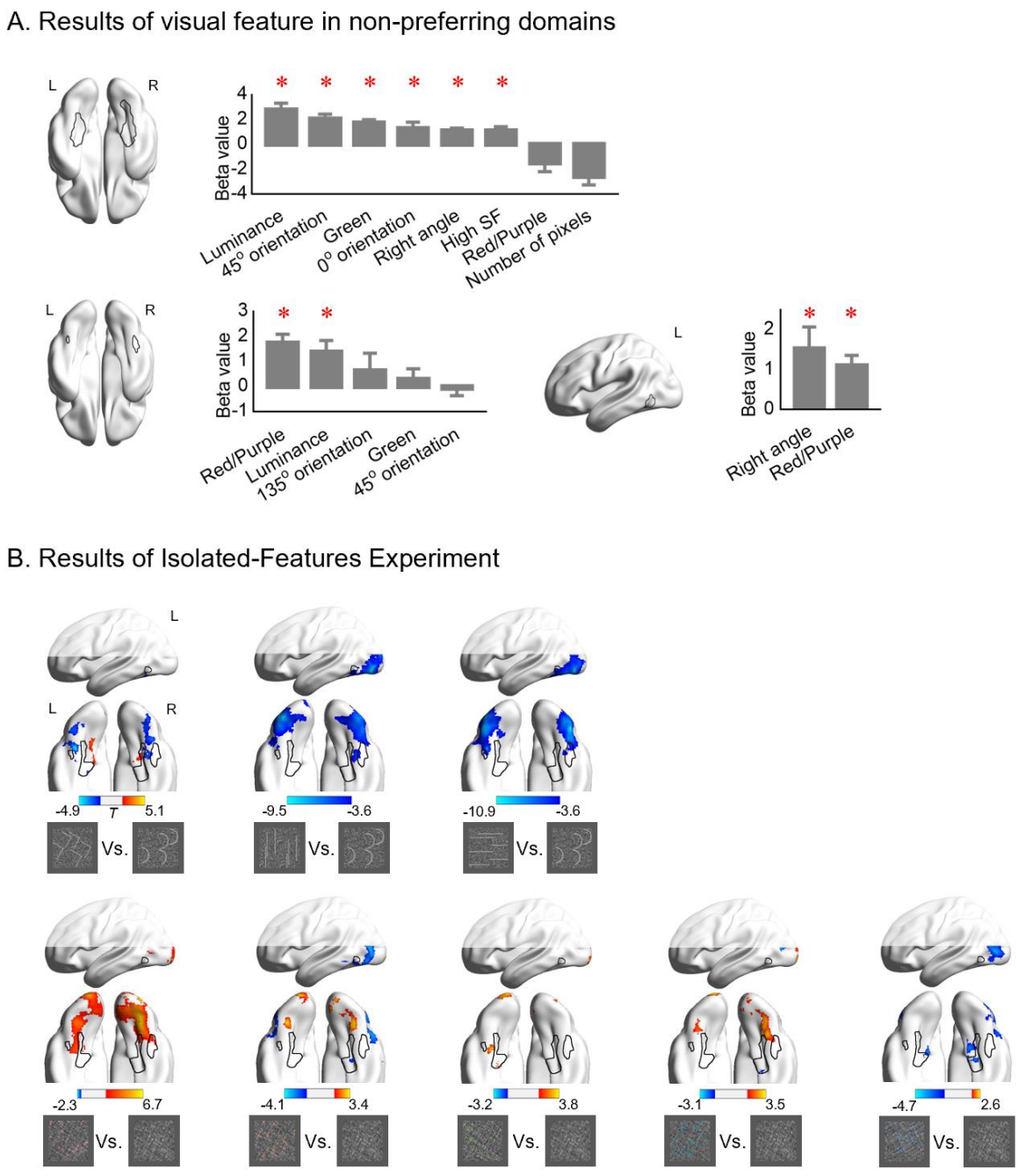

Figure 4. Visual feature effects independent of object domains. **(A)** Results of visual features in the non-preferring domains. In the parametric modulation analyses to obtain the visual feature effects, only object pictures belonging to the non-preferring domains were used. That is, for visual features significant in the large-object-preferring PPA, the parametric modulation model used pictures of animal and small manipulable objects. For visual features significant in animal-preferring latFG, the

parametric modulation model used pictures of large and small objects. For visual features in small-object-preferring LOTC, the parametric modulation results used pictures of animals and large objects. Beta values were extracted from the clusters showing overlap between domain-preferring regions and each visual feature map obtained in full-model parametric modulation analysis (see **Fig. 3B**). The bar plots show the averaged beta values of visual features across subjects; error bars indicate standard error. Asterisks indicate that beta values were significantly greater than zero (one-tailed one-sample *t* tests, FDR corrected $q < .05$). **(B)** Results of Isolated-Features Experiment. In the actual experiment, each array contained more elements/dots than the examples shown here (see Materials and Methods for details). The brain activity maps for the contrasts between horizontal/vertical lines and curvature were thresholded at cluster-level FWE-corrected $p < .05$ within the VOTC mask, voxel-wise $p < .001$ and other contrast maps were thresholded at voxel-wise $p < .05$, uncorrected. The outlines show the domain-preferring clusters for large objects/scenes (bilateral PPA), animals (bilateral latFG), and small manipulable objects (left LOTC) localized by contrasting each domain with the other two domains in Isolated Features Experiment.

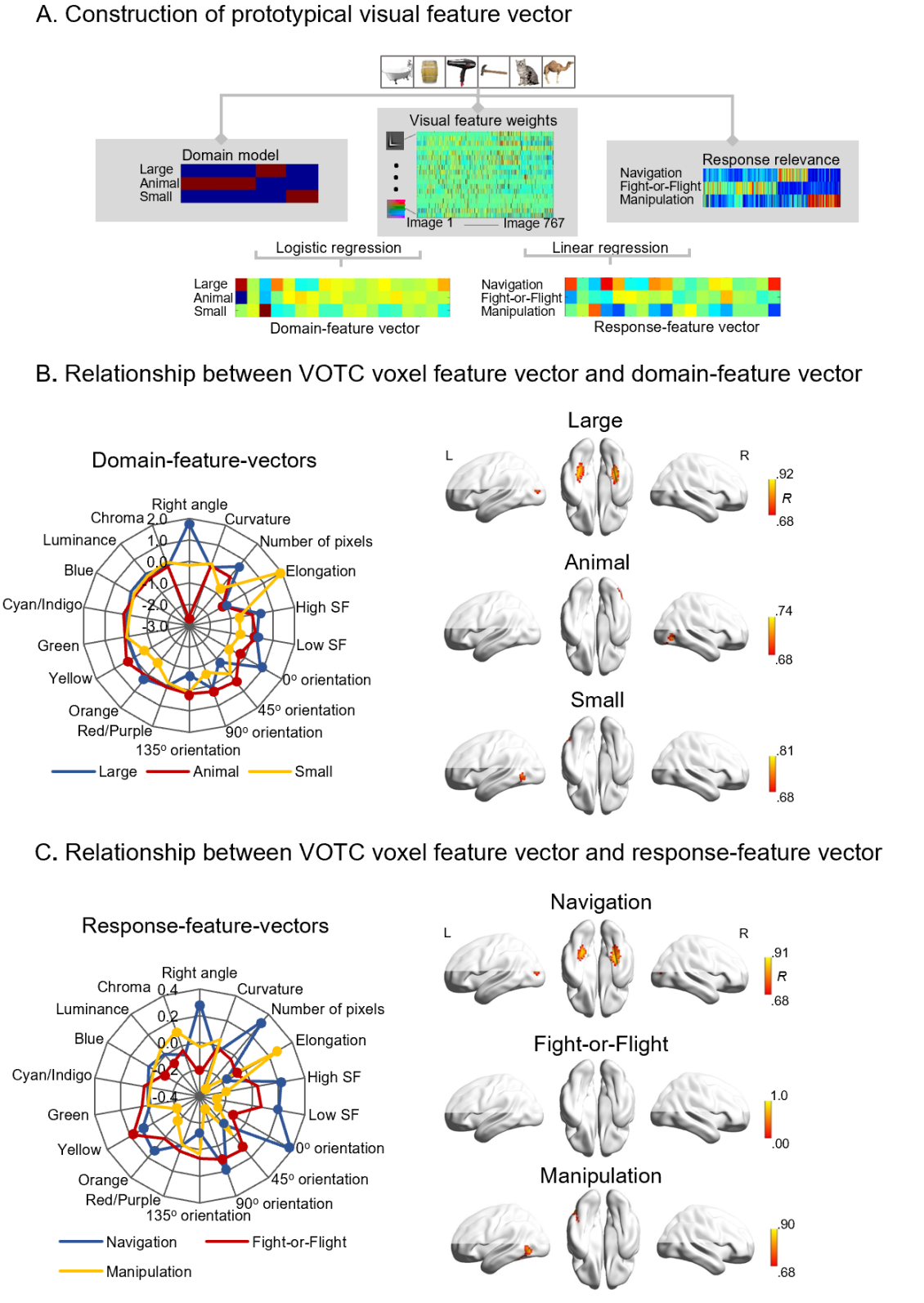

Figure 5. Relationship between domain-model, response-model, and visual featural effects. **(A)** Illustration of the construction scheme of prototypical vectors for the three domains and for the three response systems. In an image set of 767 images, visual feature weights for each image were obtained using computational vision models. For domain-driven "prototypical" feature vectors, we constructed three binary domain vectors (e.g., animal vector: animals = 1, others = 0) and performed logistic

regression analyses between each domain vector and each of the 18 visual feature weights, resulting in 3 domain-feature beta vectors. For response-driven "prototypical" visual feature vectors, we examined 3 theorized response systems (Navigation, Fight-or-Flight, and Manipulation) by asking 20 participants to rate how strongly each object associates with each of the three response systems. Linear regression was conducted between each response vector and each visual feature weight, resulting in 3 response-feature beta vectors. In **(B)** and **(C)**, the left column shows the "prototypical" visual feature vectors associated with each object domain (large nonmanipulable objects, animals, and small manipulable objects) or with each response system (Navigation, Fight-or-Flight, and Manipulation). Dots indicate that beta values were significant at FDR corrected $q < .05$ for 54 comparisons. The right column shows the Pearson correlation coefficients between each of these "prototypical" vectors and VOTC voxels' visual feature vector obtained from the fMRI parametric modulation analyses in independent models. The correlation maps are thresholded at cluster-level FWE corrected $p < .05$, voxel-wise $p < .001$.