

Category Trumps Shape as an Organizational Principle of Object Space in the Human Occipitotemporal Cortex

Elahe' Yargholi and  Hans Op de Beeck

Department of Brain and Cognition, Leuven Brain Institute, Faculty of Psychology & Educational Sciences, KU Leuven, 3000 Leuven, Belgium

The organizational principles of the object space represented in the human ventral visual cortex are debated. Here we contrast two prominent proposals that, in addition to an organization in terms of animacy, propose either a representation related to aspect ratio (stubby-spiky) or to the distinction between faces and bodies. We designed a critical test that dissociates the latter two categories from aspect ratio and investigated responses from human fMRI (of either sex) and deep neural networks (BigBiGAN). Representational similarity and decoding analyses showed that the object space in the occipitotemporal cortex and BigBiGAN was partially explained by animacy but not by aspect ratio. Data-driven approaches showed clusters for face and body stimuli and animate-inanimate separation in the representational space of occipitotemporal cortex and BigBiGAN, but no arrangement related to aspect ratio. In sum, the findings go in favor of a model in terms of an animacy representation combined with strong selectivity for faces and bodies.

Key words: animacy; aspect ratio; deep neural networks; fMRI; object recognition; occipitotemporal cortex

Significance Statement

We contrasted animacy, aspect ratio, and face-body as principal dimensions characterizing object space in the occipitotemporal cortex. This is difficult to test, as typically faces and bodies differ in aspect ratio (faces are mostly stubby and bodies are mostly spiky). To dissociate the face-body distinction from the difference in aspect ratio, we created a new stimulus set in which faces and bodies have a similar and very wide distribution of values along the shape dimension of the aspect ratio. Brain imaging (fMRI) with this new stimulus set showed that, in addition to animacy, the object space is mainly organized by the face-body distinction and selectivity for aspect ratio is minor (despite its wide distribution).

Introduction

Visual object recognition is crucial for human image understanding, with the lateral and ventral occipitotemporal cortex (OTC) being particularly important. This large cortical region has large-scale maps for distinctions, such as animate versus inanimate, selective areas for specific categories, such as faces, bodies, and scenes, and further maps for visual and semantic features (Konkle and Oliva, 2012; Grill-Spector and Weiner, 2014; Nasr et al., 2014; Bracci and Op de Beeck, 2016; Kalfas et al., 2017). However, it is unclear how these different aspects of this functional architecture can be put together in one comprehensive model.

Different proposals suggest either low-level visual properties (Coggan et al., 2016; Long et al., 2018) or higher-level features (Kriegeskorte et al., 2008; Carlson et al., 2014; Bracci and Op de Beeck, 2016; Bryan et al., 2016; Bracci et al., 2017; Peelen and Downing, 2017) could explain the functional organization of OTC.

Recently, Bao et al. (2020) proposed a comprehensive map of object space in monkeys' inferior temporal (IT) cortex based on data from monkey fMRI, single-neuron electrophysiology, and deep neural network (DNN). They suggested that the IT cortex is organized as a map with two main dimensions: aspect ratio (stubby-spiky) and animate-inanimate. This map would explain the location of specific regions in the IT cortex. In particular, the anatomic location of face- and body-selective regions would be constrained by the visual properties of these stimuli, with faces being stubby and bodies being spiky. A similar space was observed in DNNs.

This proposal of a 2D animacy \times aspect ratio space is appealing as a comprehensive model. At first sight, it fits with earlier but somewhat dispersed findings of both animacy and spiky-stubby selectivity in human and monkey (Kriegeskorte et al., 2008; Op de Beeck et al., 2008a; Op de Beeck and Bracci, 2022). In addition, the 2D object space as proposed by Bao et al. (2020)

Received Nov. 24, 2022; revised Feb. 22, 2023; accepted Mar. 3, 2023.

Author contributions: E.Y. and H.O.d.B. designed research; E.Y. and H.O.d.B. performed research; E.Y. analyzed data; E.Y. wrote the first draft of the paper; E.Y. and H.O.d.B. edited the paper; E.Y. and H.O.d.B. wrote the paper.

This work was supported by Fonds voor Wetenschappelijk Onderzoek 12A6122N, G0D3322N, and G073122N; and Excellence of Science Grant G0E8718N. Stimuli and processed data files will be available in the Open Science Framework repository for this project. We thank Artem Platonov for help with setting up the study.

The authors declare no competing financial interests.

Correspondence should be addressed to Elahe' Yargholi at elahe.yargholi@kuleuven.be.

<https://doi.org/10.1523/JNEUROSCI.2179-22.2023>

Copyright © 2023 the authors

was mostly confirmed in the human brain with the same stimulus set (Coggan and Tong, 2021). Nevertheless, the evidence for the model is fundamentally flawed. None of the experiments properly controlled the two dimensions. In Bao et al. (2020), the stimulus set overall dissociated stubby-spiky from animate-inanimate, but this was not true within stimulus classes: Face stimuli were all stubby and body stimuli were mostly spiky. Much of the evidence for aspect ratio as an overall dimension and as a dimension underlying face and body selectivity and determining the location of selectivity for these categories, might be because of this major confound between aspect ratio and stimulus category (face vs body). It is possible that, alternatively, face versus body selectivity has a primary role in the organization of OTC and that aspect ratio is at best a dimension of secondary importance. This is an alternative view of how object representations are organized, why they have this structure, and what constraints might determine where specific category-selective regions emerge in OTC.

We designed a new experimental paradigm to dissociate two alternative hypotheses: An object space organized in terms of the dimensions of animacy and aspect ratio, or instead an object space organized in terms of a primary dimension of animacy and a further distinction between faces and bodies that is NOT related to the aspect ratio of these categories. To be able to properly distinguish among these alternatives, we designed a novel stimulus set in which the face and body categories are explicitly dissociated from aspect ratio (being stubby or spiky). We obtained neural responses using fMRI and examined object representations in human OTC as well as a state-of-the-art DNN. We observed no tuning representation for aspect ratio in human face and body regions. A representation for aspect ratio was restricted to right object-selective regions and only present for inanimate objects. Our findings go against the 2D animacy \times aspect ratio model, and in favor of a model in terms of an animacy representation combined with strong selectivity for faces and bodies.

Materials and Methods

Participants

Seventeen subjects (12 females, 19–46 years of age) took part in the experiment. All subjects were healthy and right-handed with normal visual acuity. They gave written informed consent and received payment for their participation. The experiment was approved by the Ethics Committee on the use of human subjects at the Universitair Ziekenhuis/Katholieke Universiteit Leuven. Two subjects were later excluded because of excessive head motion (for more details, see Preprocessing), and the final analyses included the remaining 15 subjects.

Stimuli

Stimuli in the main experiment included 52 images from four categories: animal body, animal face, manmade, and natural object (see Fig. 1A). We selected 13 exemplar images from each of these four categories that were selected to provide a comparably wide range of aspect ratios for different categories. In the literature on shape description, aspect ratio is usually defined as the ratio of principle axes in a shape, but following Bao et al. (2020), we defined aspect ratio as a function of perimeter P and area A ; aspect ratio = $P^2/(4\pi A)$. Using this formula, the ranges of aspect ratio for body and face stimuli in Bao et al. (2020) were 1.38–6.84 and 0.91–1.19, respectively, while we had ranges of 1–9.66 for bodies and 1.51–10.89 for faces. In each category, we sorted stimuli according to the aspect ratio, determined the category median of aspect ratio, and labeled stimuli with aspect ratios larger than the category median as spiky and those with aspect ratios smaller than the category median as stubby.

All images were grayscale with a gray background, cropped to 700 \times 700 pixels (subtended 10 degrees of visual angle in the MRI scanner).

We equalized the luminance histogram and the average energy at each spatial frequency across all the images using the shine toolbox (Willenbockel et al., 2010). Examples of finalized stimuli used for fMRI experiments and computational analyses are provided in Figure 1B.

To compare stimuli with respect to animacy, aspect ratio, face-body, and low-level shape properties, we computed representational dissimilarity matrices (RDMs) and quantified pair-wise resemblance of images for these properties. For constructing a dissimilarity matrix based on animacy, we assigned scores +1 to animate (face and body) and –1 to inanimate (manmade and natural) stimuli and computed the absolute difference of scores for each stimulus pair (see Fig. 1C). The aspect ratio model included the pairwise absolute difference of aspect ratios (see Fig. 1C). Face-body model assigned a score of +1 to the stimulus pairs where one stimulus included a face and the other a body, and 0 to all other pairs (see Fig. 1C). As measures of low-level shape properties, we investigated pixel-wise dissimilarity (Op de Beeck et al., 2008b) and dissimilarities based on outputs of hmax model (S1-C1, C2) (Serre et al., 2007). We computed correlation distances between pixels/S1-C1/C2 for each pair of stimuli to build dissimilarity matrices. Comparison of model RDMs (Spearman correlation and p values from one-sided permutation tests) showed nonsignificant similarity between animacy and aspect ratio models ($r = -0.0201$, $p = 0.8800$), animacy and low-level models (pixels: $r = 0.0358$, $p = 0.1150$, S1-C1: $r = 0.0412$, $p = 0.0540$, C2: $r = -0.0052$, $p = 0.5270$), and aspect-ratio and low-level models (pixels: $r = -0.0057$, $p = 0.5460$, S1-C1: $r = 0.0148$, $p = 0.3600$, C2: $r = 0.0473$, $p = 0.2080$). The face-body model showed significantly negative correlations with animacy ($r = -0.3898$, $p < 0.001$) and S1-C1 ($r = -0.1477$, $p = 0.007$) models and no significant correlations with aspect ratio ($r = 0.0045$, $p = 0.4730$), pixels ($r = -0.0499$, $p = 0.06$), and C2 ($r = -0.0094$, $p = 0.3850$) models.

Experimental design and statistical analyses

Sample size

The number of participants was chosen to be sufficient to have a power of ~ 0.95 with reliable data that guarantee an effect size of $d = 1$. According to previous studies (i.e., Bracci and Op de Beeck, 2016; Bracci et al., 2019), we know that the distinction between animals and other objects as revealed by representational similarity analysis (RSA) has a very high effect size (Cohen's d of 1–4) even in small ROIs. The effect size was larger than expected in the power calculation: All participants showed a larger correlation with the animacy model compared with the aspect ratio. We also confirmed the high between-subject reliability of the neural data by calculating the noise ceiling for RSA correlations in our ROIs.

Scanning procedures

Data recording consisted of eight experimental runs, three localizer runs, and one anatomic scan, all completed in one session for each participant.

For the experimental runs, we used a rapid event-related design. Each experimental run included a random sequence of 138 trials: two repetitions of each stimulus image plus 34 fixation trials. Each trial was 3 s. Stimulus trials began with the stimulus presentation for 1500 ms and were followed by 1500 ms of the fixation point. Each experimental run lasted 7 min 14 s. The fixation point was presented at the center of the screen continuously throughout each run. Subjects were instructed to fixate on the fixation point and, on each trial, press a button to indicate whether they preferred looking at the current image or the previous one (same task as in Ritchie et al., 2021).

For the localizer runs, we used a block design with six stimulus types: body, face, stubby object, spiky object, word, and box-scrambled version of the object images. Each run included 24 blocks with four blocks for each stimulus type. The presentation order of the stimulus types was counterbalanced across runs. Each block lasted 16 s. There was a 10 s blank period at the beginning and end, and three 12 s blank periods between stimulus blocks of each repetition. Each localizer run lasted 7 min 20 s. In each stimulus block, there were 18 images of the same stimulus type and two different randomly selected images were repeated.

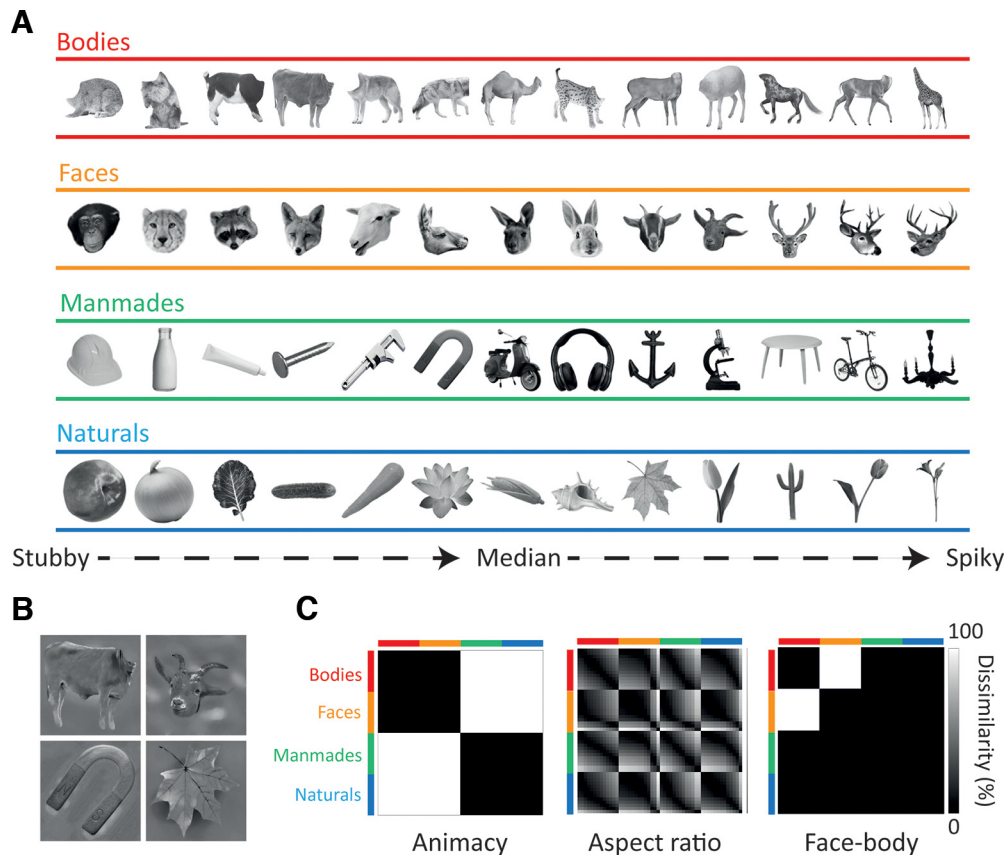


Figure 1. Stimulus design. **A**, All 52 images used to prepare stimuli, in the order of increasing aspect ratio from left to right and color-coded based on the category. **B**, Examples of finalized stimuli (gray background, equalized luminance histogram, etc.) used for fMRI experiment and computational analysis. **C**, The main model RDMs used throughout the research. The axes of RDMs are color-coded based on stimulus category.

Each image in a block appeared for 400 ms followed by 400 ms fixation. A fixation point was presented at the center of the screen continuously throughout each run. Subjects were instructed to fixate on the fixation point, detect the stimulus one-back repetition, and report it by pressing a response key with their right index finger. For one subject, the data for one of three localizer runs were discarded since there were uncorrectable artifacts in the data.

Subjects viewed the visual stimuli through a back-projection screen, and the tasks were presented using MATLAB and Psychtoolbox-3 (Brainard and Vision, 1997).

Acquisition parameters

MRI data were collected at the Department of Radiology of the Universitair Ziekenhuis Leuven University Hospitals using a 3T Philips scanner, with a 32-channel head coil. Functional images were obtained using a 2D multiband T2*-weighted EPI sequence with a multiband of 2, TR of 2 s, TE 30 ms, 90° flip angle, 46 transverse slices, and a voxel size of $2 \times 2 \times 2 \text{ mm}^3$. A high-resolution T1-weighted structural scan was also acquired from each participant using an MPRAGE pulse sequence ($1 \times 1 \times 1 \text{ mm}^3$ isotropic voxels).

Preprocessing

fMRIPrep (Esteban et al., 2019) was used for preprocessing anatomic and functional data, using default settings unless otherwise noted. The T1-weighted image was corrected for intensity nonuniformity, skull-stripped, and went through nonlinear volume-based registration to ICBM 152 nonlinear Asymmetrical template version 2009c. Each of the bold runs was motion-corrected, coregistered to the individual's anatomic scan, and normalized into standard space MNI152NLin2009cAsym. Subjects with excessive head motion (frame-wise displacement $> 2 \text{ mm}$; > 1 voxel size) were excluded (2 subjects). Subjects either did not have frame-wise displacements greater than the predefined threshold or had it repeated

multiple times. The rest of the analyses were conducted with SPM12 software (version 6906). As the last step of preprocessing, all functional volumes were smoothed using a Gaussian kernel, 4 mm FWHM. After preprocessing, a run-wise GLM analysis was performed to obtain the β values for each stimulus image of the experimental and each stimulus type of the localizer runs in each voxel. For the experimental runs, the GLM consisted of the 52 stimulus regressors (boxcar functions at the stimulus onsets with a duration of 1500 ms convolved with a canonical HRF) and six motion correction parameters (translation and rotation along the x , y , and z axes). For the localizer runs, the GLM included the six stimulus regressors (boxcar function at the block onsets with a duration of 16 s convolved with a canonical HRF) and the same six motion correction parameters.

Defining ROIs

We used contrasts from localizer runs intersected with masks from functional (Julian et al., 2012) or anatomic atlases (Anatomy Toolbox) (Eickhoff et al., 2005) to define a maximum of 26 ROIs covering lateral and ventral OTC in each subject. Since the functional atlas included few ROIs in the ventral cortex, we used the intersection of contrast maps with FG1-FG4 (FGs) from the Anatomy Toolbox to define more ventral ROIs. All lateral and ventral body-, face-, and object-selective ROIs in each hemisphere were merged to produce a large ROI representing left or right OTC. We also used the anatomic V1 mask to define EVC. Details on localizing ROIs (using functional contrasts and masks) are provided in Table 1. To ensure that ROIs with the same selectivity were independent, we examined their overlap and took overlapping voxels away from the larger ROIs. This mostly happened for LOC and FGs in the object, spiky, and stubby areas; and overlapping voxels were removed from LOC. This removal of overlap did not meaningfully change any of

Table 1. Functional contrasts and masks used to localize category-selective ROIs

| ROIs | Contrast | Mask |
|---------------------------|---------------------------------------|------------------|
| Body areas | | |
| IFGs-body & rFGs-body | Body > (face + objects ^a) | Anatomic atlas |
| IEBA & rEBA | | Functional atlas |
| Face areas | | |
| IFFA & rFFA | Face > (body + object ^a) | Functional atlas |
| IOFA & rOFA | | Functional atlas |
| ISTS & rSTS | | Functional atlas |
| Object areas | | |
| IFGs-object & rFGs-object | Object ^a > scramble | Anatomic atlas |
| ILOC-object & rLOC-object | | Functional atlas |
| Spiky areas | | |
| IFGs-spiky & rFGs-spiky | Spiky > stubby | Anatomic atlas |
| ILOC-spiky & rLOC-spiky | | Functional atlas |
| Stubby areas | | |
| IFGs-stubby & rFGs-stubby | Stubby > spiky | Anatomic atlas |
| ILOC-stubby & rLOC-stubby | | Functional atlas |
| Word areas | | |
| IFGs-word & rFGs-word | Word > object ^a | Anatomic atlas |
| EVC | | |
| IV1 & rV1 | All conditions > fixation | Anatomic atlas |

^aObject condition includes both spiky and stubby objects.

the findings and statistics, and very similar findings were obtained when these voxels are not removed.

To capture gradual effects, we defined a series of 37 small consecutive ROIs along the anteroposterior axis that forms the ventral visual stream. These ROIs were localized using the functional contrast of All conditions > Fixation intersected with anatomic masks, including V1, V2, V3v, V4v, and FGs (Spriet et al., 2022).

All ROIs included at least 25 voxels that surpassed the statistically uncorrected threshold $p < 0.001$ in the relevant functional contrast and were included in the relevant mask (Table 1). If the number of surviving voxels was < 25 , a more liberal threshold of $p < 0.01$ or $p < 0.05$ was applied.

Figures 4 and 7, respectively, show category-selective ROIs and the ventral visual stream in one representative subject mapped onto the inflated cortex using FreeSurfer (<https://surfer.nmr.mgh.harvard.edu>).

Mean sizes (SEs) for category-selective ROIs were as follows: IEBA = 823 (54), rEBA = 1064 (77), IFGs-body = 145 (28), rFGs-body = 138 (26), IFFA = 90 (13), rFFA = 139 (19), IOFA = 54 (7), rOFA = 122 (13), ISTS = 39 (7), rSTS = 91 (29), IFGs-object = 326 (50), rFGs-object = 208 (31), ILOC-object = 294 (42), rLOC-object = 284 (44), IFGs-spiky = 99 (28), rFGs-spiky = 77 (16), ILOC-spiky = 248 (61), rLOC-spiky = 207 (74), IFGs-stubby = 50 (8), rFGs-stubby = 53 (6), ILOC-stubby = 83 (31), rLOC-stubby = 70 (9), IFGs-word = 207 (38), rFGs-word = 50 (6), IV1 = 335 (24), and rV1 = 467 (22).

Statistical analysis of the main fMRI experiment

Multivariate (RSA and classification) and univariate analyses were used to study the role of animacy and aspect ratio in organizing OTC's object space. In-house MATLAB code and the CoSMoMvPA (Oosterhof et al., 2016) toolbox were used for the following analyses.

RSA

We compared the RDMs based on neural activity in different ROIs with model RDMs based on animacy, aspect ratio, and face-body. Neural RDM for each ROI included the pairwise Mahalanobis distance between activity patterns (β weights) of the ROI for different stimuli (Walther et al., 2016; Ritchie and Op de Beeck, 2019). The off-diagonal of the neural and model RDMs was vectorized, and the Spearman's correlation between dissimilarity vectors was then calculated and compared. We tested the significance of correlation values between neural and model RDMs across subjects using one-sample t tests and the significance of differences between correlation

Table 2. Train and test sets for within- and cross-category classification

| Classifier | | Train set | Test set |
|-------------------|---------------------|-----------|-----------|
| Animate-inanimate | Within aspect ratio | Spiky | Spiky |
| | | Stubby | Stubby |
| | Cross aspect ratio | Spiky | Stubby |
| Spiky-stubby | Within animacy | Stubby | Spiky |
| | | Animate | Animate |
| | Cross animacy | Inanimate | Inanimate |
| | | Animate | Inanimate |
| | | Inanimate | Animate |

values obtained for animacy RDM and aspect ratio RDM, animacy RDM and face-body RDM, or aspect ratio RDM and face-body RDM across subjects using paired t tests. Then, p values were corrected for multiple comparisons across all ROIs (false discovery rate [FDR]) (Benjamini and Hochberg, 1995).

For body-, face-, and object-selective ROIs, we repeated the RSA for a subset of our stimuli limited to their favorite category; neural and aspect ratio RDMs were computed using body, face, or object (manmade and natural) stimuli separately, the off-diagonal of the neural and model RDMs were vectorized, and the Spearman's correlation between dissimilarity vectors was calculated and tested for significance using one-sample t tests using correction for multiple comparisons across ROIs (FDR) (Benjamini and Hochberg, 1995).

When correlating neural and model dissimilarity, we can compare this correlation with the reliability of the neural dissimilarity matrix. This reliability can be interpreted as the estimate of the highest possible correlation given the noise in the data (Op de Beeck et al., 2008a). For each region, it was computed as the cross-validated correlation of each subject's neural dissimilarity matrix with the mean of the remaining subjects' neural dissimilarity matrices. These reliabilities, called the noise ceiling for fMRI data, are provided as gray background bars in all the relevant figures.

For OTC, we averaged the neural dissimilarity matrices across subjects to obtain a group-level neural dissimilarity matrix, applied multidimensional scaling (MDS) analysis, and used the first two dimensions that explained most of the variance to produce an MDS diagram in which the distance among stimuli expresses the similarity among their neural representation.

Within- and cross-category classification. We performed category classification (i.e., animate vs inanimate and spiky vs stubby classification). In a leave-one-run-out cross-validation procedure, samples (β weights) were partitioned to train and test sets and linear discriminant analysis was used to perform brain decoding classification in individual brains. The animate versus inanimate classification was performed within and cross aspect ratio; for within aspect ratio, we trained and tested classifiers with the same category of stimuli (either spiky or stubby); while for cross aspect ratio, classifiers were trained and tested with stimuli from different categories (Table 2). The spiky versus stubby classification was also performed within and cross animacy; for within animacy, we trained and tested classifiers with the same category of stimuli (either animate or inanimate); while for cross animacy, classifiers were trained and tested with stimuli from different categories (Table 2). Using median as a criterion to label stimuli as "stubby" or "spiky," we had six stubby bodies, six spiky bodies, six stubby faces, ... so we would have the same number of stimuli from bodies, faces, manmades and naturals in all classifications. We tested whether the classification accuracies for within- or cross-category classifications across subjects were significantly above chance level (50% chance level for two-category classification) using one-sample t tests and the significance of differences between within- and cross-category classification accuracies across subjects using paired t tests. Then, p values were corrected for multiple comparisons across ROIs (FDR) (Benjamini and Hochberg, 1995). Comparing within- and cross-category classification accuracies for these two dimensions, we determined the ROIs' capability of generalizing distinction along one dimension over changes in the other dimension, which specifies the ROIs in which animacy and aspect ratio contents are independently represented.

Univariate analysis. We investigated the mean response of each category-selective ROI to each of the animacy-aspect ratio conditions: animate-spiky, animate-stubby, inanimate-spiky, and inanimate-stubby. For each stimulus, β weights were averaged across all runs, and then all voxels within each ROI in individual subjects. Then, responses for animate-spiky were compared with animate-stubby and responses for inanimate-spiky were compared with inanimate-stubby using paired *t* tests across participants, and *p* values were corrected for multiple comparisons across ROIs (FDR) (Benjamini and Hochberg, 1995).

Computational simulation with the artificial neural network: BigBiGAN

BigBiGAN is a state-of-the-art artificial neural network for unconditional image generation regarding image quality and visual plausibility. It is a large-scale bidirectional generative adversarial network, pretrained on ImageNet that converts images into a 120-dimensional latent space. This unified latent space captures all properties of objects, including high-level image attributes and categories (Donahue and Simonyan, 2019; Mozafari et al., 2020). In order to study features that characterize the identity of our stimuli for BigBiGAN, we passed our stimuli through the network (<https://tfhub.dev/deepmind/bigbigan-revnet50x4/1>) and acquired corresponding latent vectors. Then, for obtaining the main factors, principal component analysis (PCA) was applied to the latent vectors and the first two principal components (PCs) were used for visualization. We computed an RDM based on the correlation distances of the first two PCs (called BigBiGAN RDM), vectorized its off-diagonal, and calculated Spearman's correlation between this dissimilarity vector and the vectors from the animacy and aspect ratio model. The significance of these correlations is also verified by using permutation tests.

Results

In this study, we aimed to investigate whether animacy and aspect ratio are the two main dimensions organizing object space in OTC using a stimulus set that properly dissociated the type of visual stimuli, and in particular face and body, from the aspect ratio. A stimulus set of body, face, manmade, and natural objects with a comparably wide range of aspect ratios in each of these four categories was presented (Fig. 1A), and fMRI responses (β weights) were extracted in each voxel of the brain for each stimulus. Then, we used fMRI multivariate (RSA and classification) and univariate analysis. Representations of our stimuli in a DNN's latent space were also inspected for the animacy and aspect ratio contents.

To what extent do animacy, aspect ratio, and face-body models explain the organization of OTC's object space?

To examine the neural representation of animacy and aspect ratio in different ROIs, we compared neural dissimilarity in the multivoxel patterns elicited by individual images with predictions from animacy and aspect ratio models (Fig. 1C). The result for OTC (combination of body-, face-, and object-selective ROIs) is shown in Figure 2A. Both left and right OTC showed

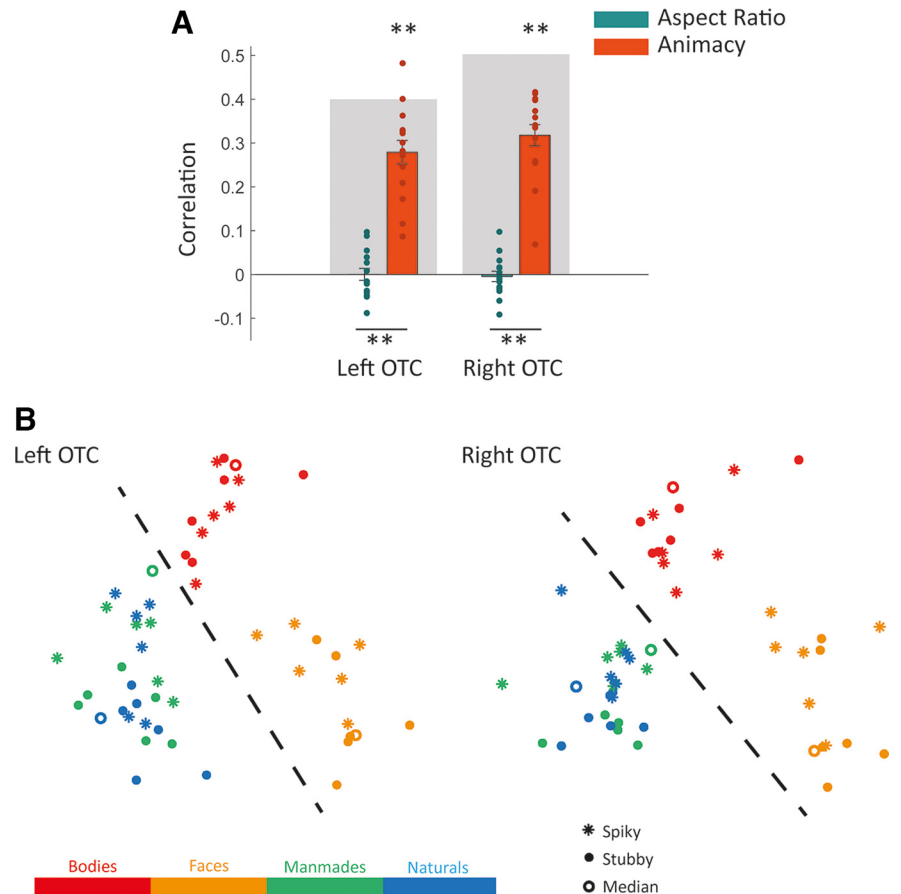


Figure 2. Representational similarity and effects of aspect ratio and animacy in the occipitotemporal cortex. **A**, Bar plot represents the mean Spearman's correlations between neural RDMs for individual subjects (dots) and model RDMs. Error bars indicate SEM. Gray background bars represent the noise ceiling. ****** $p < 0.001$; ***** $p < 0.01$; one-sided one-sample or two-sided paired *t* test. **B**, Two-dimensional representational space as obtained by applying MDS to the OTC's neural RDM. Points are color-coded based on stimulus category and are dots, rings, or asterisks based on aspect ratio. Dashed line indicates the separation of animates from inanimates.

significant positive correlations with the animacy model (one-sided one-sample *t* test, left OTC: $t_{(14)} = 10.39$, $p < 0.001$, right OTC: $t_{(14)} = 13.09$, $p < 0.001$) and not with the aspect ratio model (one-sided one-sample *t* test, left OTC: $t_{(14)} = 0.0054$, $p > 0.5$, right OTC: $t_{(14)} = -0.38$, $p > 0.5$). The direct comparison of the correspondence with neural dissimilarity between animacy and aspect ratio showed significant differences in both hemispheres (two-sided paired *t* test, left OTC: $t_{(14)} = 10.02$, $p < 0.001$, right OTC: $t_{(14)} = 11.38$, $p < 0.001$), revealing a far stronger correspondence for the animacy model than for the aspect ratio model. The large effect size of this difference is further illustrated by the observation that all 15 participants showed a larger correspondence with animacy compared with the aspect ratio, and this is in each hemisphere. The MDS plots for left and right OTC (Fig. 2B) reveal clusters for faces and bodies and a clear separation of animates from inanimates, but spiky and stubby stimuli are intermingled and not at all separated in clusters.

Strong selectivity for the distinction between faces and bodies is an organizational principle that is an alternative for aspect ratio. To test this alternative, we compared the neural dissimilarity in OTC with the face-body model (Fig. 1C). Both left and right OTC showed significant positive correlations with the face-body model, Figure 3 (one-sided one-sample *t* test, left OTC: $t_{(14)} = 4.08$, $p < 0.001$, right OTC: $t_{(14)} = 6.51$, $p < 0.001$). The direct comparison of the correspondence with neural dissimilarity between face-

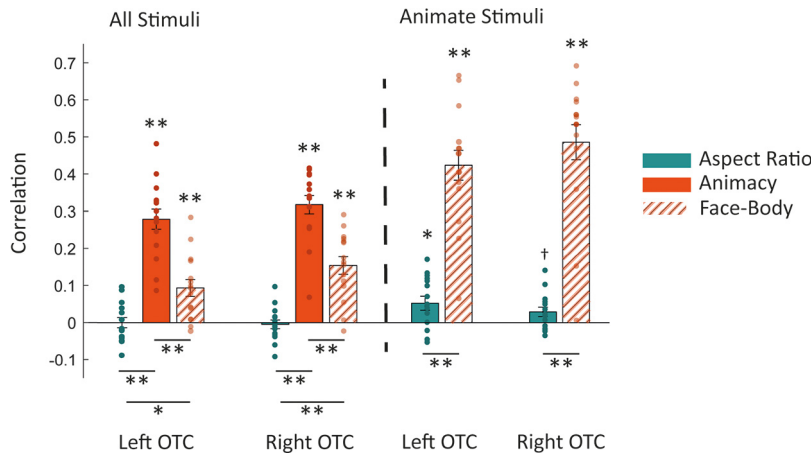


Figure 3. Representational similarity and effects of aspect ratio, animacy, and face-body in the occipitotemporal cortex for all stimuli and animate stimuli. Bar plot represents the mean Spearman's correlations between neural RDMs for individual subjects (dots) and model RDMs. Error bars indicate SEM. ** $p < 0.001$; * $p < 0.01$; † $p < 0.05$; one-sided one-sample or two-sided paired t test.

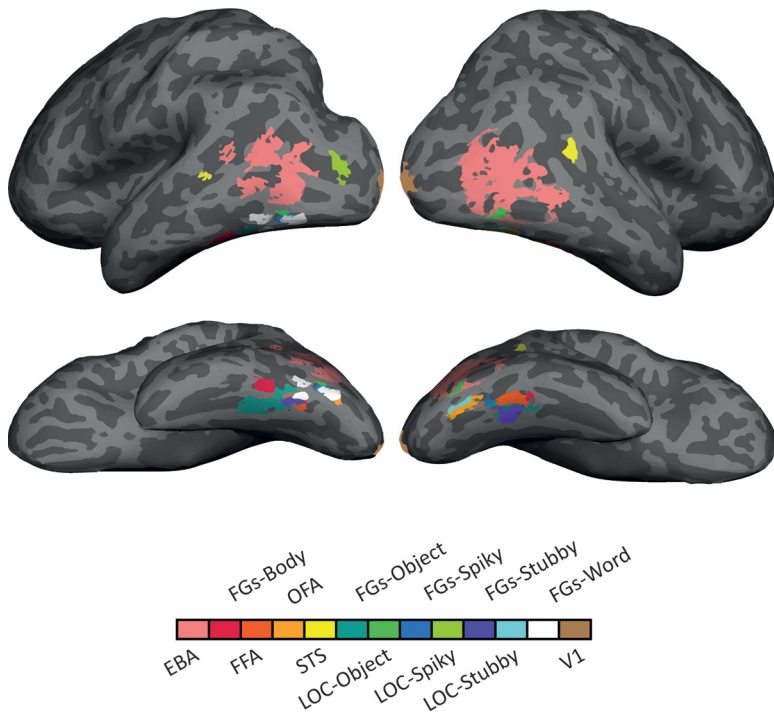


Figure 4. Inflated brain surface from an individual participant illustrating the category-selective ROIs examined, in lateral and ventral views.

body and animacy (two-sided paired t test, left OTC: $t_{(14)} = 5.16$, $p < 0.001$, right OTC: $t_{(14)} = 6.43$, $p < 0.001$) or face-body and aspect ratio (two-sided paired t test, left OTC: $t_{(14)} = 3.34$, $p < 0.005$, right OTC: $t_{(14)} = 6.31$, $p < 0.001$) showed significant differences in both hemispheres, revealing a correspondence for face-body model weaker than animacy, but stronger than aspect ratio. We should be careful with interpreting the smaller effect for face-body relative to animacy because the animacy dimension splits the stimulus set evenly in two subsets. This explains the discrepancy with the findings of Ritchie et al. (2021) who found stronger face-body than animacy selectivity, in that case using a design where the distinction of faces and bodies split the stimulus set most evenly.

To have a more balanced design for the face-body distinction, we considered a reduced face-body model based on 26 animate

stimuli (i.e., upper left quadrant of the face-body model in Fig. 1C). This reduced face-body model was not correlated with a reduced version of the aspect ratio model (upper left quadrant of aspect ratio model, $r = 0.0750$, $p = 0.0710$). Both left and right OTC showed significant correlations between neural dissimilarity and the reduced face-body model (one-sided one-sample t test, left OTC: $t_{(14)} = 10.49$, $p < 0.001$, right OTC: $t_{(14)} = 10.25$, $p < 0.001$) and minor correlation with the aspect ratio model (one-sided one-sample t test, left OTC: $t_{(14)} = 2.75$, $p < 0.01$, right OTC: $t_{(14)} = 2.29$, $p < 0.05$, not corrected for multiple comparisons) (Fig. 3). The direct comparison of the correspondence with neural dissimilarity between the reduced face-body model and aspect ratio (two-sided paired t test, left OTC: $t_{(14)} = 8.57$, $p < 0.001$, right OTC: $t_{(14)} = 9.05$, $p < 0.001$) showed significant differences in both hemispheres (Fig. 3). This strong selectivity for faces versus bodies is widely distributed within OTC, and not restricted to the face and body areas. A test of the reduced face-body model in the four object-selective ROIs (IFGs, ILOC, rFGs, rLOC) revealed a strong correlation ($r > 0.2$, $p < 0.001$) in each ROI. Overall, the face-body distinction trumps aspect ratio when it comes to explaining neural dissimilarity in OTC.

We followed the same approach for category-selective regions (Fig. 4) and the result showed significant correlations between neural dissimilarity and the animacy model in most ROIs (Fig. 5, one-sided one-sample t test, FDR-corrected across 26 ROIs; * $t > 4.05$, $p < 0.01$; ** $t > 5.31$, $p < 0.001$), while no ROI had a significant correlation with aspect ratio. The direct comparisons of correlations with the two models also showed significant differences in most ROIs (Fig. 5, two-sided paired t test, FDR-corrected across 26 ROIs; * $t > 4.32$, $p < 0.01$; ** $t > 5.22$, $p < 0.001$).

The total absence of predictive power of the aspect ratio model is surprising and was scrutinized further. We hypothesized that it might exist only for the preferred category of an ROI. For body-, face-, and object-selective ROIs, we examined the correlation between neural dissimilarity and the aspect ratio model limiting stimuli to their favorite/selected category. The result of this analysis showed an effect of aspect ratio that was restricted to the right rFGs-object and rLOC-object (Fig. 6, one-sided one-sample t test, FDR-corrected across four ROIs, rFGs-object: $t_{(14)} = 2.97$, $p = 0.0212$, right OTC: $t_{(14)} = 3.18$, $p = 0.0212$). Looking more closely at objects (manmade and natural) in the MDS plot for the right OTC (Fig. 2B), we could also see a modest degree of arrangement in spiky and stubby stimuli. Thus, in accordance with previous work on shape representations with

artificial shapes (e.g., Op de Beeck et al., 2008a), the aspect ratio of inanimate objects is represented in the object-selective cortex. However, as the main results above show, aspect ratio is not an encompassing dimension when we also include animate stimuli (faces & bodies), and is thus unlikely to be an explanation for the location of face- and body-selective regions. The computations of the noise ceiling provide an estimate based on empirical data. This explains why this estimate can be slightly <0 , particularly in analyses focusing on stimuli that do not differ in animacy or face versus body.

We also considered that the relative role of the aspect ratio might shift along the ventral processing pathway. To explore this possibility, we investigated the similarity of neural and model dissimilarity in small consecutive ROIs along the ventral visual cortex (Fig. 7). As we move from posterior to anterior ROIs in both hemispheres, correlations between neural and animacy RDMs increased to significant levels and then decreased (Fig. 7, one-sided one-sample t test, FDR-corrected across 74 ROIs, thin line: $t_{(14)} > 3.78$, $p < 0.01$, thick line: $t_{(14)} > 5.19$, $p < 0.001$). In contrast, we did not observe significant correlations between neural and aspect ratio RDMs. Comparing the correlations with the two models also showed a significance pattern like that of correlations between neural and animacy RDMs.

In sum, these results suggest that object space in the whole OTC and most of OTC's category-selective ROIs was much better explained by the animacy rather than the aspect ratio model and restricting stimuli to selected categories does not change this correlation pattern. Furthermore, this animacy content increases and then decreases as we move along the anatomic posterior-to-anterior axis in VOTC. All analyses confirm a major role for animacy as a primary dimension that determines the functional organization of object space. In contrast, the aspect ratio does not function as a fundamental dimension characterizing the full extent of object space and its anatomic organization.

Animacy distinction generalizes over aspect ratio, but not vice versa

In order to further investigate how each dimension is represented relative to the other, we examined the capability of neural representations to generalize distinctions along one dimension over the other dimension. With this aim, we compared within- and cross-category classification accuracies for animate versus inanimate and spiky versus stubby dimensions. For animate versus inanimate within-aspect-ratio classification, we trained and tested classifiers with either spiky or stubby stimuli; while for cross aspect ratio, classifiers were trained with spiky and tested with stubby stimuli and vice versa (Table 1). For spiky versus stubby within-animacy-classification, we trained and tested classifiers with either animate or inanimate stimuli while for cross animacy, classifiers were trained with animate and tested with inanimate stimuli and vice versa (Table 2).

As Figure 8 shows most category-selective ROIs had classification accuracies significantly above chance level (classification accuracies minus chance level were significantly >0) for animate

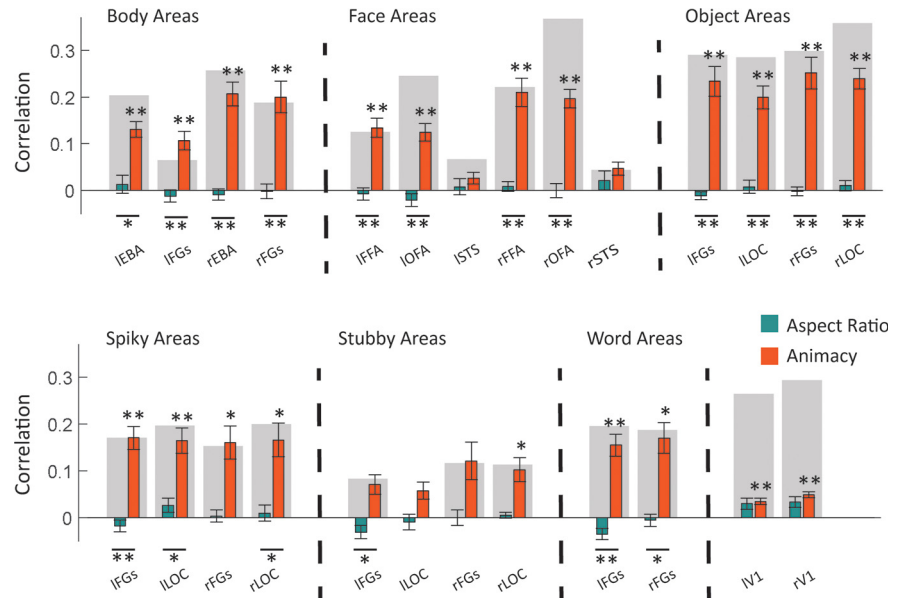


Figure 5. Effect of aspect ratio and animacy in specific category-selective ROIs. Bar plots represent the mean Spearman's correlations between neural RDMs for individual subjects and model RDMs. Error bars indicate SEM. Gray background bars represent the noise ceiling. $**p < 0.001$; $*p < 0.01$; one-sided one-sample or two-sided paired t test, FDR-corrected.

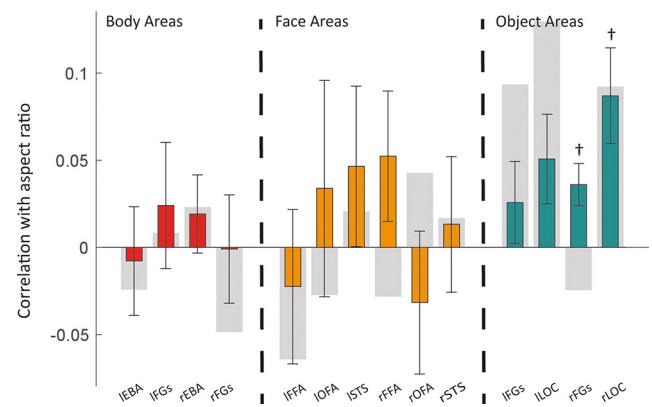


Figure 6. Effect of aspect ratio for the preferred stimulus category in body-, face-, and object-selective ROIs. Bar plot represents the mean Spearman's correlations between neural RDMs for individual subjects and the aspect ratio model RDM. Error bars indicate SEM. Gray background bars represent the noise ceiling. $†p < 0.05$ (one-sided one-sample t test, FDR-corrected).

versus inanimate-within and cross aspect ratio (Fig. 8, classification accuracies minus chance level, one-sided one-sample t test, FDR-corrected across 26 ROIs; $*t > 3.9$, $p < 0.01$; $**t > 4.98$, $p < 0.001$). In contrast, only a few ROIs had classification accuracies significantly above the chance level for spiky versus stubby-within animacy, and they were not able to generalize the classification across animacy (Fig. 9, classification accuracies minus chance level, one-sided one-sample t test, FDR-corrected across 26 ROIs; $*t > 3.9$, $p < 0.01$).

As could already be expected from the initial representational similarity findings, accuracies were larger for animate versus inanimate classification than spiky versus stubby in most OTC regions. The invariance shown in the current section in addition substantiates the claim that this salient representation of animacy stands orthogonal to aspect ratio in a strong sense of the word: It allows the determination of animacy independent from large changes in aspect ratio.

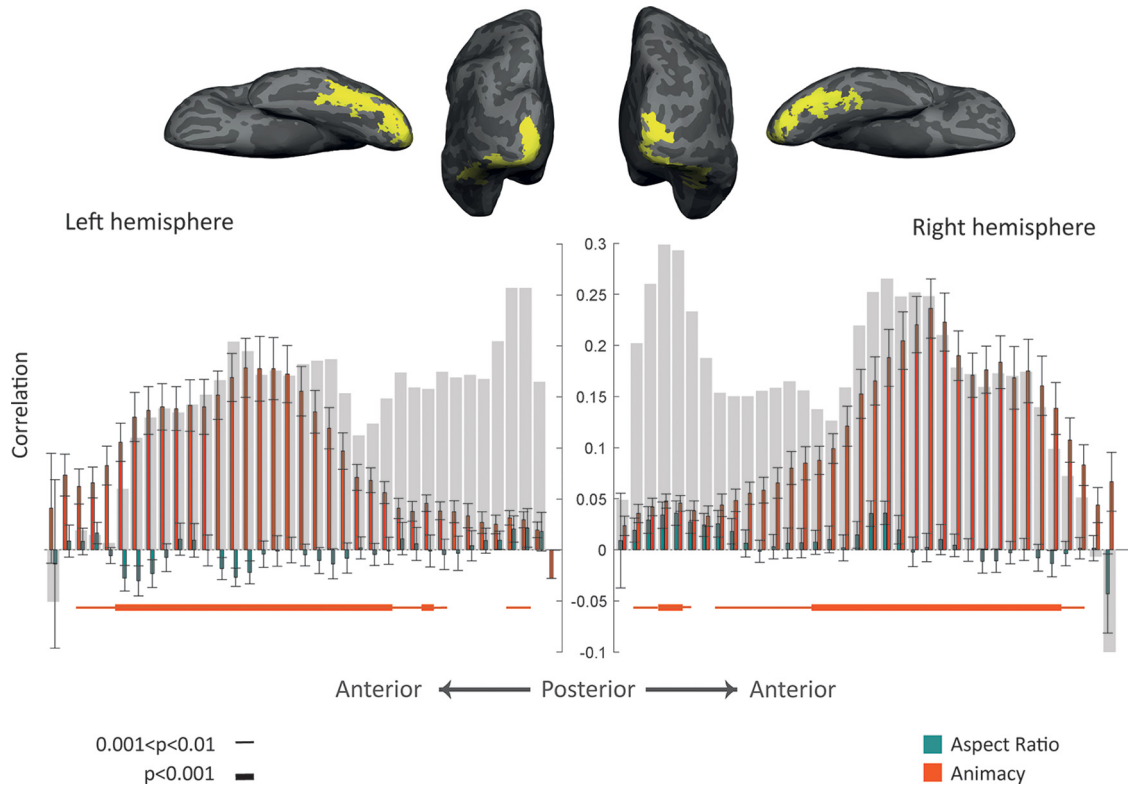


Figure 7. Gradual progression of effects of animacy and aspect ratio along the posterior-anterior gradient in ventral visual cortex. Top, Inflated brain surface from a representative participant showing the ventral visual cortex (union of small consecutive ROIs) in posterior and ventral views. Bottom, Comparing the model RDMs with the neural RDM of ROIs along the ventral visual cortex. Bar plot represents the mean Spearman's correlations between neural RDMs for individual subjects and model RDMs (for the left hemisphere, the bar plot is mirrored to have a consistent direction with the anatomic image at the top). Error bars indicate SEM. Gray background bars represent the noise ceiling. Thin line, $p < 0.01$; thick line, $p < 0.001$; one-sided one-sample t test, FDR-corrected.

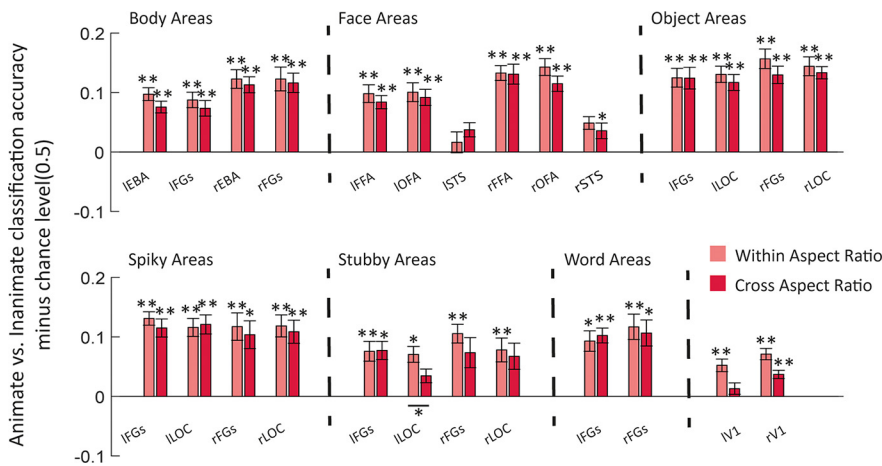


Figure 8. Animate versus inanimate classification in category-selective ROIs. Bar plot represents the mean classification accuracy minus chance level (0.5) for animate versus inanimate classification when aspect ratio is similar (within) or very different (across) in the training and the test images. Error bars indicate SEM. $*p < 0.01$; $**p < 0.001$; one-sided one-sample or two-sided paired t test, FDR-corrected.

Univariate responses reflect differences in animacy and not the aspect ratio

Category-selective regions were recognized and defined based on univariate responses, and there are discussions over the dependency of these selectivities on different features. In the literature at large, it has been suggested that face-selective cortex has a preference for curved and concentric shapes (Kosslyn et al., 1995; Wilkinson et al., 2000; Tsao et al., 2006; Yue et al., 2014) and

body-selective cortex has a preference for shapes with a comparable physical form to bodies (Popivanov et al., 2012, 2014). More specifically based on Bao et al. (2020), one would expect that the face-selective cortex would prefer stubby stimuli, and the body-selective cortex would have a stronger response to spiky stimuli. With our stimulus design, we can test whether this hypothesis holds while stimulus class (e.g., face and body) and aspect ratio are properly dissociated. To evaluate the univariate effect of animacy and aspect ratio, we obtained the mean response of each category-selective ROI to each of the animacy-aspect ratio conditions (animate-spiky, animate-stubby, inanimate-spiky, and inanimate-stubby). The result (Fig. 10) clearly shows that most category-selective ROIs had larger responses for animates than inanimates regardless of aspect ratio. However, comparing responses for animate-spiky with animate-stubby and responses for inanimate-spiky with inanimate-stubby, we found no significant differences in any of the category-selective ROIs, such as face- and body-selective ROIs. This refutes the idea that somehow the category selectivity of these regions is related to aspect ratio and is a difficult finding for arguments that the region would have developed in that cortical location because of the existence of an aspect ratio map early in development.

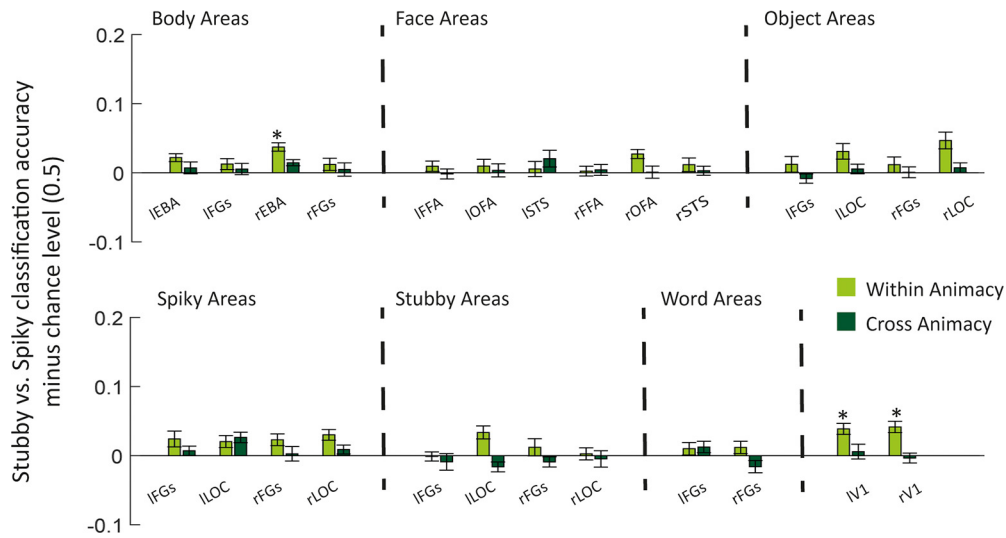


Figure 9. Spiky versus stubby classification in category-selective ROIs. Bar plot represents the mean classification accuracy minus chance level (0.5) for spiky versus stubby classification when animacy is similar (within) or very different (across) in the training and the test images. Error bars indicate SEM. * $p < 0.01$ (one-sided one-sample or two-sided paired t test, FDR-corrected).

Nevertheless, there are some small regions in the occipital and occipitotemporal cortex with a replicable preference for spiky or stubby objects. Several of our ROIs were selected to have such preference based on the localizer data, and some of these preferences were partially replicated with our experimental stimuli. Specifically, ILOC-spiky showed a higher response for animate-spiky than for animate-stubby (Fig. 10, two-sided paired t test, FDR-corrected, $t_{(14)} = 6.55$, * $p < 0.01$), and IFGs-stubby for inanimate-stubby! compared with inanimate-spiky (Fig. 10, two-sided paired t test, FDR-corrected across 52 tests, $t_{(14)} = 6.54$, † $p < 0.05$). Nevertheless, this is only a partial replication of the selectivity seen in the localizer data, which already hints at how weak this selectivity is. Furthermore, while aspect ratio tuning exists in some small subregions of the ventral visual pathway, it falls short from explaining the large-scale organization of object representations; and, in particular, it fails to capture the selectivity of face- and body-selective regions.

Animacy, and not the aspect ratio model, explains the latent space of BigBiGAN

BigBiGAN has shown great success in generating high-quality and visually plausible images. To test whether animacy and aspect ratio had key roles in determining object identity for BigBiGAN, we obtained the mapping of our stimuli to its latent space and applied PCA. Figure 11A shows the plot of the first two PCs computed for latent vectors. The separation of animate from inanimate stimuli is clearly seen, but no similar arrangement based on aspect ratio is perceivable. RSA quantifies this observation; the BigBiGAN RDM (see Materials and Methods) was significantly correlated with the animacy model ($r = 0.6096$, one-sided permutation test $p < 0.001$) and not the aspect ratio model ($r = 0.003$, one-sided permutation test $p = 0.418$). Using the

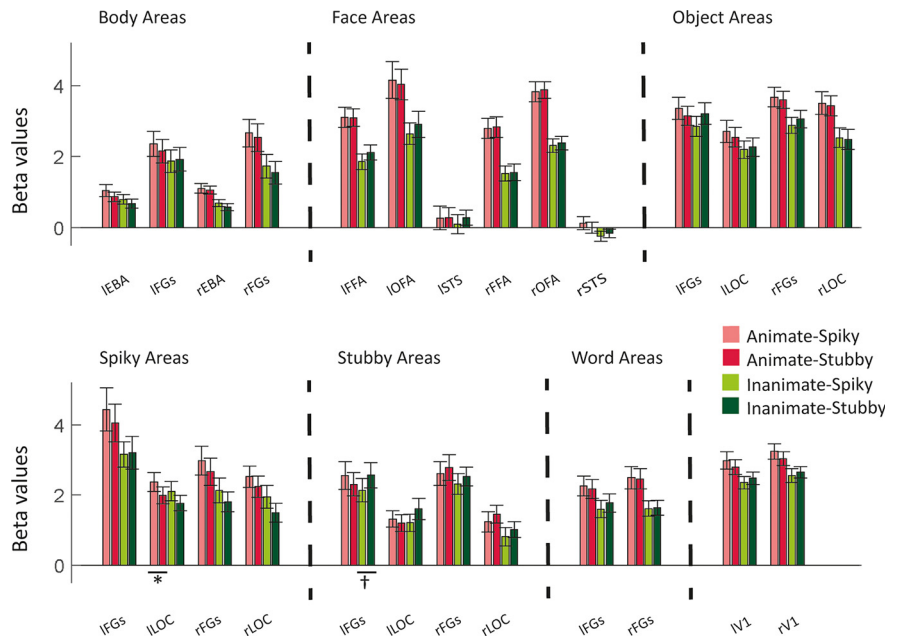


Figure 10. Univariate responses of category-selective ROIs to combinations of animacy and aspect ratio. Bar plots represent mean β values for animate-spiky, animate-stubby, inanimate-spiky, and inanimate-stubby. Error bars indicate SEM. * $p < 0.01$; † $p < 0.05$; two-sided paired t test, FDR-corrected.

full pattern of representational similarity (without the data reduction step of PCA), RSA resulted in significant correlations between BigBiGAN representational similarity and both the animacy model ($r = 0.3658$, one-sided permutation test $p < 0.001$) or aspect ratio model ($r = 0.1019$, one-sided permutation test $p < 0.01$), but the correlation with the animacy model was almost 3.5 times larger than the correlation with the aspect ratio model. These data showed that animacy was the dominant factor in shaping the latent space of BigBiGAN.

We considered the possibility that we fail to find a convincing representation of aspect ratio because our stimulus set would include a too restricted range of aspect ratios. Aspect ratio ranges from ~ 1 to ~ 11 in our stimulus set, while the range was higher in the set of Bao et al. (2020), ~ 1 to ~ 15 , mostly because of a few outlier shapes with high aspect ratios. Therefore, to more

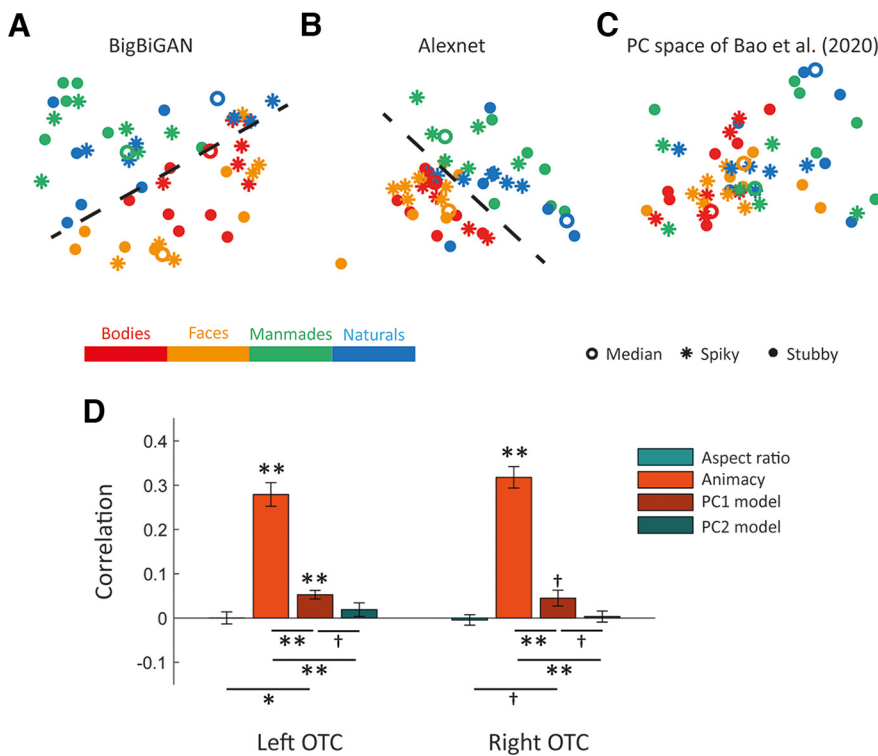


Figure 11. Two-dimensional representations of our stimuli in different spaces and RSA. Top panels, Points are color-coded based on stimulus category and are dots, rings, or asterisks based on aspect ratio. Dashed line indicates the separation of animate from inanimate. **A**, The first two PCs computed for the BigBiGAN's latent vectors of our stimulus set. **B**, The first two PCs computed for the Alexnet-fc6 activation of our stimulus set. **C**, Our stimuli mapped to the PC space originally defined by Bao et al. (2020). **D**, Representational similarity and effects of aspect ratio, animacy, and PC1 and PC2 models in the occipito-temporal cortex. Bar plot represents the mean Spearman's correlations between neural RDMs and model RDMs. Error bars indicate SEM. ** $p < 0.001$; * $p < 0.01$; † $p < 0.05$; one-sided one-sample or two-sided paired t test.

severely test the effect of the aspect ratio range, we performed further analyses on how BigBiGAN's representation relates to the range of included aspect ratios. We reduced the range of aspect ratio in the Bao et al. (2020) stimulus set to our range by removing the outliers in their set (here referred to as Bao-restricted). An aspect ratio model was computed for the Bao-restricted as the pairwise absolute difference of aspect ratio for stimuli in this set. We also obtained the mapping of Bao-restricted stimuli to BigBiGAN latent space and applied PCA. The BigBiGAN RDM using the first two PCs was significantly correlated with the Bao-restricted aspect ratio model ($r = 0.3134$, one-sided permutation test $p < 0$). Using the full pattern of representational similarity (without the data reduction step of PCA), RSA resulted again in a significant correlation between BigBiGAN representational similarity and Bao-restricted aspect ratio model ($r = 0.4586$, one-sided permutation test $p < 0$). We conclude that the main experimental design property that determines the occurrence of aspect ratio clustering in BigBiGAN is whether the stimulus set confounds the distinction of faces and bodies with aspect ratio and not the included range of aspect ratios.

To compare our results more directly with Bao et al. (2020), we passed our stimuli through the same network, Alexnet (Krizhevsky et al., 2012), and acquired corresponding values at the fully connected 6 (fc6) layer. Then, for obtaining the main factors, PCA was applied to the fc6 activation and the first two PCs were used for visualization. The separation of animate from inanimate stimuli is clearly seen in Figure 11B, but no similar arrangement based on aspect ratio is perceivable, like Figure 11A

(based on BigBiGAN's latent variables). The Alexnet RDM using the first two PCs was not significantly correlated with the aspect ratio model ($r = 0.0322$, one-sided permutation test $p = 0.7080$). Similarly, using the full pattern of representational similarity (without the data reduction step of PCA), RSA did not result in a significant correlation between Alexnet representational similarity and aspect ratio model ($r = 0.0963$, one-sided permutation test $p = 0.0750$).

One could also speculate that the definition of aspect ratio might be most appropriate as a label for characterizing the PC space originally defined by Bao et al. (2020), not as a semantic label of spikiness and stubbiness, which is only approximated by the quantitative definition of aspect ratio. Following this reasoning, the difference between our results and Bao et al. (2020) might be because of differences in the definition of aspect ratio. To test this hypothesis, we obtained the PC space originally defined by Bao et al. (2020) using their stimuli and their DNN (Alexnet), and mapped our stimuli into this space (Fig. 11C). Then, we derived RDMs based on the value of our stimuli in their first two PCs (PC1 and PC2) and compared them with our animacy and aspect ratio models. The PC1 model was significantly correlated with the animacy model ($r = 0.1619$, one-sided permutation test $p < 0.001$), while other correlations were very small: PC1 model and aspect ratio model ($r = 0.0024$,

one-sided permutation test $p = 0.4390$), PC2 model and animacy model ($r = 0.0506$, one-sided permutation test $p = 0.0390$), and PC2 model and aspect ratio model ($r = -0.0721$, one-sided permutation test $p = 0.0370$). We reran the main analysis with PC1 and PC2 models; and as Figure 11D shows, the animacy model is by far the best model describing responses in OTCs. Therefore, using RDMs derived from the DNN PCs of Bao et al. (2020) does not lead to results similar to Bao et al. (2020), indicating that it is not about the definition of aspect ratio, but the stimulus set, and how it dissociates aspect ratio from face/body might be crucial. Finally, it is striking that all the different implementations of DNN representations in Figure 11 also suggest a lack of clustering of faces apart from bodies. Interestingly, there was a small negative correlation between the RDM matrix of BigBiGAN and the full face-body model ($r = -0.10$, $p = 0.0350$), and a positive correlation for the reduced face-body model restricted to animate stimuli ($r = 0.20$, $p = 0.005$). Taking into account that the full face-body model is correlated negatively with the animacy model, it seems that DNN representations are more dominated by the animacy dimension compared with the face/body distinction than we find in the human fMRI data.

Discussion

We designed a novel stimulus set that dissociates object category (face, body, manmade, and natural) from aspect ratio, and recorded fMRI responses in human OTC and DNN representations for these stimuli to examine two alternative hypotheses: An

object space organized with two main dimensions of animacy and aspect ratio, or an object space with a primary dimension of animacy and a further distinction between faces and bodies that is NOT related to the aspect ratio of these categories. Using fMRI RSA, results show that whole OTC and most of OTC's category-selective ROIs represent object animacy and not aspect ratio. Studying the correlation between neural dissimilarity and the face-body model by analyzing all stimuli or by limiting the stimuli to animates showed a strong effect for face-body compared with no or only a very weak effect of aspect ratio. Restricting RSA based on selectivity for one category reveals that object-selective ROIs in the right hemisphere show a relatively weak representation of the aspect ratio of objects (manmade and natural). The represented animacy content is not constant but increases and then decreases along the anatomic posterior-to-anterior axis in VOTC. This dominant effect of animacy is reflected in larger classification accuracies for animate versus inanimate compared with stubby versus spiky. The animate versus inanimate classifiers also show invariance to the aspect ratio that could express the independent representation of animacy in human OTC. Consistent with MVPA results, univariate responses of category-selective ROIs show an obvious effect of animacy, but they are not influenced by aspect ratio. Similar to human OTC, BigBiGAN (the most advanced DNN in capturing object properties) represents animacy, as has been observed previously (Khaligh-Razavi and Kriegeskorte, 2014; Jozwik et al., 2017; Bracci et al., 2019; Zeman et al., 2020), but not aspect ratio. Finally, results of data-driven approaches clearly show clusters for face and body stimuli and separation of animate from inanimate stimuli in representational space of OTC and BigBiGAN, but there is not any arrangement related to aspect ratio. In conclusion, these findings rule out the 2D animacy \times aspect ratio model for object space in human OTC while an object space with a primary dimension of animacy and a further distinction between faces and bodies could clearly explain the results.

We find that the organization of object space in human OTC is not related to aspect ratio. This is in contrast to the proposed model by Bao et al. (2020), which considers stubby-spiky as one of the two main dimensions of object space in the IT cortex. There are two major differences between the present study and Bao et al. (2020) that could explain the contrasting results: stimulus sets and subject species. In the stimulus set of Bao et al. (2020), animate-inanimate and stubby-spiky were dissociated overall, but it was not true between stimulus categories, particularly for faces and bodies: Face stimuli were all stubby and body stimuli were mostly spiky. Therefore, there was a major confound between aspect ratio and stimulus category. To overcome this problem, we had a comparatively wide range of aspect ratios for different categories and explicitly dissociated aspect ratio from these two categories. Our result showed no significant representation of aspect ratio in OTC, suggesting that much of the evidence for aspect ratio as an overall dimension and as a dimension underlying face and body selectivity and the location of selectivity for these categories were because of this major confound between aspect ratio and stimulus category (face vs body). An effect of aspect ratio was absent or very weak, depending on the exact stimuli that were included in the analyses (e.g., only body stimuli; only inanimate stimuli; ...). This suggests that a possible bias in terms of the selection of stimuli does not impact our conclusions to a large degree. The main point of attention is the degree to which aspect ratio is properly dissociated from the

face/body distinction, which is the most salient difference between our stimulus set and the set of Bao et al. (2020).

The other major difference is species; results in Bao et al. (2020) were mostly based on electrophysiology data from monkeys while we investigated fMRI data from human subjects. Aspect ratio might be more important for object representations in monkeys. This suggestion is supported by earlier work, although it was never investigated specifically. In particular, Kriegeskorte et al. (2008) compared monkey electrophysiology data and human fMRI for the same set of stimuli and reported a common organization of object space in humans and monkeys. However, when looking more closely at their stimulus arrangement (Fig. 2A) (Kriegeskorte et al., 2008), we can recognize an aspect ratio effect in monkey (spiky stimuli are closer to each other) that is not present in human (see also Hong et al., 2016). However, Coggan and Tong (2021) reported in a conference abstract to partially replicate the findings of Bao et al. (2020) in humans, using the same stimulus set as Bao et al. (2020). For that reason, we doubt that species is the main explanation for the divergence between our findings and Bao et al. (2020).

Despite the absence of an overall dimension of aspect ratio, our findings are still in line with earlier reports of shape selectivity and even selectivity for aspect ratio. Previous studies on shape representation provided evidence of a relationship between perceived shape similarity and neural representation in object-selective ROIs, investigating both single-unit recordings in monkeys (Op de Beeck et al., 2001) and fMRI data in humans (Op de Beeck et al., 2008b; Drucker and Aguirre, 2009; Peelen and Caramazza, 2012; Chen et al., 2018). Recent work has illustrated that many tens of dimensions contribute to the perception of the visual form (Morgenstern et al., 2021), and several of these dimensions are correlated with aspect ratio (which in the specialized literature is referred to as “compactness” or “circularity”). A previous study that dissociated animacy from shape also suggested a role for aspect ratio in the representation of shape (Bracci and Op de Beeck, 2016; Chen et al., 2018). Nevertheless, we show that the face and body selectivity take priority when we properly dissociate these categories from aspect ratio. Tuning for aspect ratio is restricted to part of OTC, generally object-selective regions and a few small spots with univariate preference, and mostly restricted to inanimate objects. While aspect ratio is one of many dimensions by which object shape is represented, it does not have a special status for explaining the large-scale object space and how it is organized neuroanatomically in selective patches (e.g., faces and bodies).

In face and body regions, we find no evidence for tuning for aspect ratio. There is a hypothesis that category selectivity for faces and bodies is partly explained by the aspect ratio of stimuli (Kosslyn et al., 1995; Wilkinson et al., 2000; Tsao et al., 2006; Popivanov et al., 2012, 2014; Yue et al., 2014). Following this hypothesis, stronger responses for stubby stimuli in face-selective and spiky stimuli in body-selective regions are anticipated. Our stimulus set, including faces and bodies with wide ranges of aspect ratios, provides a touchstone for testing this hypothesis. We compared responses for animate-spiky with animate-stubby and responses for inanimate-spiky with inanimate-stubby and found no significant differences in any of the category-selective ROIs, such as face- and body-selective ROIs. Probably, other stimulus features are more prominent when it comes to representing faces and bodies.

While our findings put doubt on the role of aspect ratio, we confirm a strong selectivity for animacy. We found that animacy is a determining factor in organizing object space in both human

OTC and BigBiGAN, consistent with Bao et al. (2020) and earlier reports (Kriegeskorte et al., 2008; Bracci and Op de Beeck, 2016). This representation of animacy was prominent for a long distance along the anatomic posterior-to-anterior axis in VOTC. Following a cross-decoding approach, we found that animacy is represented independently of aspect ratio. Previous studies have disclosed that animacy and category distinctions are correlated with low/high-level visual features (Cohen et al., 2014; Rice et al., 2014; Jozwik et al., 2016; Zachariou et al., 2018), but there is enough evidence that animacy and category structure remain even when dissociated from shape or texture features (Bracci and Op de Beeck, 2016; Kaiser et al., 2016; Proklova et al., 2016; Bracci et al., 2017; Chen et al., 2018). The strength of animacy selectivity and lack of aspect ratio tuning were consistent across smaller ROIs within OTC. This indicates a high level of analogy in multivariate patterns across category-selective regions, as previously reported (Cohen et al., 2017) while they studied multivariate responses of category-selective regions in OTC to different categories (bodies, buildings, cars, cats, chairs, faces, hammers, and phones).

The occipitotemporal cortex has a key role in visual object recognition, but the organization of object space in this region is still unclear. To examine hypotheses considering animacy, aspect ratio, and face-body as principal dimensions characterizing object space in the occipitotemporal cortex, we devised a novel stimulus set that dissociates these dimensions. Investigation of human fMRI and DNN responses to this stimulus set shows that a 2D animacy \times aspect ratio model cannot explain object representations in either occipitotemporal cortex or a state-of-the-art DNN while a model in terms of an animacy dimension combined with strong selectivity for faces and bodies is more compatible with both neural and DNN representations.

References

- Bao P, She L, McGill M, Tsao DY (2020) A map of object space in primate inferotemporal cortex. *Nature* 583:103–108.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300.
- Bracci S, Op de Beeck H (2016) Dissociations and associations between shape and category representations in the two visual pathways. *J Neurosci* 36:432–444.
- Bracci S, Ritchie JB, Op de Beeck H (2017) On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia* 105:153–164.
- Bracci S, Ritchie JB, Kalfas I, Op de Beeck HP (2019) The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. *J Neurosci* 39:6513–6525.
- Brainard DH, Vision S (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
- Bryan PB, Julian JB, Epstein RA (2016) Rectilinear edge selectivity is insufficient to explain the category selectivity of the parahippocampal place area. *Frontiers in Human Neuroscience* 10:137.
- Carlson TA, Simmons RA, Kriegeskorte N, Slevc LR (2014) The emergence of semantic meaning in the ventral temporal pathway. *Journal of Cognitive Neuroscience* 26:120–131.
- Chen J, Snow JC, Culham JC, Goodale MA (2018) What role does ‘elongation’ play in ‘tool-specific’ activation and connectivity in the dorsal and ventral visual streams? *Cereb Cortex* 28:1117–1131.
- Coggan D, Tong F (2021) Maps of object animacy and aspect ratio in human high-level visual cortex. *J Vis* 21:2811.
- Coggan DD, Liu W, Baker DH, Andrews TJ (2016) Category-selective patterns of neural response in the ventral visual pathway in the absence of categorical information. *Neuroimage* 135:107–114.
- Cohen MA, Konkle T, Rhee JY, Nakayama K, Alvarez GA (2014) Processing multiple visual objects is limited by overlap in neural channels. *Proc Natl Acad Sci USA* 111:8955–8960.
- Cohen MA, Alvarez GA, Nakayama K, Konkle T (2017) Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *J Neurophysiol* 117:388–402.
- Donahue J, Simonyan K (2019) Large scale adversarial representation learning. *Adv Neural Information Process Syst* 32. Vancouver, Canada.
- Drucker DM, Aguirre GK (2009) Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cereb Cortex* 19:2269–2280.
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25:1325–1335.
- Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, Kent JD, Goncalves M, DuPre E, Snyder M, Oya H, Ghosh SS, Wright J, Durnez J, Poldrack RA, Gorgolewski KJ (2019) fMRIPrep: a robust pre-processing pipeline for functional MRI. *Nat Methods* 16:111–116.
- Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci* 15:536–548.
- Hong H, Yamins DL, Majaj NJ, DiCarlo JJ (2016) Explicit information for category-orthogonal object properties increases along the ventral stream. *Nat Neurosci* 19:613–622.
- Jozwik KM, Kriegeskorte N, Mur M (2016) Visual features as stepping stones toward semantics: explaining object similarity in IT and perception with non-negative least squares. *Neuropsychologia* 83:201–226.
- Jozwik KM, Kriegeskorte N, Storrs KR, Mur M (2017) Deep convolutional neural networks outperform feature-based but not categorical models in explaining object similarity judgments. *Front Psychol* 8:1726.
- Julian JB, Fedorenko E, Webster J, Kanwisher N (2012) An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* 60:2357–2364.
- Kaiser D, Azzalini DC, Peelen MV (2016) Shape-independent object category responses revealed by MEG and fMRI decoding. *J Neurophysiol* 115:2246–2250.
- Kalfas I, Kumar S, Vogels R (2017) Shape selectivity of middle superior temporal sulcus body patch neurons. *eNeuro* 4:ENEURO.0113-17.2017.
- Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput Biol* 10:e1003915.
- Konkle T, Oliva A (2012) A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74:1114–1124.
- Kosslyn SM, Hamilton SE, Bernstein JH (1995) The perception of curvature can be selectively disrupted in prosopagnosia. *Brain Cogn* 27:36–58.
- Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60:1126–1141.
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *Adv Neural Information Process Syst* 25:1097–1105.
- 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* 115:E9015–E9024.
- Morgenstern Y, Hartmann F, Schmidt F, Tiedemann H, Prokott E, Maiello G, Fleming RW (2021) An image-computable model of human visual shape similarity. *PLoS Comput Biol* 17:e1008981.
- Mozafari M, Reddy L, VanRullen R (2020) Reconstructing natural scenes from fMRI patterns using BigBiGAN. *2020 International Joint Conference on Neural Networks (IJCNN)*, pp 1–8. Glasgow, United Kingdom.
- Nasr S, Echavarria CE, Tootell RB (2014) Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. *J Neurosci* 34:6721–6735.
- Oosterhof NN, Connolly AC, Haxby JV (2016) CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Front Neuroinform* 10:27.
- Op de Beeck H, Bracci S (2022) Understanding human object vision: a picture is worth a thousand representations. *Annu Rev Psychol* 74:113–135.
- Op de Beeck H, Wagemans J, Vogels R (2001) Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat Neurosci* 4:1244–1252.
- Op de Beeck HP, Deutsch JA, Vanduffel W, Kanwisher NG, DiCarlo JJ (2008a) A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. *Cereb Cortex* 18:1676–1694.

- Op de Beeck HP, Torfs K, Wagemans J (2008b) Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *J Neurosci* 28:10111–10123.
- Peelen MV, Caramazza A (2012) Conceptual object representations in human anterior temporal cortex. *J Neurosci* 32:15728–15736.
- Peelen MV, Downing PE (2017) Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia* 105:177–183.
- Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2012) Stimulus representations in body-selective regions of the macaque cortex assessed with event-related fMRI. *Neuroimage* 63:723–741.
- Popivanov ID, Jastorff J, Vanduffel W, Vogels R (2014) Heterogeneous single-unit selectivity in an fMRI-defined body-selective patch. *J Neurosci* 34:95–111.
- Proklova D, Kaiser D, Peelen MV (2016) Disentangling representations of object shape and object category in human visual cortex: the animate-inanimate distinction. *J Cogn Neurosci* 28:680–692.
- Rice GE, Watson DM, Hartley T, Andrews TJ (2014) Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. *J Neurosci* 34:8837–8844.
- Ritchie JB, Op de Beeck H (2019) A varying role for abstraction in models of category learning constructed from neural representations in early visual cortex. *J Cogn Neurosci* 31:155–173.
- Ritchie JB, Zeman AA, Bosmans J, Sun S, Verhaegen K, Op de Beeck HP (2021) Untangling the animacy organization of occipitotemporal cortex. *J Neurosci* 41:7103–7119.
- Serre T, Oliva A, Poggio T (2007) A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci USA* 104:6424–6429.
- Spriet C, Abassi E, Hochmann JR, Papeo L (2022) Visual object categorization in infancy. *Proc Natl Acad Sci USA* 119:e2105866119.
- Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311:670–674.
- Walther A, Nili H, Ejaz N, Alink A, Kriegeskorte N, Diedrichsen J (2016) Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* 137:188–200.
- Wilkinson F, James TW, Wilson HR, Gati JS, Menon RS, Goodale MA (2000) An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Curr Biol* 10:1455–1458.
- Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW (2010) Controlling low-level image properties: the SHINE toolbox. *Behav Res Methods* 42:671–684.
- Yue X, Pourladian IS, Tootell RB, Ungerleider LG (2014) Curvature-processing network in macaque visual cortex. *Proc Natl Acad Sci USA* 111: E3467–E3475.
- Zachariou V, Del Giacco AC, Ungerleider LG, Yue X (2018) Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization. *J Vis* 18:3.
- Zeman AA, Ritchie JB, Bracci S, Op de Beeck H (2020) Orthogonal representations of object shape and category in deep convolutional neural networks and human visual cortex. *Sci Rep* 10:1–12.