Behavioral/Cognitive

Untangling the Animacy Organization of Occipitotemporal Cortex

J. Brendan Ritchie,[1] Astrid A. Zeman,[1] Joyce Bosmans,[2] Shuo Sun,[1] Kirsten Verhaegen,[1] and Hans P. Op de Beeck[1]

[1]Laboratory of Biological Psychology, Department of Brain and Cognition, Leuven Brain Institute, Katholieke Universiteit Leuven, 3000 Leuven, Belgium, and [2]Faculty of Medicine and Health Sciences, University of Antwerp, 2000 Antwerp, Belgium

Some of the most impressive functional specializations in the human brain are found in the occipitotemporal cortex (OTC), where several areas exhibit selectivity for a small number of visual categories, such as faces and bodies, and spatially cluster based on stimulus animacy. Previous studies suggest this animacy organization reflects the representation of an intuitive taxonomic hierarchy, distinct from the presence of face- and body-selective areas in OTC. Using human functional magnetic resonance imaging, we investigated the independent contribution of these two factors—the face-body division and taxonomic hierarchy—in accounting for the animacy organization of OTC and whether they might also be reflected in the architecture of several deep neural networks that have not been explicitly trained to differentiate taxonomic relations. We found that graded visual selectivity, based on animal resemblance to human faces and bodies, masquerades as an apparent animacy continuum, which suggests that taxonomy is not a separate factor underlying the organization of the ventral visual pathway.

Key words: **animacy; bodies; category selectivity; deep learning; faces; occipitotemporal cortex**

Significance Statement

Portions of the visual cortex are specialized to determine whether types of objects are animate in the sense of being capable of self-movement. Two factors have been proposed as accounting for this animacy organization: representations of faces and bodies and an intuitive taxonomic continuum of humans and animals. We performed an experiment to assess the independent contribution of both of these factors. We found that graded visual representations, based on animal resemblance to human faces and bodies, masquerade as an apparent animacy continuum, suggesting that taxonomy is not a separate factor underlying the organization of areas in the visual cortex.

Introduction

One of the most fascinating examples of functional specialization in the human brain is the presence of areas in the lateral and ventral occipitotemporal cortex (OTC) that preferentially respond to a small number of ecologically important visual categories. These areas tend to spatially cluster in a manner that respects the superordinate dichotomy between animate objects that are capable of volitional self-movement and inanimate objects that are not (Behrmann and Plaut, 2013; Grill-Spector and Weiner, 2014; Bao et al., 2020). In particular, face and body areas are known to cluster separately from those for scenes and tools, and this functional organization has been taken to show that at a broader spatial scale the OTC represents the animate-inanimate division (Kriegeskorte et al., 2008b; Grill-Spector and Weiner, 2014). Recent studies suggest the OTC also represents stimulus animacy in a continuous, or even hierarchical, fashion (Sha et al., 2015; Thorat et al., 2019). In these studies stimuli consist of animal images groups based on an intuitive taxonomy in which some animals rank high on the animacy scale (e.g., primates), some are intermediary (e.g., birds), and others (e.g., insects) are low (Connolly et al., 2012, 2016; Sha et al., 2015; Nastase et al., 2017). These results introduce the possibility that the OTC represents conceptual relations among categories that do not easily reduce to their diagnostic visual properties (Bracci et al., 2017; cf. Fairhall and Caramazza, 2013).

The familiar face-body division and proposed intuitive taxonomy may both be factors that help explain animacy organization in the OTC. These factors are not mutually exclusive (nor are they exhaustive), but they are not the same. Images of the face

Received Oct. 11, 2020; revised Apr. 20, 2021; accepted May 20, 2021.

Author contributions: J.B.R. and H.P.O.d.B. designed research; J.B.R., A.A.Z., J.B., S.S., and K.V. performed research; J.B.R., A.A.Z., J.B., S.S., and K.V. analyzed data; J.B.R. wrote the paper.

J.B.R. was supported by the Fonds Wetenschappelijk Onderzoek (FWO) and the European Union Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 665501, via an FWO [PEGASUS]^2 Marie Skłodowska-Curie Fellowship (12T9217N). A.A.Z. and H.P.O.d.B. were supported by Grant C14/16/031 of the Katholieke Universiteit Leuven Research Council. Neuroimaging was funded by the Flemish Government Hercules Grant ZW11_10. This project (G0E8718N, Human Visual Categorization) has received funding from the FWO and Fonds de la Recherche Scientifique under the Excellence of Science program.

The authors declare no competing financial interests.

Correspondence should be addressed to J. Brendan Ritchie at j.brendan.w.ritchie@gmail.com.

https://doi.org/10.1523/JNEUROSCI.2628-20.2021

and (face cropped) body of a person are at the same (high) taxonomic level, yet they are distinct, resulting in clearly dissociable neural responses. To date, studies providing evidence of intuitive taxonomic organization in the OTC have not factored the face/body division into their study design and so have failed to disentangle these two factors in several ways. First, these studies have used images of whole animal bodies and ignored the face-body division in general and more specific issues such as the fact that we may be more accustomed to looking at the faces of some animals and the bodies of others. Second, these studies have focused on large swaths of the OTC and so are unable to determine whether the taxonomic organization is exhibited more narrowly in category-selective areas, like those for faces and bodies. Third, these studies equate the idea of a continuous, graded organization in the OTC with the representation of a taxonomic hierarchy. Therefore, they do not allow for the possibility that apparent animacy continuum may simply be coding for diagnostic visual features of objects, which is an important factor in the category-selective organization of the OTC more generally (Jozwik et al., 2016; Bracci et al., 2017). In particular, as face and body areas are already known to be preferentially selective to images of humans, the claimed animacy continuum may simply reflect the relative visual similarity of animal faces and bodies to those of humans. Finally, these studies do not rule out the possibility that a system without any consideration of intuitive taxonomy might nonetheless show a similar continuity in its representation of animal stimuli to the OTC.

In light of these shortcomings, we performed a functional magnetic resonance imaging (fMRI) experiment to disentangle these two factors and their contributions to explaining the animacy organization of the OTC. First, we designed a stimulus set that allowed us to evaluate whether the taxonomic hierarchy explains the relationship among activity patterns in the OTC when controlling for the face-body division and vice versa. Second, we investigated the influence of both factors in explaining the relationship among activity patterns in more circumscribed, category-selective areas of the OTC. Third, we investigated whether the relative visual similarity of animal faces and bodies, and in particular, their similarity to the faces and bodies of humans, might better explain the relationship among activity patterns than taxonomy. Finally, we assessed whether an intuitive taxonomic organization might also be reflected in the patterns of activation weights of layers of multiple deep neural networks (DNNs), although they have not been trained on taxonomic relations between animal classes.

Materials and Methods

Participants

The fMRI experiment included 15 adult volunteers (10 women; mean age, 24.2 years; age range, 21–33 years). A total of 40 volunteers participated in the different similarity judgment tasks with subjects randomly selected to participate in the pairwise body ($N = 10$; seven women; mean age, 21.8 years; age range, 19–26 years), pairwise face ($N = 10$; seven women; mean age, 22.6 years; age range, 19–28 years), human body ($N = 10$; 10 women; mean age, 19.1 years; age range, 18–23 years), and human face ($N = 10$; seven women; mean age, 21.1 years; age range, 18–32 years) similarity tasks. All volunteers were predominantly right-handed, had normal or corrected vision, and provided written informed consent for participation in the experiments. All experiments were approved by the Ethics Committee of Universitair Ziekenhuis/Katholieke Universiteit Leuven, and all methods were performed in accordance with the relevant guidelines and regulations.

Stimuli

Stimuli consisted of 54 natural images of objects (Fig. 1*A*) and included a body and face image of 24 animals, as well as 2 images each of 3 natural objects, resulting in 48 animal and 6 natural object images. The animals depicted were selected to cluster into six levels of an intuitive taxonomic hierarchy based on stimulus design and results of previous studies on the intuitive taxonomic organization of the OTC (Sha et al., 2015; Nastase et al., 2017): mammal cluster 1 and mammal cluster 2 (birds, reptiles/amphibians, fish, and exoskeletal invertebrates). We refer to this as an intuitive taxonomic hierarchy because of the following: (1) it reflects commonsense divisions of types of animals based directly on behavioral and neural findings from previous studies, and (2) it involves an ordering where mammals are on one end and invertebrates on the other. This notion of taxonomy tends to group biologically distinct classes. For example, using similar divisions, for convenience Connolly et al. (2016) refer to the grouping of reptiles and amphibians as simply "reptiles" and the group of exoskeletal invertebrates as "bugs." The class of mammal is quite broad, and previous studies also make a point of distinguishing primates as a separate class (Sha et al., 2015). Here, the distinction between the two mammal clusters was intended to distinguish high intelligence mammals, including primates (gorilla and lemur) and trainable aquatically mobile mammals (dolphin and seal), from comparatively less intelligent mammals that are terrestrial (leopard and kangaroo) or capable of flight (flying fox and bat). We note that as the taxonomy structure we employ here is based directly on previous studies on the taxonomic organization of the OTC, it may differ from other ways of measuring the rich conceptual connections between different animal classes, for example, verbal fluency (Goñi et al., 2011). Natural object images were of orchids, fruit/vegetables, and mushrooms (two images each). All images were cropped to 700 × 700 pixels (subtending ~10 degrees of visual angle in the scanner), converted to grayscale, focus blurred in the background regions, and then filtered using the SHINE toolbox (Willenbockel et al., 2010) to equate the luminance histogram and the average energy at each spatial frequency. The full-sized stimulus images are available at https://osf.io/xcpw6/.

Experimental design and statistical analyses

Similarity judgment experiments. Participants were tasked with making visual similarity judgments based on the sequential presentation of pairs of either face or body stimulus images. For the pairwise face and pairwise body similarity tasks, subjects responded using a 6-point scale (1 = highly similar, 6 = highly dissimilar) on how visually similar the two animal faces/bodies were to each other. For the human face and body similarity tasks, the subjects responded whether they considered that the first (press 1) or second (press 2) animal face/body looked visually more similar to an imagined face/body of a human. Previous studies have found that such overt similarity judgment tasks can successfully dissociate information about visual features from categorical information about the type of object observers are looking at (Bracci and Op de Beeck, 2016; Zeman et al., 2020). Across all similarity tasks, the trial structure was virtually identical: the fixation cross appeared (1000 ms), then the first image (1000 ms), followed by an interstimulus interval (1000 ms), and then the second stimulus (1000 ms). After the second stimulus disappeared from the screen, text appeared reminding participants of either the 6-point scale or the pairwise choice options. The next trial did not start until subjects made a response. All possible sequential pairwise combinations of images were presented during the experiment in random order with five rest breaks evenly spaced throughout. Stimulus presentation and control were performed via PC computers running PsychoPy2 (Peirce, 2007).

Naming task. Before both the similarity and fMRI experiments, all participants conducted a naming task with the following trial structure: a fixation cross appeared for 500 ms after which an image appeared for 1000 ms, then subjects typed in English or Dutch the name of the animal or natural object depicted in the image (e.g., "duck"/"eend"). If participants recognized the animal/natural object but could not remember the name, they were instructed to type "y," and if they did not recognize it at all, to respond with "n." After typing their response, subjects pressed enter, and the image and the correct English and Dutch labels appeared

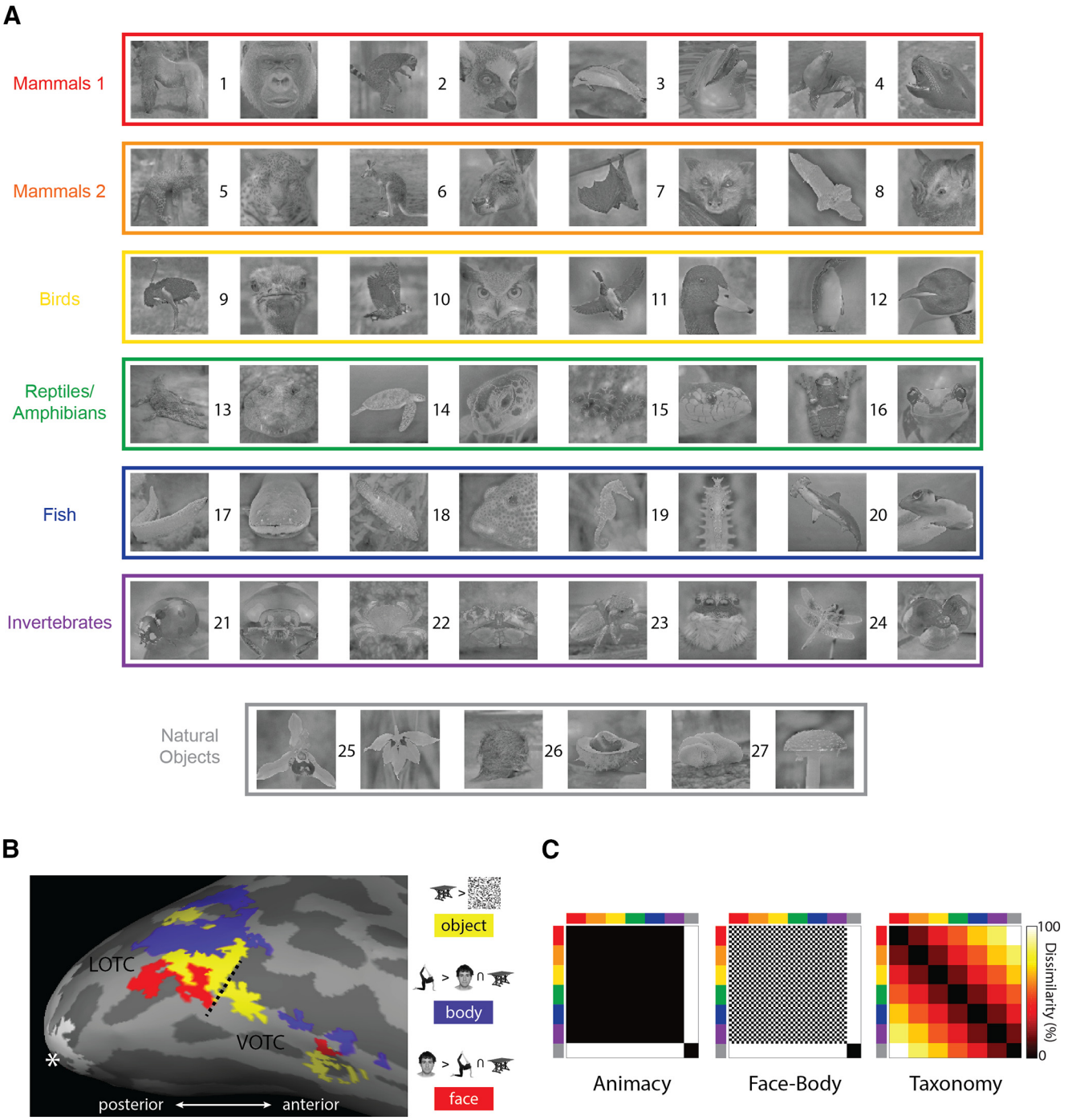

Figure 1. Core features of the experimental design. ***A***, All 54 natural image stimuli, color coded based on the taxonomic hierarchy. Numbers identify individual animals/natural object types: (1) gorilla, (2) lemur, (3) dolphin, (4) seal, (5) leopard, (6) kangaroo, (7) flying fox, (8) bat, (9) ostrich, (10) owl, (11) duck, (12) penguin, (13) crocodile, (14) turtle, (15) snake, (16) frog, (17) eel, (18) reef fish, (19) seahorse, (20) shark, (21) ladybug, (22) crab, (23) spider, (24) dragon fly, (25) orchid, (26) fruit, and (27) mushroom. ***B***, The results of the functional contrasts used in the study to define the ROIs, for one representative participant, mapped onto the inflated cortex using FreeSurfer (Fischl, 2012). The hashed line indicates the boundary between the lateral and ventral masks defined using the Anatomical Toolbox (Eickhoff et al., 2005). The white asterisk indicates the occipital pole, and the white patch is the EVC. ***C***, The three main model RDMs used throughout the study. The axes of the RDMs are color coded to reflect the taxonomic hierarchy for the stimuli.

for 3000 ms before the next trial began. Correct responses were coded based on predetermined labels, with any discrepancy marked as incorrect. For example, if a subject correctly labeled the body image for animal 24 as "dragonfly," but the corresponding face image as simply "fly," the latter response would be coded as incorrect. Stimulus presentation and control were performed via PC computers running PsychoPy2 (Peirce, 2007).

Only the naming data from fMRI participants were coded and analyzed for comparison with their neural data. The mean number of images for which participants responded "n" was 3.5 (maximum = 12), and the mean number of correctly coded labels was 44.4 (minimum = 36). The stimulus with the highest number of "n" responses was the coconut (9/15). Only five images were incorrectly labeled by a majority of subjects and three of these were invertebrate face images.

fMRI experiment sample size. For the fMRI experiment the number of participants was sufficient to have a power above 0.95 with reliable data that guarantee an effect size of $d = 1$. Based on previous studies with similar amounts of data per stimulus and subject (i.e., two long scan

sessions per participant; Bracci and de Beeck, 2016; Bracci et al., 2019), we know that the distinction between animals and other objects as revealed by representational similarity analysis (RSA), described below, has a very high effect size, typically with a Cohen's *d* of 1–4 even in smaller regions of interest (ROIs). We also assessed the between-subject reliability of the neural data by calculating the noise ceiling for RSA correlations for each of the ROIs we considered (see below).

Scanning procedures. The fMRI experiment consisted of two sessions of eight experimental runs followed by two localizer runs, for a total of 16 experimental and four localizer runs per subject, with one or two anatomic scans also collected for each participant. Using a rapid event-related design, each experimental run consisted of a random sequence of trials including two repeats of each of the 54 images and 18 fixation trials, for a total of 144 trials per run. Each stimulus trial began with the stimulus being presented for 1500 ms, followed by 1500 ms of the fixation bull's-eye. Subjects performed a one-back task in which on each trial they indicated with a button press whether they preferred looking at the current image or the previous one. Experimental runs had a total duration of 7 min and 30 s. A fixation bull's-eye was centrally presented continuously throughout each run.

For the localizer runs a block design was used with four stimulus types: bodies, faces, objects, and box-scrambled versions of the object images, with 18 images of each stimulus type. Each image in a block appeared for 400 ms followed by 400 ms fixation with four repeats of each stimulus block type per run. All four series of image types were presented sequentially in each stimulus block in pseudorandom order, followed by a 12 s fixation block. Localizer runs had a duration of 8 min and 0 s. To maintain their attention during a run, participants indicated when one of the images was repeated later in an image series, which occurred once each for two different randomly selected images per image type per block. A fixation bull's-eye was centrally presented continuously throughout each run. For one subject the data for three of four localizer runs were used as the data file for the remaining run was corrupted and unusable.

For both types of runs, stimulus presentation and control were performed via a PC computer running the Psychophysical Toolbox package (Brainard, 1997), along with custom code, in MATLAB (MathWorks).

Acquisition parameters. Data acquisition was conducted using a 3T Philips scanner, with a 32-channel coil, at the Department of Radiology of the Universitair Ziekenhuis Leuven university hospitals. Functional MRI volumes were acquired using a two-dimensional (2D) multiband (MB) T2*-weighted echo planar imaging sequence: MB = 2; repetition time, 2000 ms; echo time, 30 ms; flip angle, 90°; field of view = 216; voxel size = 2 × 2 × 2 mm; matrix size = 108 × 108. Each volume consisted of 46 axial slices (0.2 mm gap) aligned to encompass as much of the cortex as possible and all of the occipital and temporal lobes. Typically this resulted in the exclusion of the most superior portions of the parietal and frontal lobes from the volume. The T1-weighted anatomic volumes were acquired for each subject using an MPRAGE sequence, 1 × 1 × 1 mm resolution.

fMRI preprocessing and analysis. Preprocessing and analysis of the MRI data were conducted with SPM12 software (version 6906) using default settings unless otherwise noted. For each participant, fMRI volumes were combined from the two sessions (while preserving run order) and slice-time corrected (indexing based on slice acquisition time relative to 0 ms, not slice order), motion corrected using the realign operation (using fourth-degree spine interpolation), and coregistered to the individual anatomic scan. For all these steps the transformations were estimated and saved to the image header files before a single reslicing was conducted using the coregistration (reslice) operation. Functional volumes were then normalized to standard MNI space by first aligning the SPM tissue probability map to the individual subject anatomic scan and then applying the inverted warp to the functional volumes. Finally, all functional volumes were smoothed using a Gaussian kernel, 4 mm FWHM.

After preprocessing, the BOLD signal for each stimulus, at each voxel, was modeled separately for the experimental and localizer runs using GLMs. For the experimental runs, the predictors for the GLM consisted of the 54 stimulus conditions and six motion correction parameters (translation and rotation along the *x*-, *y*-, and *z*-axes). The time course for each predictor was characterized as two boxcar functions at the two stimulus onsets (duration = 1500 ms) convolved with the canonical hemodynamic response function. The GLM analysis produced one parameter estimate for each voxel for each stimulus predictor for each run. For the localizer runs the same modeling procedure was conducted for the four localizer conditions, with a single onset at the beginning of each image series (duration = 16 s) for each image type for each of the four stimulus blocks.

Defining ROIs. Three contrasts were used to specify separate functional ROIs for each subject based on the four conditions from the localizer runs (Fig. 1*B*): body > face + objects; face > body + object; and object > scrambled. We used conjunctions of masks from the Anatomy Toolbox to isolate lateral (conjunction of bilateral hOc4lp, hOc4la, and hOc4v) and ventral (conjunction of bilateral FG1–FG4) components of the OTC (Eickhoff et al., 2005). Within the two masked areas we used a threshold of FWE = 0.05, and then lowered the threshold to uncorrected $p = 0.001$ if no activity was detected or if it was detected in only one hemisphere. This procedure resulted in six functionally defined ROIs: lateral occipitotemporal cortex (LOTC)-body, LOTC-face, LOTC-object, ventral occipitotemporal cortex (VOTC)-body, VOTC-face, and VOTC-object. To define the early visual cortex (EVC), we used the posterior (i.e., most foveal) ~2/5 of the V1 mask as defined by the Anatomy Toolbox. Thus, unlike the other ROIs, which were defined functionally for each individual participant (although constrained by the masks), identical V1 mask coordinates were applied across participants without further functional feature selection.

Representational similarity analysis. RSA was used to compare the activity patterns from the different ROIs to the stimulus models, similarity judgments, and the layers of a suite of DNNs (Kriegeskorte and Kievet, 2013; Kriegeskorte et al., 2008a). For each comparison, representational dissimilarity matrices (RDMs) were constructed, which are matrices that are symmetrical around the diagonal and reflect the pairwise dissimilarities among all stimulus conditions. RDMs from different data modalities can be directly compared in order to evaluate the second order isomorphisms of the dissimilarities between conditions. RSA was conducted using CoSMo Multivariate Pattern Analysis, along with custom code (Oosterhof et al., 2016).

Neural RDMs for the different ROIs for each subject were constructed using the (non-cross-validated) Mahalanobis distance as the dissimilarity metric, characterized as the pairwise distance along the discriminant between conditions for the β weight patterns in an ROI (Walther et al., 2016; Ritchie and Op de Beeck, 2019). To assess the between-subject reliability of the RDMs for each ROI, the RDM of one subject was left out and those of the remaining subjects were averaged, and Pearson's *r* correlated with the left-out subject's RDM. This was conducted for all subjects, and the resulting coefficients were averaged. The resulting region-specific average value was used as an estimate of the noise ceiling when correlating individual neural RDMs for each ROI with the RDMs from the other data modalities. Visualization of group-averaged neural RDMs included multidimensional scaling (MDS) with stress 1 as the criterion.

Model RDMs were constructed in a number of different ways. For the main model RDMs (Fig. 1*C*), a 54-value vector was coded based on whether a stimulus was animate or not (Animacy), a face-body or not (Face-Body), or its rank (mammal 1 = 1; invertebrates = 6) in the intuitive taxonomic hierarchy (Taxonomy). The values of the model RDMs were then filled based on the absolute difference in the pairwise values in these coding vectors.

We also constructed an RDM from the GIST descriptors of the images (Oliva and Torralba, 2001). GIST was included in the analysis to control for potential low-level visual confounds (see below). Each image was segmented into a 4 × 4 grid, and Gabor filters (eight orientations and four spatial frequencies) were applied to each block in the grid. For each image, the values for each filter were converted to a vector, and all pairwise 1–*r* Pearson's correlations between these vectors were used to fill the cells of the RDM.

For the different DNNs (described below), layer-specific RDMs were constructed based on the 1–*r* pairwise Pearson's correlation between the

vectors of unit responses for each image. For the similarity judgments with a 6-point scale, RDMs were constructed based on populating a matrix with all pairwise judgments.

For the other similarity tasks, participants' responses resulted in a ranking vector, averaged across runs in the case of the in-scanner preference task. For example, the gorilla face might always be chosen as visually more similar to that of a human relative to all other animal faces and so would have the highest rank. The absolute pairwise differences in this ranking vector were used as the dissimilarity metric to construct the matrices. Individual behavioral RDMs were averaged to create a group-averaged matrix for comparison with individual neural RDMs.

For the naming task, a single matrix was constructed based on the proportion of subjects who correctly labeled an image with the absolute difference in proportion correct as the dissimilarity metric. To assess the reliability of the group-averaged human similarity RDMs, the individual RDMs were split into two groups, then averaged and correlated. This was done for all possible half splits of the data, and the resulting average coefficient value was transformed using the Spearman–Brown formula. This resulting value gives an estimate of the reliability of the group-averaged data, based on the full sample size (DiCarlo and Johnson, 1999; Op de Beeck et al., 2008).

To compare RDMs from different data modalities, the bottom half of each matrix was converted to a vector, and the Spearman rank-order correlation was calculated between matrices. The median correlations across subjects were tested for significance using the Wilcoxon signed-rank test. Because the test statistic W can be computed exactly for $N \leq 15$, it is reported along with the p value. Effect sizes for these tests are reported as the rank-biserial correlation (r_{bc}). Following Kerby (2014), this was calculated as $r_{bc} = W/S$, where S is the summed rank of N. When multiple similar statistical tests were conducted, for brevity we report the lower/upper bounds of the test statistics, effect sizes, and p values. In the case of the DNNs, for which several comparisons of the same type were made, the false discovery rate (FDR) adjusted p values are reported to correct for multiple comparisons.

Commonality analysis. Commonality analysis is a method for determining whether multiple predictors uniquely or jointly explain variance in the dependent variable (Newton and Spurell, 1967; Seibold and McPhee, 1979). This method has increasingly been used in conjunction with RSA and other multivariate pattern analysis methods and is also sometimes known as "variance partitioning" (Lescroart et al., 2015; Groen et al., 2018; Hebart et al., 2018). In the present case of three predictors (a, b, c) for some dependent variable y, there will be seven coefficients of determination (R^2) for all possible combinations of predictors in a linear regression model: $R^2_{y \cdot a}$, $R^2_{y \cdot b}$, $R^2_{y \cdot c}$, $R^2_{y \cdot ab}$, $R^2_{y \cdot ac}$, $R^2_{y \cdot bc}$, $R^2_{y \cdot abc}$. The last of these is the full model, for which the variance is partitioned based on differential weighting of the R^2 of the different models. When there are only three predictors, the partitioning can be performed using a simple weighting table (Nimon and Reio, 2011), in which the vector of coefficients is multiplied with row vectors of weights for each of the unique and common variance components. These results were visualized with EulerAPE (Micallef and Rodgers, 2014), which can be used to plot overlapping ellipses proportional to the variance partition of the total explained variance (Groen et al., 2018). Although in principle, negative variance can reflect informative relationships among predictors (Capraro and Capraro, 2001), in the present context these values were typically so small they are negligible (e.g., −0.1% of the total explained variance). They were therefore excluded from the visualization. Notably, the exclusion of these negative values means that the displayed values in the Euler plots depicting the commonality analyses that were performed will not sum to exactly 100% as is normally the case when all unique and common variance components are combined (Nimon and Reio, 2011).

To carry out the multiple regression necessary for commonality analysis, RDMs were converted to vectors, and the group-averaged neural dissimilarity values were regressed on the different model or behavioral dissimilarity vectors. Because this application of linear regression violates standard assumptions of independence between samples and normality, significance for the full model was determined by using a permutation test. For each individual subject, the rows of the bottom half of their neural RDM were independently randomly shuffled, and the resulting random vector was averaged across subjects and then fit with the full model. This procedure was conducted 1000 times for each application of multiple regression. The resulting proportion of R^2–values greater than that observed for the full model (fit to the unshuffled dissimilarities) provides the p value for the test.

Univariate analysis. For each subject the β weight values were averaged across all runs, and then all voxels, for each stimulus condition. Univariate RDMs were then constructed based on the pairwise absolute differences in the β weight values for each condition. The individual subject univariate RDMs were then correlated with the main model matrices as with the other forms of RSA described above. To further visualize the univariate results, average β weight values were calculated for each of the clusters of four face and body images for each level of the Taxonomy model and for all six of the natural object images.

DNNs. Networks consisted of stacked multiple convolutional (conv) layers that were intermittently followed by pooling operations, which fed into fully connected (FC) layers before output. Each DNN was pretrained on the ImageNet dataset (Russakovsky et al., 2015). To generate the response vectors for RSA we passed each image through the networks, with the activation weights of each layer as outputs. Softmax classification layers, which reflect the 1000 ImageNet labels commonly used for training DNNs, were excluded from the analysis. The networks used in our analysis are well known for their performance in the ImageNet Large Scale Visual Recognition Challenge (ISLVRC) competitions, or in the case of CORnet, have been promoted as superior in brain predictability, in contrast to image classification performance, using Brain-Score (Schrimpf et al., 2018).

CaffeNet is an implementation of the AlexNet architecture within the Caffe deep learning framework (Krizhevsky et al., 2012; Jia et al., 2014). The network includes five conv layers followed by three FC layers, for a total network depth of eight layers. The VGG-16 consists of 13 conv layers interspersed with five pooling layers, followed by three FC layers (Simonyan and Zisserman, 2014). GoogLeNet, or InceptionNet, is a 22-layer deep network, when counting only the parameterized layers (Szegedy et al., 2015). As with the majority of standard networks, the initial layers are conv layers followed by maximum pooling operations, and the final layers involve (average) pooling followed by a single FC layer. The main architectural point of difference from standard serial networks is that in GoogLeNet intermediary layers consist of stacked inception modules, which are themselves miniature networks containing parallelized conv and maximum pooling layers with convolutions of different sizes. ResNet50 is a deeper network (50 layers), with 48 convolutional layers and two pooling layers (He et al., 2015). The unique feature of ResNet50 is the implementation of shortcut connections that perform identity (residual) mappings between every three layers. CORnet is a family of architectures that include recurrent and skip connections, with all networks containing four layers that are premapped onto the areas of the ventral visual pathway in the primate brain: V1, V2, V4, and IT (Kubilius et al., 2018). We used CORnet-S, which combines skip connections with within-area recurrent connections and performed best "overall" on Brain-Score (Schrimpf et al., 2018), and so in principle should potentially be superior at matching the dissimilarity structure of OTC.

Results

Face-body and intuitive taxonomy models both uniquely explain neural similarity in the OTC

Because previous studies of the animacy organization in the OTC have tended to focus on large areas of the lateral and ventral OTC rather than category-selective ROIs (Connolly et al., 2012; Sha et al., 2015; Thorat et al., 2019), we first investigated whether the face-body and taxonomy model RDMs would correlate with the neural RDMs for the LOTC and VOTC, broadly construed. The ROIs for lateral and ventral ROIs in OTC (face, body, and object) were grouped into two ROIs, the LOTC-all and VOTC-all, which together encompass much of the large

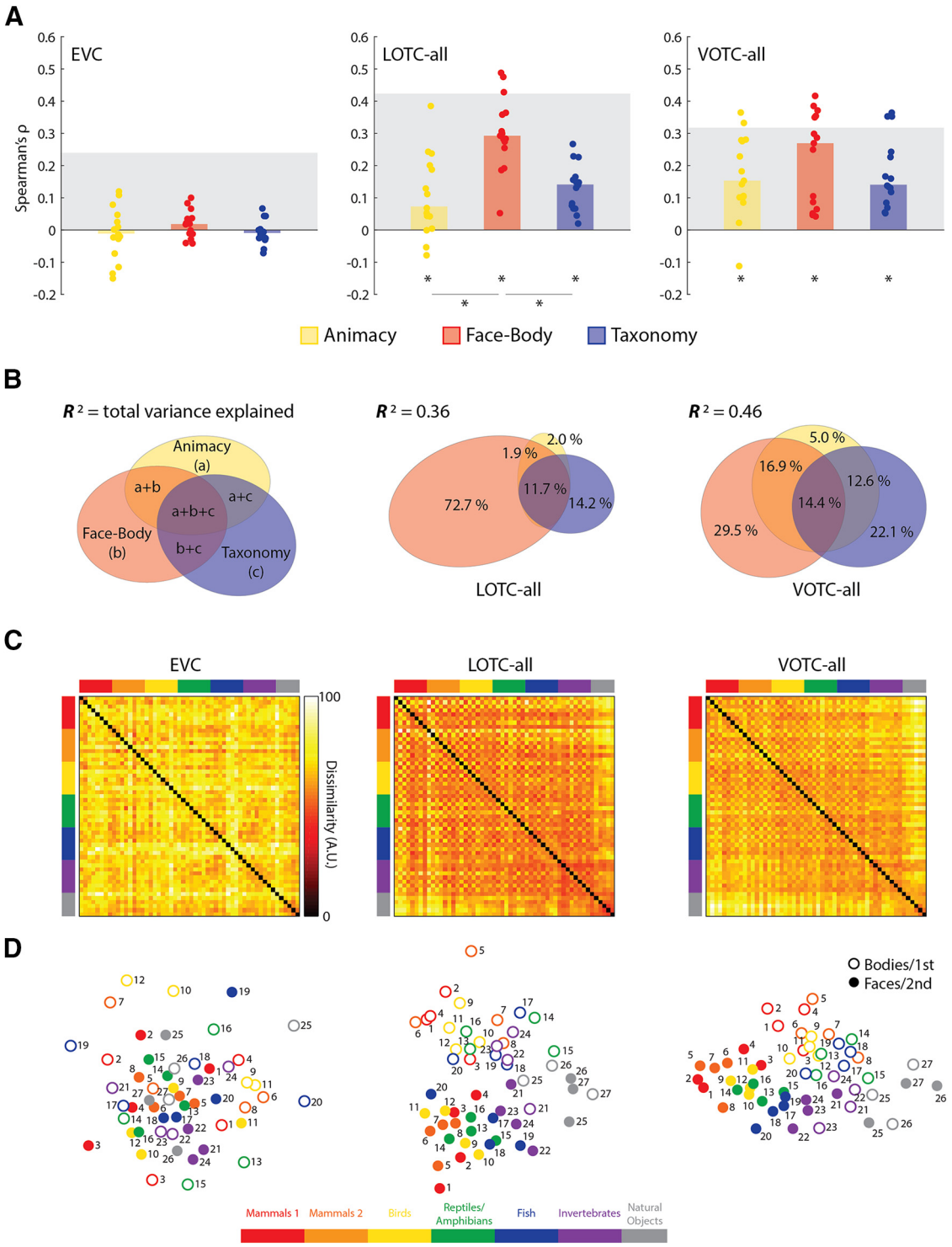

Figure 2. Comparing the main model RDMs to the EVC, LOTC-all, and VOTC-all. ***A***, Bar charts indicating the median Spearman's ρ rank-order correlations between neural RDMs for individual subjects (dots) and model RDMs. Gray blocks indicate the noise ceiling. *$p < 0.05$. ***B***, Results of commonality analysis for the LOTC-all and VOTC-all, depicted with Euler diagrams. A schema of the unique and common variance components is also depicted. Coefficients of determination (R^2) indicate the total proportion of explained variance for the full model. Regions of the Euler plots indicate percentages of the explained, and not total, variance accounted for by each component. ***C***, Group-averaged neural RDMs with the axes color coded based on the taxonomic hierarchy. Dissimilarity values are scaled to range 0–100. ***D***, Two-dimensional multidimensional scaling applied to the dissimilarity matrices. Points are color coded to reflect the taxonomic hierarchy and are either rings or dots to reflect the face/body division, or the first/second item for each natural object type. Numbers indicate individual animals or natural object type based on Figure 1*A*.

cortical territory typically investigated in previous studies of animacy organization in the OTC (Fig. 1*B*). We also included the early visual cortex (EVC) as a control region. In this analysis and those to follow, we also included an Animacy model that reflects the baseline distinction between animal and natural object stimuli that is common to both the Face-Body and Taxonomy models (Fig. 1*C*).

Consistent with previous studies, the median correlations of the three model RDMs (Fig. 2*A*) were significant for both the LOTC-all and VOTC-all [all: $W(15) \geq 90$, $r_{bc} \geq 0.75$, $p \leq 0.008$],

but not the EVC [all: $W(15) \leq 48$, $r_{bc} \leq 0.4$, $p \geq 0.19$]. For the LOTC-all, there were also significant differences between the median correlations for Animacy versus Face-Body and Face-Body versus Taxonomy [both: $W(15) \geq 118$, $r_{bc} \geq 0.97$, $p \leq 0.001$]. Despite similar magnitude median correlations for the VOTC-all, the differences in median correlation were not significant for Animacy versus Face-Body [$W(15) = 66$, $r_{bc} = 0.55$, $p = 0.06$] or Face-Body versus Taxonomy [$W(15) = 44$, $r_{bc} = 0.37$, $p = 0.23$], although for the Face-Body model the individual correlations have a clear bimodal distribution.

All three model RDMs are correlated with each other (Fig. 3), which raises the issue of how much of the observed effects for the three models reflect their common structure, although if the natural object images are excluded, the Animacy RDM contains no internal structure, and the Face-Body and Taxonomy RDMs are not correlated ($\rho = -0.03$, $p = 0.28$). Still, to address this, we conducted commonality analysis on the group-averaged neural RDMs for the LOTC-all and VOTC-all. The full model, containing all three predictors (Animacy, Face-Body, Taxonomy), explained a sizable amount of the variance for both ROIs (Fig. 2*B*) and was significant based on permutation tests (LOTC-all: $R^2 = 0.36$, $p < 0.001$; VOTC-all: $R^2 = 0.46$, $p < 0.001$). In the LOTC-all, most of the explained variance was uniquely accounted for by the Face-Body model and to a lesser extent the Taxonomy model. For the VOTC-all, the Face-Body and Taxonomy were qualitatively more equitable in their unique contributions. These results show that the Face-Body and Taxonomy models each account for unique and independent components of the explained variance in neural dissimilarity in the lateral and ventral OTC for our image set.

These findings can be visually summarized using the group-averaged neural RDMs (Fig. 2*C*), which can be compared with the model RDMs (Fig. 1*C*) and their 2D MDS solutions (Fig. 2*D*). When this is done, one can see that the RDMs for the LOTC-all and VOTC-all, but not the EVC, show structural similarity to the main model RDMs. Multidimensional scaling also makes more salient the differences in the commonality analysis results of the LOTC-all and VOTC-all (Fig. 2*D*). On the one hand, the face-body division is much more pronounced in the 2D space for the LOTC-all. On the other hand, the taxonomic hierarchy is more apparent in the VOTC-all, whereas the face-body division is also still clearly present. Notably, in the VOTC-all the patterns for the body images tended to be more similar to those for the natural objects than the face images as reflected in the clustering in Figure 2*D* and the bands through the hot bars on the bottom and ride side of the matrix in Figure 2*C*. This could reflect greater shape similarity between the body and natural object images.

Nuisance models do not explain neural similarity in the OTC

Three nuisance models were considered that might also account for the structure of the neural RDMs for the EVC, LOTC-all, and VOTC-all. First, it has been suggested that low-level and midlevel image properties may explain away apparent effects of object category in the OTC, including animacy (Rice et al., 2014; Andrews et al., 2015; Coggan et al., 2016; Long et al., 2018). To determine whether such properties might also account for the present results we used GIST, a model which captures spatial frequency and orientation information of images (Oliva and Torralba, 2001). Second, to ensure that familiarity with the animals or recognizability of the images was not a confound, the subjects performed a naming task before all experiments. Finally, during scanning, subjects performed a one-back image preference task. RDMs for these two tasks and

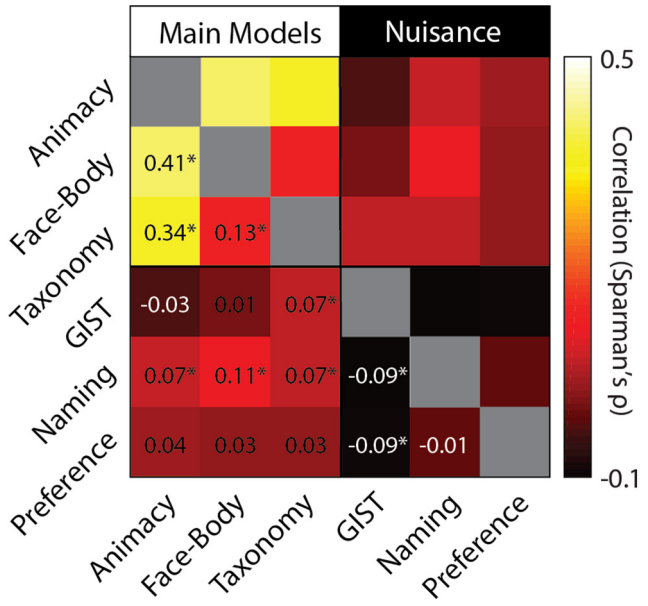

Figure 3. Correlation between main model and nuisance predictor RDMs. Cells in the matrix correspond to the pairwise correlations (Spearman's ρ) between matrices. Text for negative correlations appear in white. *$p < 0.05$.

GIST were weakly correlated with the main model RDMs (Fig. 3) and could not account for the results observed for the OTC (Fig. 4). There was a significant median correlation between individual RDMs for the EVC and the GIST model [$W(15) = 106$, $r_{bc} = 0.88$, $p = 0.001$] and also negatively correlates with the Naming model [$W(15) = 70$, $r_{bc} = 0.58$, $p = 0.048$], which also exhibited a significant median correlation with the VOTC-all [$W(15) = 74$, $r_{bc} = 0.62$, $p = 0.04$]. As the Naming RDM was correlated with the Face-Body and Taxonomy RDMs, we also conducted partial correlations with the individual neural RDMs for the VOTC-all. In both cases, the median correlations decreased slightly after controlling for the naming RDM [both: Δ median $\rho < 0.01$; $W(15) = 120$, $r_{bc} = 1.0$, $p = 6.10e\text{-}05$].

Face-body and intuitive taxonomy models both explain neural similarity in face-and body-selective and object-selective areas of the OTC

OTC is well known to contain regions that show preferential selectivity for face and body images in both the lateral and ventral OTC (Kanwisher et al., 1997; McCarthy et al., 1997; Downing et al., 2001; Peelen and Downing, 2005; Tsao et al., 2006) and object images more generally (Grill-Spector et al., 1999). However, previous investigations of an intuitive taxonomy organization in the OTC have generally not functionally isolated these regions. Therefore, we next assessed whether the results observed in the LOTC-all and VOTC-all might be maintained in subordinate face- and body-selective areas; in particular, we sought to assess whether the Taxonomy model would also show effects in these areas and not just the lateral and ventral OTC more generally. To this end we conducted the same analysis as before, that is, correlating individual neural RDMs for LOTC/VOTC-body/face areas with the three model RDMs, followed by commonality analysis. These analyses were then also conducted for LOTC/VOTC-object areas to assess whether they were restricted to face- and body-selective portions of the OTC.

For all three model RDMs (Fig. 5*A*) the median correlations across subjects were all highly significant for all four ROIs [all: *W*

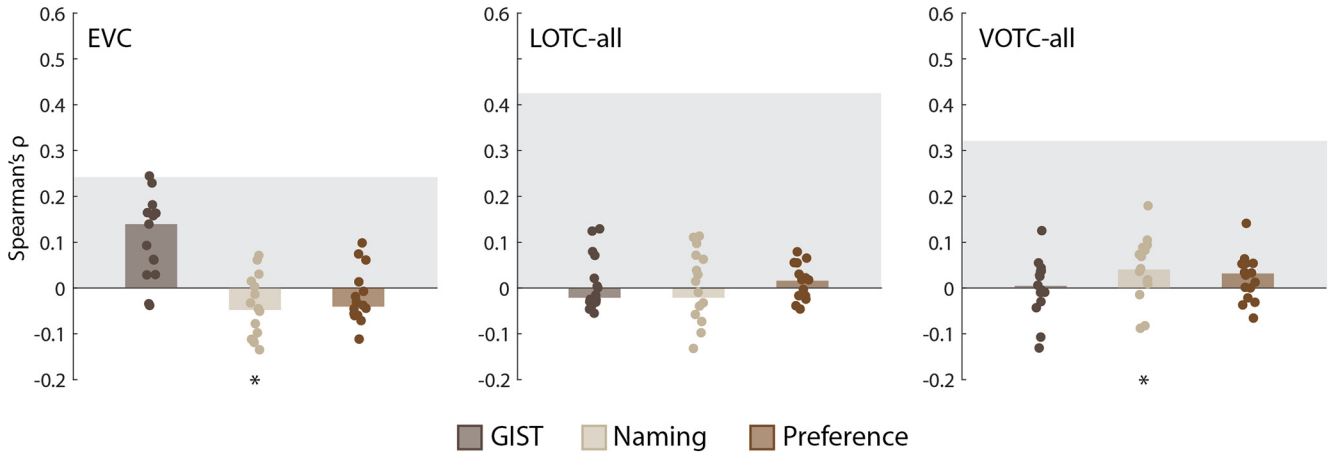

Figure 4. Comparing the nuisance model RDMs to the EVC, LOTC-all, and VOTC-all. Bar charts indicating the median Spearman's ρ rank-order correlations between neural RDMs for individual subjects (dots) and the model RDMs. Gray blocks indicate the noise ceiling. *$p < 0.05$ based on Wilcoxon signed-rank tests.

(15) $\geq$ 92, $r_{bc} \geq 0.77$, $p \leq 0.007$], with the exception being the median Animacy correlation in the LOTC-face, which was the only weakly significant effect [$W(15) = 78$, $r_{bc} = 0.65$, $p = 0.03$]. For both the LOTC-body and LOTC-face there were also significance differences in the median correlations [all: $W(15) \geq 88$, $r_{bc} \geq 0.73$, $p \leq 0.01$]. In contrast in the VOTC-body neither comparison was significant [both: $W(15) \leq 28$, $r_{bc} \leq 0.23$, $p \geq 0.45$], and only Face-Body versus Taxonomy [$W(15) = 88$, $r_{bc} = 0.73$, $p = 0.01$], and not Animacy versus Face-Body [$W(15) = 42$, $r_{bc} = 0.35$, $p = 0.25$] were significantly different in the VOTC-face. Although the effect of the Face-Body model is to be expected based on the category selectivity of these areas, these findings also show that an intuitive taxonomy effect can be localized separately to both face- and body-selective areas.

When all three predictors were regressed on the group-averaged neural dissimilarities (Fig. 5*B*), the full model explained significant variance in the group neural RDMs (all: $R^2 > 0.25$, $p < 0.001$). There was a notable difference, however, in the portioning of the explained variance between the lateral and ventral ROIs (Fig. 5*B*). In the LOTC-body, both the Face-Body and Taxonomy models uniquely explained large portions of the variance. In the LOTC-face virtually all the explained variance was uniquely accounted for by the Face-Body model. For the ventral areas, both Face-Body and Taxonomy, and their common components, were substantive contributors to the explained variance. Some of the characteristics of the group-averaged RDMs for the face- and body-selective areas include the hot bands for the division between animal and natural objects and the face-body checkering (Fig. 5*C*); and MDS plots again show a clear face-body division as well as an intuitive taxonomic continuum (Fig. 5*D*).

Similar results were also found for object-selective areas of the OTC. The median correlations for all three model RDMs (Fig. 6*A*) were very significant for both the LOTC-object and VOTC-object [all: $W(15) \geq 94$, $r_{bc} \geq 0.78$, $p \leq 0.006$], with the exception of a weaker median effect for the animacy model in the VOTC-object [$W(15) = 82$, $r_{bc} = 0.68$, $p = 0.02$]. For the LOTC-object, there were also significant differences in the median correlations for Animacy versus Face-Body [$W(15) = 82$, $r_{bc} = 0.68$, $p = 0.02$] and Face-Body versus Taxonomy [$W(15) = 108$, $r_{bc} = 0.90$, $p = 8.54\text{e-}04$]. This was not the case for the VOTC-object [both: $W(15) \leq 28$, $r_{bc} \leq 0.23$, $p \geq 0.45$]. The full model also explained a significant amount of the variance in the group-averaged RDMs for both ROIs (both: $R^2 > 0.15$, $p < 0.001$). Commonality analysis revealed that the vast majority of the explained variance was unique to the Face-Body model in the LOTC-object, whereas the explained variance was more equitable between the unique and shared components in the VOTC-object (Fig. 6*B*). Visualizations of the group-averaged results were also similar to the face and body areas (Fig. 6*C*,*D*). These results show that the effect of the intuitive taxonomy model is also independent of that for the Face-Body model outside of the face- and body-selective cortex.

Taken as a whole, these findings show that the independent effect of the Taxonomy model relative to the Face-Body and Animacy models is not a result of analyzing large portions of OTC but is also consistently observed in category-selective subregions.

Neural similarity in the OTC may partially reflect the univariate responses in face- and body-selective areas

Previous investigations of the intuitive taxonomic organization of the OTC have relied on multivariate methods, like RSA. However, given that the category-selective ROIs were defined by the univariate response to human face and body images, we next asked whether differences in the magnitude of this response might predict the position in the intuitive taxonomic hierarchy of the stimuli.

Figure 7*A* shows the univariate results by averaging across the four face/body images for each level of the intuitive taxonomic hierarchy. This served to verify that face and body animal stimuli showed greater activity in the face- and body-selective areas, respectively, and in each case, greater activity than for the natural objects across levels of the hierarchy. As can be seen, there is a very robust difference between face images and the other stimuli for both face-selective regions. However, the difference is weaker, and less reliable, for the LOTC-body, and is wholly absent in the VOTC-body. For the LOTC-face and VOTC-face, an intuitive taxonomy effect is suggested, at least for the face stimuli. To quantify these univariate differences, we correlated individual univariate RDMs with the three main model RDMs (Fig. 7*B*).

For the LOTC-body, a significant median correlation was observed for Face-Body [$W(15) = 86$, $r_{bc} = 0.72$, $p = 0.01$]. For the LOTC-face, the median correlations for the Face-Body model were highly significant [$W(15) = 120$, $r_{bc} = 1.0$, $p = 6.10\text{e-}05$], but only marginally significant for the Animacy [$W(15) = 70$, $r_{bc} = 0.58$, $p = 0.048$] and Taxonomy [$W(15) = 76$, $r_{bc} = 0.63$, $p = 0.03$] models. There were also significant differences in the median correlations [both: $W(15) \geq 100$, $r_{bc} \geq 0.83$, $p \leq 0.003$]. For the

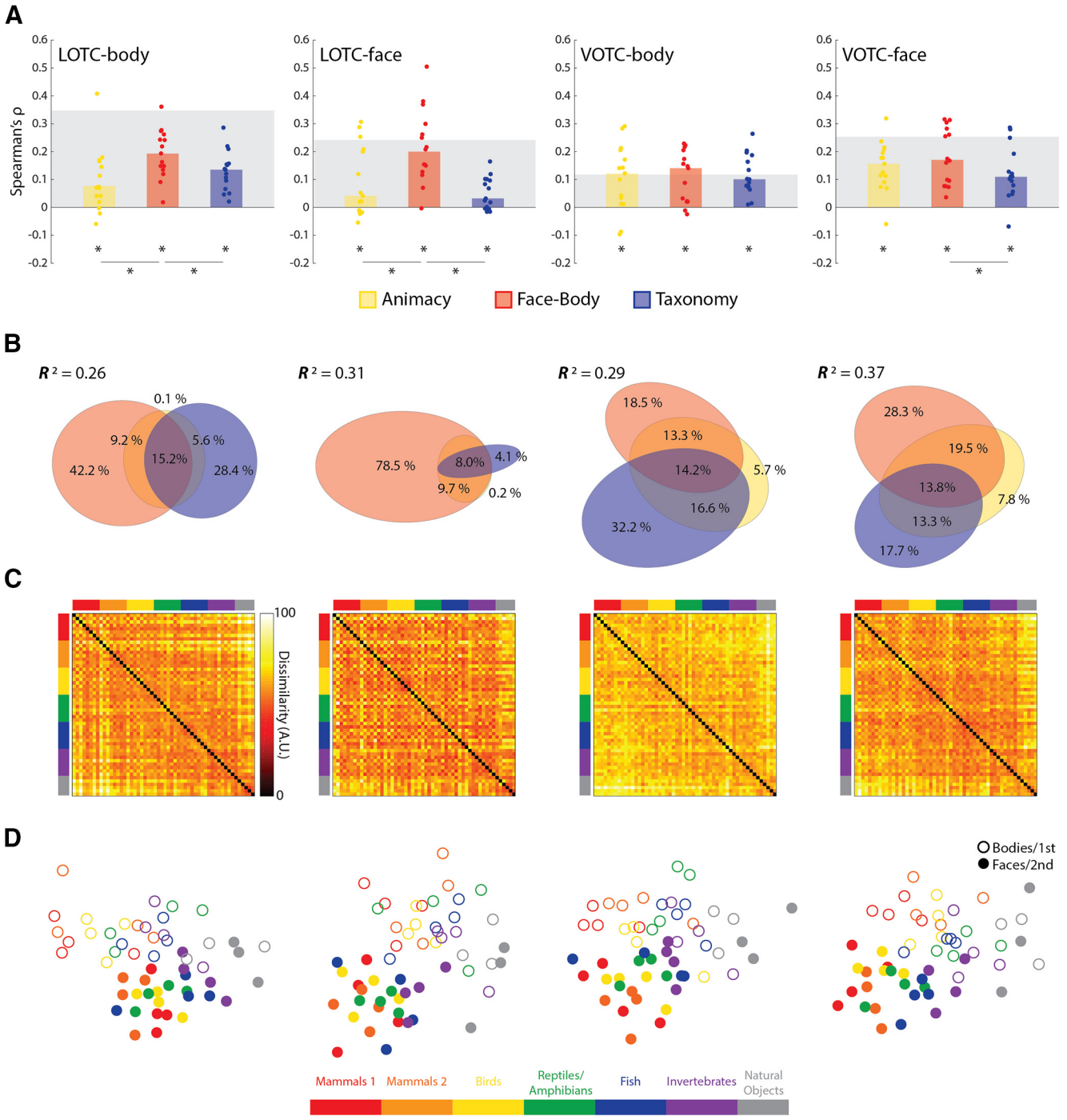

Figure 5. Results of model RDM comparisons to face- and body-selective regions of OTC. ***A***, Bar charts indicating the median Spearman's ρ rank-order correlations between neural RDMs for individual subjects and the three main model RDMs, for all four ROIs. ***B***, Results of commonality analysis for all ROIs, visualized with Euler diagrams. ***C***, Group-averaged neural RDMs for the four ROIs. ***D***, Two-dimensional multidimensional scaling applied to the dissimilarity matrices for each ROI. Conventions follow those of Figure 2.

VOTC-body, there were significant median correlations with the Animacy [$W(15) = 80$, $r_{bc} = 0.67$, $p = 0.02$] and Taxonomy [$W(15) = 106$, $r_{bc} = 0.88$, $p = 0.01$] models. There was also a significant difference in the median correlations for Animacy versus Face-Body [$W(15) = 74$, $r_{bc} = 0.62$, $p = 0.035$]. Finally, for the VOTC-face, there were significant median correlations with all three model RDMs [all: $W(15) \geq 82$, $r_{bc} \geq 0.68$, $p \leq 0.01$]. There was also a significant difference in median correlations for Face-Body versus Taxonomy [$W(15) = 72$, $r_{bc} = 0.60$, $p = 0.04$]. As to be expected from the results in Figure 5*A*, the Face-Body RDM correlations were substantially higher in the face-selective ROIs.

Most notably, these results show that to the extent there is an ordering of the univariate responses that is equivalent to an intuitive taxonomy of the stimuli, the ordering effect is far weaker than what is revealed using multivariate methods.

Visual similarity to human templates, not intuitive taxonomy, best explains multivariate neural similarity in face- and body-selective areas of the OTC

The results so far suggest that the effect of intuitive taxonomy in regions of the OTC is independent of that for faces versus bodies, which implies that patterns of activity for the faces and bodies

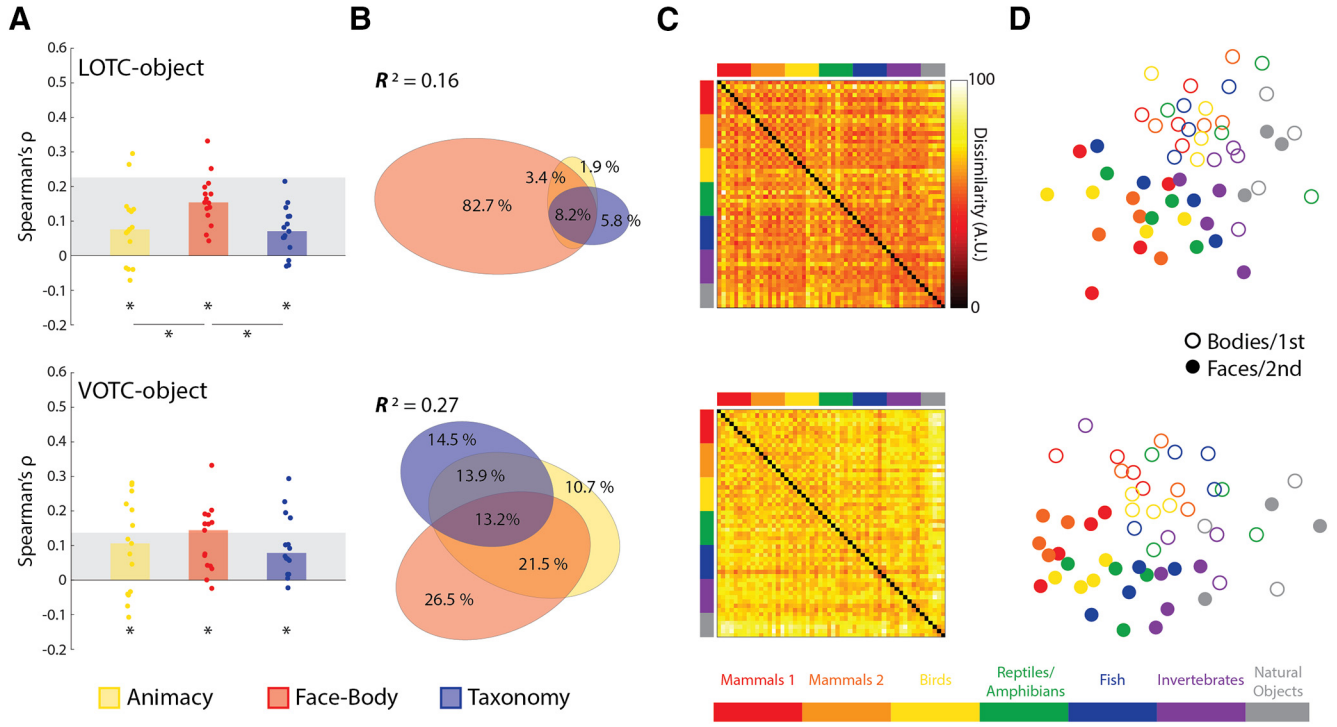

Figure 6. Results of model RDM comparisons to object-selective regions of OTC. ***A***, Bar charts indicating the median Spearman's ρ rank-order correlations between neural RDMs for individual subjects and the three main model RDMs for all four ROIs. ***B***, Results of commonality analysis for all ROIs, visualized with Euler diagrams. ***C***, Group-averaged neural RDMs for the four ROIs. ***D***, Two-dimensional multidimensional scaling applied to the dissimilarity matrices for each ROI. Conventions follow those of Figure 2.

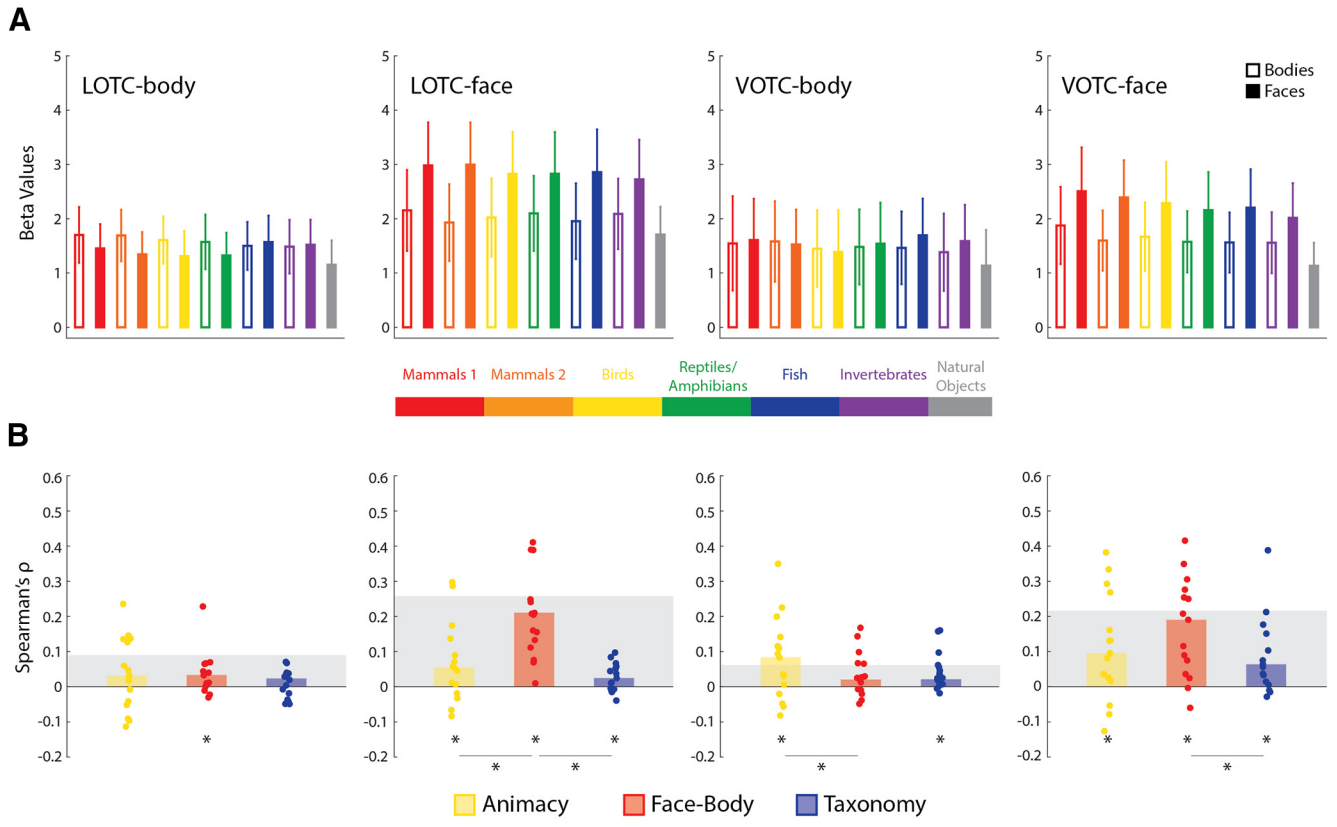

Figure 7. Results of univariate RSA for the face- and body-selective ROIs. ***A***, The average univariate responses for face and body images, at each level of the taxonomic hierarchy, are depicted against the average data of natural object images. Error bars indicate SEM. ***B***, Bar charts indicating the median Spearman's ρ rank-order correlations between univariate RDMs for individual subjects (dots) and the three main model RDMs for all four ROIs. Gray blocks indicate the noise ceiling. *$p < 0.05$ based on Wilcoxon signed-rank tests.

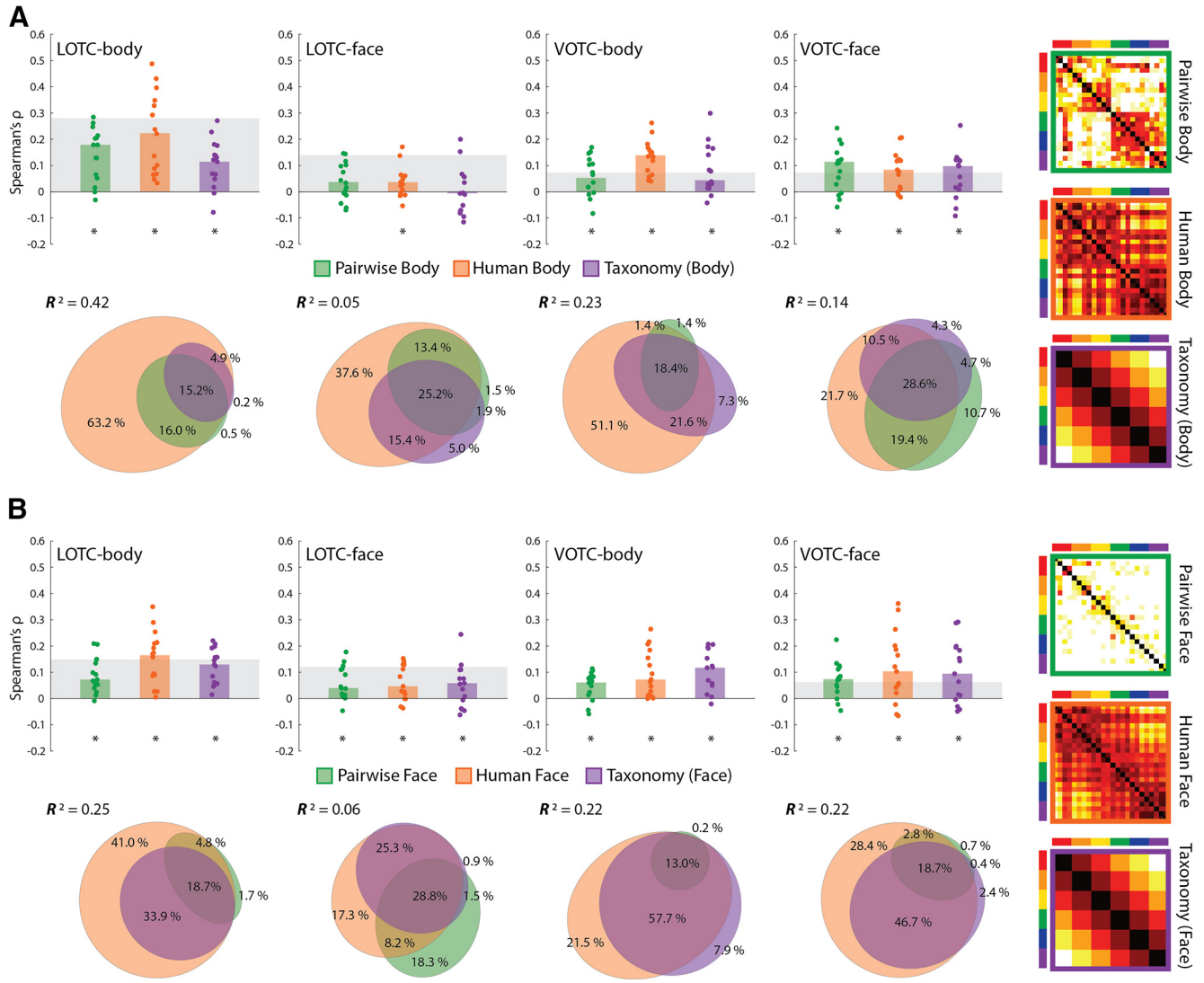

Figure 8. Results of comparing model RDMs for face and body stimuli to face- and body-selective regions of the OTC. ***A***, Bar charts indicating the median Spearman's ρ rank-order correlations between neural RDMs for individual subjects and three model RDMs for the body images (also depicted); pairwise body similarity judgments (Pairwise Body), human similarity judgments (Human Body), and a (6-level) taxonomy model for 24 images. ***B***, Bar charts and commonality analysis for the three model RDMs for the face images (also depicted). All conventions are the same as in Figure 2.

both separately exhibit an intuitive taxonomic structure. Several lines of evidence suggest that regions of the OTC code for visual features that are diagnostic of different object categories (Jozwik et al., 2016; Bracci et al., 2017). Thus, we next sought to assess the patterns of neural dissimilarity of face and body images separately in each of the face- and body-selective ROIs and how their apparent taxonomic organization relates to the perceptual similarity of the stimuli. For the images, two kinds of visual similarity judgments were collected. First, separate groups of subjects judged the pairwise visual similarity of face or body images using a 6-point scale. These judgments provided a general measure of perceived visual similarity of any two images. Second, other groups of subjects judged how visually similar the animal face or body images were to a particular reference—an imagined human face or body. The inclusion of these human similarity tasks was inspired by the finding that relative similarity to humans can also generate an animacy continuum (Contini et al., 2020) and the conjecture that this continuum may in fact reflect gradation in selectivity for the visual characteristics of animal faces and bodies. Both types of judgments were converted to group-averaged behavioral RDMs, which were compared with a Taxonomy model RDM constructed for just the 24 face and body images (Fig. 8*A*,*B*). These models allowed us to assess the extent to which taxonomic effects for the face and body images in the OTC might be accounted for by forms of perceived visual similarity.

The two behavioral RDMs and truncated Taxonomy model RDM were all highly correlated with each other for both the face and body images (all $\rho > 0.4$, $p < 0.0001$). As seen in Figure 8, *A* and *B*, the human body RDM suggests a grouping of the mammals and birds separate from the reptiles/amphibians, fish, and invertebrates, with a similar division for the human face RDM, although with greater similarity between the judgments for the bird and reptile/amphibian faces. These aspects that differentiate the human similarity matrices from the intuitive taxonomy model are reliable, given that our estimate of the split-half reliability of these matrices was $r = 0.92$ (human body) and $r = 0.89$ (human face) and so was much higher than the respective correlations with the Taxonomy RDM.

We correlated the pairwise similarity, human visual similarity, and Taxonomy RDMs with the separate neural RDMs for the face and body images across the face- and body-selective ROIs. For the body images (Fig. 8*A*), there were significant median correlations between each of the model RDMs and individual neural RDMs for all but one ROI [all: $W(15) \geq 86$, $r_{bc} \geq 0.72$, $p \leq 0.01$]. The exception was the LOTC-face where a significant median correlation was only observed for the human body RDM [$W(15) = 82$, $r_{bc} = 0.68$, $p = 0.02$], but not the Pairwise [$W(15) = 62$, $r_{bc} = 0.52$, $p = 0.08$] or Taxonomy [$W(15) = 12$, $r_{bc} = 0.10$, $p = 0.76$] models. For the face images (Fig. 8*B*), all three predictor RDMs showed significant median correlations across all four ROIs [all: $W(15) \geq 88$, $r_{bc} \geq 0.73$, $p \leq 0.01$], with the median correlation of the Taxonomy model only barely in the LOTC-face [$W(15) = 70$, $r_{bc} = 0.58$, $p = 0.048$]. In light of the correlations between the human visual similarity and Taxonomy RDMs for both the face and body images, we also conducted partial rank-order correlations between the Taxonomy RDM and the individual neural RDMs, controlling for the human visual similarity models. When this was done, there was no significant median correlation of the Taxonomy model for either image type, in any of the four ROIs [all: median $\rho \leq 0.05$, $W(15) \leq 58$, $r_{bc} \leq 0.48$, $p \geq 0.11$]. This result suggests that intuitive taxonomy is not a substantial predictor of neural dissimilarity in regions of the OTC once controlling for the human visual similarity of the images.

To further determine the unique versus common contributions of the three predictors, we conducted commonality analysis across image types and ROIs (Fig. 8*A*,*B*). The full model explained a significant amount of the variance across ROIs for the body images (all: $0.04 < R^2 < 0.42$, $p < 0.002$) and face images (all: $0.05 < R^2 < 0.26$, $p < 0.003$) based on permutation tests. Although, notably, the explained variance was considerably lower for the LOTC-face for both stimulus types. Such a result is somewhat expected in light of the comparatively low individual model correlations, which are anticipated by the lower Taxonomy correlations in the ROI for the full stimulus set (Fig. 5*A*). For the body images, across all ROIs, the human body similarity judgments were consistently the best unique predictor of variance with both pairwise similarity and Taxonomy predicting very little of the remaining variance uniquely. For the face images, the same picture emerged, with the exception of the LOTC-face where pairwise face judgments uniquely predicted slightly more of the variance. Taken as a whole, these results suggest that any observed effect of the Taxonomy model is almost entirely a reflection of commonly explained variance with the human face and body visual similarity judgments.

Face-body and intuitive taxonomy, but not animacy, models explain activation dissimilarity in DNNs

The preceding results suggest that the apparent intuitive taxonomic organization of the OTC may in fact reflect gradation in the neural representation of animal faces and bodies based on their visual similarity to humans. Another way to assess this possibility is to evaluate how our stimulus set is represented in a model system, which in contrast to humans has no knowledge of intuitive taxonomic relations between animal categories. If a similar organization is revealed, then this would suggest that the effects observed in the OTC can be explained without positing the representation of an intuitive taxonomic hierarchy. Here, DNNs provide a useful foil.

DNNs are typically trained solely to carry out a first-order classification of animal types without explicit instructions to represent superordinate relationships between animal categories. Nevertheless, several studies have reported correlations between FC layers of DNNs and category-selective areas in the human OTC, suggesting that the animacy division is represented in these networks when trained to discriminate classes of natural images, including many types of animals (Khaligh-Razavi and Kriegeskorte, 2014; Jozwik et al., 2017; Bracci et al., 2019; Zeman et al., 2020). These previous studies did not assess whether DNNs exhibit an intuitive taxonomic organization similar to what we describe in our study for the OTC. If present, then such a representational structure would be wholly dependent on between-category generalizations learned from comparing the visual properties of exemplar images of different animal categories. This would provide further evidence that one need not suggest an overt representation of a taxonomic hierarchy to account for gradation in the animacy organization of the OTC. Therefore, we correlated the layer RDMs for five well-known DNNs to the main model RDMs, the broadly defined ROIs, and finally the similarity judgment RDMs for the face and body images.

Across all five networks, the correlations with the Face-Body model increased and peaked close to the first FC layers, or CORnet's decoding layer, followed by the Taxonomy model correlations peaking at the final FC/decoding layers (Fig. 9*A*). Unlike previous studies, the layer RDMs tended to be negligibly, or even negatively, correlated with the Animacy model. These findings show that the trained networks exhibit an organization in their final layers that is highly correlated with an intuitive taxonomic hierarchy, although they have not been trained to compare superordinate relationships for any of the categories they have learned to classify.

To verify that the networks also showed some correspondence to patterns of responses in the OTC, the layers of the network were also correlated with individual neural RDMs for the three initial ROIs (Fig. 9*B*). The peak median correlations with the EVC tended to occur for middle conv layers, whereas those for the LOTC-all and VOTC-all tended to occur at the later conv or FC/decoding layers. These findings are consistent with the claim that FC/decoding layers better reflect the structure of regions of the OTC. Visualization of the final layer RDMs (Fig. 9*C*) and MDS plots (Fig. 9*D*) provide some insight into these findings. For each network, the layer activity patterns for the mammal and bird face images tend to cluster separately, whereas those for the other intuitive taxonomy groups and the body images cluster with those for the natural images.

We also investigated whether the dissimilarity structure of the network layers might be better captured by pairwise and human similarity judgment RDMs for the face and body images. For the body image models (Fig. 10*A*), we found that for all but one network (CaffeNet), the human body RDM tended to show the highest correlations across layers of networks, with the pairwise body RDM peaking at a similar or higher level at the final layers. In contrast, the Taxonomy model only showed a significant correlation with the final layer of the VGG-16. For the face image models (Fig. 10*B*), all three models showed a consistent increase in correlation effect sizes with network depth. Again, as with the neural data, the human face model tended to show the highest correlations. The Taxonomy model also consistently correlated with many of the layers of the different networks. These results suggest that correlation with the Taxonomy model for the full

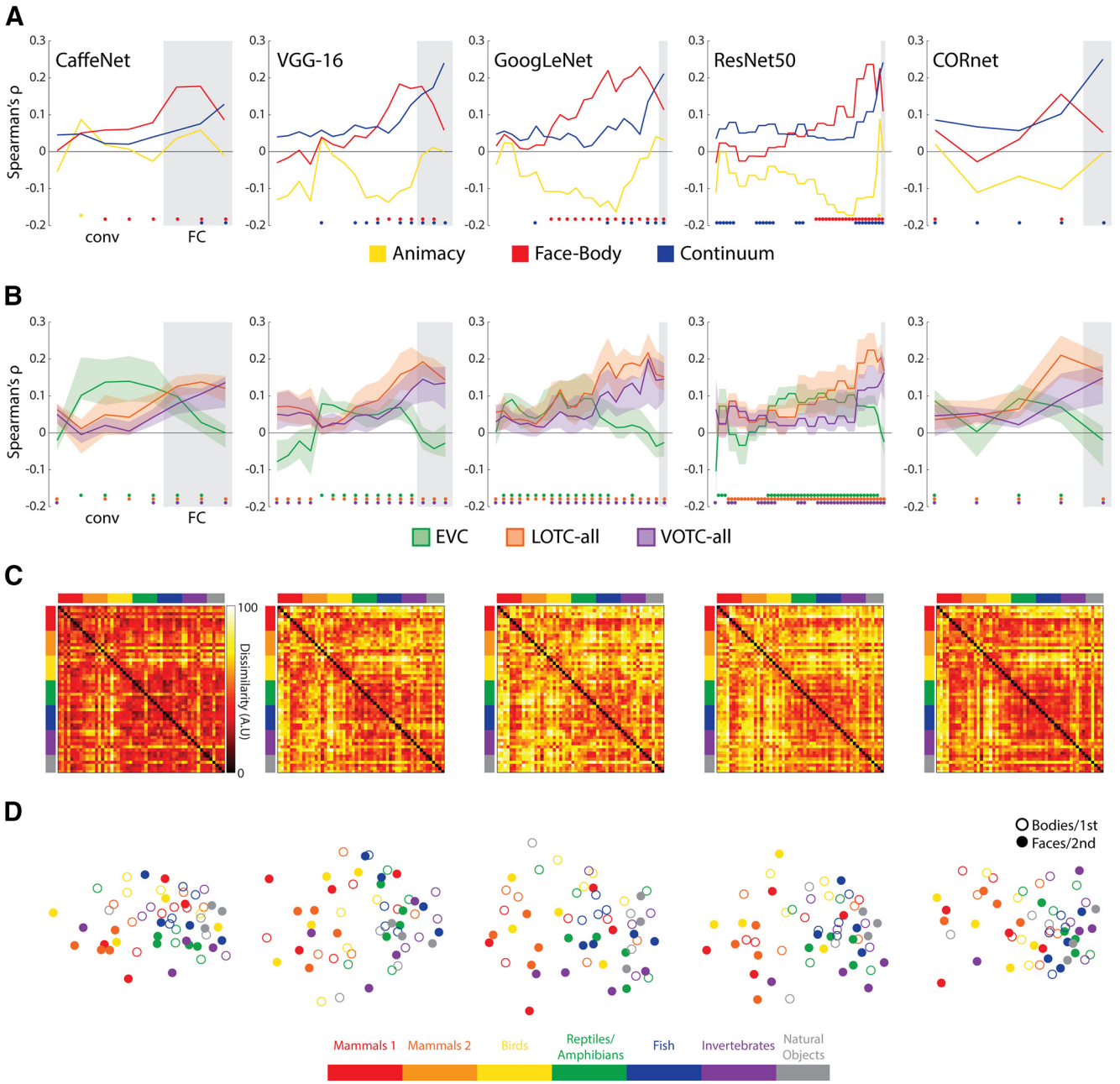

Figure 9. Results of RSA with DNNs for all 54 images. ***A***, The layers of each of the five networks correlated with each of the three model RDMs. ***B***, The layers of each of the five networks correlated with the individual subject neural RDMs for the three initial ROIs considered. The solid lines indicate the median correlations at each layer, and the transparent regions range from the first quartile to the third quartile. For both ***A*** and ***B***, gray areas indicate FC/decoding layers for each network and are preceded by the conv/area layers. Color-coded dots indicate significant (positive) correlations at $p < .05$ (FDR adjusted), based on two-sided permutation tests. ***C***, RDMs for the final layers of each of the DNNs. ***D***, Two-dimensional multidimensional scaling applied to the dissimilarity matrices for each DNN. Conventions follow those of Figure 2.

stimulus set (Fig. 9*A*) was likely being driven by the face images (compare Fig. 9*D*). This result is notable because faces are not a category the networks were trained to classify.

We next conducted commonality analysis across image types and the final layers of the DNNs, which consistently showed the peak correlation with the Taxonomy model for the full image set (Fig. 10*A*,*B*). This allowed us to assess, like with face- and body-selective regions of the OTC, whether the effects of the Taxonomy model in these layers can be accounted for by the perceptual similarity of the stimuli. For the body images, the full model explained a significant amount of the variance for the final layers of all networks ($0.05 < R^2 < 0.3$, all $p < 0.004$). For all but CaffeNet (for which the R^2 was low), the human body RDM was consistently the best unique predictor, followed by the pairwise body RDM. For the face images, the full model explained a significant amount of the variance for the final layers of all networks ($0.1 < R^2 < 0.23$, all $p = 0.001$). For all networks, either the pairwise or human face RDMs were the best unique predictor of variance, whereas the Taxonomy explained virtually no unique variance.

These findings are broadly consistent with those observed for the face- and body-selective areas (Fig. 8). There are two notable differences. First, the pairwise similarity RDMs consistently rivaled, or surpassed, the human similarity RDMs as predictors of

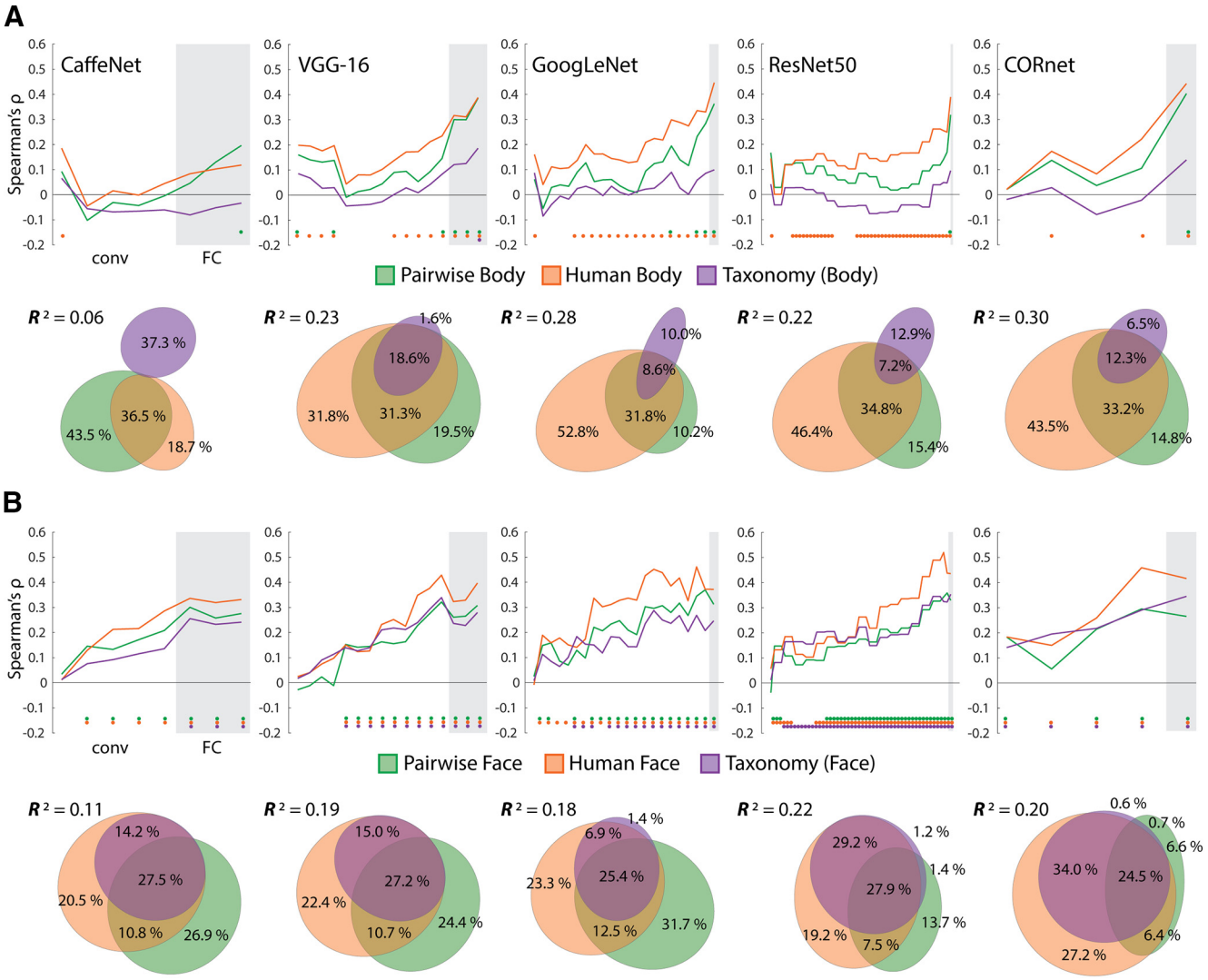

Figure 10. Results of comparing models for face and body images to DNNs. ***A***, The layers of each of the five networks were correlated with each of the body image RDMs. Color-coded dots indicate significant (positive) correlations at $p < .05$ (FDR adjusted), based on two-sided permutation tests. Results of commonality analysis for the final layers of all networks are shown with Euler diagrams. ***B***, Correlations between the layers of the five networks and the three model RDMs for the face images. All conventions are the same as in ***A***. For both ***A*** and ***B*** Gray areas indicate FC/decoding layers for each network and are preceded by the conv/area layers.

layer dissimilarity values; second, the Taxonomy RDM was only a significant predictor for layer dissimilarity values for the face images. Crucially, when it came to the final network layers, any effect of the intuitive taxonomy model was enveloped by the variance components shared with the human face RDM.

In sum, the apparent effect of Taxonomy in these networks cannot be the result of an overt representation of superordinate relationships among animal classes, which further suggests that correlations of the Taxonomy model with neural RDMs from regions of the OTC likewise do not suffice to provide evidence of an intuitive taxonomic organization. Similarly, the observed representational structure of the final FC layers of the networks, which also correlated with the perceptual similarity judgments, further suggests that gradation in the representations of the faces and bodies of animals of different species can emerge simply from a first-order representation of image classes.

Discussion

Animacy is an important organizing principle in the OTC (Behrmann and Plaut, 2013; Grill-Spector and Weiner, 2014). Two conflicting explanations for this organization are suggested in the literature, the face-body division and an intuitive taxonomy continuum. Each factor has been studied extensively. The presence of face versus body selectivity in the OTC is well established, and more recently a number of studies have investigated whether the animacy organization of the OTC might reflect a more nuanced intuitive taxonomic structure. However, ours is the first in which these two important factors have been studied together and dissociated, which is crucial to understand their relationship and relative importance (Bracci et al., 2017). We found that both factors independently explained variance in the dissimilarities between activity patterns in the OTC. When the OTC was partitioned, the same result was also observed in the face-, body-, and even object-selective areas. However, human visual similarity judgments were better predictors than taxonomy when data for face and body images were analyzed separately. Finally, the later layers of DNNs also correlated with the Face-Body and Taxonomy, but not Animacy, models, and the pairwise and human similarity judgments for the face and body images. These results have important implications for the following: (1) the claimed taxonomic organization of the OTC,

(2) whether OTC in fact represents animacy, (3) and whether DNNs distinguish the animacy of objects in images.

No evidence that the occipitotemporal cortex represents an intuitive taxonomy

Previous studies suggest that animacy organization of the OTC may reflect the representation of a continuum, rather than a dichotomy, between animate and inanimate objects. For example, Sha et al., (2015) found that neural dissimilarity in the LOTC showed no such division. Their stimuli formed an intuitive taxonomic hierarchy with humans/primates as the most animate compared with invertebrates, other mammals, birds, and fish in between. Other studies have also found that pattern dissimilarity in the OTC exhibits an intuitive taxonomy for similar collections of animals (Connolly et al., 2012, 2016; Nastase et al., 2017). Similarly, Thorat et al. (2019) report that neural dissimilarity in the VOTC was well captured by judgments of the relative capacity for thoughts and feelings, or agency, although notably these are properties of entity subjectivity not agency (Gray et al., 2007). Still, the resulting ranking of the images was very similar to what one would predict based on an intuitive taxonomic grouping of the object images.

Our results differ from these previous studies in a few ways. First, we consistently found a robust animate-inanimate dichotomy in neural dissimilarity across ROIs, in direct contrast to the results of Sha et al. (2015), Second, no previous studies controlled for the face-body division in the stimulus designs. Therefore, it is significant that the Taxonomy model independently contributes to explaining the neural dissimilarity in the OTC for both face and body images. This finding on its own can be interpreted as evidence of an intuitive taxonomic continuum in the OTC. However, third, we also considered the possibility that the apparent continuum does not reflect representation in the OTC of intuitive taxonomic relations per se but rather graded responses to the images based on coding for diagnostic visual features in face- and body-selective areas of the OTC. We compared the neural dissimilarity of face- and body-selective ROIs to pairwise and human similarity judgments for the face and body images and found that human visual similarity was the dominant predictor over and above the Taxonomy model. Furthermore, layers of DNNs correlated with the Taxonomy model, yet clearly do not represent intuitive hierarchical taxonomic relations among animal stimuli, and the representational structure of the final layers was also better captured by our behavioral measures of visual similarity.

Taken as a whole, these results suggest that the apparent animacy continuum may not reflect the representation of an intuitive taxonomy. Instead, they point to an alternative hypothesis, that is, perhaps there is no intuitive taxonomic organization in the OTC at all. As the OTC is well known to display distributed and differential selectivity for faces and bodies, the apparent continuum effect reflects variation in response of animal faces and bodies based on visual similarity to human faces and bodies, which are considered the most preferential stimuli for the ROIs. Such a possibility is acknowledged by Thorat et al. (2019), and the present results provide some support for this alternative proposal concerning the OTC. Of course, what accounts for the morphologic differences between faces and bodies between species is their evolutionary history, and our intuitive grouping of animals based on their visual features in part reflects this. Nor does our study rule out the possibility that intuitive taxonomy might capture the organization of other portions of the OTC or in the same regions if alternative intuitive taxonomy models are used as could be constructed from a semantic feature model (Clarke and Tyler, 2014; Jozwik et al., 2016). Still, the most parsimonious explanation at present is that the OTC represents the resemblance of animal faces/bodies to human faces/bodies, based on the principle that face and body areas code for diagnostic visual features of region-defining categories and not an intuitive taxonomic continuum per se.

Does the occipitotemporal cortex represent animacy?

Given that the apparent animacy continuum may reflect in part gradation in the face- and body-selectivity throughout the OTC, our results also raise the question of whether the OTC represents animacy at all. It is well known that category-selective areas spatially pool depending on whether they represent animate stimuli with face- and body-selective areas in the lateral VOTC contrasting with medial portions selective for scenes. Based on this division, one popular hypothesis is that animacy is represented at a broader spatial scale, which subsumes areas that represent more specific animate or inanimate categories such as faces or scenes (Grill-Spector and Weiner, 2014; Bao et al., 2020). However, alternatively it is possible that animacy may not be represented at all, although the spatial layout of the areas respects the animate-inanimate division. It might be tempting to consider animacy as a parsimonious explanation for why face and body regions end up being close together, but there are other explanations for this proximity that do not refer to animacy. For example, faces and bodies co-occur in nearby positions in a visual image, and such spatial relationships could result in anatomic proximity (Orlov et al., 2010). Based on our results, it is worth asking whether the apparent representation of animacy in the OTC may simply be a by-product of (principally) strong selectivity for faces and bodies.

Although this deflationary hypothesis is in need of further study, it gains some support from the results of Bracci et al., (2019), who selected trios of animals, artefacts, and look-alike artefacts (e.g., duck, kettle, and duck-shaped kettle) as stimuli. They found that neural dissimilarity in the VOTC correlated with the object appearance (animals and look-alikes vs artefacts) and not the object identity (animals vs look-alikes and artefacts) and proposed that the VOTC does not represent the animate-inanimate division but selectivity for diagnostically important visual features. We would further suggest these features are specifically diagnostic for faces and bodies, as part of a feature-based neural code for object categories (Bracci et al., 2017). Indeed, in a recent study Proklova and Goodale (2020) found that the VOTC did not exhibit a robust animate-inanimate division between patterns of activity for faceless animals versus artifact objects, in further support of our conjecture. At the same time, other studies suggest that animacy organization may not simply reduce to the representation of faces and bodies in the OTC. Neuropsychological results suggest that lesions to the visual cortex can cause selective deficits in naming animals but not body parts (Caramazza and Shelton, 1998). Furthermore, converging behavioral and neural evidence suggests that the OTC may subserve the holistic representation of whole persons, which does not reduce to separate selectivity for faces and bodies (Hu et al., 2020). Such a framework may also run counter to the deflationary hypothesis we propose.

Deep neural networks do not represent object animacy

DNNs are increasingly being used as models of visual processing (Kriegeskorte, 2015; Cichy and Kaiser, 2019; Serre, 2019). The use of DNNs in this capacity has in part been motivated by

similarities in the activity patterns of the later FC layers to regions of the OTC for networks trained on the ImageNet dataset. In particular, several studies using ImageNet-trained networks have reported that FC layers exhibit a similar animacy organization to the OTC (Khaligh-Razavi and Kriegeskorte, 2014; Jozwik et al., 2017; Bracci et al., 2019; Zeman et al., 2020). In light of these previous findings, it is striking that we did not observe a consistent animacy organization across five ImageNet-trained DNNs. Yet, when the data for the face and body images were analyzed separately, we found that the pairwise and human similarity RDM explained most of the variance in the final layer RDMs. So the ImageNet-trained DNNs plausibly do not represent either an intuitive taxonomic continuum or a categorical division between animate and inanimate objects but rather graded representations of face- and body-related visual features. This is also consistent with the training history of these networks with still images, which do not contain direct information about animacy, agency, or self-initiated motion, and it further confirms the visual nature of the human face/body similarity judgments.

Summary and conclusion

The animacy organization of the OTC may reflect either the representation of an intuitive taxonomic hierarchy or selectivity for faces and bodies. We attempted to disentangle these factors. Our results suggest that graded visual selectivity for faces and bodies in the OTC may masquerade as an animacy continuum and that intuitive taxonomy may not be a separate factor underlying the organization of the OTC. In this respect, our results provide new insights into the functional organization of the ventral visual pathway more generally.

References

Andrews TJ, Watson DM, Rice GE, Hartley T (2015) Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. J Vis 15:3.

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

Behrmann M, Plaut DC (2013) Distributed circuits, not circumscribed centers, mediate visual recognition. Trends in cognitive sciences 17:210–219.

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, Op de Beeck H (2016) Dissociations and associations between shape and category representations in the two visual pathways. J Neurosci 36:432–444.

Bracci S, Ritchie JB, Kalfas I, Op de Beeck HP (2019) The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. J Neurosci 39:6513–6525.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Capraro RM, Capraro MM (2001) Commonality analysis: understanding variance contributions to overall canonical correlation effects of attitude toward mathematics on geometry achievement. Multiple Linear Regression Viewpoints 27:16–23.

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

Cichy RM, Kaiser D (2019) Deep neural networks as scientific models. Trends Cogn Sci 23:305–317.

Clarke A, Tyler LK (2014) Object-specific semantic coding in human perirhinal cortex. J Neurosci 34:4766–4775.

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.

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

Connolly AC, Sha L, Guntupalli JS, Oosterhof N, Halchenko YO, Nastase SA, di Oleggio Castello MV, Abdi H, Jobst BC, Gobbini MI, Haxby JV (2016) How the human brain represents perceived dangerousness or "predacity" of animals. J Neurosci 36:5373–5384.

Contini EW, Goddard E, Grootswagers T, Williams M, Carlson T (2020) A humanness dimension to visual object coding in the brain. NeuroImage 1:117–139.

DiCarlo JJ, Johnson KO (1999) Velocity invariance of receptive field structure in somatosensory cortical area 3b of the alert monkey. J Neurosci 19:401–419.

Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. Science 293:2470–2473.

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.

Fairhall SL, Caramazza A (2013) Brain regions that represent amodal conceptual knowledge. J Neurosci 33:10552–10558.

Fischl B (2012) FreeSurfer. Neuroimage 62:774–781.

Goñi J, Arrondo G, Sepulcre J, Martincorena I, Vélez de Mendizábal N, Corominas-Murtra B, Bejarano B, Ardanza-Trevijano S, Peraita H, Wall DP, Villoslada P (2011) The semantic organization of the animal category: evidence from semantic verbal fluency and network theory. Cogn Process 12:183–196.

Gray HM, Gray K, Wegner DM (2007) Dimensions of mind perception. Science 315:619–619.

Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron 24:187–203.

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

Groen II, Greene MR, Baldassano C, Fei-Fei L, Beck DM, Baker CI (2018) Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. Elife 7:e32962.

He K, Zhang X, Ren S, and Sun J (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 IEEE International Conference on Computer Vision (ICCV), pp 1026–1034, Santiago, Chile.

Hebart MN, Bankson BB, Harel A, Baker CI, Cichy RM (2018) The representational dynamics of task and object processing in humans. Elife 7: e32816.

Hu Y, Baragchizadeh A, O'Toole AJ (2020) Integrating faces and bodies: psychological and neural perspectives on whole person perception. Neurosci Biobehav Rev 112:472–486.

Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: Convolutional architecture for fast feature embedding. Paper presented at the 22nd Association for Computing Machinery International Conference on Multimedia, Orlando, Florida, November.

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.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Kerby DS (2014) The simple difference formula: An approach to teaching nonparametric correlation. Comprehensive Psychology 3:11-IT.

Khaligh-Razavi SM, Kriegeskorte N (2014) Deep supervised, but not unsupervised, models may explain IT cortical representation. PLoS Comput Biol 10:e1003915.

Kriegeskorte N (2015) Deep neural networks: a new framework for modeling biological vision and brain information processing. Annu Rev Vis Sci 1:417–446.

Kriegeskorte N, Kievit RA (2013) Representational geometry: integrating cognition, computation, and the brain. Trends Cogn Sci 17:401–412.

Kriegeskorte N, Mur M, Bandettini PA (2008a) Representational similarity analysis-connecting the branches of systems neuroscience. Front Syst Neurosci 2:4.

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

Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Communications of the ACM 60:84–90.

Kubilius J, Schrimpf M, Nayebi A, Bear D, Yamins DL, DiCarlo JJ (2018) Cornet: modeling the neural mechanisms of core object recognition. BioRxiv 408385.

Lescroart MD, Stansbury DE, Gallant JL (2015) Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. Front Comput Neurosci 9:135.

Long B, Yu CP, Konkle T (2018) Mid-level visual features underlie the high-level categorical organization of the ventral stream. Proc Natl Acad Sci USA 115:E9015–E9024.

McCarthy G, Puce A, Gore JC, Allison T (1997) Face-specific processing in the human fusiform gyrus. J Cogn Neurosci 9:605–610.

Micallef L, Rodgers P (2014) eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. PloS One 9:e101717.

Nastase SA, Connolly AC, Oosterhof NN, Halchenko YO, Guntupalli JS, Visconti di Oleggio Castello M, Gors J, Gobbini MI, Haxby JV (2017) Attention selectively reshapes the geometry of distributed semantic representation. Cereb Cortex 27:4277–4291.

Newton RG, Spurrell DJ (1967) A development of multiple regression for the analysis of routine data. J R Stat Soc Ser C Appl Stat 16:51–64.

Nimon K, Reio TG Jr, (2011) Regression commonality analysis: a technique for quantitative theory building. Human Resource Development Review 10:329–340.

Oliva A, Torralba A (2001) Modeling the shape of the scene: a holistic representation of the spatial envelope. Int J Comput Vis 42:145–175.

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 HP, Deutsch JA, Vanduffel W, Kanwisher NG, DiCarlo JJ (2008) A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. Cereb Cortex 18:1676–1694.

Orlov T, Makin TR, Zohary E (2010) Topographic representation of the human body in the occipitotemporal cortex. Neuron 68:586–600.

Peelen MV, Downing PE (2005) Selectivity for the human body in the fusiform gyrus. J Neurophysiol 93:603–608.

Peirce JW (2007) PsychoPy—psychophysics software in Python. J Neurosci Methods 162:8–13.

Proklova D, Goodale MA (2020) The role of animal faces in the animate-inanimate distinction in the ventral temporal cortex. bioRxiv 2020.10.08.330639.

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.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. Int J Comput Vis 115:211–252.

Schrimpf M, Kubilius J, Hong H, Majaj NJ, Rajalingham R, Issa EB, Kar K, Pouya Bashivan P, Prescott-Roy J, Schmidt K, Yamins DLK, DiCarlo JJ (2018) 3020 Brain-score: Which artificial neural network for object recognition is most brain-like? BioRxiv 407007.

Seibold DR, McPhee RD (1979) Commonality analysis: a method for decomposing explained variance in multiple regression analyses. Human Comm Res 5:355–365.

Serre T (2019) Deep learning: the good, the bad, and the ugly. Annu Rev Vis Sci 5:399–426.

Sha L, Haxby JV, Abdi H, Guntupalli JS, Oosterhof NN, Halchenko YO, Connolly AC (2015) The animacy continuum in the human ventral vision pathway. J Cogn Neurosci 27:665–678.

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556.

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke D, Rabinovich A (2015) Going deeper with convolutions. Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, June.

Thorat S, Proklova D, Peelen MV (2019) The nature of the animacy organization in human ventral temporal cortex. Elife 8:e47142.

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.

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.

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.