This is a repository copy of *Disentangling Representations of Object Shape and Object Category in Human Visual Cortex : The Animate-Inanimate Distinction*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/152954/

Version: Published Version

Article:

Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate–Inanimate Distinction

Daria Proklova*, Daniel Kaiser*, and Marius V. Peelen

Abstract

■ Objects belonging to different categories evoke reliably different fMRI activity patterns in human occipitotemporal cortex, with the most prominent distinction being that between animate and inanimate objects. An unresolved question is whether these categorical distinctions reflect category-associated visual properties of objects or whether they genuinely reflect object category. Here, we addressed this question by measuring fMRI responses to animate and inanimate objects that were closely matched for shape and low-level visual features. Univariate contrasts revealed animate- and inanimate-preferring regions in ventral and lateral temporal cortex even for individually matched object pairs (e.g., snake–rope). Using representational similarity analysis, we mapped out brain regions in which the pairwise dissimilarity of multivoxel activity patterns (neural dissimilarity) was predicted by the objects' pairwise visual dissim-

ilarity and/or their categorical dissimilarity. Visual dissimilarity was measured as the time it took participants to find a unique target among identical distractors in three visual search experiments, where we separately quantified overall dissimilarity, outline dissimilarity, and texture dissimilarity. All three visual dissimilarity structures predicted neural dissimilarity in regions of visual cortex. Interestingly, these analyses revealed several clusters in which categorical dissimilarity predicted neural dissimilarity after regressing out visual dissimilarity. Together, these results suggest that the animate–inanimate organization of human visual cortex is not fully explained by differences in the characteristic shape or texture properties of animals and inanimate objects. Instead, representations of visual object properties and object category may coexist in more anterior parts of the visual system. ■

INTRODUCTION

Large-scale patterns of fMRI activity spanning the ventral temporal cortex (VTC) distinguish animate from inanimate object categories (e.g., Kriegeskorte, Mur, Ruff, et al., 2008), with animate objects evoking higher BOLD responses in lateral VTC and inanimate objects evoking higher BOLD responses in medial VTC (e.g., Mahon et al., 2007; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Chao, Haxby, & Martin, 1999). Within these broader regions, focal regions exhibit selective responses to more specific categories, including regions selective for buildings and scenes, faces, tools, body parts, and words (Peelen & Downing, 2005; Cohen & Dehaene, 2004; Downing, Jiang, Shuman, & Kanwisher, 2001; Chao et al., 1999; Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 1997). Although the selectivity for object categories in VTC has been widely replicated, particularly the animate–inanimate distinction, the factors driving this selectivity are still under debate (Andrews, Watson, Rice, & Hartley, 2015; Grill-Spector & Weiner, 2014; Mahon & Caramazza, 2011; Op de Beeck, Haushofer, & Kanwisher, 2008; Martin, 2007).

One of the key questions is whether category-specific patterns of brain activity reflect genuine categorical distinctions (Caramazza & Shelton, 1998) or whether these can be alternatively explained by factors that covary with category membership, such as shape properties. Because of the close association between certain visual properties and category membership, it is to be expected that category-selective regions are optimized for processing these visual properties and/or that these regions are located in parts of the visual system that have visual and retinopic biases that are optimal for processing the visual features that are characteristic of the category. However, although specific visual properties often characterize object categories, these two dimensions (visual, categorical) are not identical and can indeed be experimentally dissociated. For example, although most tools are elongated, this shape property can be dissociated from the conceptual properties associated with tools (e.g., that tools are manipulable and used as effectors; Bracci & Peelen, 2013). For a visually more homogenous category such as animals, this distinction is more challenging but may still be addressed by testing responses to visually less typical examples (e.g., snakes) and, conversely, testing responses to inanimate objects that share visual features with animals (e.g., mannequins, dolls, statues). These considerations raise the intriguing question of whether category selectivity in VTC

University of Trento
*Shared first authorship.

reflects selectivity for conceptual category or selectivity for visual properties that characterize a category.

According to the object form topology account, category-selective fMRI responses in VTC reflect the activation of object form representations that are mapped onto VTC in a continuous fashion (Haxby, Ishai, Chao, Ungerleider, & Martin, 2000; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999). The selective response to animals in VTC may thus arise from selectivity for characteristic animal shape(s) rather than selectivity for animacy per se. A recent monkey study provided support for this hypothesis, showing that the organization of animate and inanimate object representations in monkey inferotemporal cortex primarily reflects visual similarity rather than semantic similarity (Baldassi et al., 2013; but see Kiani, Esteky, Mirpour, & Tanaka, 2007). Further support for the visual similarity account comes from fMRI studies showing that category-selective regions in VTC respond selectively to visual properties that are characteristic of the regions' preferred categories, even for otherwise meaningless stimuli (i.e., in the absence of category recognition). For example, the fusiform face area was shown to respond more strongly to oval shapes with a greater number of black elements in the top half than to oval shapes with a greater number of elements in the bottom half, although none of these stimuli were recognized as faces (Caldara et al., 2006). Similarly, the parahippocampal place area, located within the medial inanimate-preferring VTC, was shown to respond preferentially to objects made up of cardinal orientations and right angles, features typical of manmade objects, buildings, and scenes (Nasr, Echavarria, & Tootell, 2014).

Recent evidence against a "visual properties" account of category selectivity in VTC comes from studies in congenitally blind individuals. These individuals, with no visual experience, show a categorical organization of VTC that is remarkably similar to that observed in sighted individuals (Ricciardi, Bonino, Pellegrini, & Pietrini, 2014). For example, aurally presented words describing large inanimate objects, versus animals, activate medial VTC in both blind and sighted groups (He et al., 2013; Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009). Using a variety of presentation methods, most of the category-selective VTC regions found in sighted individuals have now also been reported in blind individuals, often at nearly identical anatomical locations in the two groups (Striem-Amit & Amedi, 2014; Peelen et al., 2013; Reich, Szwed, Cohen, & Amedi, 2011; Wolbers, Klatzky, Loomis, Wutte, & Giudice, 2011; Buchel, Price, & Friston, 1998). These studies show that the processing of visual features is not necessary for some category-selective responses to develop. However, they do not exclude the possibility that category selectivity in VTC nevertheless reflects shape properties of objects. This is because VTC has been shown to extract object shape from nonvisual input modalities (Amedi et al., 2007; Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005), with VTC activity patterns reflecting the shape similarity of objects in both

blind and sighted groups (Peelen, He, Han, Caramazza, & Bi, 2014).

This study was designed to investigate the contribution of visual similarity in the representation of animate and inanimate object categories in VTC. Participants viewed pictures of a variety of animals that systematically differed in their shape, grouping into four shape clusters (Figure 2A, right). Importantly, inanimate control objects were selected to closely match the animals in terms of their shape, following the same four shape clusters. This design allowed us to test whether animate- and inanimate-preferring regions (localized with a standard functional localizer) maintain their selectivity for carefully matched animate–inanimate pairs (e.g., snake vs. rope) and whether this is true for a variety of animals (e.g., birds, insects, reptiles) and inanimate objects (e.g., plane, rope, pine cone). In addition to analyses measuring activation differences, we used representational similarity analysis (RSA) to map out regions in which neural similarity reflected the objects' visual and/or categorical similarity (animate/inanimate). For this purpose, we quantified pairwise visual similarity using visual search tasks designed to measure different aspects of visual similarity (overall visual similarity, outline similarity, and texture similarity; Figure 2).

METHODS

Participants

Eighteen participants (seven men; mean age = 25 years, $SD = 2.4$ years) were scanned at the Center for Mind/Brain Sciences of the University of Trento. All participants gave informed consent. All procedures were carried out in accordance with the Declaration of Helsinki and were approved by the ethics committee of the University of Trento. One participant was excluded from all analyses because of excessive head movement.

Stimuli

The stimuli of the main experiment were organized into four sets of four objects. The four objects within each set all had a roughly similar shape. Two objects of each set were animate, and two objects were inanimate (see Figures 2A and 4). In addition, there were four exemplars of each object (e.g., four images of a snake), resulting in 16 stimuli per set and a total of 64 stimuli. All images were gray scaled, placed on gray background and matched for luminance and contrast using the SHINE toolbox (Willenbockel et al., 2010). Stimulus presentation was controlled using the Psychtoolbox (Brainard, 1997). Images were back-projected on a translucent screen placed at the end of the scanner bore. Participants viewed the screen through a tilted mirror mounted on the head coil. Stimuli were presented foveally and subtended a visual angle of approximately 4.5°.

Visual Search Experiments

To provide a measure of pairwise visual similarity of the stimulus set, a series of three behavioral visual search experiments was conducted. In these experiments, participants searched for an oddball target surrounded by identical distractor objects (Figure 2). The response time in this task is a measure of visual similarity (Mohan & Arun, 2012): the longer the response time for locating the oddball stimulus, the more visually similar are the target and the distractor object. Experiment 1 measured overall visual similarity, Experiment 2 measured outline visual similarity, and Experiment 3 measured texture visual similarity.

Experiment 1

To quantify overall pairwise visual similarity of the stimulus set, 18 new participants were tested in a behavioral experiment (two men; mean age = 22.5 years, SD = 2.97 years). Stimuli were presented on a 17-in. CRT monitor, and presentation was controlled using Psychtoolbox (Brainard, 1997). Each search display contained 16 objects placed in a 4×4 grid, with one oddball target and 15 identical images of the distractor object. The location of the 16 objects in the grid was randomized. The size of the target and seven of the distractors was 100×100 pixels, which corresponded to 2.9° visual angle. The remaining distractors differed in size, with four being 120% of the target size (3.4° visual angle) and four being 80% of the target size (2.3° visual angle). Participants had to indicate whether the oddball target appeared on the left side or on the right side of the screen. No information about the category of the oddball target was provided. The search display remained on the screen until the response, followed by 500-msec fixation, after which the next trial started. The experiment consisted of four blocks. In each block, only one of the four exemplars of each object was used (e.g., always the same snake within one block), resulting in 16 unique objects and 240 trials per block (for all possible target–distractor pairings of the 16 stimuli). Accuracy was high (97.2%) and was not further analyzed.

RTs were averaged across corresponding target–distractor object pairs and across blocks. The data from the visual search experiment served to create a matrix of overall pairwise visual dissimilarity, to be used as a predictor in the fMRI analysis. For this purpose, we took the inverse of these RTs (1/RT) for each stimulus pair as a measure of dissimilarity. The resulting visual dissimilarity matrix consisted of one visual dissimilarity value for every pairwise combination of objects (Figure 2A, center). Multidimensional scaling analysis (using cmdscale function in MATLAB) revealed that stimuli from the same shape sets clustered together (Figure 2A, right), whereas there was no apparent categorical organization. Furthermore, within each shape set, there was no evidence for categoricality, with the average visual dissimilarity within categories

(e.g., snake–snail) being equal to the average visual dissimilarity across categories (e.g., rope–snail), $t(17)$ = 1.22, p = .239. These results confirm our intuitive shape sets and show that there were no obvious visual properties that covaried with category membership.

Experiments 2 and 3

To measure pairwise similarity of the outline shape of the stimuli, 18 new participants (three men; mean age = 23.3 years, SD = 3.4 years) were tested in Experiment 2. The experiment was identical to Experiment 1 except that the stimulus set consisted of outline drawings of the stimuli, created by automatically tracing the outline contours of binarized silhouette versions of the original stimuli (see Figure 2B). Accuracy was high (98.1%) and was not further analyzed.

To measure pairwise texture similarity of the stimulus set, 18 participants (eight men; mean age = 25.7 years, SD = 5.5 years) were tested in Experiment 3. Two of the participants had also participated in Experiment 2. The experiment was identical to Experiments 1 and 2 except that the stimulus set consisted of circular texture patches, created by masking the original images with a circular aperture that covered about 20% of the image (see Figure 2C). The aperture was centered on the mean pixel coordinate of the image. Any blank spaces were filled in using the clone stamp tool in Photoshop. This circular masking abolishes outline shape information, while leaving inner features and texture properties of the stimuli largely intact. It should be noted that these patches still contained some local structure (e.g., the inner contour of the snake) and may thus capture some internal shape features in addition to texture properties. Accuracy was high (98.1%) and was not further analyzed.

Data for Experiments 2 and 3 were analyzed as in Experiment 1, resulting in dissimilarity matrices representing outline and texture dissimilarity of the stimulus set (Figure 2B and C, center). Further analyses showed that overall visual dissimilarity could be nearly perfectly predicted by a linear combination of outline and texture dissimilarity, with the optimal weights being 0.75 and 0.25, respectively. Linear combinations of dissimilarity matrices were computed by using data from different single participants (e.g., 0.75 × outline of one participant + 0.25 × texture of another participant), each of which was then correlated with the average visual dissimilarity matrix (with one participant left out). The resulting average correlation (r = .77) approached the noise ceiling of visual dissimilarity (r = .82, computed by correlating each participant's visual dissimilarity matrix with the group-averaged visual dissimilarity matrix, leaving out this participant). Finally, the combined model was significantly more strongly correlated with visual dissimilarity than was either outline or texture alone (p < .001, for both comparisons). These analyses show that overall visual dissimilarity is influenced both by outline and texture

properties, which together almost fully explain overall visual dissimilarity.

Main fMRI Experiment Procedure

The main fMRI experiment consisted of eight runs. Each run consisted of 80 trials that were composed of 64 object trials and 16 fixation-only trials. In object trials, a single stimulus was presented for 300 msec, followed by a 3700-msec fixation period (Figure 1A). In each run, each of the 64 images appeared exactly once. In fixation-only trials, the fixation cross was shown for 4000 msec. Trial order was randomized, with the constraints that there were exactly eight 1-back repetitions of the same category (e.g., two snakes in direct succession) within the object trials and that there were no two fixation trials appearing in direct succession. Each run started and ended with 16-sec fixation period, leading to a total run duration of 5.9 min. Participants were instructed to press a button whenever they detected a 1-back repetition.

Functional Localizer Experiment Procedure

In addition to the main experiment, participants completed one run of a functional localizer experiment. During the localizer, participants viewed grayscale pictures of 36 animate and 36 inanimate stimuli in a block design (Figure 1B). Animate stimuli included five different types of animals (mammals, birds, fish, reptiles, and insects). Inanimate stimuli included five types of inanimate objects (cars, chairs, musical instruments, tools, and weapons). These stimuli were not matched for their shape (thus, this design resembled the standard animate–inanimate contrast used in previous studies). Each block lasted 16 sec, containing 20 stimuli that were each presented for

400 msec, followed by 400-msec blank interval. There were eight blocks of each stimulus category and four fixation-only blocks per run. The order of the first 10 blocks was randomized and then mirror reversed for the other 10 blocks. Participants were asked to detect 1-back image repetitions, which happened twice during every nonfixation block.

fMRI Acquisition

Imaging data were acquired using a Bruker Biospin MedSpec 4-T head scanner (Bruker Biospin), equipped with an eight-channel head coil. For functional imaging, T2*-weighted EPIs were collected (repetition time = 2.0 sec, echo time = 33 msec, 73° flip angle, $3 \times 3 \times 3$ mm voxel size, 1-mm gap, 34 slices, 192-mm field of view, 64×64 matrix size). A high-resolution T1-weighted image (magnetization prepared rapid gradient echo; $1 \times 1 \times 1$ mm voxel size) was obtained as an anatomical reference.

fMRI Preprocessing and Modeling

The neuroimaging data were analyzed using MATLAB and SPM8. During the preprocessing, the functional volumes were realigned, coregistered to the structural image, resampled to a $2 \times 2 \times 2$ mm grid, and spatially normalized to the Montreal Neurological Institute 305 template included in SPM8. For the univariate analysis, the functional images were smoothed with a 6-mm FWHM kernel, whereas for the multivariate analysis, the images were left unsmoothed. For the main experiment, the BOLD signal of each voxel in each participant was modeled using 22 regressors in a general linear model, with 16 regressors for each of the objects (e.g., one regressor for all snakes) and six regressors for the movement parameters obtained from the realignment procedure. For the functional

Figure 1. fMRI paradigm. (A) In the main fMRI experiment, participants viewed images of animate objects and shape-matched inanimate objects (see Figure 4 for further stimulus examples). Trial order was randomized, and participants detected, by button press, 1-back object-level repetitions (here, two ladybugs). (B) In the functional localizer experiment, participants viewed blocks of animate (top sequence) and inanimate (bottom sequence) stimuli. All stimuli were different from the ones used in the main experiment. Each block lasted 16 sec, and participants detected, by button press, 1-back image-level repetitions (here, the fish image).
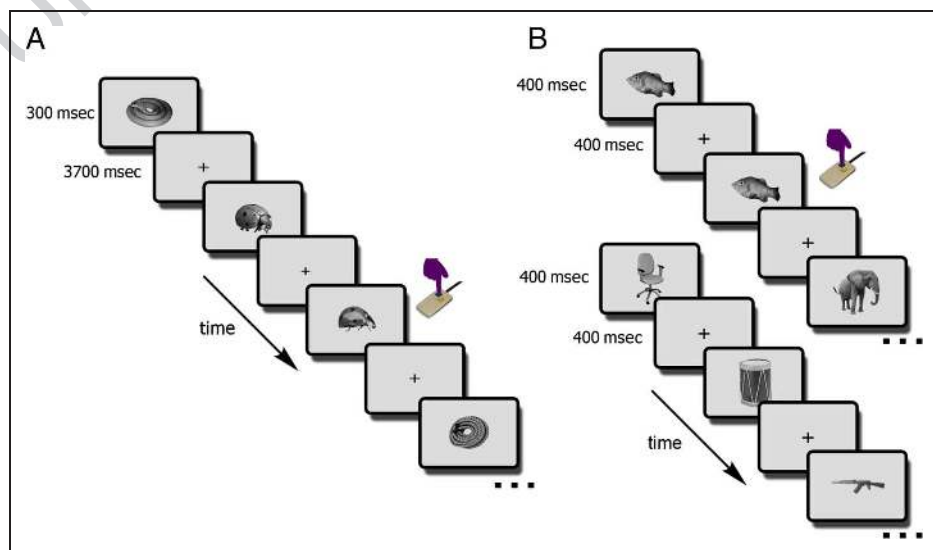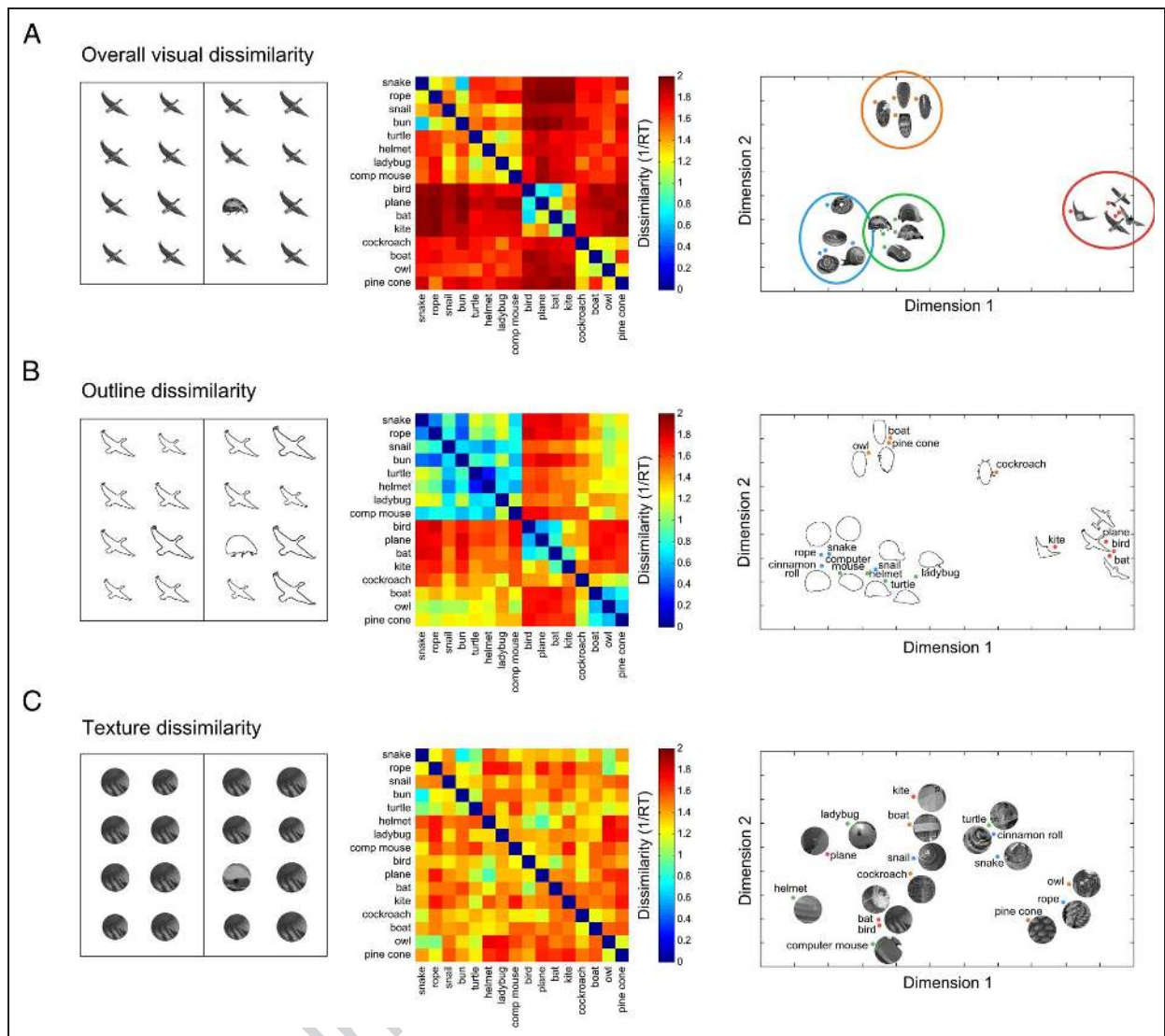
Figure 2. Visual search experiments. In three visual search experiments, participants indicated whether an oddball target in a 4 × 4 search array (left) was located to the right or to the left of the vertical display midline. No prior information about the target was given, so that participants had to rely on bottom–up visual differences to perform the task. From the RTs in these experiments, we created visual dissimilarity matrices (center), where each element represents the inverse RT for a pair of stimuli, averaged across the two respective target–distractor pairings: High dissimilarity values thus correspond to short RTs, reflecting that target and distractor were visually dissimilar. Multidimensional scaling representations of the three visual dissimilarity matrices are shown in the right. The three experiments differed only in the stimuli used: (A) the original images used in the fMRI experiment, (B) outline drawings of these images, and (C) internal textures of these images.

localizer data, the signal was modeled using two regressors (animate and inanimate objects) and six movement regressors. All models included an intrinsic temporal high-pass filter of 1/128 Hz to correct for slow scanner drifts.

Univariate Analysis

Univariate random effects whole-brain analyses were performed separately for the localizer and the main experiment, contrasting animate with inanimate objects. Statistical maps were thresholded using a voxel-level threshold of $p < .001$ (uncorrected) and a cluster-level threshold of $p < .05$ (family-wise error [FWE] corrected). In addition, regions activated in the localizer were defined as ROIs. Within these ROIs, beta estimates for the conditions of the main experiment were extracted and averaged across the voxels of each ROI. These beta values were statistically compared using ANOVAs and t tests.

RSA

RSA (Kriegeskorte, Mur, & Bandettini, 2008) was used to relate the visual and categorical similarity of our objects to neural similarity. RSA was performed throughout the

whole brain using searchlight analysis (Kriegeskorte, Goebel, & Bandettini, 2006), implemented in the CoSMoMVPA software package (www.cosmomvpa.org). Each spherical searchlight neighborhood consisted of 100 voxels, centered on every voxel in the brain. For each of these spheres, we correlated the activity (beta values) between each pair of conditions from the main experiment across the voxels of the sphere, leading to a 16×16 symmetrical correlation matrix with an undefined diagonal. This matrix was transformed into a neural dissimilarity matrix by subtracting the correlation values from 1.

In a first analysis, neural dissimilarity matrices were related to the visual dissimilarity matrix and the categorical dissimilarity matrix using multiple regression analysis (see Figure 5A). The visual dissimilarity matrix was derived from RTs in a visual search experiment (Experiment 1; Figure 2A), whereas the categorical dissimilarity matrix reflected whether two objects were from the same category (0) or from different categories (1). All dissimilarity matrices were z normalized. The multiple regression analysis yielded beta estimates for the two predictors of neural dissimilarity (visual and categorical dissimilarity), reflecting the independent contributions of these predictors in explaining neural dissimilarity. These two beta estimates were obtained for all spheres, resulting in two whole-brain maps for each participant. These maps were then tested against zero using random effects analyses (t tests), thresholded using a voxel-level threshold of $p < .001$ (uncorrected) and a cluster-level threshold of $p < .05$ (FWE corrected). In the second analysis, neural dissimilarity matrices were related to outline dissimilarity (Figure 2B), texture dissimilarity (Figure 2C), and categorical dissimilarity (Figure 6A). In all other respects, the analysis was the same as the first analysis described above.

RESULTS

Univariate Results

Whole-brain Analysis

The contrast between animate and inanimate objects in the functional localizer experiment revealed a characteristic medial-to-lateral organization in VTC (Figure 3A). In line with previous findings, animate stimuli more strongly activated regions around the lateral fusiform gyrus (left hemisphere [LH]: 1936 mm³, peak Montreal Neurological Institute coordinates: $x = -40$, $y = -48$, $z = -22$; right hemisphere [RH]: 3424 mm³, peak coordinates: $x = 42$, $y = -52$, $z = -20$), and inanimate stimuli preferentially activated more medial regions around parahippocampal gyrus (LH: 3760 mm³, peak coordinates: $x = -28$, $y = -46$, $z = -12$; RH: 3184 mm³, peak coordinates: $x = 34$, $y = -42$, $z = -12$). In addition to these ventral regions, animate stimuli preferentially activated a more posterior and lateral region, around middle temporal gyrus (LH: 3528 mm³, peak coordinates: $x = -48$, $y =$

-80, $z = 0$; RH: 7800 mm³, peak coordinates: $x = 52$, $y = -74$, $z = -2$).

The same animate–inanimate comparison was performed for the main experiment, in which the animate and inanimate stimuli were closely matched for shape and low-level visual features (see Methods). This contrast revealed a significant animacy organization: Animate stimuli more strongly activated two clusters in the RH, again around fusiform gyrus (808 mm³, peak coordinates: $x = 42$, $y = -52$, $z = -20$) and middle temporal gyrus (1120 mm³, peak coordinates: $x = 50$, $y = -74$, $z = 0$). Similar to the functional localizer results, inanimate-preferring regions were found around bilateral parahippocampal gyrus (LH: 4736 mm³, peak coordinates: $x = -26$, $y = -52$, $z = -16$; RH: 3456 mm³, peak coordinates: $x = 24$, $y = -40$, $z = -16$). Figure 3A shows the animate- and inanimate-preferring clusters from the localizer and the main experiment as well as their overlap.

These results indicate that the medial-to-lateral animacy organization is also found when controlling for shape differences of animate and inanimate objects. As can be seen in Figure 3A, activity was stronger in the localizer than in the main experiment. This effect is hard to interpret, however, given the many differences between the localizer and the main experiment (e.g., block design vs. event-related design, stimulus duration, the specific animals and objects included, etc.). Indeed, the purpose of this analysis was not to compare the strength of activity between localizer and main experiment directly but to show that the medial-to-lateral organization is remarkably similar in both experiments. Finally, although some of the LH clusters did not survive multiple comparisons correction in the main experiment, the functionally localized ROIs maintained their selectivity in the main experiment in both hemispheres, as reported in the next section.

ROI Analysis

ROI analyses were used to test for selectivity for the conditions in the main experiment within each of the six clusters of the functional localizer (Figure 3A; for coordinates and cluster sizes, see the whole-brain analysis section above). A three-way ANOVA with the factors Animacy (animate, inanimate), Region (animate lateral, animate ventral, inanimate ventral), and Hemisphere (right, left) revealed a critical Region × Animacy interaction ($F(2, 32) = 42.2$, $p < .001$). Because there were no interactions with hemisphere ($F < 2.49$, $p > .10$, for all tests), data were collapsed across hemispheres for all follow-up analyses. Separate Region × Animacy ANOVAs for every pair of regions revealed that the animacy preferences of the two animate-preferring regions were each significantly different from that of the inanimate-preferring region (ventral animate region: $F(1, 16) = 89.24$, $p < .001$; lateral animate region: $F(1, 16) = 45.65$, $p < .001$). There was no significant difference in animacy preference between the two animate regions ($F(1, 16) = 0.59$, $p = .45$). As expected,

both animate-preferring regions showed an increased response to the animate stimuli in the main experiment ($t(16) > 3.53$, $p < .003$, for both tests), whereas the inanimate-preferring region showed a significant preference in favor of the inanimate stimuli ($t(16) = 5.25$, $p < .001$). These results show that all regions defined in the functional localizer maintained their selectivity in the main experiment (see Figure 3B for results in separate hemispheres).

To explore whether the animate–inanimate organization in the main experiment was driven by some of the stimuli preferentially, we next compared the responses to each of the eight individual animate stimuli and their shape-matched inanimate counterparts. For simplicity and for optimal statistical power, the conditions were recoded into "preferred" (e.g., a snake for an animate-peferring region) and "nonpreferred" conditions (e.g., a snake for an inanimate-preferring region), and responses were then averaged across ROIs, so that each ROI contributed equally. A two-way ANOVA with the fac-

tors Preference (preferred, nonpreferred) and Object pair (the eight different pairs) revealed a significant interaction ($F(7, 112) = 3.68$, $p = .001$), indicating that different object pairs differentially contributed to the observed animacy organization (Figure 4). A significant category preference was observed for six of the eight pairs ($t(16) > 2.16$, $p < .047$, for all tests), with one pair (ladybug–computer mouse) showing a trend ($t(16) = 2.10$, $p = .052$) and one pair (snail–bun) not reaching significance ($t(16) = 0.22$, $p = .83$).

RSA

In addition to showing overall differences in focal regions, objects of different categories evoke distinct multivoxel activity patterns in visual cortex: Within VTC, activity patterns to objects from the same category are more similar than activity patterns to objects from different categories (Haxby et al., 2001). An open question is
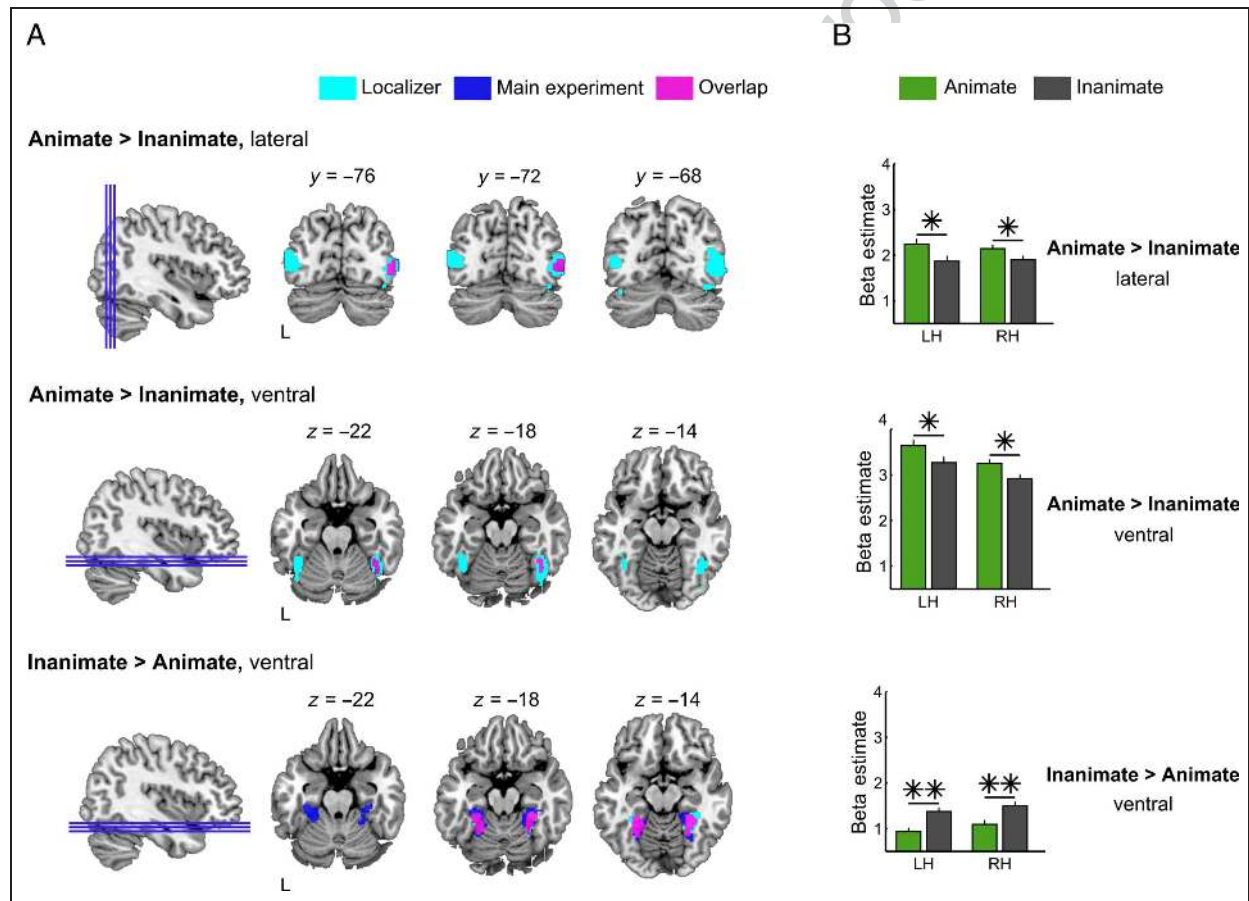

Figure 3. Univariate results. (A) Results of univariate whole-brain group analysis comparing animate and inanimate conditions in the functional localizer experiment (cyan) and main experiment (blue). Overlapping regions are displayed in purple. Top row shows lateral animate-preferring clusters, middle row shows ventral animate-preferring clusters, and bottom row shows ventral inanimate-preferring clusters. Statistical maps were thresholded at $p < .05$ (FWE corrected) and overlaid on a structural brain template (MRIcron). (B) Results of ROI analyses. The bar graphs show the responses to the conditions in the main experiment within the animate- and inanimate-preferring clusters defined based on the functional localizer experiment (corresponding to the cyan regions in A), separately for each hemisphere. Green bars indicate average response to animate objects; gray bars indicate average response to inanimate objects. Error bars reflect *SEM* difference. *$p < .01$, **$p < .001$.

Figure 4. Individual pair analysis. Category preference analyzed for each shape-matched pair separately. Responses were combined across all regions defined in the localizer experiment (cyan regions in Figure 3A) by recoding responses according to the preference of the region (e.g., snake response in animate-preferring regions and rope response in inanimate-preferring region both contribute to the "preferred" condition, shown in red). Error bars reflect *SEM* difference. †p = .052, *p < .05, **p < .001.
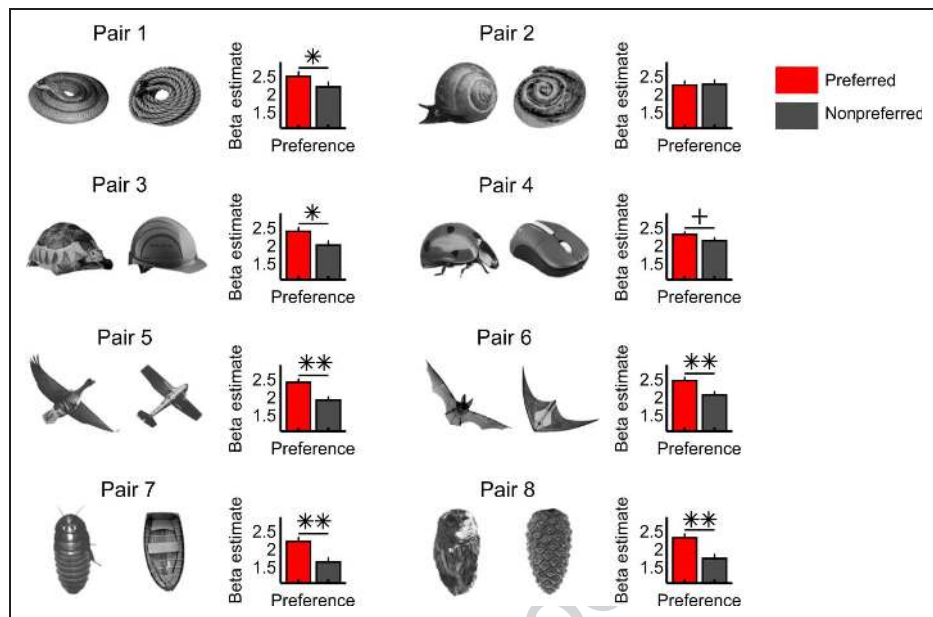
whether these effects reflect visual differences and/or categorical differences between objects. To address this question, we used RSA to relate neural dissimilarity (based on correlations between multivoxel activity patterns) to visual and categorical dissimilarity. In this analysis, for every spherical neighborhood (100 voxels) of the brain, the pairwise neural dissimilarity structure between the 16 objects was modeled as a linear combination of their pairwise categorical dissimilarity and overall visual dissimilarity (Figure 5A; see Methods). Overall visual dissimilarity was quantified using response times in a visual search task (Figure 2A), capturing all contributing factors to visual discriminability (potentially beyond the ones we explicitly matched in the univariate comparisons). This analysis yielded two beta maps reflecting the independent contributions of visual and categorical variables in accounting for neural dissimilarity.

Random effects group analysis revealed widespread clusters of voxels in which neural dissimilarity was significantly related to overall visual dissimilarity, including primary visual cortex and large parts of extrastriate visual cortex, extending into VTC (51,864 mm³ in total; peak in the lingual gyrus, $x = -18, y = -92, z = -8$). These clusters partly overlapped with the animate- and inanimate-preferring regions of the functional localizer experiment (cyan regions in Figure 3A), with visual dissimilarity being significant in 64% of the ventral animate-preferring voxels, 24% of the lateral animate-preferring voxels, and 3% of the ventral inanimate-preferring voxels. Interestingly, categorical dissimilarity was independently reflected in two clusters: one in the right ventral visual cortex (3512 mm³; peak in the fusiform gyrus, $x = 42, y = -60, z = -18$) and one in the LH (4400 mm³ in total, including a lateral visual cortex part with a local peak in middle occipital gy-

rus, $x = -42, y = -80, z = 6$, and a ventral visual cortex part with a local peak in fusiform gyrus, $x = -44, y = -52, z = -16$). As can be seen in Figure 5B, these clusters partly overlapped with the anterior end of the overall visual dissimilarity clusters. These clusters also partly overlapped with the animate- and inanimate-preferring regions of the functional localizer experiment, with categorical dissimilarity being significant in 33% of the ventral animate-preferring voxels, 10% of the lateral animate-preferring voxels, and 2% of the ventral inanimate-preferring voxels.

It is possible that the measure of overall visual dissimilarity captured some visual features better than others. For example, the overall visual dissimilarity structure could be driven more by the outline shape than by texture properties of the stimuli. In this case, if animate and inanimate stimuli consistently differed in their texture, the category-selective regions revealed in the previous analysis could in principle reflect texture information rather than category information. To address this possibility, we quantified the pairwise dissimilarity structure for outline shape and texture independently from each other in two further visual search experiments (see Methods). Using these data, we repeated the RSA, this time modeling pairwise neural dissimilarity using the combination of three predictors: outline shape dissimilarity (Figure 2B), texture dissimilarity (Figure 2B), and categorical dissimilarity (see Figure 6A). This analysis resulted in three beta maps reflecting the independent contributions of outline shape, texture, and category membership to the neural dissimilarity structure.

The results of random effects group analyses on these three maps are shown in Figure 6B. Four widespread clusters of voxels in which neural dissimilarity reflected

outline dissimilarity were found in primary visual cortex extending to extrastriate visual cortex (22,584 mm³, occipital and temporal lobe; peak around the left lingual gyrus, $x = -22$, $y = -92$, $z = -8$), right superior parietal cortex (peak $x = 16$, $y = -50$, $z = 60$), right ventral visual cortex (1552 mm³; peak $x = 44$, $y = -64$, $z = -12$), and left temporal lobe around fusiform gyrus (1240 mm³; peak in BA 37, $x = -28$, $y = -46$, $z = -18$). Texture dissimilarity was related to neural dissimilarity in one cluster of voxels in left early visual cortex (2456 mm³; peak in the lingual gyrus, $x = -20$, $y = -84$, $z = -14$). Crucially, categorical dissimilarity was independently reflected in two clusters: one in right ventral visual cortex (2928 mm³; peak in the fusiform gyrus, $x = 42$, $y = -60$, $z = -18$) and one in left lateral visual cortex (3960 mm³; peak in the middle occipital gyrus, $x = -42$, $y = -80$, $z = 6$). Adding overall visual dissimilarity

to this analysis as fourth predictor revealed nearly identical category clusters, as would be expected based on the finding that a linear combination of outline and texture dissimilarity nearly perfectly captured overall visual dissimilarity (see Methods), thus making this variable redundant as additional predictor.

In summary, these results reveal distinct but overlapping representations for outline shape and texture in early visual cortex. Importantly, categorical representations in higher level visual cortex were still present even after regressing out both outline and texture dissimilarity.

DISCUSSION

In this study, we asked whether the animate–inanimate organization of object responses in human VTC reflects characteristic visual properties of animate and inanimate
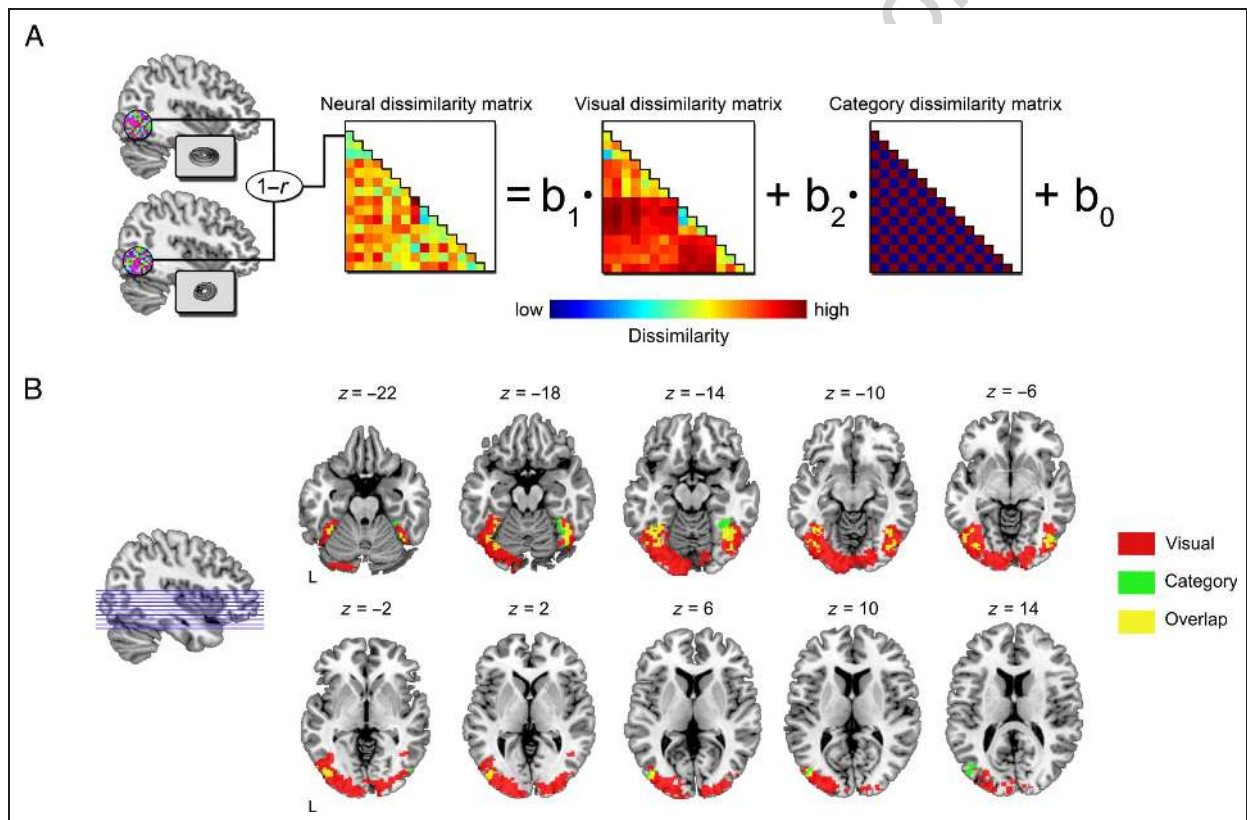

Figure 5. Multivariate searchlight analysis: overall visual dissimilarity and categorical dissimilarity. (A) Schematic of the linear modeling approach. For every spherical searchlight neighborhood, a 16 × 16 neural dissimilarity matrix was constructed using pairwise correlations of multivoxel activity patterns. These pairwise dissimilarity values ($1 - r$) were modeled by a linear combination of two predictors: One predictor was derived from the RTs in the visual search experiment that served as a proxy for overall visual dissimilarity of the stimuli (Figure 2A), and the other predictor reflected pairwise categorical (animate vs. inanimate) dissimilarity. This procedure tested the extent to which the neural dissimilarity structure in a given sphere reflected overall visual dissimilarity (while regressing out categorical dissimilarity) and/or categorical dissimilarity (while regressing out overall visual dissimilarity). (B) Results of whole-brain group-averaged analyses testing the value of each predictor versus zero. The analysis identified a large cluster of voxels where neural dissimilarity reflected overall visual dissimilarity (red and yellow voxels), spanning early visual cortex and extrastriate regions up to VTC. In addition, clusters were identified where neural dissimilarity was independently predicted by the category of the stimuli (yellow and green voxels): These clusters partly overlapped with the overall visual dissimilarity clusters (yellow voxels) and showed local peaks in right and left fusiform gyrus and left middle occipital gyrus. Statistical maps were thresholded at $p < .05$ (FWE corrected) and shown as binary color maps overlaid on a structural brain template (MRIcron).
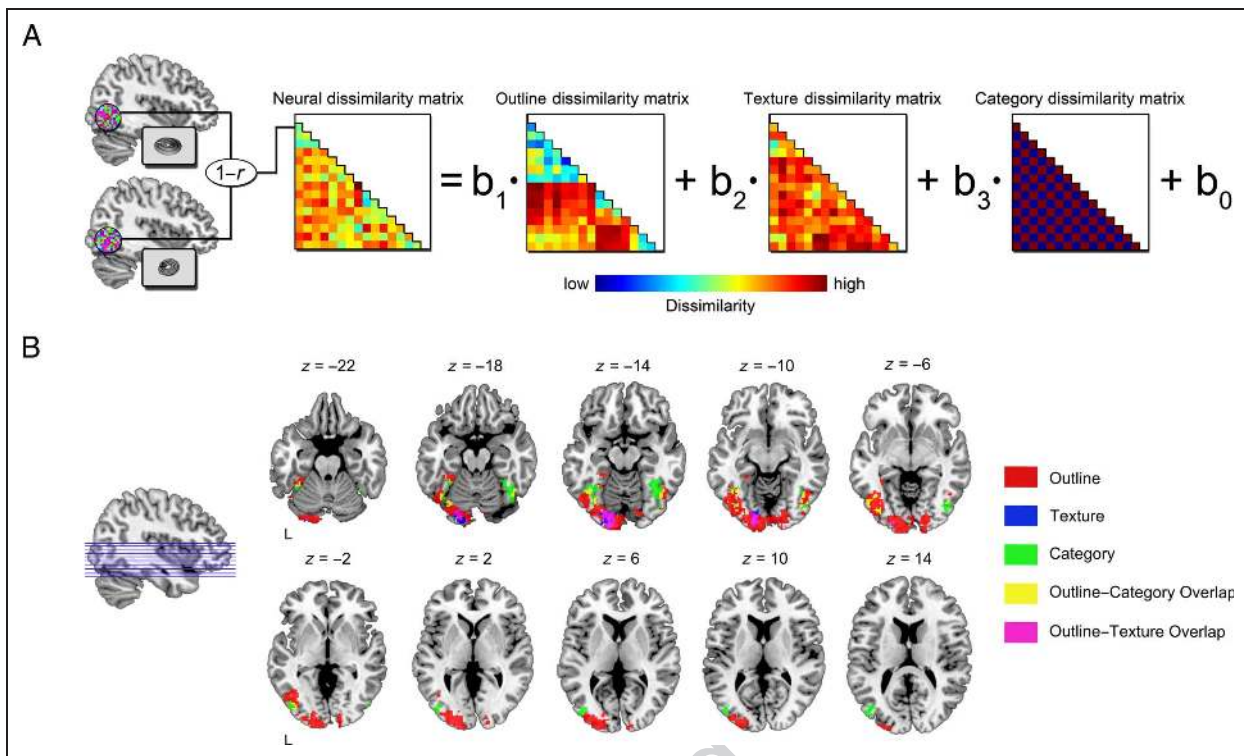
Figure 6. Multivariate searchlight analysis: outline, texture, and categorical dissimilarity. (A) Schematic of the linear modeling approach similar to the one shown in Figure 5 (see the caption for details). In this analysis, neural dissimilarity was modeled using the combination of three predictors: outline dissimilarity, texture dissimilarity, and categorical dissimilarity. (B) Results of whole-brain group-averaged analyses testing the value of each predictor versus zero. Clusters of voxels in which neural dissimilarity was predicted by outline dissimilarity (red, yellow, and pink voxels) were found in primary visual cortex, extending into extrastriate visual cortex. Texture dissimilarity predicted neural dissimilarity in a cluster in left early visual cortex (blue and pink voxels), partly overlapping with an outline dissimilarity cluster (overlap indicated in pink). Finally, this analysis revealed clusters of voxels in which neural dissimilarity was predicted by categorical dissimilarity (green and yellow voxels), with local peaks in right fusiform gyrus and left middle occipital gyrus. These clusters partly overlapped with outline dissimilarity clusters (overlap indicated in yellow). There was no overlap between texture dissimilarity clusters and categorical dissimilarity clusters. Statistical maps were thresholded at $p < .05$ (FWE corrected) and shown as binary color maps overlaid on a structural brain template (MRIcron).

objects (e.g., characteristic animal shapes or textures) or whether it (partly) reflects a true categorical organization. We approached this question by testing whether the animate–inanimate organization can still be observed when controlling for visual similarity of objects from animate and inanimate domains. A standard functional localizer experiment contrasting activity to a variety of animals with activity to a variety of inanimate objects replicated previous studies, showing animate- and inanimate-preferring regions in VTC and animate-preferring regions in lateral occipitotemporal cortex. Importantly, all of these regions, in both hemispheres, remained selective for their preferred category in the main experiment in which animate and inanimate objects were carefully matched for shape as well as for low-level features such as luminance and contrast. Results were consistent across all but one of the eight animate–inanimate pairs. (We speculate that the lack of preference for the snail condition may relate to snail shells frequently being experienced as inanimate objects, because they are often viewed without an animal inside.)

These pairs varied widely in terms of their shape (e.g., snake vs. bird), further supporting the claim that specific shape properties (e.g., presence of limbs) do not fully account for the animate–inanimate organization in VTC. Finally, the inanimate objects also varied widely on various conceptual dimensions that have been linked to inanimate-preferring regions, such as real-world size (Konkle & Oliva, 2012) and manipulability (Mahon et al., 2007). The consistency of results across the pairs suggests that the animate–inanimate organization revealed here is not fully explained by such alternative conceptual properties.

The objects that were contrasted in the univariate analyses were matched for visual similarity by the experimenters. This approach is subjective and assumes that visual similarity can be accurately judged through visual inspection. An important additional aspect of our study was therefore the use of behavioral visual search tasks to quantify different aspects of visual similarity in a naive group of participants. On each trial, participants simply

had to indicate the location of the unique stimulus in an array of identical distractors; that is, there was no predefined target category. Visual differences are the only source of information to locate the target in this task, such that performance (RT) closely reflects the visual similarity of the target and distractor stimuli (Mohan & Arun, 2012). In the first experiment, we measured the visual similarity of the same images used in the fMRI experiment. These data potentially capture not only differences in outline shape but also any other visual property (e.g., texture, extent, spatial frequency) that helps to visually distinguish the target from the distractor, rendering it a measure of overall visual similarity. Moreover, in two additional experiments, we specifically measured outline similarity (using outline drawings) and texture similarity (using texture patches).

RSA with overall visual similarity and category similarity as predictors revealed that activity patterns throughout visual cortex reflected visual similarity, confirming that visual similarity is a dominant organizing principle of both low- and high-level visual cortex (e.g., Andrews et al., 2015). The overall visual similarity matrix derived from the visual search task was additionally used to regress out variance in the neural similarity matrices, testing whether any remaining variance can be attributed to categorical similarity. This analysis revealed clusters in VTC in which this was the case. Similar results were obtained when modeling outline and texture similarity separately as predictors of the neural similarity structure. Although both texture and outline shape where represented independently in visual cortex, category information was still present in some regions. In both analyses, these category clusters were found in the vicinity of animate- and inanimate-preferring regions.

Although our results provide evidence for an animate–inanimate organization of VTC that is not explained by outline shape or texture, we should consider the possibility that there may be remaining visual features that distinguished animals from objects. Clearly, in the absence of other cues, there have to be visual properties that allow the observer to recognize the objects and to distinguish, for example, between a snake and a rope—we do not claim that there are no visual differences between the two objects of each pair. However, it seems unlikely that there were visual features that consistently covaried with category membership. Furthermore, such consistent features would likely be reflected in the visual similarity measures (Mohan & Arun, 2012) and thus regressed out in the representational similarity analyses. Nevertheless, we acknowledge that we cannot fully exclude that there may be residual visual differences between animals and inanimate objects that do not affect visual similarity as measured in the visual search experiments. For example, it is possible that certain category-specific shape features are not visually salient (and may not even be visible in the image) but become represented once an object is recognized as an animal (e.g., eyes).

Interestingly, clusters representing categorical similarity partly overlapped with clusters representing outline shape similarity at higher levels of the visual system (yellow clusters in Figures 5B and 6B). This suggests that a shape-based organization coexists with a category-based organization, with neither of these two reducible to the other. This coexistence suggests close mutual interactions between shape and categorical representations. In one direction, shape properties strongly inform category membership in most real-world situations. For example, the set of midlevel visual features that characterize animals allows for efficiently detecting the presence of an animal in a natural scene (Ullman, Vidal-Naquet, & Sali, 2002; Thorpe, Fize, & Marlot, 1996). In the other direction, category membership provides information about likely visual properties of an object, such as the structure of its parts, and allows for making perceptual predictions, for example, related to characteristic motion patterns of animals. The close proximity and partial overlap of shape and category representations may thus be optimal for real-world behavior in which these levels of representation need to closely interact.

Previous findings have shown that the degree to which a stimulus evokes an "animate" response in VTC depends on the degree to which it shares characteristics with the animate prototype—humans (Sha et al., 2014). This is consistent with earlier findings of strong selectivity for human faces and bodies at the approximate locations of the ventral and lateral animacy clusters in our study (Peelen & Downing, 2005; Downing et al., 2001; Kanwisher et al., 1997). These regions show a graded response profile, responding most strongly to human faces and bodies, followed by mammals, birds, reptiles, and insects (Downing et al., 2006). These findings are consistent, however, both with a visual similarity interpretation (i.e., differences in visual typicality; Mohan & Arun, 2012) and a conceptual similarity interpretation (e.g., differences in agency; Sha et al., 2014). Therefore, future work is needed to independently manipulate the degree to which animals share visual and semantic properties with humans to test whether graded animacy effects are primarily reflecting one or both of these properties. Interestingly, our current results show that a reliable animal preference exists even for animals that are visually and conceptually distinct from humans (e.g., snakes, insects).

Together, the present results suggest that the animate–inanimate organization of VTC is not fully explained by local biases for visual features. Instead, we interpret this organization as reflecting the recognition of an object as belonging to a particular domain. In daily life, visual properties are of course an important cue for categorizing objects, but many other cues also contribute. These cues include information from other modalities (e.g., audition, touch) and, more generally, our expectations, knowledge, goals, and beliefs. Rather than following the visual features falling on the retina, category-selective activity in VTC appears to partly reflect the

interpretation, based on all available cues, that the object we look at is animate or inanimate. On this account, category-specific activity that is independent of visual features would reflect a relatively late stage in the object recognition process. Future work could use multivariate analysis of magnetoencephalography data (e.g., Cichy, Pantazis, & Oliva, 2014; Carlson, Tovar, Alink, & Kriegeskorte, 2013) to reveal the temporal dynamics of object categorization using carefully designed stimuli that allow for disentangling visual and categorical similarity. One prediction consistent with our results would be that the initial response in VTC primarily reflects visual similarity, with later stages additionally reflecting category membership.

If not visual features, then what property might drive the animate–inanimate distinction? One proposal is that this distinction reflects agency: the potential of an object to perform self-initiated, complex, goal-directed actions (Sha et al., 2014; Caramazza & Shelton, 1998; Premack, 1990). For example, studies have shown that activity in the right fusiform gyrus—at the approximate location of the ventral animate-preferring region—can be evoked by simple geometric shapes that, through their movements, are interpreted as social agents (Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007; Martin & Weisberg, 2003; Schultz et al., 2003; Castelli, Happe, Frith, & Frith, 2000). Other work consistent with this account has shown that animal selectivity in VTC is strongest for animals, such as mammals, that are perceived as having relatively more agentic properties (Sha et al., 2014).

In summary, the present results suggest that the animate–inanimate organization of VTC may not fully reflect visual properties that characterize animals and objects. Results from RSA indicate that visual and categorical representations coexist in more anterior parts of the visual system. Clearly, future work is needed to further exclude the possibility of confounding visual features, to define exactly what dimensions drive the animate–inanimate distinction, to reveal the time course of visual and categorical representations, and to test how these interact to allow for efficient object categorization in our daily life environments.

Acknowledgments

REFERENCES

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., et al. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience, 10,* 687–689.

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research, 166,* 559–571.

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision, 15,* 3.

Baldassi, C., Alemi-Neissi, A., Pagan, M., Dicarlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLoS Computational Biology, 9,* e1003167.

Bracci, S., & Peelen, M. V. (2013). Body and object effectors: The organization of object representations in high-level visual cortex reflects body–object interactions. *Journal of Neuroscience, 33,* 18247–18258.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Buchel, C., Price, C., & Friston, K. (1998). A multimodal language region in the ventral visual pathway. *Nature, 394,* 274–277.

Caldara, R., Seghier, M. L., Rossion, B., Lazeyras, F., Michel, C., & Hauert, C. A. (2006). The fusiform face area is tuned for curvilinear shapes with more high-contrasted elements in the upper part. *Neuroimage, 31,* 313–319.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience, 10,* 1–34.

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision, 13,* pii: 1.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage, 12,* 314–325.

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

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience, 17,* 455–462.

Cohen, L., & Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *Neuroimage, 22,* 466–476.

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 16,* 1453–1461.

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

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

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience, 19,* 1803–1814.

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience, 15,* 536–548.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Haxby, J. V., Ishai, I. I., Chao, L. L., Ungerleider, L. G., & Martin, I. I. (2000). Object-form topology in the ventral temporal

lobe. Response to I. Gauthier (2000). *Trends in Cognitive Sciences, 4,* 3–4.

He, C., Peelen, M. V., Han, Z., Lin, N., Caramazza, A., & Bi, Y. (2013). Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *Neuroimage, 79,* 1–9.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences, U.S.A., 96,* 9379–9384.

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

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology, 97,* 4296–4309.

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. *Neuron, 74,* 1114–1124.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 3863–3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2,* 4.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron, 60,* 1126–1141.

Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not require visual experience. *Neuron, 63,* 397–405.

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

Mahon, B. Z., Milleville, S. C., Negri, G. A., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron, 55,* 507–520.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58,* 25–45.

Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology, 20,* 575–587.

Mohan, K., & Arun, S. P. (2012). Similarity relations in visual search predict rapid visual categorization. *Journal of Vision, 12,* 19.

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

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

Peelen, M. V., Bracci, S., Lu, X., He, C., Caramazza, A., & Bi, Y. (2013). Tool selectivity in left occipitotemporal cortex develops without vision. *Journal of Cognitive Neuroscience, 25,* 1225–1234.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology, 93,* 603–608.

Peelen, M. V., He, C., Han, Z., Caramazza, A., & Bi, Y. (2014). Nonvisual and visual object shape representations in occipitotemporal cortex: Evidence from congenitally blind and sighted adults. *Journal of Neuroscience, 34,* 163–170.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36,* 1–16.

Reich, L., Szwed, M., Cohen, L., & Amedi, A. (2011). A ventral visual stream reading center independent of visual experience. *Current Biology, 21,* 363–368.

Ricciardi, E., Bonino, D., Pellegrini, S., & Pietrini, P. (2014). Mind the blind brain to understand the sighted one! Is there a supramodal cortical functional architecture? *Neuroscience & Biobehavioral Reviews, 41,* 64–77.

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., et al. (2003). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 358,* 415–427.

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., et al. (2014). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience, 27,* 665–678.

Striem-Amit, E., & Amedi, A. (2014). Visual cortex extrastriate body-selective area activation in congenitally blind people "seeing" by using sounds. *Current Biology, 24,* 687–692.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520–522.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5,* 682–687.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42,* 671–684.

Wolbers, T., Klatzky, R. L., Loomis, J. M., Wutte, M. G., & Giudice, N. A. (2011). Modality-independent coding of spatial layout in the human brain. *Current Biology, 21,* 984–989.