

Comparative fMRI reveals differences in the functional organization of the visual cortex for animacy perception in dogs and humans

Reviewed Preprint

v1 • March 11, 2025

Not revised

Eszter Borbála Farkas , Raúl Hernández-Pérez, Laura Veronica Cuaya, Eduardo Rojas-Hortelano, Márta Gácsi, Attila Andics

Neuroethology of Communication Lab, Department of Ethology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary • MTA-ELTE 'Lendület' Neuroethology of Communication Research Group, Eötvös Loránd University, Budapest, Hungary • ELTE NAP Canine Brain Research Group, Budapest, Hungary • Social, Cognitive and Affective Neuroscience Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria • Facultad de Medicina, Universidad Nacional Autónoma de México, Santiago de Querétaro, Mexico • Department of Ethology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary • HUN-REN-ELTE Comparative Ethology Research Group, Budapest, Hungary

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

eLife Assessment

This **useful** study presents a comparative investigation of category selectivity in dogs and humans. The study compares brain representations of animate and inanimate objects, replicating and extending previous reports in this nascent field of dog fMRI. The methods and results seem to lack sufficient detail, appropriate controls, or statistical evidence, so at this stage of the review process, the strength of evidence is deemed **incomplete**.

<https://doi.org/10.7554/eLife.104525.1.sa3>

Abstract

The animate-inanimate category distinction is one of the general organizing principles in the primate high-level visual cortex. Much less is known about the visual cortical representations of animacy in non-primate mammals with a different evolutionary trajectory of visual capacities. To compare the functional organization underlying animacy perception of a non-primate to a primate species, here we performed an fMRI study in dogs and humans, investigating how animacy structures neural responses in the visual cortex of the two species. Univariate analyses identified animate-sensitive bilateral occipital and temporal regions, non-overlapping with early visual areas, in both species. Multivariate tests confirmed the categorical representations of animate stimuli in these regions. Regions sensitive to different animate stimulus classes (dog, human, cat) overlapped less in dog than in human brains. Together, these findings reveal that the importance of animate-inanimate distinction is reflected in the organization of higher-level visual cortex, also beyond primates. But a key species difference, that neural representations for animate stimuli are less

concentrated in dogs than in humans suggests that certain underlying organizing principles that support the visual perception of animacy in primates may not play a similarly important role in other mammals.

Introduction

General principles that govern the organization of the high-level visual cortex as well as several functional specializations were described in visually oriented, highly social primate species (1–5). The animate-inanimate category distinction is one of the general organizing principles (1, 2, 6, 7), resulting in widely distributed animacy representations in the human high-level visual cortices. These representations show various diagnostic features of animacy, including mid-level features such as curvature, face-like and body-like features, and features reflecting capacity for movement and agency (8–10). The animate-inanimate distinction is also represented in homologous cortical areas of non-human primates (11).

Much less is known about the functional organization of non-primary visual areas of non-primate mammals, for which vision is secondary to other sensory modalities, with its significance not having increased as dramatically during evolution as for certain primates (3). For non-primates that are known to use vision for social interaction processing, behavioral evidence suggests that animacy cues are relevant. Cats and dogs distinguish biological motion from other forms of motions (3, 12, 13) and other animacy cues such as self-propulsion and speed changes can lead to an orienting response in dogs (14). When viewing natural scenes, dogs focus their gaze on animate entities, especially on their heads, more than on inanimate entities (15). Although recent neuroscientific works suggest a handful of high-level visual specializations in those non-primate species that primarily use vision for social interaction processing (3), data on how the representations of animate and inanimate entities are organized in non-primate visual cortices remains scarce.

Neuroimaging studies of non-primates directly comparing the visual processing of animate and inanimate entities have so far only been conducted with dogs. Dogs that live as companions constitute a special case for comparison to humans, even with their lower visual acuity and different color vision (16): dogs' visual social environment is shared with humans, and they are the only non-human species that can be tested in an awake, unrestrained state (17). Three recent fMRI studies compared dog' and humans' high-level visual functions related to different aspects of animacy but remained inconclusive. Phillips and colleagues (18) found that a classifier distinguishing representations of dynamic animate and inanimate objects in humans failed to make the same distinction in dogs. In contrast, using static images, Boch and colleagues (19) identified body sensitivity as a factor accounting for animacy organization in the dog visual cortex. Finally, looking for representations of another diagnostic feature of animacy, faces, Bunford, Hernández-Pérez and colleagues (20) found areas exhibiting sensitivity for dog and human faces in humans but not in dogs. Together, these studies indicated that, similarly to the human brain, the dog brain exhibits sensitivity for certain aspects of animacy, but its extent and similarity to animacy-sensitivity in humans is still unclear.

To better understand the similarities and differences in how animacy structures higher-level visual perception in evolutionarily distant mammal species, we conducted a directly comparative dog-human fMRI study, in which participants viewed short videos of moving animate (dog, human, cat) and inanimate (car) entities. We selected stimuli to be therefore relevant to both species, and dynamic to involve motion-based features of animacy. To test animacy-sensitivity, we measured and compared neural response patterns elicited by these different stimulus classes. We predicted that if the animate-inanimate distinction is one of the general organizing principles of the dog visual cortex as it is in humans, we will find stronger responses to animate stimuli, and distinct response patterns to animate and inanimate stimuli, in non-primary visually responsive

areas in the dog brain. Furthermore, we expected that if specific animate stimulus classes share some animacy features that are central for the functional organization of the human but not of the dog visual cortex – perhaps because dogs’ visual-social perception is far less detailed than humans’ –, then the neural responses for dog, human and cat stimuli would be less similar to each other and therefore overlap less in dogs than in humans.

Methods

Participants

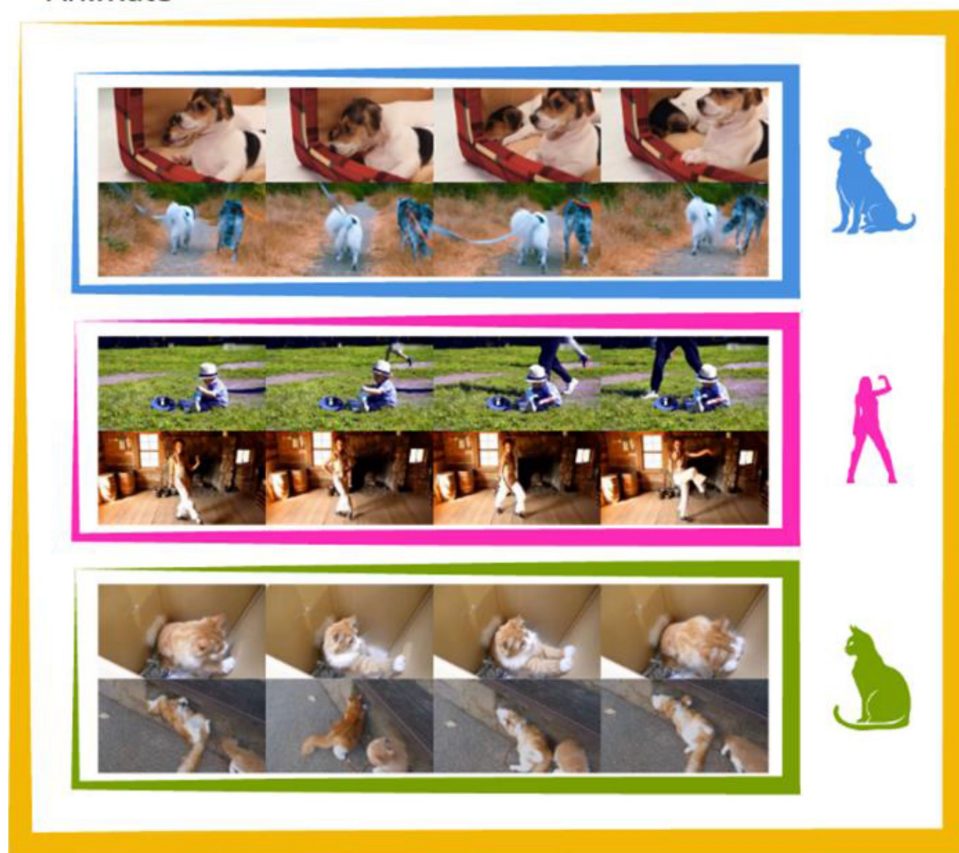
15 family dogs (average (range) age of 7.1 (3-12.6); 7 spayed females, 1 intact-, 7 neutered males), and 13 humans (average age (range) of 31.5 (21-42), 7 females, 6 males) participated in the experiment. All dogs (7 Border Collies, 2 Golden Retrievers, 1 Australian Shepherd, 1 Labradoodle, and 4 mixed breeds) were mesocephalic and were trained to remain still inside the scanner (21 [↗](#)). All human participants had normal or corrected to normal visual acuity and participated voluntarily. Dog owners and humans were recruited from the participant pool of the Department of Ethology at Eötvös Loránd University in Budapest, Hungary. All procedures were in accordance with relevant guidelines and regulations. Procedures involving dogs were approved by the Food Chain Safety and Animal Health Directorate Government Office, Hungary. Procedures involving humans were approved by the Committee of Scientific and Research Ethics (ETT-TUKEB), Hungary. Dog owners and human participants signed an informed consent.

Experimental Design

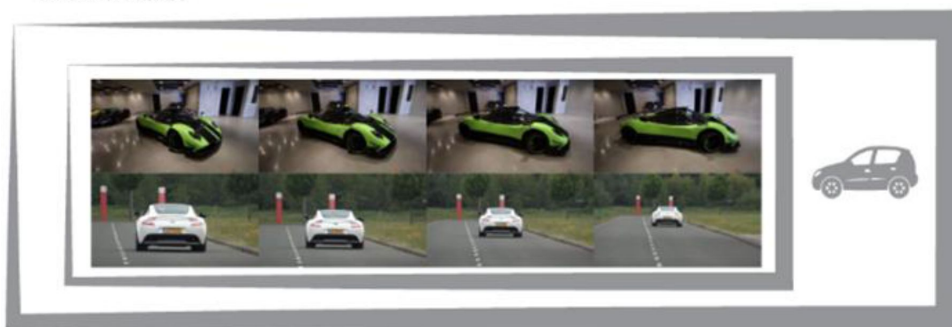
The stimuli consisted of natural videos of dogs, humans, cats, and cars, some showing the whole body or object, some only parts, through close-up or wider shots (**Figure 1** [↗](#) and Supplementary Video S1). The dog (D), human (H), and cat (C) videos constituted the animate (A) stimuli. We selected these animate stimulus classes to have an identical stimulus set for dogs and humans that involves familiar species to maximize the attention of dog subjects: the conspecific and two heterospecific species. Cars were selected as inanimate (iA) stimuli that move, vary in color and shape and all dog participants had ample experience with them. A reason to use a restricted number of stimulus classes was ensuring enough repetitions of each condition to elicit robust neural response in dogs where representations of categories are not established. All the videos were downloaded from YouTube (available under a Creative Commons License), cut to match the length required and resized to a final resolution of 1024 × 576 pixels. The videos displayed various situations and backgrounds using natural colors to minimize the possibility of systematic differences in visual properties across conditions and maximize ecological validity. To balance the differences in visual properties, brightness, contrast, hue, saturation, and motion for every video was calculated and videos were selected so that there were no significant differences in these properties between conditions in any of the runs (one-way ANOVA, $p > 0.05$). To account for other low-level visual properties, captured by Gabor filters, the within- and between-category correlation of the outputs of the HMAX model’s C1 units (22 [↗](#)) were compared and no differences were found for any of the categories (two-sample t test, $p > 0.05$).

We used an event-related design (**Figure 1** [↗](#)). Each run contained different stimuli, 12 exemplars from each of the four conditions. The length of the videos varied between 3.5 and 5.5 s (mean 4.5 s). Each video was preceded by a grey screen with variable duration of 0.7 - 2.5 s (mean 1.7 s). The length and variability of stimulus duration and interstimulus interval were chosen to minimize predictability but maximize signal quality (23 [↗](#), 24 [↗](#)). Runs ended with an additional 10 s of grey screen and lasted 310 s. The order of the videos within a run was randomized for every participant separately. Thirteen dogs completed six runs and two dogs completed only three, all humans completed six runs. Dogs were tested in two runs per session with rest periods of at least 20 min between sessions and a maximum of four runs per day. The dogs’ eyes were observed through an MR-compatible camera during the run to confirm they were awake during scanning.

Animate



Inanimate



Mean video
duration = 4.5 s

Mean interval
duration = 1.7 s



Figure 1.

Experimental paradigm.

Still images from two sample video stimuli representing each of the three animate (yellow) conditions: dog (blue), human (pink), cat (green) and the inanimate condition: car (grey) and the design of a single run. Each 310 s long run was composed of 12 videos (3.5–5.5 s, mean 4.5 s) of each of the four conditions followed by an interstimulus interval (0.7–2.5 s, mean 1.7 s). Stimulus design was identical for dog and human subjects. See also Video S1.

Four runs were discarded and repeated due to the dog falling asleep. Humans were tested in one or two days and in one session with 3 to 6 runs per day depending on the availability of the scanner and the participant.

Videos were presented using an MR compatible NordicNeuroLab LCD Monitor and controlled using Psychophysics Toolbox. Dogs viewed the screen directly, while humans viewed it through a mirror. The screen was set up at 155 cm from the head of the participants; the videos covered the entire height (49.8 cm) and width (88.6 cm) of the screen.

fMRI Data Acquisition and Preprocessing

Data acquisition was performed at the Medical Imaging Centre at the Semmelweis University on a Philips 3T Ingenia scanner. For dogs, a Philips dStream Pediatric Torso 8ch coil and for humans, a Philips dStream Head 32ch coil was used. Functional data were collected using a gradient-echo-planar imaging (EPI) sequence (dogs and humans: TR=2500 ms, TE=20 ms, flip angle=90°, 2.5 mm-thick slices with 0.5 mm gap; dogs: field of view: 200×150×120 mm, acquisition matrix 80×58; 40 axial slices; humans: field of view: 240×240×122 mm, acquisition matrix 96×94; 41 axial slices). 124 volumes were acquired in each run. Anatomical data was collected using a T1-weighted 3D TFE sequence, with 1×1×1 mm resolution (dogs: at a separate session; humans: the end of the last functional imaging session). We acquired 180 slices covering the whole brain for anatomical localization.

Image preprocessing was conducted with FSL version 4.19 (Jenkinson et al., 2012). Statistical analysis of MRI data was performed using FEAT (fMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmriv.ox.ac.uk/fsl/), time-series statistical analysis using FILM (25), and representational similarity analysis using PyMVPA software package (26) and MATLAB (2018b) custom code.

Volumes were motion-corrected and filtered using a 128 s high-pass filter. Scan-to-scan movement was calculated using framewise displacement (FD) (27), all the volumes where FD exceeded 1 mm were excluded (dogs 0.17%; humans 0%). The mean FD across subjects and runs was 0.07 mm for dogs and 0.05 mm for humans. Dog and human images were skull-stripped.

For dogs, we calculated a mean functional image that comprised all the volumes of all the runs for a given subject. This mean functional image was then manually transformed into a dog anatomic template (28) using 12 degree-of-freedom affine transformation, and the transformation matrix was used to transform all the volumes into the template space. Human images were first registered to the anatomical scan using a 7 degree-of-freedom linear transformation, then to the Montreal Neurologic Institute MNI template using 12 degree-of-freedom affine transformation. Next, images were smoothed (Gaussian kernel, FWHM 5 mm) for both dogs and humans.

fMRI Statistical Analysis

RSA Comparing Neural and an EVC Model

To map neural responses to low-level visual properties, we assessed, using RSA, where the representational geometry of the stimuli in the dog and human brains was similar to how an early visual cortex (EVC) model represent them.

First, representational dissimilarity matrices (DSM) of the stimulus set were calculated to characterize the pairwise dissimilarity of the response patterns elicited by the stimuli (Kriegeskorte, 2009). For this, each stimulus event was modelled convolving its times series with the canonical hemodynamic response function and the resulting, normalized β values were used to describe every voxel by how dissimilar a sphere of voxels ($r=2$ voxels for dogs and $r=3$ voxels

for humans) around it represents each pair of the stimuli. Dissimilarity was measured as the correlation distance (1 - Pearson correlation). DSMs were calculated this way for each run of every participant and then were averaged across runs and participants. In the case of dogs, this process was repeated for the whole brain. In the case of humans, we restricted the analysis to the occipital and temporal lobes (delineated based on MNI atlas) where the visual areas are, to reduce the computation load.

Second, the stimuli were presented to the computational model HMAX (22 [↗](#)). In its first and second layers, the HMAX model simulates the selective tuning of neurons in primary visual cortex (V1) and has been shown to match with the activity patterns of the EVC of the primate brain (11 [↗](#), 29 [↗](#), 30 [↗](#)). Specifically, we used the second layer of the model (referred here as EVC model), which corresponds to V1 complex cells. This layer yields a response vector of a V1 unit by performing a 2-dimensional convolution of a Gabor filter with an image. The algorithm was applied to each frame of each video using a set of Gabor filters with four different orientations (-45° , 0° , 45° , 90°) and 17 sizes (7 to 39 pixels in 2-pixel steps) per orientation, modeling 68 V1 units for each frame of a video. To obtain a unidimensional response vector for every video stimulus, principal component analyses were performed within the response vectors of every V1 unit across all frames and for all stimuli. The first principal components were used in the following step, because those contained the most explained variance for the given video. Next, the DSM of response vectors was calculated to assess the representational geometry expected based on our stimuli's low-level visual features.

Third, to compare the brain representation and the model, correlation maps were created, in which each voxel was assigned the Pearson correlation between neural DSM and model DSM. To account for the spatial dependency of the searchlight approach (31 [↗](#)) and to reduce the number of assumptions on the maps, we used a random permutation method (32 [↗](#)) for group-level analysis and statistical inference.

The stimulus labels were randomly swapped, a random DSM map and then the correlation map between the random DSM map and the DSM of the model were created. This process was repeated 1,000,000 times. Next, the maps were averaged to calculate the correlations expected by chance and their standard deviations. These parameters were then used to transform the original correlation maps to Z-score maps. The maps were then filtered to maintain only the voxels with a Z-score > 3.1 (equivalent to $p < 0.001$). The random correlation maps were also filtered at Z-score > 3.1 . Using these maps, the cluster size distribution expected by chance was calculated and this distribution was used to filter out the clusters with a size smaller than the size expected by chance ($p < 0.05$).

General Linear Model (GLM)

To identify regions that are sensitive to our conditions, univariate analysis with the four conditions (D, H, C, iA) as regressors were conducted using the FSL's implementation of the General Linear Model.

The first-level individual models included the movement parameters as regressors of no interest. The regressors were convolved with the canonical hemodynamic response function, then using all six runs, condition effects were estimated for each participant. At the group level, a random-effects analysis was performed for each species.

For further analyses, the visually responsive regions of the dog and human cortex were identified by comparing all conditions to the implicit baseline. To determine the brain regions sensitive to animate stimuli, the videos from all three animate conditions (D, H, C) were grouped and

contrasted with the inanimate condition (A>iA contrast). To further characterize the animacy-sensitivity of the dog and human brains, each of the three animate conditions was contrasted with the inanimate condition separately (D>iA, H>iA, C>iA, specific contrasts).

For all contrasts, the resulting Z-statistic (Gaussianised T/F) images were thresholded at $Z > 3.1$ and the threshold for cluster-correction was $p = 0.05$ (Worsley, 2001).

Response profiles

To further characterize the A>iA contrast results, for every subject and for each group level-derived peak, first the parameter estimates (β weights) for each condition in a sphere (radius = 3 voxels) centered around the peak was assessed using FSL's Featquery tool. Second, these parameter estimates were compared on the group level, using t tests.

Overlap calculation

To determine and compare across species the extent to which specific contrast maps overlap, first, the visually response cortex was determined in both species by contrasting all conditions to baseline using a cluster threshold of $Z > 3.1$ and a corrected cluster significance threshold of $p = 0.05$. Then, the number of suprathreshold voxels were assessed for each specific contrast as well as the number of voxels in the visually responsive cortex in dogs and in humans. Then, the percentage of those visually responsive voxels that were suprathreshold in (1) all three specific contrasts (3-overlap) or (2) two of the three specific contrasts (2-overlap), were calculated, and the percentages of both 3-overlapping and 2-overlapping voxels were compared between dogs and humans, using t tests. *Species-preference*. We also assessed the proportions of animate-sensitive voxels that responded strongest to conspecific stimuli or to any of the heterospecific stimulus classes. For that, each animate condition (D, H, C) was contrasted with the implicit baseline, and the average Z-scores per contrast were calculated across all dog and across all human participants in every animate-sensitive voxel. The condition from the contrast with the highest average Z-score was assigned as the preferred species for each voxel. To determine whether the percentage of voxels differed above chance (0.33) for the three conditions, permutation test was performed by randomly swapping the labels of the stimuli, calculating the preference for each voxel and their percentages 1,000,000 times.

Category- and Class-boundary Effect Tests

To identify regions with greater representational dissimilarity for between-category (i.e., animate-inanimate) than within-category (i.e., animate-animate) stimulus pairs (category boundary effect test as described in Kriegeskorte et al., 2008), we calculated DSMs per run, per participant for both between-category and within-category pairs across the brain (for dogs: whole brain, for humans: temporal and occipital cortices), using a searchlight approach (sphere with $r = 2$ voxels for dogs, and $r = 3$ voxels for humans). DSMs were then averaged across runs and participants, and a difference map was calculated. To determine whether the difference was significant at the voxel level, we repeated the process 1,000,000 times but randomly swapping stimulus labels, thus generating 1,000,000 permuted difference maps. We transformed each voxel of the original difference map to Z-score by comparing it with the corresponding voxel in the permuted difference maps, generating a Z-score map. To determine cluster size threshold, we repeated the same process for the permuted difference maps, thus generating 1,000,000 permuted Z-score maps. All Z-score maps were thresholded at $Z > 3.1$. We calculated the cluster size distribution in the permuted Z-score maps and used it to determine the cluster size expected by chance in the original Z-score map. All clusters with a size smaller than expected by chance ($p < 0.05$) were excluded. Category-boundary tests were performed in a similar way for between-class (e.g., dog-car) vs. within-class (e.g., dog-dog) stimulus pairs.

Overlap calculation

To determine and compare across species the extent to which class-boundary tests overlap, first, the results were binarized. Then, the percentage of the voxels active in (1 [↗](#)) all three classes (3-overlap) or (2 [↗](#)) two of the three classes (2-overlap), were calculated, and the percentages of both 3-overlapping and 2-overlapping voxels were compared between dogs and humans, using t tests.

Results

RSA Comparing Neural and an EVC Model

For results in dogs and humans, see **Figure 2** [↗](#) and Supplementary Table S1. We found a significant correlation with the early visual cortex model in clusters in the occipital cortex of dogs and humans. In dogs, the cluster in the right hemisphere was centered in the marginal gyrus (MG) and the cluster in the left hemisphere was centered in the splenial gyrus (SpG). In humans, the cluster extended bilaterally, with main peaks in the bilateral calcarine fissure (CAL).

GLM

For visual responsiveness (all stimuli > implicit baseline) results in dogs and humans, see Supplementary Table S2. For GLM results for each contrast (A>iA, D>iA, H>iA, C>iA) in dogs and humans, see **Figure 3** [↗](#) and Supplementary Table S3. In both species, these contrasts revealed clusters that extended mainly through the temporal and occipital lobes. Specifically, in dogs, for A>iA, we found bilateral clusters in the mid suprasylvian gyrus (mSSG), extending caudally in the left hemisphere, and bilateral clusters in the ectomarginal gyrus (EMG); for D>iA, left SpG extending to mSSG, and a right cluster in the suprasylvian gyrus (SSG); including mid and caudal portions; for H>iA, left clusters including the mid ectosylvian gyrus (mESG), the mSSG, and the SpG, and a right cluster including mid and caudal portions of the SSG; and for C>iA, a left cluster in the caudal part of the SSG. In humans, for all four contrasts, we found clusters bilaterally, involving portions of the inferior temporal gyrus (ITG), middle temporal gyrus (MTG), inferior occipital gyrus (IOG) and the fusiform gyrus (FFG), typically extending more to the temporal lobe in the right hemisphere. Activity response profiles for GLM-derived A>iA peaks in dogs and humans are shown in **Figure 3B** [↗](#).

Overlap calculation

Comparing the proportion of overlapping activities within the visually responsive cortex, we found that in the dog brain there was a significantly lower proportion of voxels than in the human brain in which all three (D, H, C) contrast maps overlapped at $z=3.1$ (dogs: 2%, humans: 31%) ($t(14.781) = 12.044$, $p < 0.01$). The proportion of overlapping voxels between at least two categories out of the three was also lower in the dog brain compared to that of the human brain (dogs: 11%, humans: 51%) ($t(25.925) = 11.236$, $p < 0.01$) (**Figure 3D** [↗](#)).

Species-preference. Analyses of the extent to which voxels in the animate-inanimate contrast responded stronger to dog or human or cat stimuli indicated that in dogs, the majority of animate-preferring voxels (87%) responded strongest to conspecific (dog) stimuli (likelihood of obtaining the observed proportions by chance, using permutation testing: $p = 0.004$). In humans, the majority of animate-preferring voxels (67%) responded strongest to conspecific (human) stimuli (permutation testing confirmed a marginally significant effect, $p = 0.062$) (**Figure 3E** [↗](#)).

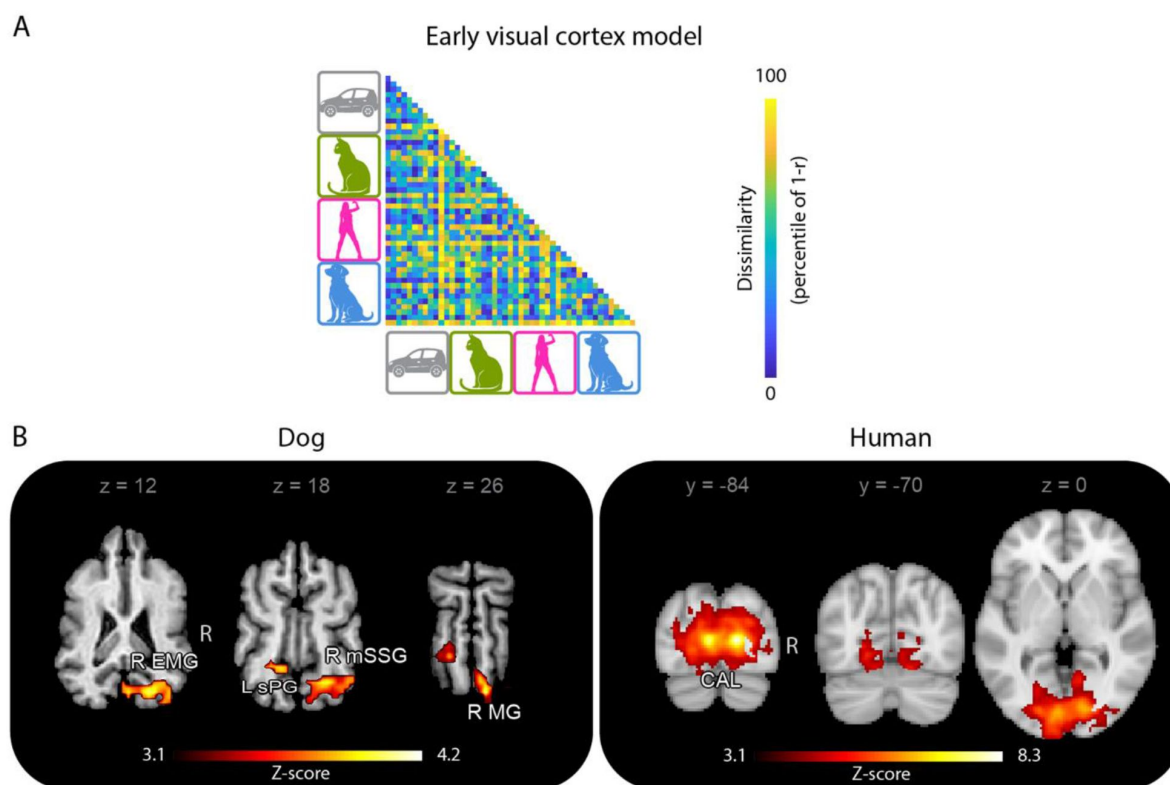


Figure 2.

RSA comparing neural and EVC model representations.

A, DSM for the EVC model. B, Results for correlation with the EVC model in dogs ($n = 15$) and humans ($n = 13$) superimposed on a template brain (for dogs: Czeibert et al., 2019; for humans: MNI template brain), using a cluster threshold of $Z > 3.1$ and a corrected cluster significance threshold of $p = 0.05$ in a searchlight analysis (sphere radius = 3 voxels for dogs and 3 voxels for humans). L=left; R=right; EMG=ectomarginal gyrus; SpG=splenial gyrus; mSSG=mid suprasylvian gyrus; MG=marginal gyrus; CAL=calcarine fissure and surrounding cortex.

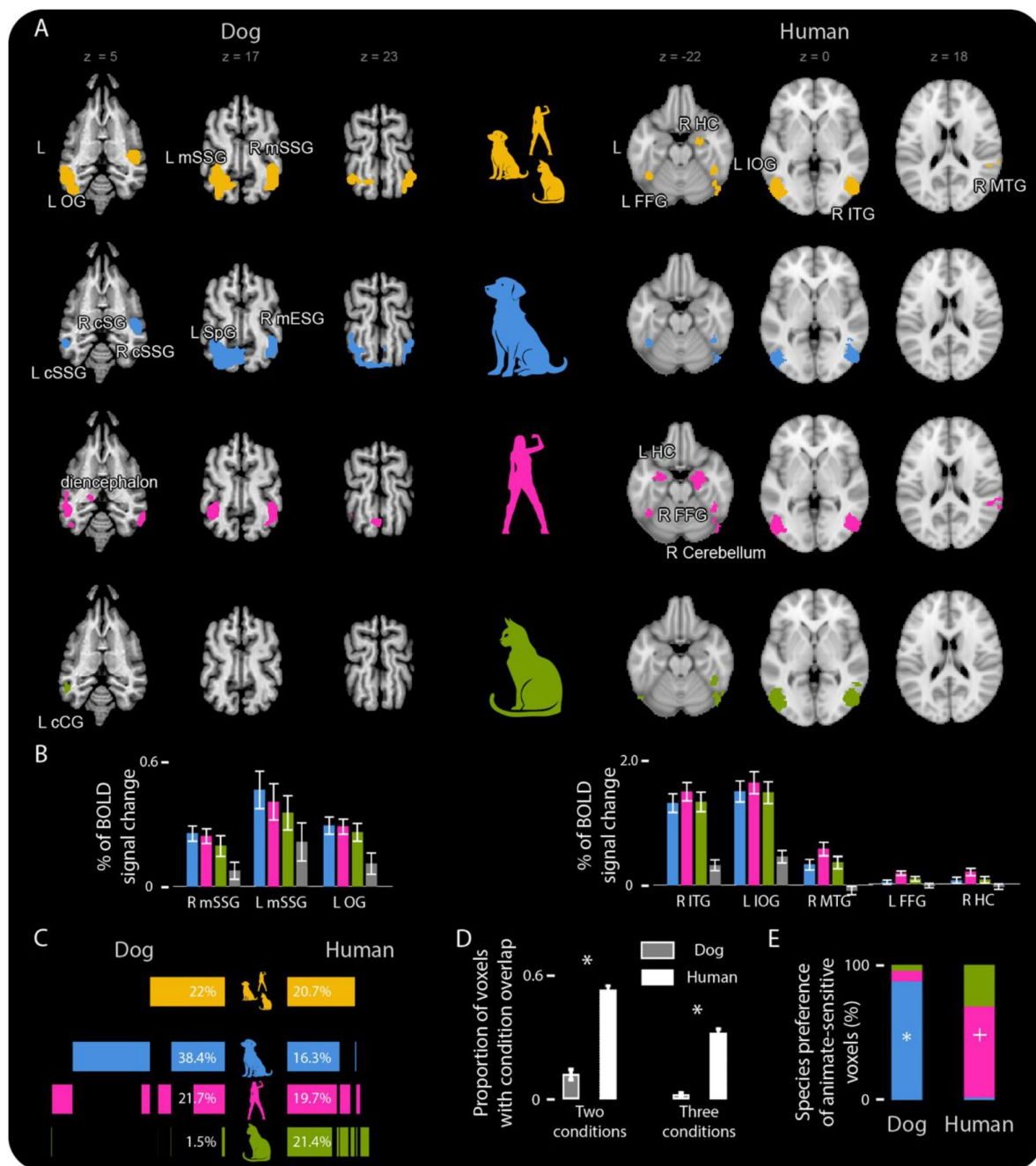


Figure 3.

GLM results in dogs and humans.

A, Dog and human contrast maps superimposed on a template brain, using a cluster threshold of $Z > 3.1$ and a corrected cluster significance threshold of $p = 0.05$. Upper row: A > iA (yellow), lower rows: D > iA (blue), H > iA (pink), C > iA (green). B, Response profiles. Bar graphs represent parameter estimates (beta weights) in select GLM-derived peaks (sphere radius = 3 voxels) for each condition. C, Activity proportions and overlaps. Horizontal bars (and percentages) represent the proportions of suprathreshold voxels per condition, relative to the total voxel number, within the visually responsive cortex; vertical overlaps represent activity overlaps. D, Overlap calculation results. The percentage of suprathreshold voxels within the visually responsive cortex for which at least two conditions overlap and in which all three overlap. E, Species-preference. Proportion of animate-prefering voxels for which the strongest response was found for the given condition. L=left; R=right; OG = occipital gyrus; mSSG = mid suprasylvian gyrus; cSSG = caudal suprasylvian gyrus; cSG = sylvian gyrus; SpG = splenial gyrus; mESG = mid ectosylvian gyrus; cCG = caudal composite gyrus; FFG = fusiform gyrus; HC=hippocampus; IOG = inferior occipital gyrus; ITG=inferior temporal gyrus; MTG = middle temporal gyrus. *: $p < 0.01$, +: $p = 0.062$. Error bars represent SE.

Category- and Class-boundary Effect Test

For category- and class-level boundary test results in dogs and humans, see **Figure 4** and Supplementary Table S4. We found temporal and occipital clusters, located similarly to those identified by univariate animacy-sensitivity contrasts. Specifically, in dogs, for A>iA, we found bilateral mSSG/mESG clusters; for D>iA, a left cluster centered in the caudal suprasylvian gyrus (cSSG) and an mSSG-centered right cluster; for H>iA, left clusters in cSSG and caudal ectosylvian gyrus (cESG), and right clusters in EMG and rostral sylvian gyrus (rSG); and for C>iA, a left cluster in mESG, and a right SpG cluster, both extending caudally. In humans, for all contrasts, we found large clusters in both hemispheres, extending from MTG to ITG and IOG, and ventrally to FFG, with similar peaks across contrasts.

Overlap calculation

Proportion of overlapping voxels within the visually responsive cortex were significantly lower in dogs than in humans, both in case all three (D, H, C) similarity maps overlapped (dogs: 1%, humans: 26%) and in case at least two of the three similarity maps overlapped (dogs: 14%, humans: 47%), ($t(12.217) = 8.518$, $p < 0.01$ and $t(20.981) = 9.405$, $p < 0.01$) (**Figure 4C**).

Discussion

This study investigated how animacy structures neural responses in the visually responsive cortex of dogs and humans. We found both similarities and differences across the two species. Univariate analyses identified bilateral occipital and temporal regions in both species responding stronger to animate than inanimate stimuli. These animate-sensitive regions are distinct from functionally determined early visual areas (i.e., that represented the stimuli similarly to an EVC model). Most animate-sensitive peaks preferred all animate stimulus classes over inanimate stimuli, but more animate-sensitive voxels responded stronger to conspecific than to any of the heterospecific stimuli in dogs, revealing conspecific-preference. A similar trend was observed in humans. The areas sensitive to the specific class of animate stimuli, dog, human and cat, overlapped less in dog than in human brains. Multivariate tests revealed regions in both species in which the representational geometry of stimulus pairs within animate and within inanimate categories were more similar than that of stimulus pairs crossing the animate-inanimate boundary. Such boundary effects were identified in both species for dog, human and cat stimuli as well, and these overlapped less in dogs than in humans. The regions exhibiting these categorical representations for animate stimulus classes largely overlapped with univariate animacy-sensitive clusters.

The similarities between dog and human neural response patterns, namely animacy-sensitivity and category boundary effects in visual cortical regions, demonstrate that animacy is an organizing principle in the visual perception of both species, and the representations of animate entities exhibit categorical structure. Categorical representation has already been shown across visual domains in multiple primate species. Going beyond these previous results, the present work suggests that animacy and the subordinate categorical feature of the input may determine the location and the structure of the neural response not only in primates', but at least in some non-primate mammals' visual perception too. We also note, however, that as animacy is multidimensional in nature (8–10), dogs' and humans' animacy sensitivity, as reported here, do not necessarily originate from the representation of the same animacy dimensions. Nevertheless, the present findings provide evidence that dogs' visual cortex contains functionally organized areas, as primates' does.

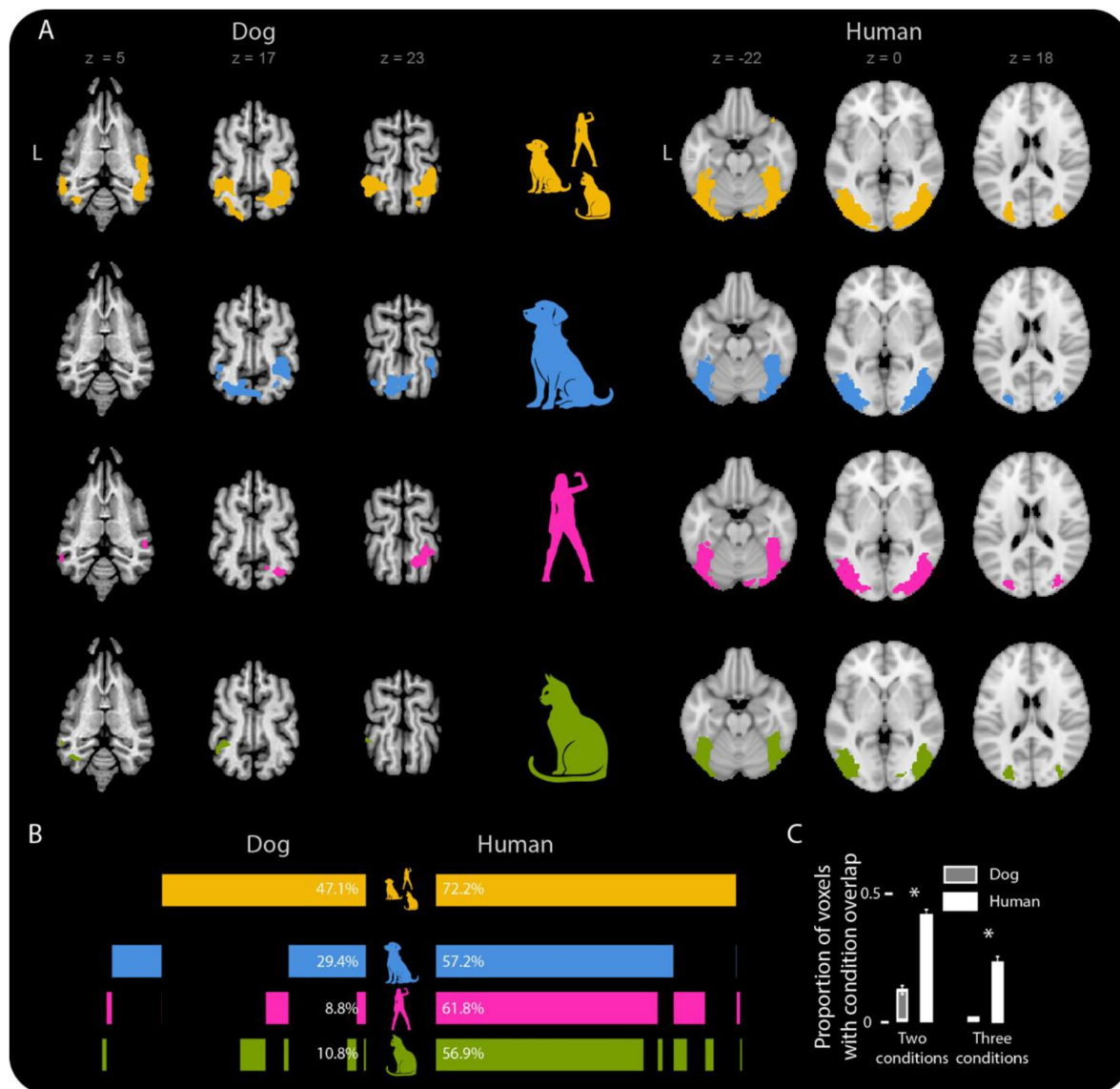


Figure 4.

Category boundary test results in dogs and humans.

A, Dog and human results superimposed on a template brain using a cluster threshold of $Z > 3.1$ and a corrected cluster significance threshold of $p = 0.05$. B, Activity proportions and overlaps. Horizontal bars (and percentages) represent the proportions of suprathreshold voxels per condition, relative to the total voxel number, within the visually responsive cortex; vertical overlaps represent activity overlaps. C, Overlap calculation results. The percentage of suprathreshold voxels within the visually responsive cortex for which at least two conditions overlap and in which all three overlap. L=left. *: $p < 0.01$. Error bars represent SE.

The animacy-sensitive regions reported here correspond to but extend beyond those reported previously (8, 9, 19). Specifically, for dogs, we found that animacy sensitivity characterizes 22% of the visually responsive cortex, including regions bilaterally in the EMG and mSSG. Using static stimuli, Boch et al. (2023) found similarly located, although less extensive regions as sensitive for animate (specifically, body and face) stimuli. Various factors may have contributed to this difference between studies, including the use of dynamic stimuli in the present study, as these stimuli also carried motion-related animacy cues (33, 34), but also scanning-technical factors. Prior results on dog neural responses to (human) faces vs. objects, potentially also driven by animacy sensitivity, also suggested the involvement of bilateral temporal regions (35, 36). For humans, animacy sensitivity characterized 20.7% of the visually responsive cortex and included extended regions in the higher-level visual cortex. Whereas previous works often focused on specific aspects of animacy (such as face, body, humanness, agency, biological motion), our highly natural stimuli carried multiple aspects of animacy at the same time, and this could have resulted in the finding of more extensive regions here (9, 11).

Beyond similarities, differences in animacy representations were also found between dogs and humans. The main species difference, that the detected neural activity evoked by animate stimulus classes overlapped less in dogs than in humans, may reflect how evolution drove mammalian brain organization for visual perception on a global level. Adaptive visual-social behavior and underlying neural representations shaped by relevant behavioral goals differ substantially between primates and (even visually oriented) non-primate mammals. The greater emphasis on visual social communication in primates, their more nuanced social interactions and within-face movements necessitate the efficient processing of bodily detail-encoding social cues and signals beyond body postures more than in non-primates. Indeed, in non-primates the relative role of perception and interpretation of whole-body actions is arguably greater. Therefore, the behavioral relevance of our stimuli showing moving dogs, humans and cats may have differed for human and dog subjects. Dogs' focus during the visual perception of animate stimuli may be on bodily actions while humans' rather on faces and gestures. This notion is supported by previous canine neuroimaging work: On one hand, these studies found no convincing evidence for the existence of face-sensitive areas in dogs, and body sensitivity, weak and restricted to small regions, has been reported for this species in a single study so far. On the other hand, Phillips et al. (18) found that representations of different actions of animate entities were separable in the dog brain, but the representations of animate and inanimate objects were not. Therefore, the pattern observed in the present study suggests that animacy sensitivity in the dog brain does not primarily stem from the common presence of face and body across all animate stimulus classes, but rather from the representation of biological motion. And it is possible that the representations of different biological motions in dogs are less generalized across the different animate stimulus classes than face and body representations in humans. These findings are in line with the proposal that dogs' high-level visual cortex may contain distinct agent-responsive areas, and also with the specific sensitivities that have been described in these areas (Boch et al., 2024). Future research should assess the validity of this potential explanation behind the species differences in animacy perception.

The finding on conspecificity-preference in dogs, that the majority of animate-sensitive voxels responded maximally to conspecific stimuli, and a similar tendency in humans, corroborate and extend previous results. Recently, Bunford, Hernández-Pérez et al. (2020) identified conspecific-prefering clusters in the visually responsive cortex in both species. In that paper, in humans, conspecific-preference was restricted to face-sensitive regions. The present findings, using more naturalistic and more variable videos, support these previous results and suggest that conspecific-preference may be characteristic for animate-sensitive visual regions, at least in dogs. Conspecific-preference, therefore, may be present at different stages of visual processing of faces and animacy. This suggests that the bias for processing conspecific stimuli does not reflect a functional specialization but rather a more generalized preference, perhaps driven by attention or motivational relevance (37, 38).

This work has a few potential limitations. First, inanimate stimuli consisted of cars only, and animate stimuli included only a narrow but highly relevant subset of all animate entities that dogs and humans may encounter, which selections, for reasons detailed in the Methods may also have affected the generalizability of the animacy sensitivity results. Potential differences in the level of relevance between animate and inanimate conditions may have also been present: for example, human stimuli may have been more relevant than cars to dogs. The similarity of the cortical locations of our animacy sensitivity results to previous findings in both dogs (19 [8](#), [9](#)), however, suggests that it was indeed the presence of animacy in dog, human and cat stimuli (and its lack in car stimuli) that drove the measured differences in neural activity. Furthermore, although our stimuli did not involve a wide variety of animate classes, they, being videos, carried not only static but also action-related diagnostic features of animacy, in a highly natural manner.

Another potential limitation of any directly comparative visual study of dogs and humans may stem from their different visual capacities: dogs perceive less details of visual stimuli presented on a monitor than humans, due to their different color vision, lower critical flicker-fusion threshold, lower depth perception, and lower visual acuity [missing references]. As a consequence, dogs may have seen our stimuli as more similar than humans have, and this may have affected neural response patterns. While we cannot exclude such effects, we do not think that visual capacity differences between dogs and humans confounded the key species difference presented here, that is the smaller overlap of condition-specific clusters for dogs. On the one hand, it can be assumed that visual capacities influenced the perception of animate and inanimate stimuli similarly. On the other hand, the less detailed visual perception of dogs may lead to less differentiated and thus more overlapping representations of the conditions, relative to those in humans, but not to less overlapping ones, as was the case here.

In conclusion, the findings presented here suggest that the importance of the animate-inanimate distinction may be reflected in the organization not only of primate, but more generally of mammalian higher-level visual cortex. The key species difference, that neural representations for animate stimuli cluster less in dogs than in humans demonstrates that different animate stimulus classes may not form a unified category in dogs and suggests that the dog brain may lack those functional specializations that drive the unified response in humans. To understand the principles that organize the visual perception of animate objects in non-primate mammals, other factors need to be considered than those underlying the animacy organization of the primate visual cortex.

Supplementary information

Brain region	Cluster size (voxel)	Peak Z-score	Coordinates (x, y, z)
Dogs			
R MG	254	4,234	5; -35; 26
R EMG		3,675	11; -33; 10
R mSSG		3,462	21; -29; 22
L SpG	101	4,248	-9; -25; 22
Human			
R CAL	8006	6,414	14; -84; 6
L CAL		5,958	-6; -86; 6
R CUN		4,343	16; -88; 22
R LING		4,196	14; -78; -10
L LING		4,189	-20; -84; -2
S SOG		4,083	-14; -82; 20
L CAL		3,937	-8; -100; -2
R MOG		3,892	30; -88; 14
L SOG		3,841	-20; -100; 16
L CER6		3,800	-18; -76; -18
L LING		3,788	-22; -66; -4
R LING		3,714	12; -62; 0
L CUN		3,687	4; -96; 14
R FFG		3,646	30; -82; -8
L MOG		3,511	-32; -84; 14
L CAL		3,494	6; -96; -12
L LING		3,430	-2; -78; -8
L CUN		3,419	2; -88; 30
R LING		3,269	22; -98; -10
R SOG		3,243	26; -78; 30
L MOG		3,214	-28; -98; 2
L CAL		3,209	-18; -68; 12
L CUN		3,206	-12; -82; 36

RSA results for dogs and humans. Threshold for reporting: cluster threshold of $Z > 3.1$ and a corrected cluster significance threshold of $p = 0.05$. All peaks ≥ 16 mm apart are reported. L=left; R=right; MG=marginal gyrus; EMG=ectomarginal gyrus; mSSG=mid suprasylvian gyrus; SpG=splenial gyrus; CAL=calcarine fissure and surrounding cortex; CUN=cuneus; LING=lingual gyrus; SOG=superior occipital gyrus; MOG=middle occipital gyrus; CER6=lobule VI of cerebellar hemisphere; FFG=fusiform gyrus.

Supplementary Table S1

Brain region	Cluser size (voxel)	Peak Z-score	Coordinates (x, y, z)
Dogs			
R SpG	1107	5.566	3; -29; 18
L EMG		5.242	-13; -31; 20
R cSSG		4.877	19; -25; 10
L cESG		4.653	-25; -23; 2
L OG		3.678	-17; -37; 4
R OG		3.138	3; -45; 16
Humans			
L IPG	12697	6.418	-40; -64; 52
L LING		5.625	-6; -88; -16
R LING		5.605	28; -88; -16
R HIP		5.497	18; -28; -8
L IOG		5.478	-44; -76; -10
R CER6		5.435	40; -36; -30
L LING		5.280	-26; -90; -14
L SPG		5.207	-24; -70; 58
L CER6		5.203	-42; -40; -30
R LING		5.191	-20; -96; 0
R MOG		5.161	30; -92; 4
R IOG		5.122	42; -86; -8
R CERCRU1		5.117	52; -70; -22
R CER6		5.055	24; -48; -26
R CERCRU1		4.928	10; -86; -22
R LING		4.763	16; -98; -12
R CER6		4.665	36; -72; -22
L CER6		4.562	-36; -64; -22
R ITG		4.463	48; -68; -6
R CERCRU1		4.424	54; -44; -30
L SOG		4.379	-22; -80; 40
L CER4_5		4.371	-28; -32; -30
R MOG		4.369	42; -82; 8
L MOG		4.365	-34; -88; 0
R CER6		4.245	42; -54; -26
L CERCRU1		4.223	-34; -80; -24
L MOG		4.157	-28; -82; 14
R CERCRU1		4.033	42; -78; -36
L CERCRU2		4.022	-46; -48; -44
L LING		3.997	-22; -52; -26
R MOG		3.83	30; -80; 28
R PHG		3.812	32; -30; -16
L CERCRU1		3.604	-52; -70; -24
L CERCRU1		3.545	-18; -82; -26
R MTG		3.459	44; -58; 6
L MTG		3.417	-48; -64; 0
R CAL		3.357	16; -100; 4
L FFG		3.336	32; -74; -2
R CER8		3.298	32; -36; -46

Supplementary Table S2

L CER8		3.216	-32; -40; -44
L SFG	1175	5.917	-20; 60; 32
R SFGmedial		5.229	8; 66; 32
L SFGmedial		5.207	-6; 54; 46
L SFGmedial		5.053	-6; 38; 60
R SFGmedial		4.878	6; 74; 16
R SFG		4.392	22; 46; 48
L SFGmedial		4.226	-8; 70; 26
L SFG		4.174	-26; 42; 48
L SFG		3.900	-20; 72; 14
L MFG		3.599	-28; 28; 56
L MFG	1094	5.384	-42; 16; 52
L PoCG		5.179	-50; -30; 60
L PreCG		5.049	-40; -6; 66

Visual responsiveness (all stimuli > implicit baseline) results. Threshold for reporting: cluster threshold of $Z > 3.1$ and a corrected cluster significance threshold of $p = 0.05$. All peaks ≥ 16 mm apart are reported. L=left; R=right; SpG=splenial gyrus; EMG=ectomarginal gyrus; cSSG=caudal suprasylvian gyrus; cESG=caudal ectosylvian gyrus; OG=occipital gyrus; IPG=inferior parietal gyrus, excluding supramarginal and angular gyri; LING=lingual gyrus; HIP=hippocampus; IOG=inferior occipital gyrus; CER6=lobule VI of cerebellar hemisphere; SPG=superior parietal gyrus; MOG=middle occipital gyrus; CERCRU1=crus I of cerebellar hemisphere; ITG=inferior temporal gyrus; SOG=superior occipital gyrus; CER4_5=lobule IV, V of cerebellar hemisphere; CERCRU2=crus II of cerebellar hemisphere; PHG=parahippocampal gyrus; MTG=middle temporal gyrus; CAL=calcarine fissure and surrounding cortex; FFG=fusiform gyrus; CER8=lobule VIII of cerebellar hemisphere; SFG=superior frontal gyrus, dorsolateral; SFGmedial=superior frontal gyrus, medial; MFG=middle frontal gyrus; PoCG=postcentral gyrus; PreCG=precentral gyrus.

Supplementary Table S2 (continued)

Contrast	Brain region	Cluster size (voxel)	Peak Z-score	Coordinates (x, y, z)
Dogs				
A>iA	R mSSG	176	4.668	17; -21; 14
	R EMG		3.305	15; -35; 22
	L mSSG	115	4.627	-15; -23; 14
	L EMG		3.715	-15; -39; 20
D>iA	L cSSG	94	4.542	-19; -33; 4
	L SpG	351	6.877	-5; -35; 20
	L mSSG		6.305	-17; -23; 22
	R mESG	118	4.695	21; -23; 16
	R EMG		3.936	15; -37; 22
	R cSG	89	4.321	21; -17; 4
	R cSSG	78	5.694	21; -25; 2
	L cSSG	45	5.882	-25; -27; 4
H>iA	R cSSG	233	5.138	23; -21; -4
	R mSSG		4.973	19; -27; 18
	R mESG	113	4.130	19; -13; 10
	L cSSG		5.145	-21; -23; 4
	L mSSG		5.175	-17; -23; 14
	diencephalon	39	4.105	-9; -15; 0
	L SpG	34	4.743	-1; -31; 22
C>iA	L cSSG	44	4.339	-19; -31; -2
Humans				
A>iA	R ITG	1550	4.702	50; -72; -12
	R FFG		4.642	42; -38; -28
	R CERCRU1		4.329	40; -56; -28
	R MTG		4.084	54; -74; 4
	R ITG		3.957	40; -60; -8
	R MTG	816	3.402	44; -62; 8
	L IOG		4.979	-46; -72; -4
	L MTG		3.736	-38; -58; -4
	L FFG		3.584	-42; -82; -16
	L MOG		3.106	-48; -86; 4
	R MTG	221	4.449	50; -42; 10
	R STG		3.682	72; -38; 16
	R MTG		3.449	58; -54; 2
	L FFG	180	4.182	-42; -52; -24
	R HIP	164	4.45	26; -10; -22
D>iA	R PHG	1359	3.745	30; 4; -30
	R ITG		4.497	50; -72; -12
	R MTG		4.402	50; -42; 10
	R ITG		4.257	40; -58; -8
	R CER6		4.234	42; -50; -26
	R MTG		3.836	54; -74; 4
	R CER6		3.415	40; -66; -24
	R MTG		3.415	66; -44; 8
	R MTG		3.290	44; -62; 8
	L IOG		4.916	-44; -70; -4

Supplementary Table S3

H>iA	L FFG		3.784	-40; -84; -16
	L MOG		3.613	-50; -84; 2
	L FFG	149	4.245	-40; -50; -24
	R FFG	1918	5.186	42; -38; -28
	R ITG		4.592	50; -72; -12
	R CERCRU1		4.424	40; -56; -28
	R MTG		4.271	50; -42; 10
	R MTG		4.214	54; -74; 4
	R STG		3.955	72; -38; 16
	R MTG		3.848	46; -60; -2
	R MTG		3.385	64; -52; 12
	L IOG	656	4.761	-46; -72; -4
	L FFG		3.418	-42; -84; -14
	L MTG		3.114	-52; -58; 2
C>iA	R HIP	517	6.069	26; -6; -18
	R FFG		4.573	32; 4; -32
	R PHG		3.831	24; -22; -20
	R PHG		3.242	12; -8; -26
	L HIP	219	4.927	-28; -6; -20
	L FFG		4.394	-28; -10; -38
	L FFG	197	3.958	-42; -52; -24
	R ITG	1781	4.785	50; -72; -12
	R FFG		4.625	42; -36; -26
	R CERCRU1		4.124	42; -54; -28
	R MTG		4.061	48; -70; 4
	R ITG		3.759	40; -60; -8
	R MTG		3.722	56; -52; 2
	R ITG		3.534	56; -60; -22
	L FFG	1183	4.969	-34; -66; -8
	L IOG		4.619	-48; -74; -4
	L CERCRU1		4.167	-50; -66; -22
	L MOG		4.016	-36; -74; 10
	L MTG		3.842	-50; -58; -2
	L IOG		3.718	-38; -84; -12
	L MOG		3.578	-42; -90; 8

GLM results for dogs and humans. Threshold for reporting: cluster threshold of $Z > 3.1$ and a corrected cluster significance threshold of $p = 0.05$. All peaks ≥ 16 mm apart are reported. A=animate; iA=inanimate; D=dog; H=human, C=cat; L=left; R=right; mSSG=mid suprasylvian gyrus; EMG=ectomarginal gyrus; cSSG=caudal suprasylvian gyrus; SpG=splenic gyrus; mESG=mid ectosylvian gyrus; cSG=caudal sylvian gyrus; ITG=inferior temporal gyrus; FFG=fusiform gyrus; CERCRU1=crus I of cerebellar hemisphere; MTG=middle temporal gyrus; IOG=inferior occipital gyrus; MOG=middle occipital gyrus; STG=superior temporal gyrus; HIP=hippocampus; PHG=parahippocampal gyrus; CER6=lobule VI of cerebellar hemisphere.

Supplementary Table S3 (continued)

Contrast	Brain region	Cluser size (voxel)	Peak Z-score	Coordinates (x, y, z)
Dogs				
A>iA	R mSSG	487	11.721	13; -31; 16
	R cESG		9.162	23; -19; 4
	R cCG		7.058	21; -7; -8
	R mSSG		7.010	15; -15; 22
	L mESG	399	14.461	-21; -25; 14
	L cESG		12.142	-25; -19; -2
	L OG		6.969	-7; -39; 14
D>iA	L cerebellar hemisphere		4.662	-15; -35; 0
	L cSSG	248	7.307	-21; -27; 14
	L MG		6.578	-7; -35; 20
	R MG		3.133	7; -25; 22
	R mSSG	159	7.438	19; -25; 14
H>iA	L cCG	35	6.624	-25; -15; -8
	R EMG	103	6.893	9; -27; 22
	L cSSG	21	5.324	-25; -25; 8
	R rSG	18	6.401	23; -13; 10
	L cESG	17	5.437	-25; -19; -2
C>iA	L mESG	129	7.021	-21; -25; 14
	L cerebellar hemisphere		5.249	-15; -33; 0
	L cESG		4.77	-25; -19; -2
	R SpG	26	5.693	11; -33; 14
	R cCG		6.007	23; -21; -6
Humans				
A>iA	R MTG	5432	109.929	52; -66; 2
	R IOG		77.278	42; -74; -8
	R LING		70.252	18; -88; -10
	R ITG		7.010	48; -52; -16
	R MOG		66.145	50; -82; 4
	R CUN		57.868	18; -94; 10
	R MOG		51.161	32; -86; 16
	R CERCRU1		48.632	50; -66; -24
	R CERCRU1		41.973	42; -80; -24
	R LING		40.262	18; -104; -6
	R CERCRU1		38.393	46; -44; -30
	R IOG		33.225	38; -92; -14
	R MOG		32.13	32; -82; 32
	R MTG		30.728	42; -52; 2
	R STG		28.811	50; -40; 10
	R FFG		28.200	28; -54; -18
	R FFG		19.127	36; -40; -18
	R FFG		14.285	30; -70; -18
	R MOG		13.498	32; -84; 0
	L CAL		13.169	4; -98; 0
	R CERCRU1		13.067	26; -88; -24
	R FFG		11.886	38; -24; -24
	L MOG	4345	125.274	-54; -78; 6

Supplementary Table S4

D>iA	L MOG		89.123	-40; -80; -2
	L IOG		72.238	-52; -72; -10
	L MOG		69.371	-26; -80; 38
	L FFG		67.717	-40; -64; -18
	L ITG		63.386	-46; -50; -24
	L MOG		63.179	-42; -64; 4
	L LING		61.51	-14; -102; -14
	L IOG		53.902	-22; -88; -10
	L FFG		47.58	-34; -80; -18
	L CERCRU2		43.786	-12; -88; -26
	L MOG		37.882	-34; -98; -4
	L SOG		34.018	-4; -86; 46
	L CER6		32.004	-30; -52; -22
	L MOG		30.463	-28; -82; 18
	L CAL		28.555	-8; -98; 2
	L CAL		18.95	-4; -82; -8
	L FFG		18.771	-38; -36; -22
	L MOG		10.326	-28; -98; 12
	R TPOsup	713	26.864	34; 26; -28
	R FFG		21.051	32; -4; -38
	R TPOsup		18.856	54; 18; -22
	R TPOmid		15.968	52; 12; -40
	R TPOmid		12.455	36; 18; -44
	R PHG		9.032	24; 12; -30
	L TPOsup	308	24.62	-42; 16; -24
	L TPOmid		20.968	-32; 24; -36
	L TPOsup		19.503	-32; 4; -30
	L FFG		4.249	-32; 0; -46
	L ITG	108	13.714	-44; 8; -38
	L FFG	49	11.166	-38; -12; -22
	R ITG	4623	31.63	42; -72; -4
	R MTG		27.897	58; -68; -2
	R FFG		24.763	46; -54; -20
	R IOG		24.080	48; -78; -18
	R MTG		23.104	46; -60; 6
	R MOG		17.039	50; -80; 8
	R CER6		13.54	36; -68; -22
	R MOG		13.378	32; -84; 0
	R ITG		11.894	54; -42; -28
	R CER6		11.486	36; -42; -26
	R IOG		10.104	30; -82; -16
	R MTG		9.106	58; -46; 6
	R MOG		8.258	34; -84; 16
	R MOG		5.983	30; -84; 34
	R CAL		4.956	24; -98; 2
	R FFG		4.617	44; -28; -26
	R FFG		3.113	30; -54; -16

Supplementary Table S4 (continued)

H>iA	L MOG	4103	32.308	-48; -84; -2
	L MTG		27.208	-54; -70; 4
	L IOG		21.574	-46; -70; -10
	L MOG		19.574	-36; -80; 8
	L FFG		17.38	-40; -60; -22
	L FFG		14.993	-36; -82; -14
	L FFG		11.243	-38; -44; -22
	L MOG		9.327	-26; -88; 18
	L MOG		7.968	-32; -92; -2
	L MTG		5.726	-44; -58; 8
	L LING		5.517	-18; -86; -8
	L SOG		5.469	-22; -84; 36
	L IOG		4.755	-20; -104; -8
	L LING		3.703	-26; -94; -20
	L MOG		3.623	-18; -96; 6
	L FFG		3.378	-26; -66; -16
	L CER4_5		3.296	-32; -30; -30
	R MTG	5416	31.986	54; -74; 6
	R MTG		30.951	44; -62; 2
	R ITG		24.622	46; -52; -22
	R IOG		23.487	52; -76; -10
	R IOG		16.741	38; -78; -2
	R ITG		16.479	56; -60; -12
	R IOG		15.305	40; -66; -14
	R MTG		14.836	56; -50; 2
	R CER6		12.093	44; -38; -30
	R FFG		10.981	26; -82; -12
	R CUN		10.035	20; -90; 8
	R MOG		9.423	40; -80; 14
	R CERCRU1		8.563	44; -78; -24
	R MOG		8.284	32; -80; 32
	R STG		7.249	44; -40; 6
	R CERCRU1		6.206	12; -88; -18
	R MOG		5.357	34; -94; 0
	R CAL		4.413	10; -96; -4
	R SOG		3.991	18; -88; 30
	L MOG	4063	36.858	-46; -76; -2
	L FFG		23.356	-46; -56; -22
	L MTG		22.769	-54; -66; 8
	L MOG		16.875	-38; -78; 12
	L IOG		14.106	-42; -88; -12
	L IOG		11.392	-52; -74; -18
	L FFG		10.296	-38; -42; -24
	L MOG		10.280	-24; -88; 18
	L FFG		9.844	-36; -72; -14
	L MTG		9.343	-44; -56; 0
	L SOG		6.966	-22; -84; 34

Supplementary Table S4 (continued)

C>iA	L IOG	4911	6.549	-20; -90; -4
	L MOG		6.056	-40; -92; 4
	L CERCRU1		5.82	-30; -86; -26
	L CER6		5.366	-30; -56; -20
	L MOG		3.301	-16; -100; 10
	R MTG		26.168	46; -62; 2
	R IOG		25.757	46; -72; -12
	R MTG		20.007	54; -76; 4
	R ITG		16.959	58; -62; -16
	R CERCRU1		16.944	46; -54; -26
	R MOG		14.037	38; -76; 2
	R IOG		11.265	44; -88; -10
	R ITG		10.112	46; -40; -18
	R MTG		9.090	48; -42; 10
	R FFG		8.84	38; -58; -12
	R SOG		8.803	24; -82; 32
	R CAL		7.929	18; -88; 6
	R FFG		7.469	30; -80; -12
	R MOG		6.677	36; -82; 20
	R IOG		5.43	26; -100; -12
	R LING		5.382	14; -86; -12
	R MOG		5.047	36; -94; 8
	R CERCRU1		4.643	44; -80; -26
	R CER6		4.398	32; -46; -24
	R FFG		3.469	44; -26; -26
	L MTG	4195	30.938	-54; -74; 4
	L MOG		26.997	-38; -78; 0
	L IOG		19.134	-52; -74; -12
	L CERCRU1		15.005	-46; -60; -24
	L IOG		12.609	-44; -88; -12
	L MOG		12.254	-42; -62; 4
	L FFG		11.011	-36; -70; -14
	L MOG		10.949	-26; -82; 18
	L FFG		10.926	-36; -48; -20
	L MOG		10.857	-28; -80; 40
	L CERCRU1		7.195	-34; -84; -24
	L CERCRU1		6.503	-48; -44; -30
	L SOG		6.185	-18; -90; 32
	L MOG		5.286	-26; -92; 2
	L LING		4.924	-14; -86; -12
	L SOG		4.746	-6; -86; 42
	L LING		4.347	-26; -100; -14

Supplementary Table S4 (continued)

Category boundary test results in dogs and humans. Threshold for reporting: cluster threshold of $Z > 3.1$ and a corrected cluster significance threshold of $p = 0.05$. All peaks ≥ 16 mm apart are reported. A=animate; iA=inanimate; D=dog; H=human; C=cat; L=left; R=right; mSSG=mid suprasylvian gyrus; cESG=caudal ectosylvian gyrus; cCG=caudal composite gyrus R; mESG=mid ectosylvian gyrus; OG=occipital gyrus; cSSG=caudal suprasylvian gyrus; MG=marginal gyrus; EMG=ectomarginal gyrus; rSG=rostral sylvian gyrus; SpG=splenial gyrus; MTG=middle temporal gyrus; IOG=inferior occipital gyrus; LING=lingual gyrus; ITG=inferior temporal gyrus; MOG=middle occipital gyrus; CUN=cuneus; CERCRU1=crus I of cerebellar hemisphere; STG=superior temporal gyrus; FFG=fusiform gyrus; CAL=calcarine fissure and surrounding cortex; CERCRU2=crus II of cerebellar hemisphere; SOG=superior occipital gyrus; CER6=lobule VI of cerebellar hemisphere; TPOsup=temporal pole: superior temporal gyrus; TPOmid=temporal pole: middle temporal gyrus; PHG=parahippocampal gyrus; CER4_5=lobule IV, V of cerebellar hemisphere.

Supplementary Table S4 (continued)

Video legend

Supplementary Video S1. fMRI stimuli. Video shows sample stimuli representing each of four conditions: dog, human, cat, car.

Video available at Open Science Framework: https://osf.io/fzt8c/?view_only=74a9c24f13dd45078e59ad2907ec807e 

Acknowledgements

This project was funded by the Hungarian Academy of Sciences [MTA Lendület (Momentum) Programme (LP2017-13/2017), National Brain Programme 3.0 (NAP2022-I-3/2022)]; the Eötvös Loránd University; the European Research Council [European Union's Horizon 2020 research and innovation program (grant number 950159)], the HUN-REN – ELTE Comparative Ethology Research Group (01 031). R.H.-P. and L.V.C. were supported by the Mexican Council of Science and Technology (CONACYT, 407590 and 409258, respectively). R.H.-P. was also supported by the Austrian Science Fund (FWF) (10.55776/ESP602).

We thank all dog and human participants and dogs' owners for their participation and contribution to these data.

Additional information

Author Contributions

R.H.-P., L.V.C., E.B.F. and A.A. performed conceptualization. R.H.-P., L.V.C. performed investigation. R.H.-P. and E.R.-H. performed formal analysis. R.H.-P., E.B.F. and A.A. performed visualization. E.B.F., R.H.-P. and A.A. performed writing – original draft. E.B.F., R.H.-P., M.G. and A.A. performed writing – review & editing.

References

1. Bao P., She L., McGill M., Tsao D. Y. 2020) **A map of object space in primate inferotemporal cortex** *Nature* **583**:103–108
2. Bracci S., Op De Beeck H. P. 2023) **Understanding Human Object Vision : A Picture is Worth a Thousand Representations** *Annu Rev Psychol* :1–33 <https://doi.org/10.1146/annurev-psych-032720-041031>
3. Leopold D. A., Mitchell J. F., Freiwald W. A. 2020) **Evolved Mechanisms of High-Level Visual Perception in Primates** Elsevier Ltd
4. Downing P. E., Jiang Y., Shuman M., Kanwisher N. 2001) **A Cortical Area Selective for Visual Processing of the Human Body** *Science* **293**:2470–3
5. Kanwisher N., McDermott J., Chun M. M. 1997) **The fusiform face area: a module in human extrastriate cortex specialized for face perception** *J Neurosci* **17**:4302–11
6. Sha L., et al. 2015) **The animacy continuum in the human ventral vision pathway** *J Cogn Neurosci* **27**:665–678
7. Grill-Spector K., Weiner K. S. 2014) **The functional architecture of the ventral temporal cortex and its role in categorization** *Nat Rev Neurosci* **15**:536–548
8. Ritchie J. B., et al. 2021) **Untangling the animacy organization of occipitotemporal cortex** *Journal of Neuroscience* **41**:7103–7119
9. Jozwik K. M., et al. 2022) **Disentangling five dimensions of animacy in human brain and behaviour** *Commun Biol* **5**
10. Robinson A. K., Quek G. L., Carlson T. A. 2023) **Visual Representations: Insights from Neural Decoding** *Annual Review of Vision Science* <https://doi.org/10.1146/annurev-vision-100120-025301>
11. Kriegeskorte N., et al. 2008) **Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey** *Neuron* **60**:1126–1141
12. Kovács K., et al. 2016) **The effect of oxytocin on biological motion perception in dogs (*Canis familiaris*)** *Anim Cogn* **19**:513–522
13. Abdai J., Ferdinandy B., Terencio C. B., Pogány Á., Miklósi Á. 2017) **Perception of animacy in dogs and humans** *Biol Lett* **13**
14. Völter C. J., Huber L. 2022) **Pupil size changes reveal dogs' sensitivity to motion cues** *iScience* **25**
15. Törnqvist H., Somppi S., Kujala M. V., Vainio O. 2020) **Observing animals and humans: Dogs target their gaze to the biological information in natural scenes** *PeerJ* **8**:e10341

16. Byosiére S. E., Chouinard P. A., Howell T. J., Bennett P. C. 2017) **What do dogs (*Canis familiaris*) see? A review of vision in dogs and implications for cognition research** *Psychon Bull Rev* :1–16 <https://doi.org/10.3758/s13423-017-1404-7>
17. Bunford N., Andics A., Kis A., Miklósi Á., Gácsi M. 2017) ***Canis familiaris* As a Model for Non-Invasive Comparative Neuroscience** *Trends Neurosci* **40**:438–452
18. Phillips E. M., Gillette K. D., Dilks D. D., Berns G. S. 2022) **Through a Dog's Eyes: fMRI Decoding of Naturalistic Videos from Dog Cortex** *J. Vis. Exp* **187**:1–18
19. Boch M., Wagner I. C., Karl S., Huber L., Lamm C. 2023) **Functionally analogous body- and animacy-responsive areas are present in the dog (*Canis familiaris*) and human occipito-temporal lobe** *Commun Biol* **6**
20. Bunford N., et al. 2020) **Comparative brain imaging reveals analogous and divergent patterns of species and face sensitivity in humans and dogs** *Journal of Neuroscience* **40**:8396–8408
21. Andics A., Gácsi M., Faragó T., Kis A., Miklósi Á. 2014) **Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI** *Current Biology* **24**:574–578
22. Serre T., Wolf L., Bileschi S., Riesenhuber M., Poggio T. 2007) **Robust Object Recognition with Cortex-Like Mechanisms** *IEEE Trans Pattern Anal Mach Intell* **29**:411–426
23. Dale A. M. 1999) **Optimal experimental design for event-related fMRI in Human Brain Mapping** Wiley-Liss Inc pp. 109–114
24. Burock M. A., Buckner R. L., Woldorff M. G., Rosen B. R., Dale A. M. 1998) **Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI** *Neuroreport*
25. Woolrich M. W., Ripley B. D., Brady M., Smith S. M. 2001) **Temporal autocorrelation in univariate linear modeling of FMRI data** *Neuroimage* **14**:1370–1386
26. Hanke M., et al. 2009) **PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data** *Neuroinformatics* **7**:37–53
27. Power J. D., et al. 2014) **Methods to detect, characterize, and remove motion artifact in resting state fMRI** *Neuroimage* **84**:320–341
28. Czeibert K., Andics A., Petneházy Ö., Kubinyi E. 2019) **A detailed canine brain label map for neuroimaging analysis** *Biol Futur*
29. Clarke A., Tyler L. K. 2014) **Object-specific semantic coding in human perirhinal cortex** *Journal of Neuroscience* **34**:4766–4775
30. Connolly A. C., et al. 2012) **The Representation of Biological Classes in the Human Brain** *Journal of Neuroscience* **32**:2608–2618
31. Kriegeskorte N., Goebel R., Bandettini P. 2006) **Information-based functional brain mapping** *Proceedings of the National Academy of Sciences* **103**:3863–3868
32. Stelzer J., Chen Y., Turner R. 2013) **Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and**

cluster size control *Neuroimage* **65**:69–82

33. Pitcher D., Ungerleider L. G. 2021) **Evidence for a Third Visual Pathway Specialized for Social Perception** *Trends Cogn Sci*
34. Robert S., Ungerleider L. G., Vaziri-Pashkam M. 2023) **Disentangling Object Category Representations Driven by Dynamic and Static Visual Input** *Journal of Neuroscience* **43**:621–634
35. Dilks D. D., et al. 2015) **Awake fMRI reveals a specialized region in dog temporal cortex for face processing** *PeerJ* **3**:e1115
36. Cuaya L. V, Hernández-pérez R., Concha L. 2016) **Our Faces in the Dog 's Brain : Functional Imaging Reveals Temporal Cortex Activation during Perception of Human Faces** *PLoS One* :1–13 <https://doi.org/10.1371/journal.pone.0149431>
37. Kay K. N., Bonnen K., Denison R. N., Acaro M. J., Barack D. L. 2023) **Tasks and their role in visual neuroscience** *Neuron* **0** <https://doi.org/10.1016/j.neuron.2023.03.022>
38. Summerfield C., Egnér T. 2009) **Expectation (and attention) in visual cognition** *Trends Cogn Sci*

Author information

Eszter Borbála Farkas[§]

Neuroethology of Communication Lab, Department of Ethology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary, MTA-ELTE 'Lendület' Neuroethology of Communication Research Group, Eötvös Loránd University, Budapest, Hungary
ORCID iD: [0009-0009-5946-1999](https://orcid.org/0009-0009-5946-1999)

For correspondence: eszter.borbala.farkas@ttk.elte.hu

[§]These Authors contributed equally

Raúl Hernández-Pérez[§]

Neuroethology of Communication Lab, Department of Ethology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary, MTA-ELTE 'Lendület' Neuroethology of Communication Research Group, Eötvös Loránd University, Budapest, Hungary, ELTE NAP Canine Brain Research Group, Budapest, Hungary, Social, Cognitive and Affective Neuroscience Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria

[§]These Authors contributed equally

Laura Veronica Cuaya

Neuroethology of Communication Lab, Department of Ethology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary, MTA-ELTE 'Lendület' Neuroethology of Communication Research Group, Eötvös Loránd University, Budapest, Hungary, Social, Cognitive and Affective Neuroscience Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria

Eduardo Rojas-Hortelano

Facultad de Medicina, Universidad Nacional Autónoma de México, Santiago de Querétaro, Mexico

Márta Gácsi

Department of Ethology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary, HUN-REN–ELTE Comparative Ethology Research Group, Budapest, Hungary

Attila Andics

Neuroethology of Communication Lab, Department of Ethology, Institute of Biology, Eötvös Loránd University, Budapest, Hungary, MTA-ELTE ‘Lendület’ Neuroethology of Communication Research Group, Eötvös Loránd University, Budapest, Hungary, ELTE NAP Canine Brain Research Group, Budapest, Hungary

Editors

Reviewing Editor

Saad Jbabdi

University of Oxford, Oxford, United Kingdom

Senior Editor

Joshua Gold

University of Pennsylvania, Philadelphia, United States of America

Reviewer #1 (Public review):**Summary**

Farkas and colleagues conducted a comparative neuroimaging study with domestic dogs and humans to explore whether social perception in both species is underpinned by an analogous distinction between animate and inanimate entities an established functional organizing principle in the primate and human brain. Presenting domestic dogs and humans with clips of three animate classes (dogs, humans, cats) and one inanimate control (cars), the authors also set out to compare how dogs and humans perceive their own vs other species. Both research questions have been previously studied in dogs, but the authors used novel dynamic stimuli and added animate and inanimate classes, which have not been investigated before (i.e., cats and cars). Combining univariate and multivariate analysis approaches, they identified functionally analogous areas in the dog and human occipito-temporal cortex involved in the perception of animate entities, largely replicating previous observations. This further emphasizes a potentially shared functional organizing principle of social perception in the two species. The authors also describe between-species divergencies in the perception of the different animate classes, arguing for a less generalized perception of animate entities in dogs, but this conclusion is not convincingly supported by the applied analyses and reported findings.

Strengths

Domestic dogs represent a compelling model species to study the neural bases of social perception and potentially shared functional organizing principles with humans and primates. The field of comparative neuroimaging with dogs is still young, with a growing but still small number of studies, and the present study exemplifies the reproducibility of previous research. Using dynamic instead of static stimuli and adding new stimuli classes, Farkas and colleagues successfully replicated and expanded previous findings, adding to the

growing body of evidence that social perception is underpinned by a shared functional organizing principle in the dog and human occipito-temporal cortex.

Weaknesses

The study design is imbalanced, with only one category of inanimate objects vs. three animate entities. Moreover, based on the example videos, it appears that the animate stimuli also differed in the complexity of the content from the car stimuli, with often multiple agents interacting or performing goal-directed actions. Moreover, while dogs are familiar with cars, they are definitely of lower relevance and interest to them than the animate stimuli. Thus, to a certain extent, the results might also reflect differences in attention towards/salience of the stimuli.

The methods section and rationale behind the chosen approaches were often difficult to follow and lacked a lot of information, which makes it difficult to judge the evidence and the drawn conclusions, and it weakens the potential for reproducibility of this work. For example, for many preprocessing and analysis steps, parameters were missing or descriptions of the tools used, no information on anatomical masks and atlas used in humans was provided, and it is often not clear if the authors are referring to the univariate or multivariate analysis.

In regard to the chosen approaches and rationale, the authors generally binarize a lot of rich information. Instead of directly testing potential differences in the neural representations of the different animate entities, they binarize dissimilarity maps for, e.g. animate entity > inanimate cars and then calculate the overlap between the maps. The comparison of the overlap of these three maps between species is also problematic, considering that the human RSA was constricted to the occipital and temporal cortex (there is now information on how they defined it) vs. whole-brain in dogs. Considering that the stimuli do differ based on low-level visual properties (just not significantly within a run), the RSA would also allow the authors to directly test if some of the (dis)similarities might be driven by low-level visual features like they, e.g. did with the early visual cortex model. I do think RSA is generally an excellent choice to investigate the neural representation of animate (and inanimate) stimuli, but the authors should apply it more appropriately and use its full potential.

The authors localized some of the "animate areas" also with the early visual cortex model (e.g. ectomarginal gyrus, mid suprasylvian); in humans, it only included the known early visual cortex - what does this mean for the animate areas in dogs?

The results section also lacks information and statistical evidence; for example, for the univariate region-of-interest (ROI) analysis (called response profiles) comparing activation strength towards each stimulus type, it is not reported if comparisons were significant or not, but the authors state they conducted t-tests. The authors describe that they created spheres on all peaks reported for the contrast animate > inanimate, but they only report results for the mid suprasylvian and occipital gyrus (e.g. caudal suprasylvian gyrus is missing). Furthermore, considering that the ROIs were chosen based on the contrast animate > inanimate stimuli, activation strength should only be compared between animate entities (i.e., dogs, humans, cats), while cars should not be reported (as this would be double dipping, after selecting voxels showing lower activation for that category). The descriptive data in Figure 3B (pending statistical evidence) suggests there were no strong differences in activation for the three species in dog and human animate areas. Thus, the ROI analysis appears to contradict findings from the binary analysis approach to investigate species preference, but the authors only discuss the results of the latter in support of their narrative for conspecific preference in dogs and do not discuss research from other labs investigating own-species preference.

The authors also unnecessarily exaggerate novelty claims. Animate vs inanimate and own vs other species perceptions have both been investigated before in dogs (and humans), so any claims in that direction seem unsubstantiated - and also not needed, as novelty itself is not a sign of quality; what is novel, and a sign of theoretical advance besides the novelty, are as said the conceptual extension and replication of previous work.

Overall, more analyses and appropriate tests are needed to support the conclusions drawn by the authors, as well as a more comprehensive discussion of all findings.

<https://doi.org/10.7554/eLife.104525.1.sa2>

Reviewer #2 (Public review):

Summary:

The manuscript reports an fMRI study looking at whether there is animacy organization in a non-primate, mammal, the domestic dog, that is similar to that observed in humans and non-human primates (NHPs). A simple experiment was carried out with four kinds of stimulus videos (dogs, humans, cats, and cars), and univariate contrasts and RSA searchlight analysis was performed. Previous studies have looked at this question or closely associated questions (e.g. whether there is face selectivity in dogs). The import of the present study is that it looks at multiple types of animate objects, dogs, humans, and cats, and tests whether there was overlapping/similar topography (or magnitude) of responses when these stimuli were compared to the inanimate reference class of cars. The main finding was of some selectivity for animacy though this was primarily driven by the dog stimuli, which did overlap with the other animate stimulus types, but far less so than in humans.

Strengths:

I believe that this is an interesting study in so far as it builds on other recent work looking at category-selectivity in the domestic dog. Given the limited number of such studies, I think it is a natural step to consider a number of different animate stimuli and look at their overlap. While some of the results were not wholly surprising (e.g. dog brains respond more selectively for dogs than humans or cats), that does not take away from their novelty, such as it is. The findings of this study are useful as a point of comparison with other recent work on the organization of high-level visual function in the brain of the domestic dog.

Weaknesses:

(1) One challenge for all studies like this is a lack of clarity when we say there is organization for "animacy" in the human and NHP brains. The challenge is by no means unique to the present study, but I do think it brings up two more specific topics.

First, one property associated with animate things is "capable of self-movement". While cognitively we know that cars require a driver, and are otherwise inanimate, can we really assume that dogs think of cars in the same way? After all, just think of some dogs that chase cars. If dogs represent moving cars as another kind of self-moving thing, then it is not clear we can say from this study that we have a contrast between animate vs inanimate. This would not mean that there are no real differences in neural organization being found. It was unclear whether all or some of the car videos showed them moving. But if many/most do, then I think this is a concern.

Second, there is quite a lot of potential complexity in the human case that is worth considering when interpreting the results of this study. In the human case, some evidence suggests that animacy may be more of a continuum (Sha et al. 2015), which may reflect

taxonomy (Connolly et al. 2012, 2016). However moving videos seem to be dominated more by signals relevant to threat or predation relative to taxonomy (Nastase et al. 2017). Some evidence suggests that this purported taxonomic organization might be driven by gradation in representing faces and bodies of animals based on their relative similarity to humans (Ritchie et al. 2021). Also, it may be that animacy organization reflects a number of (partially correlated) dimensions (Thorat et al. 2019, Jozwik et al. 2022). One may wonder whether the regions of (partial) overlap in animate responses in the dog brain might have some of these properties as well (or not).

(2) It is stated that previous studies provide evidence that the dog brain shows selectivity to "certain aspects of animacy". One of these already looked at selectivity for dog and human faces and bodies and identified similar regions of activity (Boch et al. 2023). An earlier study by Dilks et al. (2015), not cited in the present work (as far as I can tell), also used dynamic stimuli and did not suffer from the above limitations in choosing inanimate stimuli (e.g. using toy and scene objects for inanimate stimuli). But it only included human faces as the dynamic animate stimulus. So, as far as stimulus design, it seems the import of the present study is that it included a *third* animate stimulus (cats) and that the stimuli were dynamic.

(3) I am concerned that the univariate results, especially those depicted in Figure 3B, include double dipping (Kriegesorte et al. 2009). The analysis uses the response peak for the $A > iA$ contrast to then look at the magnitude of the D, H, C vs iA contrasts. This means the same data is being used for feature selection and then to estimate the responses. So, the estimates are going to be inflated. For example, the high magnitudes for the three animate stimuli above the inanimate stimuli are going to inherently be inflated by this analysis and cannot be taken at face value. I have the same concern with the selectivity preference results in Figure 3E.

I think the authors have two options here. Either they drop these analyses entirely (so that the total set of analyses really mirrors those in Figure 4), or they modify them to address this concern. I think this could be done in one of two ways. One would be to do a within-subject standard split-half analysis and use one-half of the data for feature selection and the other for magnitude estimation. The other would be to do a between-subject design of some kind, like using one subject for magnitude estimation based on an ROI defined using the data for the other subjects.

(4) There are two concerns with how the overlap analyses were carried out. First, as typically carried out to look at overlap in humans, the proportion is of overlapping results of the contrasts of interest, e.g. for face and body selectivity overlap (Schwarlose et al. 2006), hand and tool overlap (Bracci et al. 2012), or more recently, tool and food overlap (Ritchie et al. 2024). There are a number of ways of then calculating the overlap, with their own strengths and weaknesses (see Tarr et al. 2007). Of these, I think the Jaccard index is the most intuitive, which is just the intersection of two sets as a proportion of their union. So, for example, the N of overlapping $D > iA$ and $H > iA$ active voxels is divided by the total number of unique active voxels for the two contrasts. Such an overlap analysis is more standard and interpretable relative to previous findings. I would strongly encourage the authors to carry out such an analysis or use a similar metric of overlap, in place of what they have currently performed (to the extent the analysis makes sense to me).

Second, the results summarized in Figure 3A suggest multiple distinct regions of animacy selectivity. Other studies have also identified similar networks of regions (e.g. Boch et al. 2023). These regions may serve different functions, but the overlap analysis does not tell us whether there is overlap in some of these portions of the cortex and not in others. The overlap is only looked at in a very general sense. There may be more overlap locally in some portions of the cortex and not in others.

(5) Two comments about the RSA analyses. First, I am not quite sure why the authors used HMAX rather than layers of a standardly trained ImageNet deep convolutional neural

network. This strikes me also as a missed opportunity since many labs have looked at whether later layers of DNNs trained on object categorization show similar dissimilarity structures as category-selective regions in humans and NHPs. In so far as cross-species comparisons are the motivation here, it would be genuinely interesting to see what would happen if one did a correlation searchlight with the dog brain and layers of a DNN, a la Cichy et al. (2016).

Second, from the text is hard to tell what the models for the class- and category-boundary effects were. Are there RDMs that can be depicted here? I am very familiar with RSA searchlight and I found the description of the methods to be rather opaque. The same point about overlap earlier regarding the univariate results also applies to the RSA results. Also, this is again a reason to potentially compare DNN RDMs to both the categorical models and the brains of both species.

(6) There has been emphasis of late on the role of face and body selective regions and social cognition (Pitcher and Ungerleider, 2021, Puce, 2024), and also on whether these regions are more specialized for representing whole bodies/persons (Hu et al. 2020, Taubert, et al. 2022). It may be that the supposed animacy organization is more about how we socialize and interact with other organisms than anything about animacy as such (see again the earlier comments about animacy, taxonomy, and threat/predation). The result, of a great deal of selectivity for dogs, some for humans, and little for cats, seems to readily make sense if we assume it is driven by the social value of the three animate objects that are presented. This might be something worth reflecting on in relation to the present findings.

<https://doi.org/10.7554/eLife.104525.1.sa1>

Author response:

Public Reviews:

Reviewer #1 (Public review):

Summary

Farkas and colleagues conducted a comparative neuroimaging study with domestic dogs and humans to explore whether social perception in both species is underpinned by an analogous distinction between animate and inanimate entities an established functional organizing principle in the primate and human brain. Presenting domestic dogs and humans with clips of three animate classes (dogs, humans, cats) and one inanimate control (cars), the authors also set out to compare how dogs and humans perceive their own vs other species. Both research questions have been previously studied in dogs, but the authors used novel dynamic stimuli and added animate and inanimate classes, which have not been investigated before (i.e., cats and cars). Combining univariate and multivariate analysis approaches, they identified functionally analogous areas in the dog and human occipitotemporal cortex involved in the perception of animate entities, largely replicating previous observations. This further emphasizes a potentially shared functional organizing principle of social perception in the two species. The authors also describe between- species divergencies in the perception of the different animate classes, arguing for a less generalized perception of animate entities in dogs, but this conclusion is not convincingly supported by the applied analyses and reported findings.

Strengths

Domestic dogs represent a compelling model species to study the neural bases of social perception and potentially shared functional organizing principles with humans and

primates. The field of comparative neuroimaging with dogs is still young, with a growing but still small number of studies, and the present study exemplifies the reproducibility of previous research. Using dynamic instead of static stimuli and adding new stimuli classes, Farkas and colleagues successfully replicated and expanded previous findings, adding to the growing body of evidence that social perception is underpinned by a shared functional organizing principle in the dog and human occipito-temporal cortex.

Weaknesses

The study design is imbalanced, with only one category of inanimate objects vs. three animate entities. Moreover, based on the example videos, it appears that the animate stimuli also differed in the complexity of the content from the car stimuli, with often multiple agents interacting or performing goal-directed actions. Moreover, while dogs are familiar with cars, they are definitely of lower relevance and interest to them than the animate stimuli. Thus, to a certain extent, the results might also reflect differences in attention towards/salience of the stimuli.

We agree with the Reviewer and were aware that using only one class of inanimate objects but three classes of animate entities, along with the differences in complexity and relevance between the animate and the inanimate stimuli potentially elicited more attention to the inanimate condition and may have thus introduced a confound. We are revising the related limitation in the discussion to acknowledge this and to emphasize why we believe these differences do not compromise our main findings.

The methods section and rationale behind the chosen approaches were often difficult to follow and lacked a lot of information, which makes it difficult to judge the evidence and the drawn conclusions, and it weakens the potential for reproducibility of this work. For example, for many preprocessing and analysis steps, parameters were missing or descriptions of the tools used, no information on anatomical masks and atlas used in humans was provided, and it is often not clear if the authors are referring to the univariate or multivariate analysis.

We acknowledge the concerns regarding the clarity and completeness of the methods section and are significantly revising the descriptions of the methods. Of note, in humans, the Harvard-Oxford Cortical Structural Atlas (Frazier et al., 2005; Makris et al., 2006; Desikan et al., 2006; Goldstein et al., 2007), implemented within the FSL software package, was used for anatomical masks, while the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al., 2002) was used for assigning labels.

In regard to the chosen approaches and rationale, the authors generally binarize a lot of rich information. Instead of directly testing potential differences in the neural representations of the different animate entities, they binarize dissimilarity maps for, e.g. animate entity > inanimate cars and then calculate the overlap between the maps.

We thank the Reviewer for these comments and ideas. We also appreciate the second Reviewer for their related concerns and suggestions about the overlap calculation. Since the neural processing of different animate entities in the dog brain is largely unexplored, in some of our analyses we aimed to provide a straightforward and directly comparable characterization of animacy perception in the two species. We believe that a measure of how overlapping the neural representations of different animate classes are in the dog vs. the human visual cortex is a simple but meaningful and insightful characterization of how animacy perception is structured in the two species, despite the lack of spatial detail. Our decision to use binarization was based on these considerations. In response to this Reviewer's request for providing richer information, in our revised manuscript, we will present more details and additional non-binarized calculations. Specifically, we are going to use

nonbinarized data to present the response profiles of a broad, anatomically defined set of regions that have been related in other works to visual functions, to thus show where there is significant difference and overlap between the neural responses for the three animate classes in each species.

The comparison of the overlap of these three maps between species is also problematic, considering that the human RSA was constricted to the occipital and temporal cortex (there is now information on how they defined it) vs. whole-brain in dogs.

We thank this Reviewer for raising yet another relevant point about overlap calculation. We note that the overlap calculation for univariate results used the visually responsive cortex in both dogs and humans. The decision to restrict the multivariate analysis to the occipital and temporal lobes in humans, where the visual areas are, was to reduce computational load. Since RSA in dogs yielded significant voxels almost exclusively in the occipital and temporal cortices, we believe this decision did not introduce major bias in our results. This concern will also be discussed in our revised submission.

Of note, in the category- and class-boundary test, as for the other multivariate tests, the occipital and temporal cortex of humans was delineated based on the MNI atlas.

Considering that the stimuli do differ based on low-level visual properties (just not significantly within a run), the RSA would also allow the authors to directly test if some of the (dis)similarities might be driven by low-level visual features like they, e.g. did with the early visual cortex model. I do think RSA is generally an excellent choice to investigate the neural representation of animate (and inanimate) stimuli, but the authors should apply it more appropriately and use its full potential.

We thank the Reviewer for this suggestion. While this study did not aim to investigate the correlation between low-level visual features and animacy, the data is available, and the suggested analysis can be conducted in the future. This issue will also be discussed in our revised submission.

The authors localized some of the "animate areas" also with the early visual cortex model (e.g. ectomarginal gyrus, mid suprasylvian); in humans, it only included the known early visual cortex - what does this mean for the animate areas in dogs?

We thank the Reviewer for raising this point. Although the labels are the same, both EMG and mSSG are relatively large gyri, and the clusters revealed by each of the two analyses hardly overlap, with peak coordinates more than 12 mm apart for R EMG, and in different hemispheres for mSSG (but more than 11 mm apart even if projected on the same hemisphere). We will detail the differences and the overlaps in the revised submission.

The results section also lacks information and statistical evidence; for example, for the univariate region-of-interest (ROI) analysis (called response profiles) comparing activation strength towards each stimulus type, it is not reported if comparisons were significant or not, but the authors state they conducted t-tests. The authors describe that they created spheres on all peaks reported for the contrast animate > inanimate, but they only report results for the mid suprasylvian and occipital gyrus (e.g. caudal suprasylvian gyrus is missing).

We thank this Reviewer for catching these errors. The missing statistics will be provided in the revised manuscript. Also, we mistakenly named the peak in caudal suprasylvian gyrus occipital gyrus on the figure depicting the response profiles. This will also be corrected.

Furthermore, considering that the ROIs were chosen based on the contrast animate > inanimate stimuli, activation strength should only be compared between animate entities (i.e., dogs, humans, cats), while cars should not be reported (as this would be double dipping, after selecting voxels showing lower activation for that category).

We thank both Reviewers for raising this relevant point about potential double dipping. The aim of this analysis was to describe the relationship between the neural response elicited by the three animate stimulus classes, to show that the animacy-sensitive peaks are not the results of the standalone greater response to a single animate class. We conducted t-tests only to assess significant difference between these three animate conditions and no stats were performed or reported for any animate class vs. inanimate comparisons in these ROIs. In addition to providing the missing t-tests (comparing animate classes), we will present response profiles and corresponding statistics for a broad set of additional, independent ROIs, defined either anatomically or functionally by other studies in the revised version.

The descriptive data in Figure 3B (pending statistical evidence) suggests there were no strong differences in activation for the three species in dog and human animate areas. Thus, the ROI analysis appears to contradict findings from the binary analysis approach to investigate species preference, but the authors only discuss the results of the latter in support of their narrative for conspecific preference in dogs and do not discuss research from other labs investigating own-species preference.

Studying conspecific-preference was not the primary aim of this study. We only used our data to characterize the animate-sensitive regions from this aspect. The species-preference test provides an overall characterization of the entire animate-sensitive region, revealing a higher number of voxels with a maximal response to conspecific than other stimuli in dogs (and a similar tendency in humans), confirming previous evidence on neural conspecific preference in visual areas in both species. The response profiles presented so far describe only the ROIs around the main animate-sensitive peaks and, as the Reviewer points out, in most cases reveal no significant conspecific bias. We believe there is no contradiction here: the entire animate-sensitive region may weakly but still be conspecific-preferring, whereas the main animate-sensitive peaks are not; the centers of conspecific preference may be located elsewhere in the visual cortex and may be supported by mechanisms other than animacy-sensitivity. In the revised manuscript, we will elaborate more on this. Additionally, in response to other comments, and for a better and more coherent characterization of species preference (and animacy sensitivity) across the visual cortex, we will present response profiles for other, independently defined regions and explore conspecific-sensitivity in those additional regions as well. Furthermore, we will discuss related own-species preference literature in greater detail.

The authors also unnecessarily exaggerate novelty claims. Animate vs inanimate and own vs other species perceptions have both been investigated before in dogs (and humans), so any claims in that direction seem unsubstantiated - and also not needed, as novelty itself is not a sign of quality; what is novel, and a sign of theoretical advance besides the novelty, are as said the conceptual extension and replication of previous work.

We agree with this Reviewer regarding novelty claims in general, and we confirm that we had no intention to overstate the uniqueness of our results. We also did not mean to imply that this work would be the first one on animacy perception in dogs, which it obviously is not. But we understand that we could have been more explicit presenting our work as a conceptual extension and replication of previous works, and we are revising the wording of the discussion from this aspect.

Overall, more analyses and appropriate tests are needed to support the conclusions drawn by the authors, as well as a more comprehensive discussion of all findings.

We are thankful for all comments. We will revise the methods section to provide sufficient detail and ensure replicability; conduct additional analyses as detailed above; and provide a more comprehensive discussion of all findings.

Reviewer #2 (Public review):

Summary:

The manuscript reports an fMRI study looking at whether there is animacy organization in a non-primate, mammal, the domestic dog, that is similar to that observed in humans and non-human primates (NHPs). A simple experiment was carried out with four kinds of stimulus videos (dogs, humans, cats, and cars), and univariate contrasts and RSA searchlight analysis was performed. Previous studies have looked at this question or closely associated questions (e.g. whether there is face selectivity in dogs). The import of the present study is that it looks at multiple types of animate objects, dogs, humans, and cats, and tests whether there was overlapping/similar topography (or magnitude) of responses when these stimuli were compared to the inanimate reference class of cars. The main finding was of some selectivity for animacy though this was primarily driven by the dog stimuli, which did overlap with the other animate stimulus types, but far less so than in humans.

Strengths:

I believe that this is an interesting study in so far as it builds on other recent work looking at category-selectivity in the domestic dog. Given the limited number of such studies, I think it is a natural step to consider a number of different animate stimuli and look at their overlap. While some of the results were not wholly surprising (e.g. dog brains respond more selectively for dogs than humans or cats), that does not take away from their novelty, such as it is. The findings of this study are useful as a point of comparison with other recent work on the organization of high-level visual function in the brain of the domestic dog.

Weaknesses:

(1) One challenge for all studies like this is a lack of clarity when we say there is organization for "animacy" in the human and NHP brains. The challenge is by no means unique to the present study, but I do think it brings up two more specific topics.

First, one property associated with animate things is "capable of self-movement". While cognitively we know that cars require a driver, and are otherwise inanimate, can we really assume that dogs think of cars in the same way? After all, just think of some dogs that chase cars. If dogs represent moving cars as another kind of selfmoving thing, then it is not clear we can say from this study that we have a contrast between animate vs inanimate. This would not mean that there are no real differences in neural organization being found.

It was unclear whether all or some of the car videos showed them moving. But if many/most do, then I think this is a concern.

We thank this Reviewer for raising this relevant point about the potential animacy of cars for dogs and its implication for our results. Of note, two-thirds of our car stimuli showed a car moving (slow, accelerating, or fast). We acknowledge that these stimuli contained motionbased animacy cues, and in this regard, there was no clear difference between our

animate and inanimate conditions, and possibly between some of the representations they elicited. However, our animate and inanimate stimuli differed in other key factors accounting for animacy organization, such as visual features including the presence of faces, bodies, body parts, postures, and certain aspects of biological motion. So we believe that this limitation does not compromise our main conclusions. We will elaborate on this point further in the revised discussion, also considering how dogs' differential behavioral responses to cars and animate entities may provide additional insights in this regard.

Second, there is quite a lot of potential complexity in the human case that is worth considering when interpreting the results of this study. In the human case, some evidence suggests that animacy may be more of a continuum (Sha et al. 2015), which may reflect taxonomy (Connolly et al. 2012, 2016). However moving videos seem to be dominated more by signals relevant to threat or predation relative to taxonomy (Nastase et al. 2017). Some evidence suggests that this purported taxonomic organization might be driven by gradation in representing faces and bodies of animals based on their relative similarity to humans (Ritchie et al. 2021). Also, it may be that animacy organization reflects a number of (partially correlated) dimensions (Thorat et al. 2019, Jozwik et al. 2022). One may wonder whether the regions of (partial) overlap in animate responses in the dog brain might have some of these properties as well (or not).

We agree that it would be interesting to dissect which animacy-related factor(s) contribute to the observed animacy sensitivity in different regions, and although this was not the original aim of the study, we agree that we could have made better use of the variation in our stimuli to discuss this aspect. Specifically, some animacy features are shared by all three animate stimulus classes, namely the presence of biological motions, faces, and bodies. In contrast, animate classes differed in some other aspects, for example in how dogs perceived dogs, humans, and cats as social agents and in their potential behavioral goals towards them. It can therefore be argued that regions with two- and especially three-way overlapping activations are more probably involved in processing biological motion, face and body aspects, and non-overlapping ones the social agency- and behavioural goal-related aspects. In line with this, the shared animacy features are indeed ones that have been reported to be central in human animacy representation and that may have made the overlaps in human brain responses greater. We will provide a more detailed discussion of the results from this viewpoint in the revised manuscript.

*(2) It is stated that previous studies provide evidence that the dog brain shows selectivity to "certain aspects of animacy". One of these already looked at selectivity for dog and human faces and bodies and identified similar regions of activity (Boch et al. 2023). An earlier study by Dilks et al. (2015), not cited in the present work (as far as I can tell), also used dynamic stimuli and did not suffer from the above limitations in choosing inanimate stimuli (e.g. using toy and scene objects for inanimate stimuli). But it only included human faces as the dynamic animate stimulus. So, as far as stimulus design, it seems the import of the present study is that it included a *third* animate stimulus (cats) and that the stimuli were dynamic.*

We agree with this Reviewer that the findings of Dilks et al. (2015) are relevant to our study and have therefore cited them. However, the citation itself was imprecise and will be corrected in the revised manuscript.

(3) I am concerned that the univariate results, especially those depicted in Figure 3B, include double dipping (Kriegesort et al. 2009). The analysis uses the response peak for the A > iA contrast to then look at the magnitude of the D, H, C vs iA contrasts. This means the same data is being used for feature selection and then to estimate the responses. So, the estimates are going to be inflated. For example, the high magnitudes

for the three animate stimuli above the inanimate stimuli are going to inherently be inflated by this analysis and cannot be taken at face value. I have the same concern with the selectivity preference results in Figure 3E.

I think the authors have two options here. Either they drop these analyses entirely (so that the total set of analyses really mirrors those in Figure 4), or they modify them to address this concern. I think this could be done in one of two ways. One would be to do a within- subject standard split-half analysis and use one-half of the data for feature selection and the other for magnitude estimation. The other would be to do a between-subject design of some kind, like using one subject for magnitude estimation based on an ROI defined using the data for the other subjects.

We thank both Reviewers again for raising this important point about potential double dipping. We also thank this Reviewer for specific suggestions for split-half analyses – we agree that, had our original analyses involved double dipping, such a modification would be necessary. But, as we explained in our response above, this was not the case. Indeed, whereas we do visualize all four conditions in Fig. 3B, we only conducted t-tests to assess differences between the three animate conditions (the corresponding stats have been missing from the original manuscript but will be added during revision). So, importantly, we did not evaluate the magnitude of the D, H, C vs iA contrasts in any of the ROIs defined by animate-sensitive peaks; therefore, we believe that these analyses do not involve double dipping. This holds for the species preference results in Fig. 3E as well. We will clarify this in the revised manuscript. Of note, in response to a request by the other reviewer and to provide richer information about the univariate results, we will also provide response profiles and corresponding stats for a broad set of additional ROIs, defined either anatomically or functionally by other studies (e.g., Boch et al., 2023).

(4) There are two concerns with how the overlap analyses were carried out. First, as typically carried out to look at overlap in humans, the proportion is of overlapping results of the contrasts of interest, e.g, for face and body selectivity overlap (Schwarlose et al. 2006), hand and tool overlap (Bracci et al. 2012), or more recently, tool and food overlap (Ritchie et al. 2024). There are a number of ways of then calculating the overlap, with their own strengths and weaknesses (see Tarr et al. 2007). Of these, I think the Jaccard index is the most intuitive, which is just the intersection of two sets as a proportion of their union. So, for example, the N of overlapping $D > iA$ and $H > iA$ active voxels is divided by the total number of unique active voxels for the two contrasts. Such an overlap analysis is more standard and interpretable relative to previous findings. I would strongly encourage the authors to carry out such an analysis or use a similar metric of overlap, in place of what they have currently performed (to the extent the analysis makes sense to me).

We agree with this Reviewer that the Jaccard index is an intuitive and straightforward overlap measure. Importantly, for our overlap calculations we already use this measure (and a very similar one) – but we acknowledge that this was not clear from the original description. Specifically, for the multivariate overlap test, we used the Jaccard index exactly as described by this Reviewer. For the univariate overlap test, we use a very similar measure, with the only difference that there, to reference the search space, the intersection of specific animate-inanimate contrasts was divided by the total voxel number of animate-sensitive areas (which is highly similar to the union of the specific animate-inanimate contrasts). In the revised submission we will provide a more detailed explanation of the overlap calculations, making it explicit that we used the Jaccard index (and a variant of it).

Second, the results summarized in Figure 3A suggest multiple distinct regions of animacy selectivity. Other studies have also identified similar networks of regions (e.g. Boch et al.

2023). These regions may serve different functions, but the overlap analysis does not tell us whether there is overlap in some of these portions of the cortex and not in others. The overlap is only looked at in a very general sense. There may be more overlap locally in some portions of the cortex and not in others.

We thank this Reviewer for this comment, we agree that adding spatial specificity to these results will improve the manuscript. Therefore, during revision, we will assess the anatomical distribution of the overlap results, making use of a broad set of ROIs potentially relevant for animacy perception, defined either anatomically or functionally by other studies (e.g., Boch et al., 2023 for dogs).

(5) Two comments about the RSA analyses. First, I am not quite sure why the authors used HMAX rather than layers of a standardly trained ImageNet deep convolutional neural network. This strikes me also as a missed opportunity since many labs have looked at whether later layers of DNNs trained on object categorization show similar dissimilarity structures as category-selective regions in humans and NHPs. In so far as cross-species comparisons are the motivation here, it would be genuinely interesting to see what would happen if one did a correlation searchlight with the dog brain and layers of a DNN, a la Cichy et al. (2016).

We thank the Reviewer for this comment and suggestion. At the start of the project, HMAX was the most feasible model to implement given our time and expertise constraints. Additionally, the biologically motivated HMAX was also an appropriate choice, as it simulates the selective tuning of neurons in the primary visual cortex (V1) of primates, which is considered homologous with V1 in carnivores (Boch et al., 2024).

Although we agree that using DNNs have recently been extensively and successfully used to explore object representations and could provide valuable additional insights for dogs' visual perception as well, we believe that adding a large set of additional analyses would stretch the frames of this manuscript, disproportionately shifting its focus from our original research question. Also, our experiment, designed with a different, more specific aim in mind, did not provide a large enough stimulus variety of animate stimuli for a general comparison of the cortical hierarchy underlying object representations in dog and human brains and thus our data are not an optimal starting point for such extensive explorations. Having said that, we are thankful for this Reviewer for the idea and will consider using a DNN to uncover dog' visual cortical hierarchy in future studies with a better suited stimulus set. Furthermore, in accordance with eLife's data-sharing policies, we will make the current dataset publicly available so further hypothesis and models can be tested.

Second, from the text is hard to tell what the models for the class- and categoryboundary effects were. Are there RDMS that can be depicted here? I am very familiar with RSA searchlight and I found the description of the methods to be rather opaque. The same point about overlap earlier regarding the univariate results also applies to the RSA results. Also, this is again a reason to potentially compare DNN RDMS to both the categorical models and the brains of both species.

In the revised manuscript we will provide a more detailed explanation of the methods used to determine class- and category-boundary effects. In short, the analysis we performed here followed Kriegeskorte et al. (2008), and the searchlight test looked for regions in which between-class/category differences were greater than within-class/category differences. We will also include RDMS. Additionally, we will provide anatomical details for the overlap results for RSA, just as for the univariate results, using the same independently defined broad set of ROIs, defined either anatomically or functionally by other studies (e.g., Boch et al., 2023 for dogs).

(6) There has been emphasis of late on the role of face and body selective regions and social cognition (Pitcher and Ungerleider, 2021, Puce, 2024), and also on whether these regions are more specialized for representing whole bodies/persons (Hu et al. 2020, Taubert, et al. 2022). It may be that the supposed animacy organization is more about how we socialize and interact with other organisms than anything about animacy as such (see again the earlier comments about animacy, taxonomy, and threat/predation). The result, of a great deal of selectivity for dogs, some for humans, and little for cats, seems to readily make sense if we assume it is driven by the social value of the three animate objects that are presented. This might be something worth reflecting on in relation to the present findings.

We thank the Reviewer for this suggestion. The original manuscript already discussed how motion-related animacy cues involved in social cognition may explain that animacy-sensitive regions reported in our study extend beyond those reported previously and also the role of biological motion in the observed across-species differences. This discussion of the role of visual diagnostic features and features that involved in perceiving social agents will be extended in the revised discussion, also in response to the first comment of this Reviewer, to reflect on how social cognition-related animacy cues may have affected our results in dogs.

<https://doi.org/10.7554/eLife.104525.1.sa0>