Attention selectively reshapes the geometry of distributed semantic representation

Running title: Attention reshapes representational geometry

Samuel A. Nastase[1,*], Andrew C. Connolly[1,2], Nikolaas N. Oosterhof[3], Yaroslav O. Halchenko[1], J. Swaroop Guntupalli[1], Matteo Visconti di Oleggio Castello[1], Jason Gors[1], M. Ida Gobbini[1,4], James V. Haxby[1,3]

[1]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755
[2]Department of Neurology, Geisel School of Medicine at Dartmouth, Hanover, NH 03755
[3]Center for Mind/Brain Sciences, Università degli studi di Trento, 38068 Rovereto, Italy
[4]Department of Medicina Specialistica, Diagnostica e Sperimentale (DIMES), Medical School, University of Bologna, 40126 Bologna, Italy
*Corresponding author (email: samuel.a.nastase.gr@dartmouth.edu)

Keywords: attention, categorization, natural vision, neural decoding, semantic representation

Abstract

Humans prioritize different semantic qualities of a complex stimulus depending on their behavioral goals. These semantic features are encoded in distributed neural populations, yet it is unclear how attention might operate across these distributed representations. To address this, we presented participants with naturalistic video clips of animals in their natural environments while they attended to either behavior or taxonomy. We used models of representational geometry to investigate how attentional allocation affects the distributed neural representation of animal behavior and taxonomy. Attending to animal behavior transiently increased the discriminability of distributed population codes for observed actions in anterior intraparietal, pericentral, and ventral temporal cortices, while collapsing task-irrelevant taxonomic information. Attending to animal taxonomy while viewing the same stimuli increased the discriminability of distributed animal category representations in ventral temporal cortex and collapsed behavioral information. For both tasks, attention selectively enhanced the categoricity of response patterns along behaviorally relevant dimensions. These findings suggest that behavioral goals alter how the brain extracts semantic features from the visual world. Attention effectively disentangles population responses for downstream read-out by sculpting representational geometry in late-stage perceptual areas.

Significance

Humans can extract different kinds of high-level information from the visual world depending on their behavioral goals. Here, we use naturalistic stimuli and simple models of neural representation to investigate whether attention affects how the brain encodes semantic information. When paying attention to the behavior of an animal in its natural environment, for example, the neural representation of the observed action becomes more distinct, while irrelevant information about taxonomy is collapsed. Attending to taxonomy, on the other hand, has the inverse effect. These attentional effects occur primarily in late-stage sensorimotor areas rather than in early sensory areas. Overall, our behavioral goals dynamically alter how the brain processes the semantic qualities of a stimulus to better encode important information.

Introduction

The brain's information processing machinery must operate dynamically in order to accommodate diverse behavioral goals. Selective attention serves to reduce the complexity of information processing by prioritizing representational content relevant to the task at hand (1). By and large, the attention literature has focused on early vision; that is, by employing rudimentary visual stimuli and simple tasks to probe attentional changes in the representation of low-level visual information, such as orientation and motion direction (2). However, as humans, we perceive and act on the world in terms of both semantically-rich representations and complex behavioral goals. Naturalistic stimuli, although less controlled, serve to convey richer perceptual and semantic information, and have been shown to reliably drive neural responses (3–6).

The brain encodes this sort of complex information in high-dimensional representational spaces grounded in the concerted activity of distributed populations of neurons (7–9). Population encoding is an important motif in neural information processing across species (10), and has been well-characterized in early visual processing (11, 12), face and object recognition (13–16), and other sensorimotor and cognitive domains (17–20). Multivariate decoding analyses of human neuroimaging data have allowed us to leverage distributed patterns of cortical activation to provide a window into the representation of high-level semantic information (4, 9, 21–25). However, these studies generally assume that neural representations are relatively stable, rather than dynamic or context-dependent.

Electrophysiological work on attentional modulation has typically been constrained to single neurons (26–28), but more recent work has suggested that attention may alter population encoding to sharpen attended representations (29–31). In line with this, a handful of recent neuroimaging studies have examined how task demands affect multivariate pattern classification (32–36). With the exception of one recent study examining how object attention alters cortical responsivity in a natural vision paradigm (37), these studies have limited their investigation to simple visual stimuli such as oriented gratings, moving dots, and static object images. Furthermore, most of these studies have not explicitly characterized how attention alters the relationships among task-relevant and task-irrelevant neural representations.

We hypothesized that, in order to interface with distributed neural representations, attention may operate in a distributed fashion as well—that is, by selectively reshaping representational geometry (38, 39). This hypothesis was motivated by behavioral and theoretical work suggesting that attention may facilitate categorization by expanding psychological distances along task-relevant stimulus dimensions and collapsing task-irrelevant distinctions (40, 41). Here we aimed to provide neural evidence for this phenomenon by examining how attentional allocation affects the distributed neural representation of two types of semantic information thought to rely on distributed population codes: animal taxonomy (42, 43) and behavior (22, 23, 44). To expand on previous work, we used dynamic, naturalistic video clips of animals behaving in their natural environments. These stimuli not only convey information about animal form or category, but also behavior, allowing us to examine how attention affects the neural representation of observed actions (44), which has not previously been studied. Categorical models of representational geometry were employed to demonstrate that attention selectively alters distances between neural representations of both animal taxonomy and behavior along task-relevant dimensions.

Results

Twelve participants were scanned while viewing 2 s naturalistic video clips of behaving animals (Fig. 1*A*, Video S1). Stimuli comprised five folk animal taxa (birds, insects, nonhuman primates, reptiles, and ungulates) and four behaviors (eating, fighting, running, and swimming) in a fully crossed design (20 conditions total; Table S1). In half the runs, participants performed a 1-back repetition detection task requiring them to pay attention to animal taxonomy, while in the other half of the runs, they were required to attend to animal behavior (Fig. 1*A*). Surface-based searchlight whole-brain hyperalignment (3, 6), based on data collected in a separate scanning session while participants viewed approximately one hour of the *Life* nature documentary, was applied to increase inter-participant alignment of representational geometry (Fig. S1).

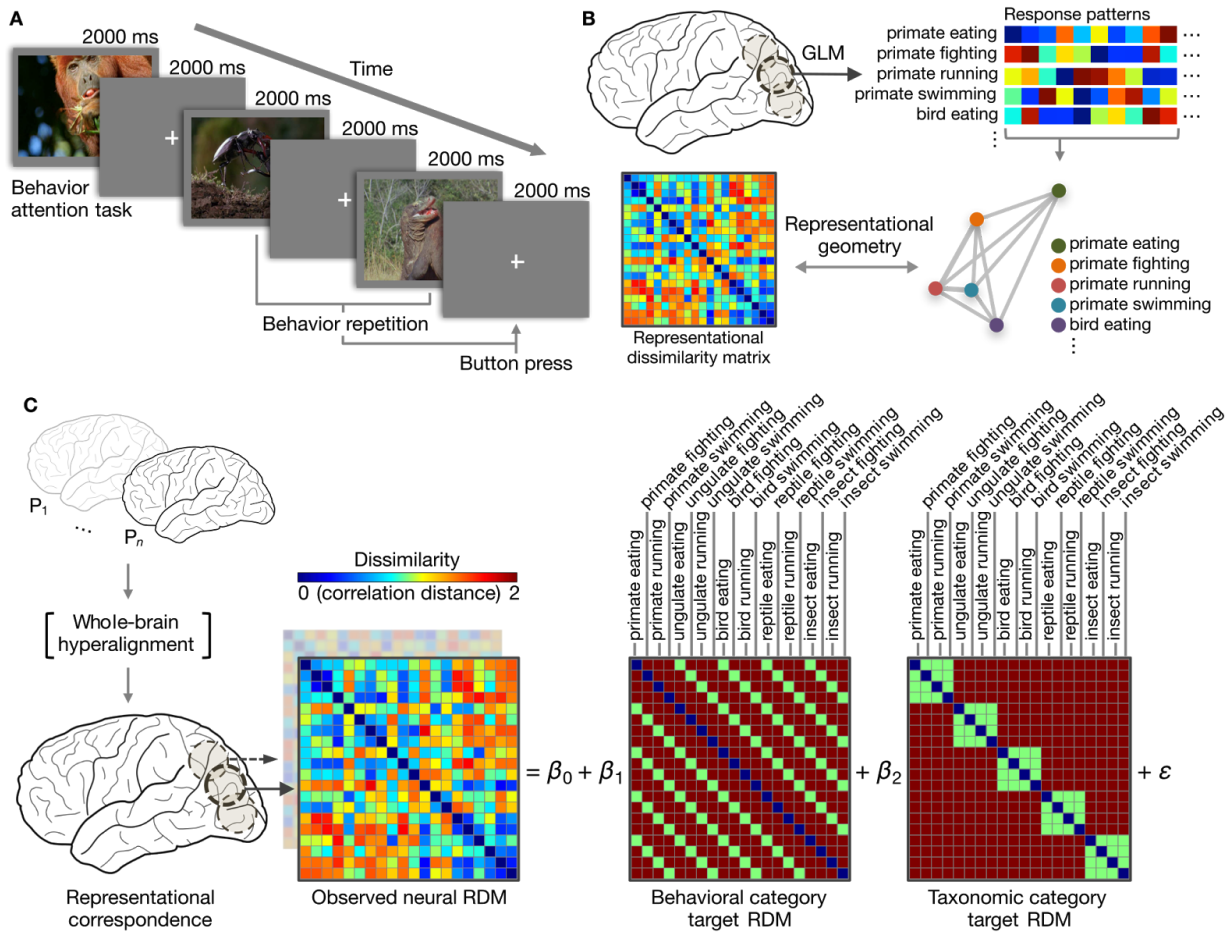

Fig. 1. Experimental procedure and analytic approach. (*A*) Schematic of event-related design with naturalistic video clips of behaving animals (Video S1, Table S1). Participants performed a repetition detection task requiring them to attend to either animal taxonomy or behavior. (*B*) Stimulus-evoked response patterns for the 20 conditions were estimated using a conventional general linear model. The pairwise correlation distances between these response patterns describe the representational geometry (representational dissimilarity matrix; RDM) for a given brain area. (*C*) Whole-brain surface-based searchlight hyperalignment was used to rotate participants' responses into functional alignment based on an independent scanning session (Fig. S1). Following hyperalignment, the neural representational geometry in each searchlight was modeled as a weighted sum of models capturing the taxonomic and behavioral categories.

Searchlight analysis. We applied representational similarity analysis of local representational geometry (45) using surface-based searchlights (46, 47). Neural representational dissimilarity matrices (RDMs) were computed based on the pairwise correlation distances between stimulus-evoked response patterns for the 20 conditions (Fig. 1*B*). The observed neural RDM for each searchlight was modeled as a weighted sum of two categorical target RDMs reflecting the experimental design, one predicting larger distances between responses to different behaviors than responses to the same behavior and the other predicting larger distances between responses to different animal taxa than between responses to the same taxon (Fig. 1*C*).

Regression coefficients for the behavioral category target RDM were strongest in lateral occipitotemporal (LO) cortex, in the dorsal visual pathway subsuming posterior parietal, intraparietal sulcus (IPS), motor and premotor areas, and in ventral temporal (VT) cortex (Fig. 2*A*). Regression coefficients for the animal taxonomy target RDM were strongest in VT, LO, and posterior parietal cortices, as well as left inferior and dorsolateral frontal cortices. Globally, attending to behavior or taxonomy increased the regression coefficients for the target RDMs corresponding to the attended categories. Attending to behavior increased the number of searchlights with significant regression coefficients for the behavioral category target RDM from 11,408 to 14,803. When considering all surviving searchlights for both attention tasks, the mean regression coefficient for the behavioral category target RDM increased significantly from 0.100 to 0.129 (p = .007, permutation test). Attending to taxonomy increased the number of searchlights with significant regression coefficients for the taxonomic category target RDM from 1,691 to 3,401, and the mean regression coefficient across these searchlights for this RDM increased significantly from 0.049 to 0.071 (p = .017). A linear SVM searchlight classification analysis cross-validated on novel stimuli resulted in qualitatively similar maps (Fig. S2), suggesting the results presented in Fig. 2 are not driven primarily by low-level visual properties of the stimuli.

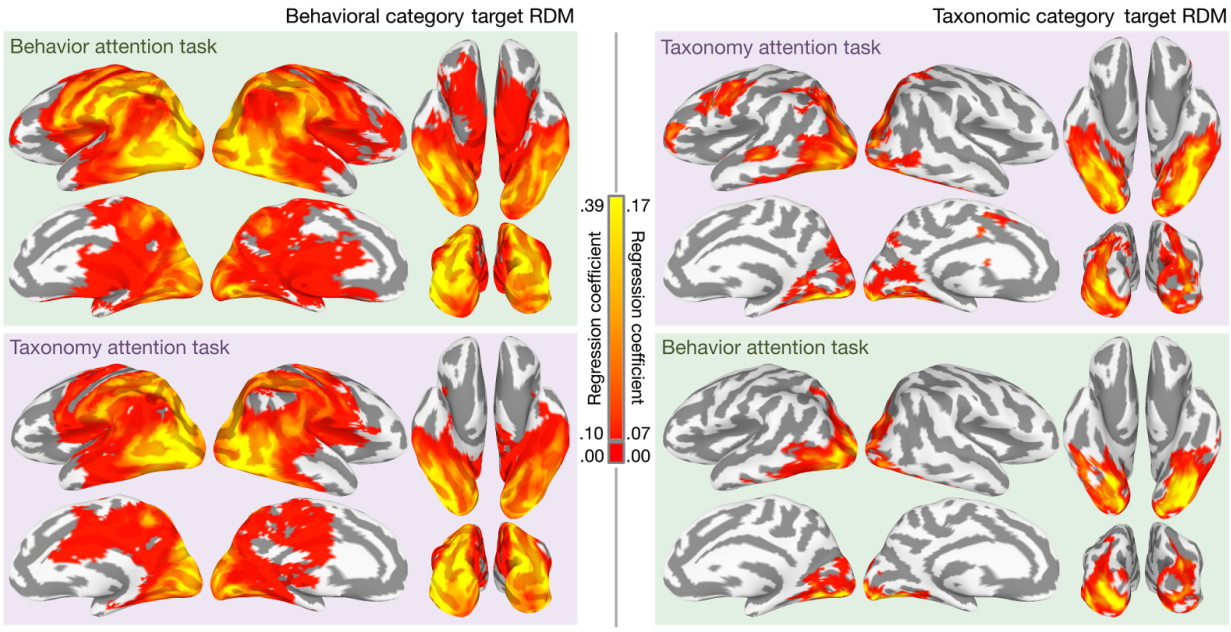

Fig. 2. Effect of attention on local representation of behavior and taxonomy. Standardized rank regression coefficients for the behavioral category target RDM (left) and the taxonomic category target RDM (right) are mapped onto the cortical surface for both attention conditions. Regression coefficients less than 0.10 for the behavioral category target RDM and less than 0.07 for the taxonomic category target RDM are plotted as red. Maps are thresholded at $p <$ .05 using TFCE (one-tailed test; see Fig. S2 for qualitatively similar searchlight classification maps, and Fig. S3 for difference maps).

Regions of interest. We tested our hypothesis in larger ROIs defined by shared representational geometry. We applied an unsupervised clustering algorithm to the searchlight representational geometries to parcellate cortex into ROIs and used a relatively reproducible parcellation with 19 areas (Fig. S4; *SI Text*). We interrogated 10 ROIs with high inter-participant similarity of searchlight representational geometry subtending the dorsal and ventral visual pathways (Figs. 3*B*, S1) . The 10 ROIs were labeled as follows: posterior early visual cortex (pEV), inferior early visual cortex (iEV), superior early visual cortex (sEV), anterior early visual cortex (aEV), lateral occipitotemporal cortex (LO), ventral temporal cortex (VT), occipitoparietal and posterior parietal cortex (OP), intraparietal sulcus (IPS), left postcentral sulcus (left PCS),

and ventral pericentral and premotor cortex (vPC/PM). ROIs were large, with a mean volume of 1,980 voxels (SD = 1,018 voxels).

For each ROI, we measured the Spearman correlation between the observed neural RDM and the two categorical target RDMs (Fig. 3*A*). A linear mixed-effects model yielded significant interactions between attention task and ROI, and attention task, target RDM, and ROI (*SI Text*), suggesting the attentional effect on Spearman correlation is more pronounced in certain ROIs than in others. Permutation tests revealed that attending to animal behavior increased correlations between the observed neural RDM and the behavioral category target RDM in vPC/PM (p = .026), left PCS (p = .005), IPS (p = .011), and VT (p = .020). A decrease in the categoricity of behavior representation was observed in sEV when participants attended to behavior (p = .032). Attending to animal taxonomy increased correlations between the observed neural RDM and the taxonomic category target RDM in VT (p = 0.10) and left PCS (p = .036). The effect in left PCS was driven by a negative correlation in the behavior attention task that was abolished when attention was directed at taxonomy.

We next evaluated how well full representational models of animal taxonomy and behavior fit the neural representational geometry in each ROI. The model RDMs used above tested our experimental hypothesis but do not capture the geometry of distances between behavioral or taxonomic categories; e.g., the animacy continuum (42, 43). To accommodate this type of geometry for behavior and taxonomy, we decomposed the categorical target RDMs into separate regressors for each between-category relationship (six regressors for behavior model, 10 for the taxonomy model). To evaluate these two flexible behavior and taxonomy models, in each ROI we estimated the coefficient of partial determination (partial R^2) and AIC separately for each model and attention task within each participant, then averaged these model fits over the two attention tasks. The six-regressor behavior model captured on average over 2 times more variance (adjusted R^2) than the single-regressor behavioral category target RDM in LO, VT, OP, IPS, left PCS, and vPC/PM, suggesting that the representations of some behaviors are more similar than others. The 10-regressor taxonomy model accounted for well over 4 times more variance than the single-regressor taxonomic category target RDM in pEV, iEV, and VT. Based on permutation tests, partial R^2 for the behavior model significantly exceeded that of the animal taxonomy model in sEV, LO, VT, OP, IPS, left PCS, and vPC/PM (Fig. 3*C*), and AIC for the behavior model was significantly lower for all 10 ROIs (*SI Text)*. Surprisingly, the behavior model accounted for over 2.5 times more variance in VT neural representational geometry than

did the taxonomy model (behavior model: 23.8% of variance; taxonomy model: 8.8% of variance).

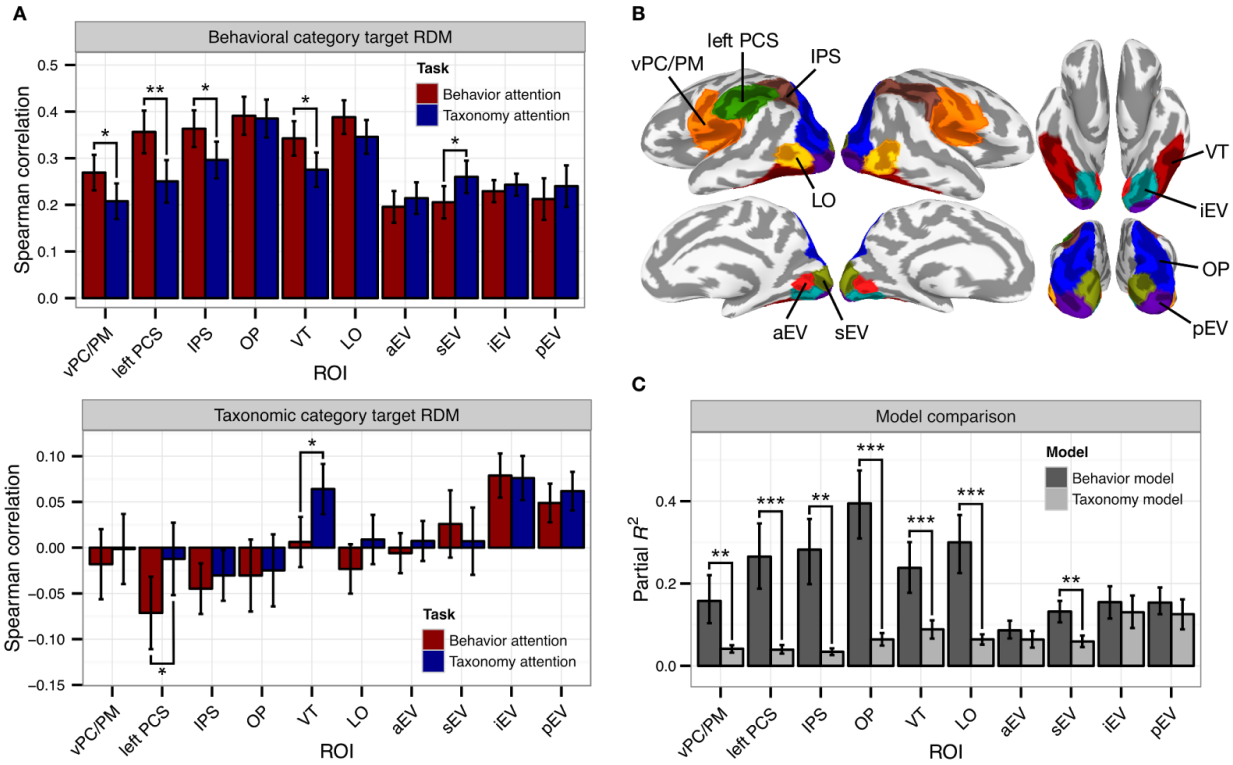

Fig. 3. Attention alters representational geometry in functionally-defined ROIs. (*A*) Attentional differences in Spearman correlation between neural RDMs and the behavioral and taxonomic category target RDMs. All error bars indicate bootstrapped 95% confidence intervals for within-participants comparisons. (*B*) Ten functional ROIs identified by parcellating the cerebral cortex based on representational geometry. ROI labels: posterior early visual cortex (pEV), inferior early visual cortex (iEV), superior early visual cortex (sEV), anterior early visual cortex (aEV), lateral occipitotemporal cortex (LO), ventral temporal cortex (VT), occipitoparietal and posterior parietal cortex (OP), intraparietal sulcus (IPS), left postcentral sulcus (left PCS), and ventral pericentral and premotor cortex (vPC/PM). (*C*) Comparison of model fit for the six-regressor behavior model and 10-regressor taxonomy model. *$p < .05$, **$p < .01$, ***$p < .001$, two-tailed permutation test.

We next isolated cells of the neural RDM capturing distances between two conditions that differed on one dimension and were matched on the other; i.e., different behaviors performed by animals from the same taxonomic category, or animals of different taxonomic categories performing the same behavior (Fig. 4*A*). Although we hypothesized that attention expands the distances between task-relevant representations and collapses the distances between task-irrelevant representations as depicted in Fig. 4*B* (40, 41), note that diagonal distances do not change; that is, the effect of attention on distances between conditions that differ on both dimensions is ambiguous. Thus, focusing on the correlation distances between pairs of conditions that differ on only one dimension affords a more unconfounded examination of the effects of attention. A significant increase in, e.g., between-taxon correlation distances within each behavior (Fig. 4*A*, red) when attending to behavior can also be interpreted as a decrease in within-taxon distances when attending to taxonomy; therefore, we refer to this effect as enhancing categoricity. The following tests were motivated by a linear mixed-effects model yielding a significant interaction between attention task, category relationship, and ROI (*SI Text*). Permutation tests indicated that attention significantly enhanced categoricity for both groups of distances in left PCS (between-taxon, within-behavior distances: $p = .002$; between-behavior, within-taxon distances: $p = .010$) and VT (between-taxon, within-behavior distances: $p = .028$; between-behavior, within-taxon distances: $p = .009$). Attention significantly enhanced the categoricity of between-taxon distances within behaviors in vPC/PM ($p = .007$), effectively collapsing taxonomic distinctions when attending to behavior. An inverted attentional effect was observed in sEV (between-taxa within-behavior distances: $p = .028$). The expansion of distances between attended category representations is illustrated with multidimensional scaling of the representational geometries in left PCS and VT (Fig. 4*C*).

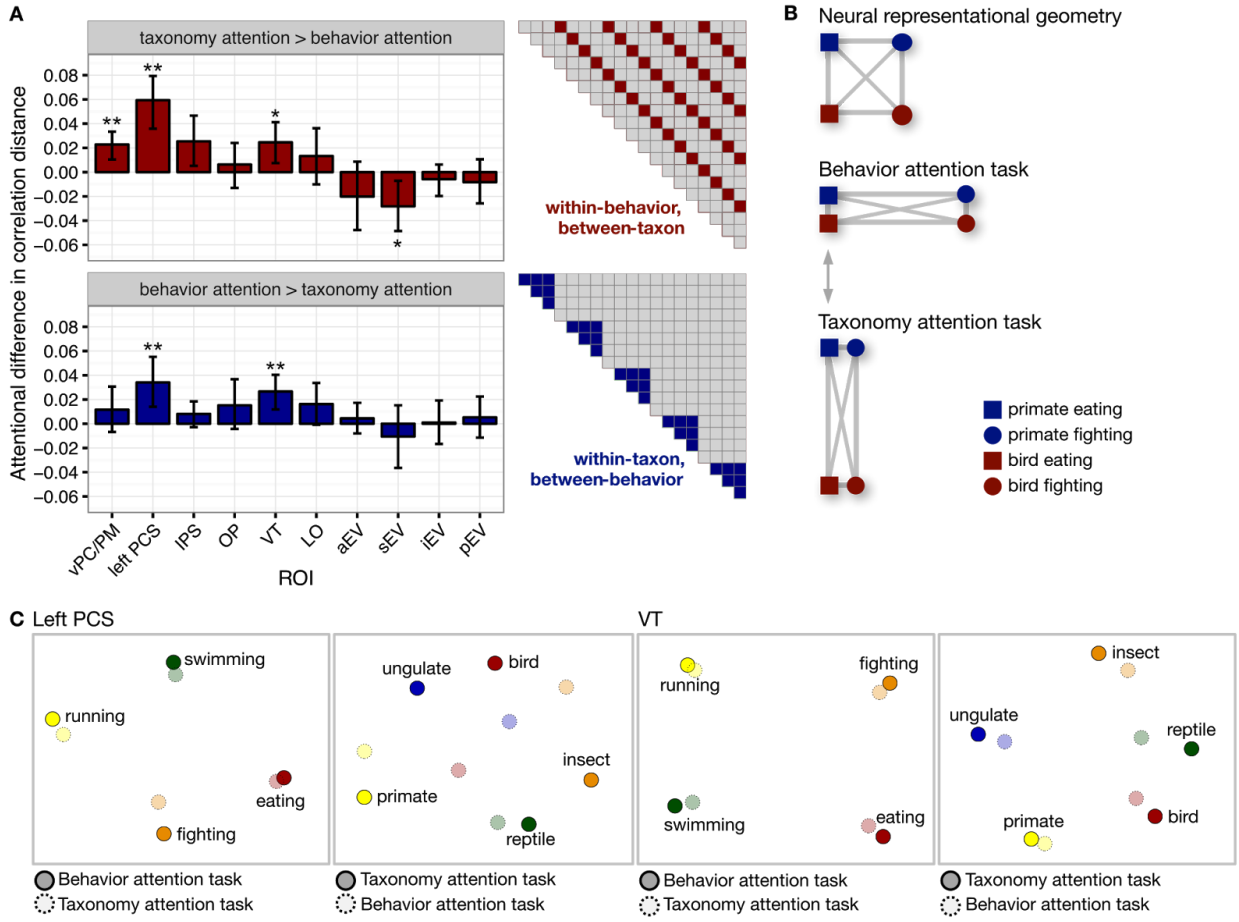

Fig. 4. Attention enhances the categoricity of neural responses patterns. (*A*) Attentional enhancement when restricted to within-category distances for both behavioral and taxonomic categories. Error bars indicate bootstrapped 95% confidence intervals. (*B*) Schematic illustrating how neural distances are expanded along the behaviorally relevant dimensions while task-irrelevant distances are collapsed (40, 41). (*C*) Multidimensional scaling (MDS) solutions for left PCS and VT depict the attentional expansion of between-category distances. *$p < .05$, **$p < .01$, two-tailed permutation test.

Discussion

The present study was motivated by the following question: How does attention prioritize certain semantic features of a complex stimulus in service of behavioral goals? We hypothesized that attention may enhance certain features of semantic information encoded in distributed neural populations by transiently altering representational geometry (48). Our

findings provide neural evidence for psychological theories of attentional deployment in categorization (40, 41) by demonstrating that attention selectively increases distances between stimulus-evoked neural representations along behaviorally relevant dimensions. To expand on prior work examining early visual (e.g., orientation, contrast, color, motion direction; 32, 36, 49) and object category (34, 35, 37) representation, we used dynamic, naturalistic stimuli to demonstrate that attention alters the representation of both animal taxonomy and observed actions according to a similar principle.

When participants attended explicitly to animal behavior, the categoricity of action representation increased most dramatically in premotor, pericentral, and postcentral somatomotor areas supporting action recognition (22, 23, 44), intraparietal areas implicated in executive control (50), and VT. Earlier areas exhibiting robust representation of animal behavior, such as LO and OP, were not strongly modulated by the attentional manipulation. Attending to animal taxonomy increased the categoricity of animal representation in VT, consistent with accounts of neural representation of animals and objects (42, 43, 51), as well as left PCS, but not in lateral occipital or early visual areas. Note that attending to behavior induced a negative correlation for the taxonomic category target RDM in left PCS, while attending to taxonomy abolished this effect. This negative correlation when attending to behavior could be driven by increased distances between behavior representations within each animal taxon.

Performing a categorization task requiring attention to either animal taxonomy or behavior enhances the categoricity of neural representations by accentuating task-relevant distinctions and reducing unattended distinctions. Attention sculpts representational geometry in late-stage sensorimotor areas, and this effect was not observed in early perceptual areas. Our results demonstrate that the representational geometry of semantic information in systems such as VT and somatomotor cortex is dynamic and actively tuned to behavioral goals, rather than being solely a reflection of static conceptual knowledge.

Numerous visual areas coded for both taxonomy and behavior, suggesting these two types of information are encoded in distributed population codes in a superimposed or multiplexed fashion (9, 51). However, the behavior model accounted for notably more variance in neural representation throughout the cortex than the taxonomy model—even in areas typically associated with animal category representation, such as VT (42, 43). This may in part be due to the heterogeneity of exemplar species within each taxon and the prevalence of motion energy

information when viewing naturalistic video stimuli (4). Work by others shows, however, that lateral fusiform cortex responds strongly to dynamic stimuli that depict agentic behavior with no biological form (52, 53), and biological motion and social behaviors drive responses in face-selective temporal areas in the macaque (54). The current findings complement a recent study by Çukur and colleagues (37) reporting that attending to a particular object category (a human or a vehicle) increases the responsivity of widely distributed cortical voxels to the presence of that object category. Our findings suggest that, beyond increasing the cortical coverage of task-relevant representation (as shown in Fig. 1), attention may enhance performance by altering representational geometry so as to make task-relevant semantic distinctions more sharply defined.

Scaling up the effects of attention on single neurons to population responses and multivoxel patterns of activity is an outstanding challenge. Top-down signals (55, 56) may bias how information is encoded at the population level by altering both neuronal responsivity and interneuronal correlations (30, 31, 57, 58) in order to optimize representational discriminability for downstream read-out systems. Our findings suggest a model whereby attention alters population encoding so as to enhance the discriminability of task-relevant representational content. At an algorithmic level (59), attention may tune a feature space of arbitrary dimensionality by dynamically altering population encoding. This would serve to disentangle (60) task-relevant representations from task-irrelevant information by increasing attended distinctions and reducing unattended distinctions, thereby enhancing behavioral performance.

Materials and Methods

Participants. Twelve right-handed adults (seven female; mean age = 25.4 ± 2.6 SD years) with normal or corrected-to-normal vision participated in the attention experiment. All participants reported no neurological conditions. Additionally, 19 adults, including the 12 from the attention experiment, participated in a separate scanning session for the purposes of hyperalignment. All participants gave written, informed consent prior to participating in the study, and the study was approved by the Institutional Review Board of Dartmouth College.

Stimuli and design. Each of the 20 conditions in the fully-crossed design comprised two unique exemplar clips and their horizontally-flipped counterparts, for a total of 40 clips and 80 total clip stimuli (Table S1). Each trial consisted of a 2 s video clip presented without sound followed by 2 s fixation period in a rapid event-related design. All 80 stimuli, as well as four

behavior repetition events, four taxon repetition events, and four null events were presented in pseudorandom order in each of 10 runs. At the beginning of each run, participants were instructed to pay attention to either taxonomy or behavior types and press the button only when they observed a repetition of that type. There were five behavior attention runs and five taxonomy attention runs presented in counterbalanced order across participants. In an independent scanning session, participants were presented with approximately 63 min of the *Life* nature documentary narrated by David Attenborough for the purpose of hyperalignment.

Preprocessing. For each participant, functional time series data were de-spiked, corrected for slice timing and head motion, normalized to the ICBM 452 template in MNI space, and spatially smoothed with a 4 mm FWHM Gaussian kernel using AFNI (61). A general linear model (GLM) was used to estimate stimulus-evoked BOLD responses for each of the 20 conditions using AFNI's 3dREMLfit. Cortical surfaces were reconstructed from structural scans using FreeSurfer, aligned according to curvature patterns on the spherical surface projection (62), and visualized using SUMA (63). Surface-based searchlight whole-brain hyperalignment (3, 6) parameters estimated from the *Life* documentary were applied to the data from the attentional experiment prior to multivariate analysis.

Multivariate pattern analysis. Representational similarity analysis (45) was applied using 100-voxel surface-based searchlights (46, 47). The pairwise correlation distances between stimulus-evoked response patterns for the 20 conditions quantified the representational geometry within a searchlight (48). Two categorical target RDMs were constructed: one of these RDMs discriminated the animal taxa invariant to behavior, the other discriminated the behaviors invariant to taxonomy. Least squares multiple regression was performed to model the observed neural RDM as a weighted sum of the two categorical target RDMs. All maps were corrected for multiple comparisons at $p = .05$ without choosing an arbitrary uncorrected threshold using threshold-free cluster enhancement (TFCE) with the recommended values (64). Global increases in regression coefficients were computed separately for each categorical target RDM. For the behavioral category target RDM, the mean regression coefficients were computed across all searchlight regression coefficients surviving TFCE in both attention conditions, and a permutation test was used to evaluate the significance of an attentional change in the mean regression coefficient across participants. This procedure was repeated for

the taxonomic category target RDM considering all searchlight regression coefficients that survived TFCE in both attention tasks.

Cluster analysis was used to identify extensive regions of the cortical surface characterized by shared representational geometry in an unsupervised manner (42). Gaussian mixture models were used to cluster searchlights according to their representational geometry at varying values of k components (clusters). We evaluated the reproducibility of parcellations across participants at values of k from 2 to 30 using a split-half resampling approach (100 iterations per k) that has previously been applied to functional parcellations based on resting-state functional connectivity (65). The reproducibility analysis indicated local maxima at k = 2, 4, 14, 19, and 23 (Fig. S4*A*), and these cluster solutions can then be mapped back to the cortical surface (e.g., Fig. S4*B*). All subsequent analyses were performed on ROIs derived from the parcellation at k = 19 based on the behavior attention data. Both the clustering algorithm and the reproducibility analysis are agnostic to any particular representational geometry or attentional effect (66).

For each of the 10 ROIs comprising early visual areas, the ventral visual pathway, the dorsal visual pathway, and motor areas, we used the stimulus-evoked patterns of activation across all referenced voxels to compute the neural RDMs for both attention conditions. We first tested for attentional differences in Spearman correlation between the observed neural RDM and the target RDMs. A linear mixed-effects model evaluating attentional changes in Spearman correlations across ROIs yielded a significant two-way interaction between attention task and ROI, and a significant three-way interaction between attention task, target RDM, and ROI (*SI Text*). Spearman correlations were Fisher transformed prior to statistical testing, and significance of the attentional effect was assessed per ROI using permutation tests (two-tailed). A linear mixed-effects model evaluating attentional changes in within-category distances yielded a significant interaction between attentional task, category relationship, and ROI, with participants and pairwise distances modeled as random effects (*SI Text*). Permutations tests were used to assess attentional differences in mean within-category correlation distances within each ROI. Model fits were evaluated for the behavior and taxonomy models using both and partial R^2 and AIC, and significance was assessed using permutation tests. To visualize attentional changes in representational geometry, we first computed 40 × 40 neural RDMs based on the 20 conditions for both attention tasks and averaged these across participants. We then averaged the between-category distances within each category of the other factor (e.g., 10 average between-taxon distances within each behavioral category; see Fig. 4).

Multidimensional scaling was then applied to these average distances. In the resulting two-dimensional space, the Procrustes transformation was used to rotate the positions of each condition from one attention task to the other to best visualize the overall attentional expansion of between-category distances. All multivariate pattern analyses were performed using the PyMVPA package (www.pymvpa.org; (67).

Acknowledgments

We thank Kelsey Wheeler for help in collecting the video stimuli and Courtney Rogers for administrative support. Funding was provided by National Institutes of Mental Health grants: F32MH085433-01A1 (Connolly), and 5R01MH075706 (Haxby); and by the National Science Foundation grant: NSF1129764 (Haxby).

Author Contributions

S.A.N. and J.V.H. designed research; S.A.N. performed research; A.C.C., N.N.O., Y.O.H., J.S.G., M.V.D.O.C., J.G., and M.I.G. contributed analytic tools; S.A.N. analyzed data; and S.A.N. and J.V.H. wrote the paper.

References

1. Tsotsos JK (2011) *A Computational Perspective on Visual Attention* (MIT Press, Cambridge, MA).
2. Carrasco M (2011) Visual attention: The past 25 years. *Vision Res* 51(13):1484–1525.
3. Haxby JV, et al. (2011) A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72(2):404–416.
4. Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76(6):1210–1224.
5. Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. *Science* 303(5664):1634–1640.
6. Guntupalli JS, et al. (2016) A model of representational spaces in human cortex. *Cereb Cortex*.
7. Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7(5):358–366.
8. Kriegeskorte N, et al. (2008) Matching categorical object representations in inferior

temporal cortex of man and monkey. *Neuron* 60(6):1126–1141.

9. Haxby JV, Connolly AC, Guntupalli JS (2014) Decoding neural representational spaces using multivariate pattern analysis. *Annu Rev Neurosci* 37:435–456.

10. Dayan P, Abbott LF (2001) *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge, MA).

11. Chen Y, Yuzhi C, Geisler WS, Eyal S (2006) Optimal decoding of correlated neural population responses in the primate visual cortex. *Nat Neurosci* 9(11):1412–1420.

12. Graf ABA, Kohn A, Jazayeri M, Movshon JA (2011) Decoding the activity of neuronal populations in macaque primary visual cortex. *Nat Neurosci* 14(2):239–245.

13. Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310(5749):863–866.

14. Kiani R, Esteky H, Mirpour K, Tanaka K (2007) Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex. *J Neurophysiol* 97(6):4296–4309.

15. Freiwald WA, Tsao DY (2010) Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science* 330(6005):845–851.

16. Rolls ET, Tovee MJ (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol* 73(2):713–726.

17. Uchida N, Takahashi YK, Tanifuji M, Mori K (2000) Odor maps in the mammalian olfactory bulb: domain organization and odorant structural features. *Nat Neurosci* 3(10):1035–1043.

18. Georgopoulos A, Schwartz A, Kettner R (1986) Neuronal population coding of movement direction. *Science* 233(4771):1416–1419.

19. Lewis JE, Kristan WB Jr (1998) A neuronal network for computing population vectors in the leech. *Nature* 391(6662):76–79.

20. Rigotti M, et al. (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497(7451):585–590.

21. Mitchell TM, et al. (2008) Predicting human brain activity associated with the meanings of nouns. *Science* 320(5880):1191–1195.

22. Oosterhof NN, Wiggett AJ, Diedrichsen J, Tipper SP, Downing PE (2010) Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. *J Neurophysiol* 104(2):1077–1089.

23. Oosterhof NN, Tipper SP, Downing PE (2012) Viewpoint (in)dependence of action representations: an MVPA study. *J Cogn Neurosci* 24(4):975–989.

24. Haxby JV, et al. (2001) Distributed and overlapping representations of faces and objects in

ventral temporal cortex. *Science* 293(5539):2425–2430.

25. Kriegeskorte N, et al. (2008) Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60(6):1126–1141.

26. Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. *Neuron* 26(3):703–714.

27. Treue S, Martínez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399(6736):575–579.

28. Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61(2):168–185.

29. Cohen MR, Maunsell JHR (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12(12):1594–1600.

30. Ruff DA, Cohen MR (2014) Attention can either increase or decrease spike count correlations in visual cortex. *Nat Neurosci* 17(11):1591–1597.

31. Downer JD, Niwa M, Sutter ML (2015) Task engagement selectively modulates neural correlations in primary auditory cortex. *J Neurosci* 35(19):7565–7574.

32. Jehee JFM, Brady DK, Tong F (2011) Attention improves encoding of task-relevant features in the human visual cortex. *J Neurosci* 31(22):8210–8219.

33. Serences JT, Boynton GM (2007) Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55(2):301–312.

34. Harel A, Kravitz DJ, Baker CI (2014) Task context impacts visual object processing differentially across the cortex. *Proc Natl Acad Sci USA* 111(10):E962–71.

35. Erez Y, Duncan J (2015) Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. *J Neurosci* 35(36):12383–12393.

36. Sprague TC, Serences JT (2013) Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nat Neurosci* 16(12):1879–1887.

37. Çukur T, Nishimoto S, Huth AG, Gallant JL (2013) Attention during natural vision warps semantic representation across the human brain. *Nat Neurosci* 16(6):763–770.

38. Kriegeskorte N, Kievit RA (2013) Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci* 17(8):401–412.

39. Edelman S (1998) Representation is representation of similarities. *Behav Brain Sci* 21(04):449–467.

40. Nosofsky RM (1986) Attention, similarity, and the identification-categorization relationship. *J Exp Psychol Gen* 115(1):39–57.

41. Kruschke JK (1992) ALCOVE: an exemplar-based connectionist model of category learning. *Psychol Rev* 99(1):22–44.

42. Connolly AC, et al. (2012) The representation of biological classes in the human brain. *J Neurosci* 32(8):2608–2618.

43. Sha L, et al. (2015) The animacy continuum in the human ventral vision pathway. *J Cogn Neurosci* 27(4):665–678.

44. Oosterhof NN, Tipper SP, Downing PE (2013) Crossmodal and action-specific: neuroimaging the human mirror neuron system. *Trends Cogn Sci* 17(7):311–318.

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

46. Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci USA* 103(10):3863–3868.

47. Oosterhof NN, Wiestler T, Downing PE, Diedrichsen J (2011) A comparison of volume-based and surface-based multi-voxel pattern analysis. *NeuroImage* 56(2):593–600.

48. Kriegeskorte N, Kievit RA (2013) Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci* 17(8):401–412.

49. Serences JT, Boynton GM (2007) Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55(2):301–312.

50. Petersen SE, Posner MI (2012) The attention system of the human brain: 20 years after. *Annu Rev Neurosci* 35:73–89.

51. Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci* 15(8):536–548.

52. Gobbini MI, Koralek AC, Bryan RE, Montgomery KJ, Haxby JV (2007) Two takes on the social brain: a comparison of theory of mind tasks. *J Cogn Neurosci* 19(11):1803–1814.

53. Grossman ED, Blake R (2002) Brain areas active during visual perception of biological motion. *Neuron* 35(6):1167–1175.

54. Russ BE, Leopold DA (2015) Functional MRI mapping of dynamic visual features during natural viewing in the macaque. *NeuroImage* 109:84–94.

55. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222.

56. Baldauf D, Desimone R (2014) Neural mechanisms of object-based attention. *Science* 344(6182):424–427.

57. Cohen MR, Maunsell JHR (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12(12):1594–1600.

58. Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. *Nat Rev Neurosci* 7(5):358–366.

59. Marr D (1982) *Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information* (MIT Press).

60. DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? *Neuron* 73(3):415–434.

61. Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29(3):162–173.

62. Fischl B, Sereno MI, Tootell RB, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8(4):272–284.

63. Saad ZS, Reynolds RC, Argall B, Japee S, Cox RW (2004) SUMA: An interface for surface-based intra- and inter-subject analysis with AFNI. *2004 IEEE International Symposium on Biomedical Imaging: From Macro to Nano* doi:10.1109/isbi.2004.1398837.

64. Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage* 44(1):83–98.

65. Yeo BTT, et al. (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* 106(3):1125–1165.

66. Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12(5):535–540.

67. Hanke M, et al. (2009) PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7(1):37–53.

Supporting Information

Attention selectively reshapes the geometry of distributed semantic representation

Samuel A. Nastase, Andrew C. Connolly, Nikolaas N. Oosterhof, Yaroslav O. Halchenko, J. Swaroop Guntupalli, Matteo Visconti di Oleggio Castello, Jason Gors, M. Ida Gobbini, James V. Haxby

SI Text

Stimuli and design. Horizontally-flipped versions of each video were created for a total of 80 visually unique stimuli. Each clip was 2 s in duration and presented without sound. See Table S1 for a description of all video clips and Video S1 for sample stimuli. Clips for the attention experiment were extracted from nature documentaries (*Life*, *Life of Mammals*, *Microcosmos*, *Planet Earth*) and YouTube videos matched for resolution. The clips used in the attention experiment were not included in the segment of the documentary presented. The *Life* documentary was presented in four runs of similar duration, and included both the visual and auditory tracks. Stimuli were presented using PsychoPy (68).

For each run, the pseudorandom trial order was first constructed such that no animal or action types were repeated. Next, four animal repetition events and four action repetition were events were pseudorandomly inserted as sparse catch trials such that a repetition event of each type fell somewhere within each quarter of the run. Four additional 2 s null events consisting of only a fixation cross were inserted into each run to effect temporal jittering. Each run consisted of 80 trials of interest, four animal repetition events, four action repetition events, and four null events. This resulted in 92 events per run, plus 12 s fixation before and after the events of interest, for a total run length of 392 s (~6.5 min). Ten unique runs were constructed and run order was counterbalanced across participants.

The same button was pressed for repetitions of both types. Button presses were only elicited by repetition events and were therefore sparse. Participants were informed that repetition events would be sparse and that they should not pay attention to or press the button if they noticed repetitions of the unattended type. Participants were only instructed to maintain fixation when the fixation cross was present, not during the presentation of the clips. In the

movie session, participants were instructed to remain still and watch the documentary as though they were watching a movie at home.

Behavioral data. Participants were highly accurate in detecting the sparse repetition events for both attention conditions with high accuracy (mean accuracy for animal attention condition = 0.993, *SD* = 0.005; mean accuracy for action attention condition = 0.994, *SD* = 0.005). There was no significant task-related difference in either accuracy ($t(11) = 0.469$, $p = 0.648$), or signal detection theoretic measures of sensitivity ($t(11) = 0.116$, $p = 0.910$) and bias ($t(11) = 0.449$, $p = 0.662$) adjusted for logistic distributions.

Image acquisition. All functional and structural images were acquired using a 3 T Philips Intera Achieva MRI scanner (Philips Medical Systems, Bothell, WA) with a 32-channel phased-array SENSE (SENSitivity Encoding) head coil. For the attention experiment, functional images were acquired in an interleaved fashion using single-shot gradient-echo echo-planar imaging with a SENSE reduction factor of 2 (TR/TE = 2000/35 ms, flip angle = 90°, resolution = 3 mm isotropic, matrix size = 80 × 80, FOV = 240 × 240 mm, 42 transverse slices with full brain coverage and no gap). Each run began with two dummy scans to allow for signal stabilization. For each participant 10 runs were collected, each consisting of 196 dynamic scans totaling 392 s (~6.5 min). At the end of each scanning session, a structural scan was obtained using a high-resolution T1-weighted 3D turbo field echo sequence (TR/TE = 8.2/3.7 ms, flip angle = 8°, resolution = 1 mm isotropic, matrix size = 256 × 256 × 220, FOV = 240 × 188 × 220 mm).

For the movie session, functional images also were also acquired in an interleaved order using single-shot gradient-echo echo-planar imaging (TR/TE = 2500/35 ms, flip angle = 90°, resolution = 3 mm isotropic, matrix size = 80 × 80, and FOV = 240 × 240 mm; 42 transverse slices with full brain coverage and no gap). Four runs were collected for each participant, consisting of 374, 346, 377, and 412 dynamic scans, totaling 935 s (~15.6 min), 865 s (~14.4 min), 942.5 s (~15.7 min), and 1030 s (~17.2 min), respectively. At the end of this session, a structural scan was obtained using a high-resolution T1-weighted 3D turbo field echo sequence (TR/TE = 8.2/3.7 ms, flip angle = 8°, resolution = 1 mm isotropic, matrix size = 256 × 256 × 220, and FOV = 240 × 188 × 220). For participants included in both the attention experiment and the movie session, structural images were registered and averaged to increase signal-to-noise ratio.

Preprocessing and model estimation. Functional time series data were de-spiked and corrected for slice timing, then functional images were then motion corrected in two passes. First, functional images were initially motion corrected, then averaged across time to create a high-contrast reference volume. Motion correction parameters were then re-estimated in a second pass using the reference volume as the target. Affine transformation parameters were then estimated to coregister the reference volume and the participant's averaged structural scans. Each participant's averaged structural scan was then normalized to the ICBM 452 template brain in MNI space. These transformation matrices were concatenated and each participant's data were motion corrected and normalized to the template via the participant's anatomical scan in a single interpolation step. All subsequent analyses were performed in MNI space. Functional images were spatially smoothed with a 4 mm Gaussian kernel. Signal intensities were normalized to percent signal change prior to applying the general linear model. Functional time series from the *Life* movie session were analyzed using the same preprocessing pipeline. Prior to hyperalignment, time series data were bandpass filtered to remove frequencies higher than 0.1 Hz and lower than 0.0067 Hz. Head motion parameters and the mean time series derived from the FreeSurfer segmentation of the ventricles were regressed out of the signal.

Stimulus-evoked BOLD responses to each event were modeled using a simple hemodynamic response function (AFNI's GAM response model) adjusted for a 2 s stimulus duration. Nuisance regressors accounting for repetition events, button presses, and head motion were included in the model. For representational similarity analyses, beta parameters were estimated over the five animal attention runs, then separately over the five action attention runs. Time points subtending abrupt head movements greater than 1 mm of displacement or 1 degree of rotation were censored when fitting the general linear model. For each of the two attention conditions, the stimulus-evoked response pattern for each action-animal condition was estimated from 20 trials presented in pseudorandom order over the course of five separate runs (interspersed with runs from the other attention condition). Therefore we expect these response patterns (and the subsequent neural RDMs) to be relatively robust to instrumental noise, temporal autocorrelation and intrinsic physiological correlations in the preprocessed time series data (69). Betas for each voxel were *z*-scored across the 20 conditions before and after hyperalignment, and prior to any multivariate analysis. Note that constructing neural RDMs by computing the correlation distance between response pattern vectors (rather than, e.g., Euclidean distance) entails that

the subsequent multivariate analyses are invariant to differences in regional-average activity levels within a searchlight or ROI (45). For searchlight classification analyses (Fig. S2), beta parameters were estimated separately for each run.

Hyperalignment. Surface-based searchlight whole-brain hyperalignment (3, 6) was performed based on data collected while participants viewed the *Life* nature documentary. Each surface-based searchlight referenced the 200 nearest voxels from the associated volume, selected based on their geodesic proximity to the searchlight center. The time series of response patterns elicited by the movie stimulus was rotated via the Procrustes transformation in order to achieve optimal functional alignment across participants and the estimated transformation matrices for each searchlight were aggregated (Fig. S1*A*). Hyperalignment transformation parameters estimated from the movie data were then applied to the independent attention experiment data. All subsequent multivariate analyses were applied to the hyperaligned data.

Searchlight representational similarity regression. Each surface-based searchlight referenced the 100 nearest voxels to the searchlight center based on geodesic distance on the cortical surface. For each searchlight, both the observed neural RDM and the target RDMs were ranked and standardized prior to regression (see (70), p. 140). Since we suspect the neural representational space does not respect the magnitude of dissimilarity specified by our models, we relax the linear constraint and ensure only monotonicity (analogous to Spearman correlation, in keeping with (45), p. 23). Although applying the rank transform prior to least squares linear regression is relatively common practice, this approach may emphasize main effects at the expense of interaction effects; however, in the current experiment, we have no a priori hypotheses corresponding to interaction terms. Intercept terms in the estimated models were negligible (ranging from -1.249×10^{-16} to 1.331×10^{-16}) across all searchlights, task conditions, and participants. The searchlight analysis was performed in the hyperaligned space, then the results were projected onto the surface reconstruction of the reference participant in the hyperalignment algorithm.

Cluster-level inference for searchlight analysis. To ascertain the statistical significance of a group map with correction for multiple comparisons, a Monte Carlo simulation permuting condition labels was used to estimate a null TFCE distribution (23, 64). First, 100 null searchlight maps were generated for each participant by randomly permuting condition labels

within each observed searchlight RDM; then 10,000 null TFCE maps were constructed by randomly sampling from these null data sets in order to estimate a null TFCE distribution. In the case of searchlight classification (Fig. S2), labels were shuffled within each run and each category of the crossed factor (e.g., the four behavior labels were permuted within each of the five taxa), then the full cross-validation scheme was applied (71). This method for multiple comparisons correction was implemented using the CoSMoMVPA package for Matlab (cosmomvpa.org). In Fig. S3, clusters surviving correction for multiple comparisons are indicated by white contours and subthreshold searchlights are displayed transparently. White contours returned by SUMA denoting clusters surviving TFCE correction were dilated by four pixels in GIMP to increase visibility.

Functional parcellation. Prior to cluster analysis, the observed neural RDMs for each surface-based searchlight were converted from correlation distances to Fisher transformed correlation values and averaged across participants. Gaussian mixture modeling is a probabilistic generalization of the *k*-means algorithm, and models the 20,484 searchlights as a mixture of *k* overlapping Gaussian distributions in a 190-dimensional feature space defined by the upper triangular of the 20 × 20 observed neural RDM. The clustering algorithm was implemented using the scikit-learn machine learning library for Python (72). To test parcellation reproducibility, for each of 100 resampling iterations, half of the participants were randomly assigned to a training set, while the other half were assigned to a test set (73). Surface-based searchlight RDMs for each participant were then meaned across participants in the separate training and test sets. Gaussian mixture models were estimated on the training set for each of *k* components ranging from 2 to 30. Test data were then assigned to the nearest mean of the model estimated from the training data. A separate mixture model was then estimated for the test data, and the predicted cluster labels (based on the training data) were compared to the actual cluster labels using adjusted mutual information (AMI; 74). AMI compares cluster solutions and assigns a value between 0 and 1, where 0 indicates random labeling and 1 indicates identical cluster solutions (robust to a permutation of labels, adjusted for greater fit by chance at higher *k*). Note that, unlike previous applications (65), we cross-validated AMI at the participant level rather than partitioning at the searchlight level. Separate cluster solutions were obtained for each attention task condition to ensure the clustering algorithm did not attenuate attentional effects. Cluster solutions yielded qualitatively similar surface parcellations for data from both the action attention task and the animal attention task, however the action attention

task tended toward more reproducible solutions at higher *k*. Note that clustering cortical searchlights according to the pairwise neural distances between a certain set of experimental conditions should not be expected to yield a generally valid cluster solution for the entire brain. Furthermore, although spatial smoothing, overlapping searchlights, and hyperalignment induce spatial correlations, there is nothing intrinsic to the clustering algorithm that ensures spatial contiguity (on the cortical surface) or bilaterality in the resulting clusters.

Attentional differences in correlation. Spearman correlations between the neural RDMs and target RDMs were Fisher transformed prior to statistical testing. Prior to permutation testing, we constructed a linear mixed-effects model to predict correlations with the categorical target RDMs using Task, Target RDM, and ROI, and their two- and three-way interactions as fixed effects, with Participant modeled as a random effect (random intercepts). The Task variable captured the two attentional task conditions, Target RDM represented the behavioral and taxonomic category target RDMs, and ROI represented the 10 regions of interest. Mixed-effects modeling was performed in R using *lme4* (75). Statistical significance was assessed using a Type III analysis of deviance. Significant main effects were observed for ROI ($\chi^2(9) = 115.690$, $p < .001$) and Target RDM ($\chi^2(9) = 69.640$, $p < .001$), but not for Task, while the Target RDM × ROI interaction was significant ($\chi^2(9) = 112.442$, $p < .001$). The Task × ROI interaction was also significant ($\chi^2(9) = 23.301$, $p = .006$), suggesting that the attentional manipulation more strongly affected correlations in certain ROIs than others. Finally, the three-way Task × Target RDM × ROI interaction was significant ($\chi^2(9) = 22.034$, $p = .009$), motivating the following within-ROI tests. To assess the statistical significance of differences in Spearman correlation as a function of attention task, exact tests were performed in which the mean difference in correlation was computed for all possible permutations of the within-participants attention task assignments (2^{12} = 4,096 permutations).

Attentional differences in representational distance. A linear mixed-effects model evaluating attentional changes in within- and between-category distances across the 10 ROIs motivated within-ROI tests for task-related changes in representational distances. To test for attentional differences in within- and between-category distances for both animal and action types, we selected the cells of the neural RDM corresponding to, e.g., the within-category distances between animal types. Correlation distances were converted to Fisher-transformed correlations prior to statistical testing. Rather than averaging the pairwise distances across cells of the target RDM within each participant, cells corresponding to particular pairwise distances were

included as a random effect (as per an items analysis; (76). We constructed a linear mixed-effects model to predict observed correlation distances based on Task, Category, and ROI, and their two- and three-way interactions as fixed effects, with Participant and Cell as random effects (random intercepts). Task represented the attentional task condition, Category represents the category relationship (within-behavior or within-taxon), ROI indicates the 10 ROIs investigated above, and Cell indicates particular cells (pairwise relationships) of the target RDM. Statistical significance was assessed using a Type III analysis of deviance. Significant main effects were observed for ROI ($\chi^2(9) = 66.850$, $p = .003$) and Category ($\chi^2(1)$ 13.047, $p < .001$), but not for Task, while the two-way Category × ROI interaction ($\chi^2(9) = 165.725$, $p < .001$) was significant. Most importantly, the three-way Task × Category × ROI interaction term was highly significant ($\chi^2(9) = 33.322$, $p < .001$), motivating the within-ROI tests. Permutation tests were then used to test the attentional differences in correlation distance for each ROI.

Visualizing representational space. To visualize attentional changes in representational geometry, we used multidimensional scaling. For a given ROI, we computed a 40 × 40 neural RDM based on the 20 conditions for both attention tasks for each participant. These RDMs were then averaged across participants within an ROI. To visualize attentional changes in observed action representation, we computed an 8 × 8 distance matrix comprising the mean between-behavior distances within each taxon (as in Fig. 4). For taxonomy representation, we computed the average between-taxon distances within each behavior to construct a 10 × 10 matrix. Distances were computed between conditions for both attention (e.g., resulting in a single 8 × 8 distance matrix rather than two separate 4 × 4 matrices for behavior representation) to ensure that distances for both attention tasks were on the same scale. These distance matrices were then submitted to metric multidimensional scaling implemented in scikit-learn (72). In the case of behavior representation, for example, this resulted in eight positions in a two-dimensional space. However, because we were interested in the overall attentional expansion between conditions (and less concerned with, e.g., the distance between one condition in one attention task and another condition in the other attention task), the positions in the resulting two-dimensional solution were then split according to attention task, and the Procrustes transformation (without scaling) was used to best align the conditions within one attention task to another. This transformation preserves the relationships between conditions within each task and captures the attentional expansion of between-category distances.

Evaluating model fit. To test for differences in fit for the taxonomy and behavior models, we computed both the partial R^2 and AIC for the six- and 10-regressor models separately for both attention task conditions within each participant (77). Partial R^2 can be interpreted as the proportion of variance accounted for by one model controlling for any variance accounted for by the other model, and was computed separately for each attention task and then averaged across tasks within participants. These average partial R^2 values were then submitted to an exact test to assess significance across participants. Similarly, we computed the difference in AIC for the six- and 10-regressor models for each attention task condition within each participant, then averaged across the attention tasks. These differences in AIC were assessed statistically using an exact test permuting the sign of the difference.

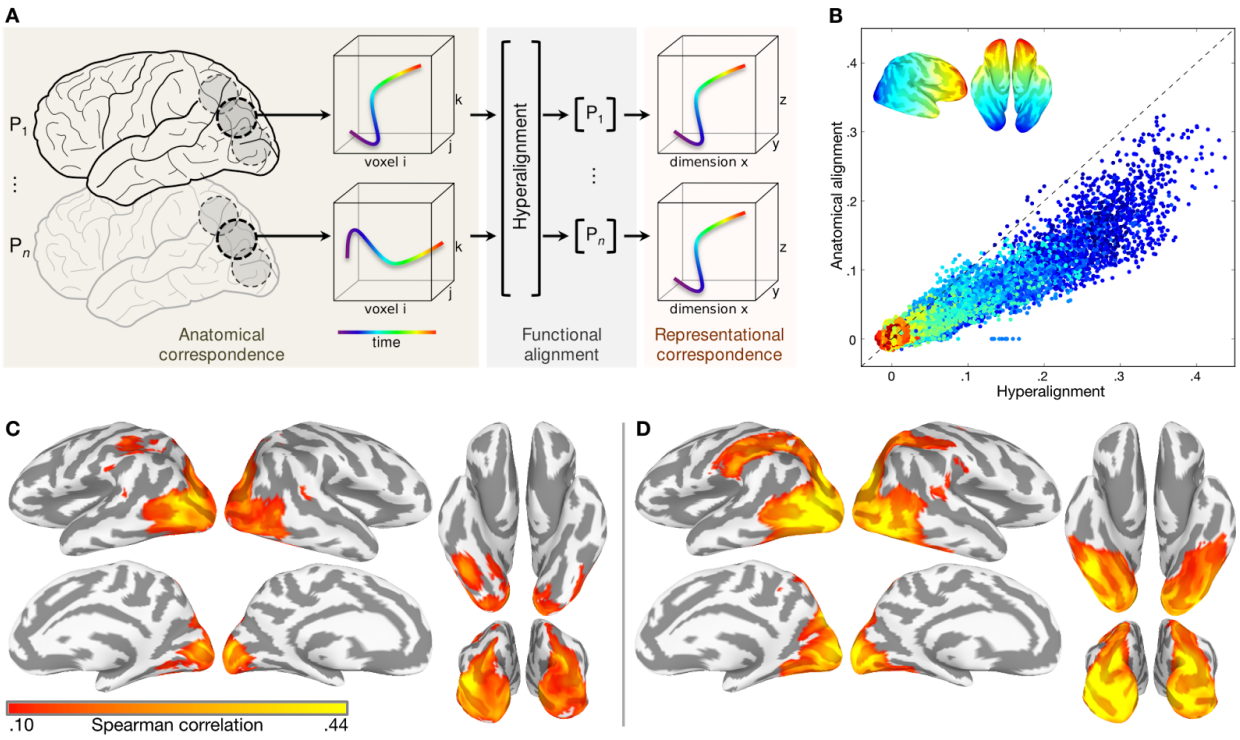

Fig. S1. Whole-brain searchlight hyperalignment enhances representational correspondence across participants. (*A*) For each surface-based searchlight, the Procrustes transformation is used to rotate each participant's time series of functional response patterns to the *Life* movie stimulus into a common space that maximizes representational correspondence across participants. These patterns are depicted as a trajectory of responses in a three-voxel space over time. (*B*) Each point in the scatterplot represents the average inter-participant Spearman correlation of RDMs for both attention tasks in a single searchlight. For each surface-based searchlight, the upper triangulars of the observed neural RDMs for both attention tasks were concatenated and pairwise Spearman correlations were computed between all participants. The vertical axis indicates Spearman correlation based on surface-based spherical alignment; the horizontal axis indicates Spearman correlation after surface-based searchlight whole-brain hyperalignment. Deviance from the identity line indicates a strong effect of alignment method on inter-participant similarity of RDMs. Searchlights are colored according their location on the posterior-anterior axis of the inflated cortical surface. (*C*) Inter-participant Spearman correlation of searchlight RDMs for both attention tasks using anatomical alignment thresholded at .10. (*D*) Average inter-participant Spearman correlation of searchlight RDMs after hyperalignment at the same threshold. Prior to hyperalignment, the maximum mean Spearman correlation was .32 in

a searchlight superior to the left lateral occipital sulcus. Following hyperalignment, the maximum mean Spearman correlation was .44 in a searchlight in the left lateral occipital sulcus.

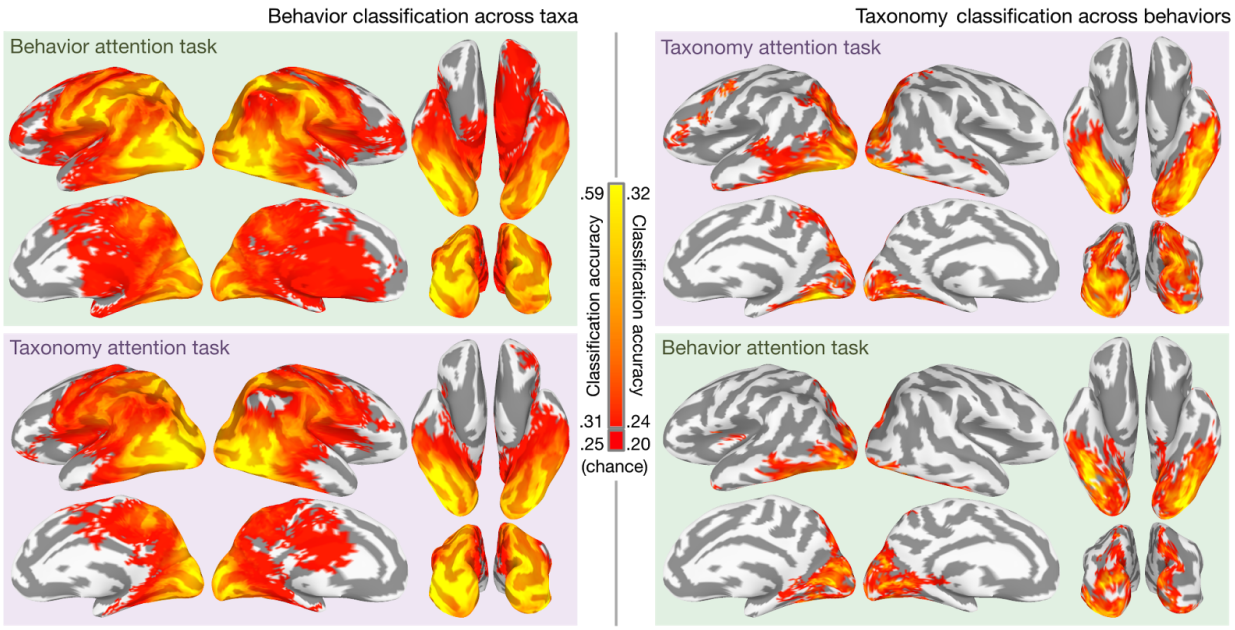

Fig. S2. Effect of attention on searchlight classification of behavior and taxonomy. Cross-validation was implemented in the following leave-one-category-out fashion: classifiers discriminating the four behaviors (left) were trained on four of the five taxa, and tested on the left-out taxon; classifiers discriminating the five animal taxa (right) were trained on three of the four behaviors and tested on the left-out behavior. This procedure ensured that any information about animal behavior generalizes across animal taxa, and vice versa. Furthermore, classifiers in this cross-validation scheme are always tested on exemplar clips not in the training set, ensuring that classification accuracy is not based on low-level visual properties idiosyncratic to particular stimuli. Prior to classification, the GLM was computed separately for each run, yielding 20 beta parameters per run. The maps are qualitatively similar to the representational similarity regression maps reported in Fig. 2, with an average correlation of .83 across conditions prior to thresholding. Chance accuracy for four-class behavior classification is .25 and chance accuracy for five-class taxonomy classification is .20. Accuracies less than 0.31 for behavior classification and less than .24 for taxonomy classification are plotted as red. Maps are thresholded at $p < .05$ using TFCE, based on a null distribution of searchlight maps generated by permuting the labels of interest within each run and within each category of the crossed factor (*SI Text*).

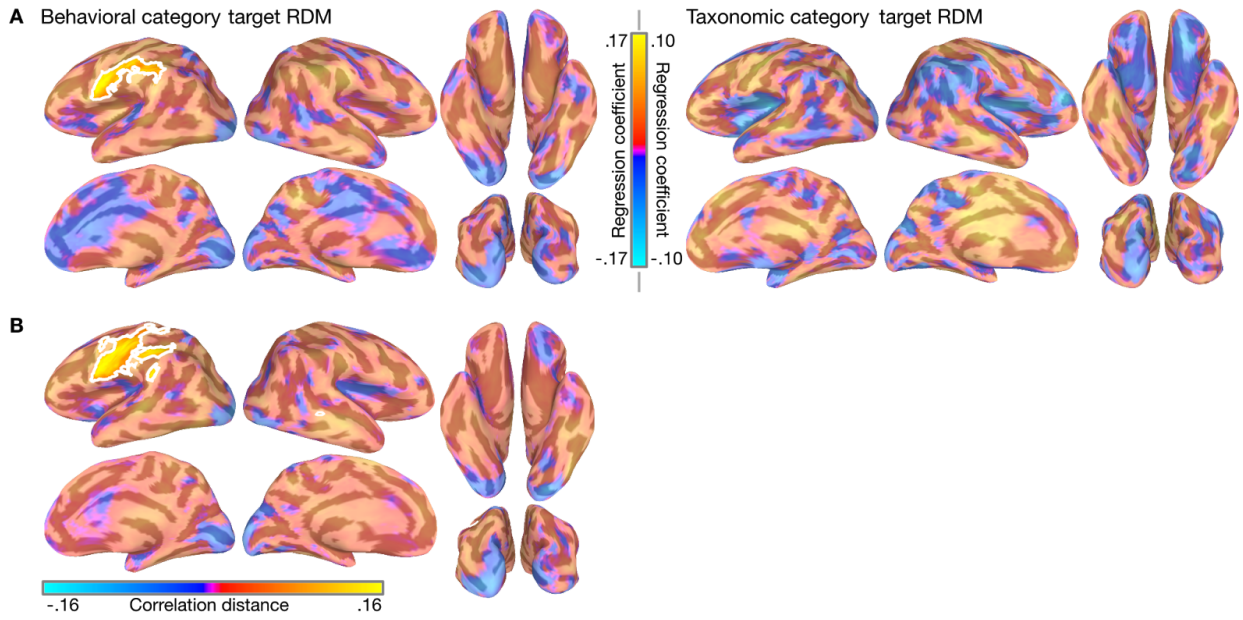

Fig. S3. Attentional differences in searchlight representational geometry. (*A*) Attention-related differences in standardized rank regression coefficients were computed for both the behavioral category and taxonomic category target RDMs. Warm colors represent attentional enhancement for the corresponding semantic information. The range of values on the color bar reflects the mean difference in the regression coefficient. (*B*) Cells of the searchlight RDMs capturing within-category distances for both animal behavior and taxonomy were isolated (see Figure 4) and tested for attentional enhancement of correlation distance. The absolute values of the within-behavior and within-taxon distances were averaged for each searchlight to compute an index of overall attentional change in within-category correlation distances. Clusters surviving TFCE-based correction for multiple comparisons at p = .05 (two-tailed test) are displayed at full opacity and outlined with a white contour, while searchlights not surviving TFCE are displayed as partially transparent. TFCE maps were estimated using a Monte Carlo simulation randomly flipping the attention task label. Note that the trend towards an effect of attention to taxonomy in VT cortex on correlation with the taxonomic RDM was not significant in this searchlight analysis but was strongly significant in the ROI analysis that used larger regions. Searchlights in this case included only 100 voxels and cannot capture the more distributed effects observed in the ROI analysis. Furthermore, searchlight analyses are subjected to strict multiple comparisons correction because of the large number of searchlights.

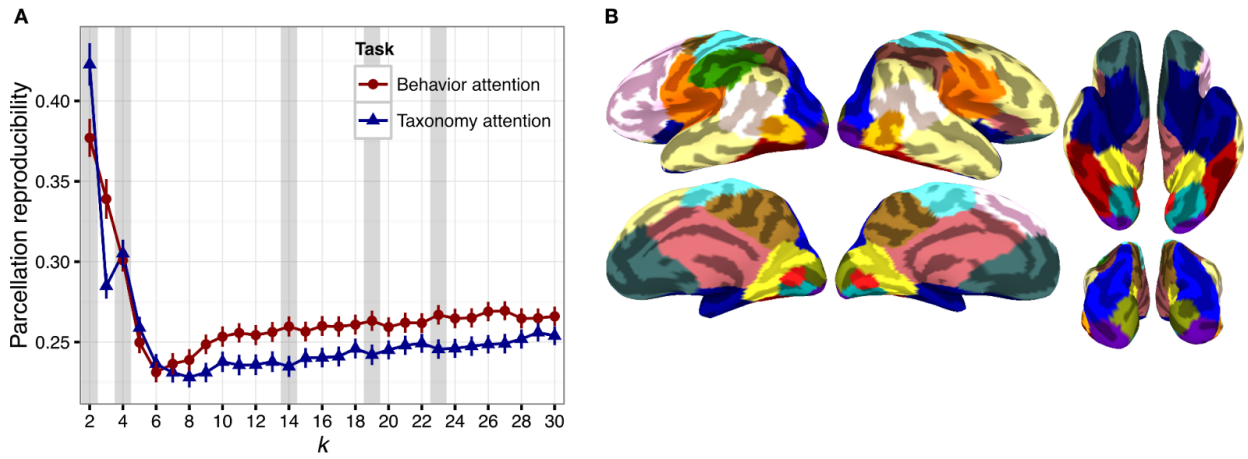

Fig. S4. Functional parcellation of the cerebral cortex based on representational geometry. (*A*) Cluster reproducibility was evaluated using split-half resampling across participants (100 partitions per *k*) separately for each attention task. The mean AMI across the 100 partitions is plotted across the values of *k*, with error bars indicating the standard error of the mean across partitions. Vertical gray bars indicate several local maxima spanning the range of *k* tested. (*B*) Full parcellation at *k* = 19. Ten parcels from this solution circumscribing the dorsal and ventral visual pathways were further interrogated in the ROI analysis.

Description	Action type	Animal taxon
Chimpanzee eating a fruit	Eating	Primate
Howler monkey eating leaves	Eating	Primate
Llama eating cactus fruits	Eating	Ungulate
Reindeer grazing on grass	Eating	Ungulate
Lammergeier eating carrion	Eating	Bird
Hummingbird drinking from flower	Eating	Bird
Chameleon eating grasshopper	Eating	Reptile
Komodo dragon eating carcass	Eating	Reptile
Caterpillar eating its own eggshell	Eating	Insect
Ladybug eating mites	Eating	Insect
Baboons fighting on rocks	Fighting	Primate
Geladas fighting amongst herd	Fighting	Primate
Bison butting heads on prairie	Fighting	Ungulate
Ibex locking horns on mountainside	Fighting	Ungulate
Seabirds fighting on rocks	Fighting	Bird
Vultures fighting in the snow	Fighting	Bird
Chameleon biting another chameleon	Fighting	Reptile
Komodo dragons fighting	Fighting	Reptile
Ant and ladybug fighting	Fighting	Insect
Stag beetles locking mandibles	Fighting	Insect
Baboon running toward water	Running	Primate
Monkey running away through tall grass	Running	Primate
Juvenile ibex running down mountainside	Running	Ungulate
Topi running through herd	Running	Ungulate
Penguin running across meadow	Running	Bird
Seagull running through cloud of insects	Running	Bird
Komodo dragon walking on rocks	Running	Reptile
Lizard running across sand	Running	Reptile
Ants traveling across sand	Running	Insect
Beatle running across dirt	Running	Insect
Macaque swimming underwater	Swimming	Primate
Snow monkey swimming in hot spring	Swimming	Primate
Deer swimming across lake	Swimming	Ungulate
Reindeer herd swimming across strait	Swimming	Ungulate
Duck swimming across stream	Swimming	Bird
Penguin swimming underwater	Swimming	Bird
Marine iguana swimming in clear water	Swimming	Reptile

Sea turtle swimming near seafloor	Swimming	Reptile
Dobsonfly larva swimming toward streambed	Swimming	Insect
Water beetle swimming underwater	Swimming	Insect

Table S1. Descriptions of video clip stimuli and condition assignments. Each of the 40 video clip exemplars is briefly described. The condition assignments are indicated for each clip. There were two exemplar clips for each condition.

Supporting References

68. Peirce JW (2007) PsychoPy—Psychophysics software in Python. *J Neurosci Methods* 162(1-2):8–13.

69. Henriksson L, Khaligh-Razavi S-M, Kay K, Kriegeskorte N (2015) Visual representations are dominated by intrinsic fluctuations correlated between areas. *NeuroImage* 114:275–286.

70. Saltelli A, Tarantola S, Campolongo F, Ratto M (2004) *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models* (John Wiley & Sons).

71. 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.

72. Pedregosa F, et al. (2011) Scikit-learn: Machine learning in Python. *J Mach Learn Res* 12:2825–2830.

73. Lange T, Roth V, Braun ML, Buhmann JM (2004) Stability-based validation of clustering solutions. *Neural Comput* 16(6):1299–1323.

74. Thirion B, Varoquaux G, Dohmatob E, Poline J-B (2014) Which fMRI clustering gives good brain parcellations? *Front Neurosci* 8:167.

75. Bates D, Maechler M, Bolker B, Walker S (2015) lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1-7). Available at: http://CRAN.R-project.org/package=lme4 [Accessed 2015].

76. Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59(4):390–412.

77. van den Berg R, Awh E, Ma WJ (2014) Factorial comparison of working memory models. *Psychol Rev* 121(1):124–149.