Decoding the neural representation of social and nonsocial human actions

Samuel A. Nastase[1,2], Rebecca E. Philip[1], Vassiki Chauhan[1], Feilong Ma[1], Morgan K. Taylor[1], Sunny Tang[1], Matteo Visconti di Oleggio Castello[1], Yaroslav O. Halchenko[1], M. Ida Gobbini[1,3], James V. Haxby[1]

[1]Dartmouth College, Hanover, NH; [2]Princeton University, Princeton, NJ; [3]University of Bologna, Bologna, Italy

Introduction

How do we understand the actions of others? The observer must extract behaviorally relevant information—such as an agent's goals and their social implications—from complex spatiotemporal patterns of visual input[1]. Action understanding relies on multiple hierarchically organized stages of processing and re-representation to disentangle behaviorally-relevant features. Neural representational spaces supporting action understanding are organized such that actions that are similar along perceptual or semantic dimensions are located nearer to each other.

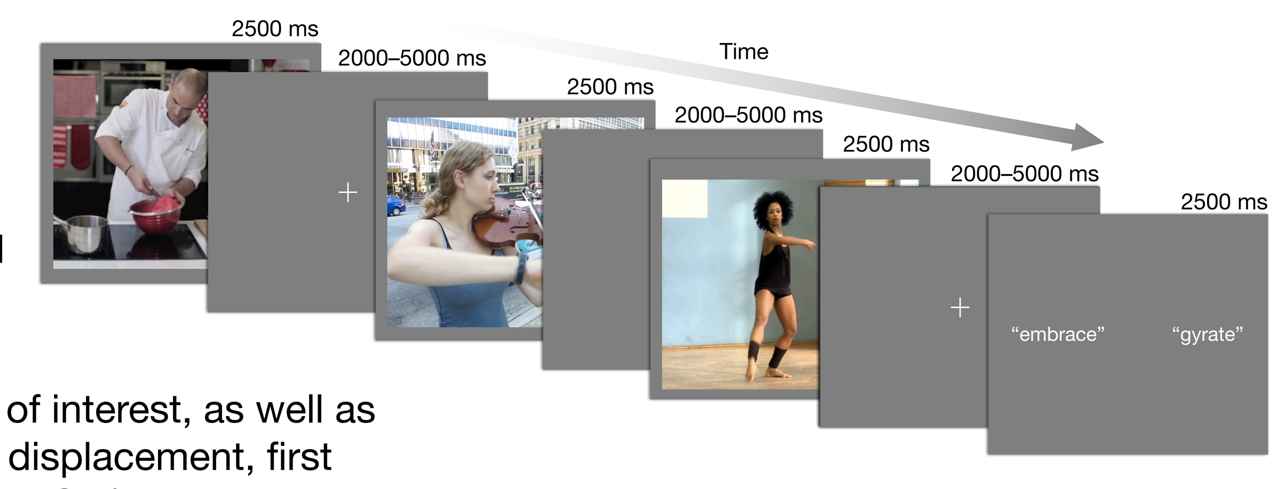
Question: What types of neural representations support action understanding, and at what stages of the processing pathway do they emerge?

Design and preprocessing

12 participants viewed 90 unique 2.5 s action clips in a condition-rich rapid event-related design[2] in two 1 hr sessions. Participants performed a semantic task in which they were intermittently asked which of two verbs best described the action in the preceding clip.

Image acquisition:
TR = 1 s, TE = 32 ms, 2.5 mm^3 voxels.

Data were preprocessed using fMRIPrep[3].



GLM with 90 regressors of interest, as well as head motion, framewise displacement, first five PCs from CSF (aCompCor).

Surface-based searchlight hyperalignment[4,5] was used to transform all data into a common response space based on a 1 hr movie session *(Raiders of the Lost Ark).*

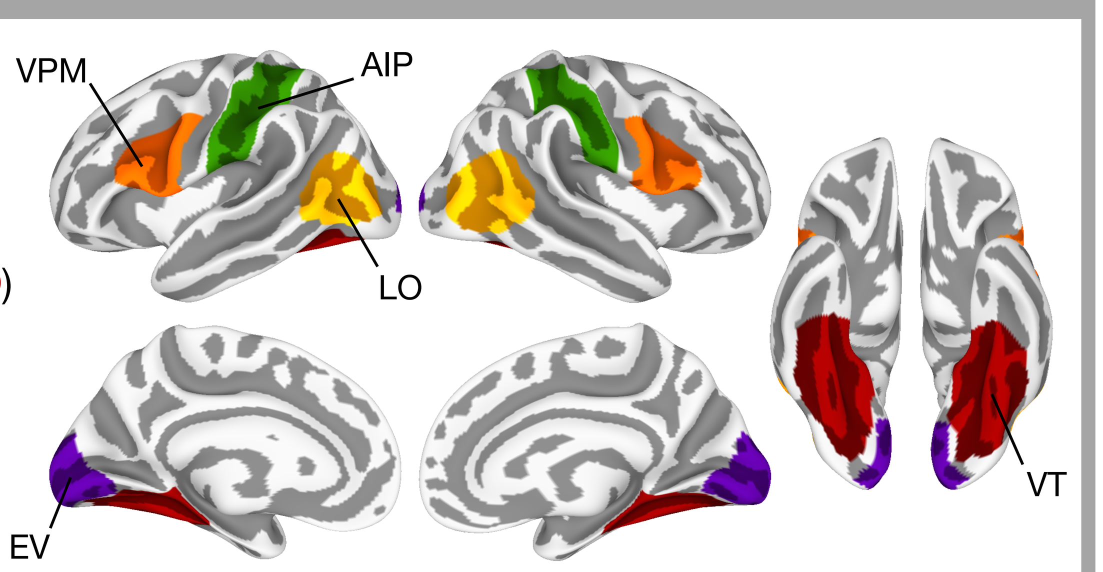
exclusively social — exclusively nonsocial — paired social and nonsocial

Action category	Sociality	Action category	Sociality
Conversation	Social	Cooking	Nonsocial
Intimacy	Social	Gardening	Nonsocial
Teaching	Social	Arts and crafts	Nonsocial
Manufacturing	Social	Musical performance	Nonsocial
Eating	Social	Eating	Nonsocial
Dancing	Social	Dancing	Nonsocial
Exercise	Social	Exercise	Nonsocial
Cosmetics and grooming	Social	Cosmetics and grooming	Nonsocial
Manual tool use	Social	Manual tool use	Nonsocial

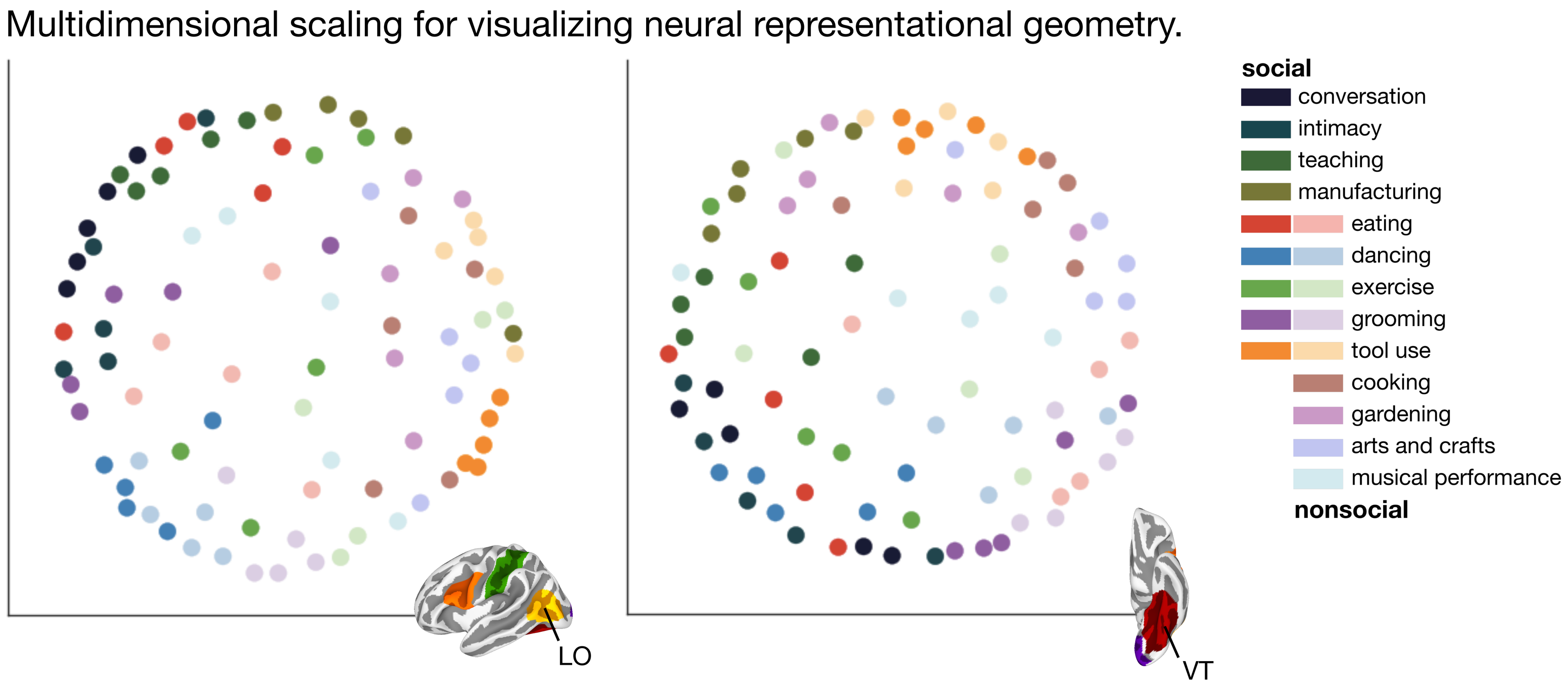
Region of interest analysis

Five anatomically-defined ROIs including three hubs of action observation network[6]:

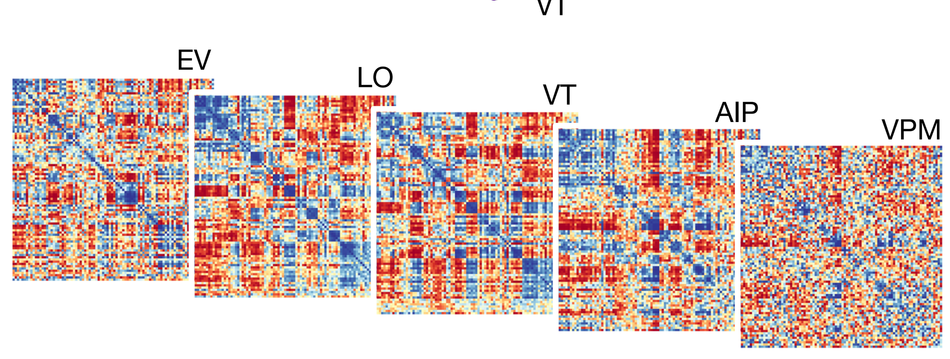
- early visual cortex (**EV**)
- lateral occipitotemporal cortex (**LO**)
- ventral temporal cortex (**VT**)
- anterior intraparietal cortex (**AIP**)
- ventral premotor cortex (**VPM**)

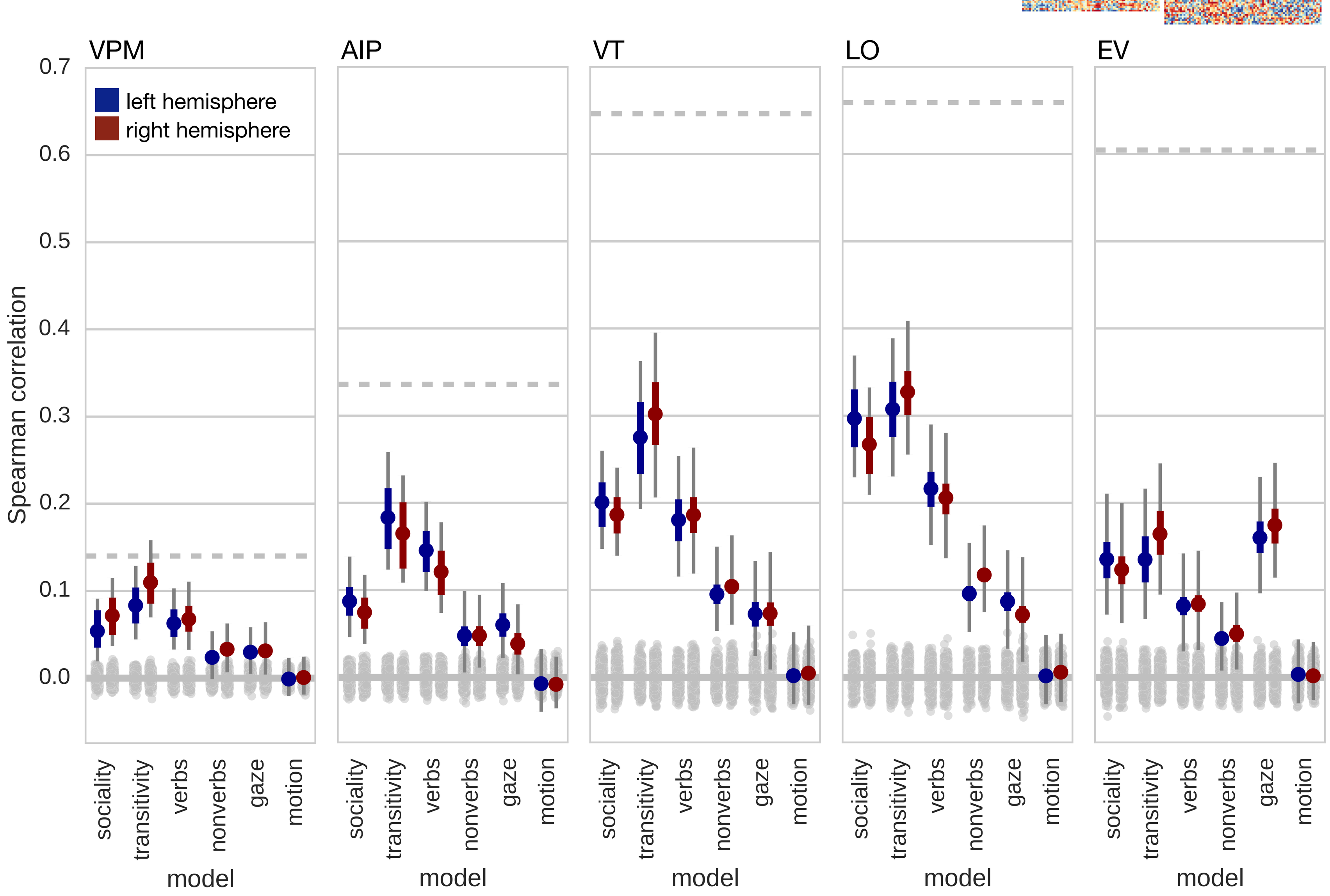


Multidimensional scaling for visualizing neural representational geometry.



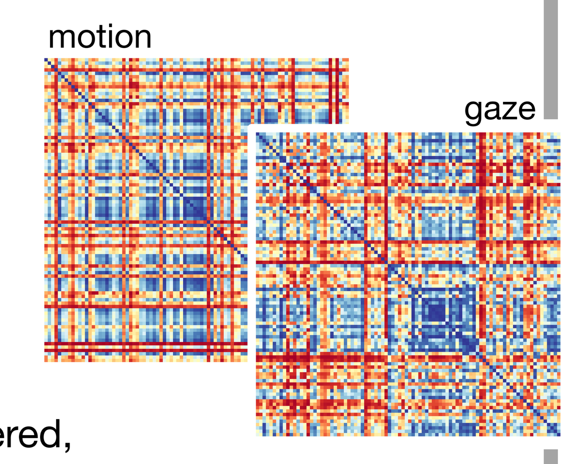
Evaluate each model in each ROI relative to noise ceiling, with permutation-based null distribution and bootstrap confidence intervals[7].





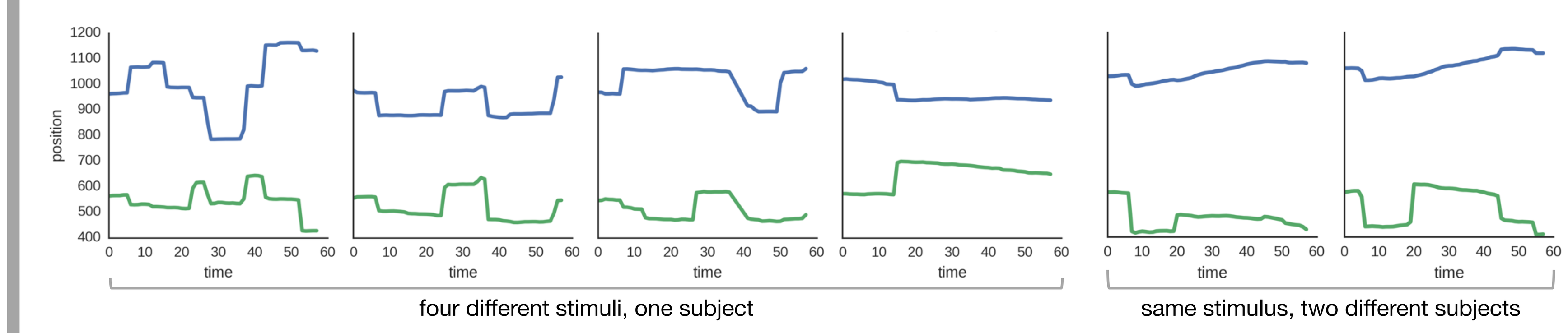
Representational models

Six representational models were constructed capturing visual features, semantic content, and behavioral judgments.

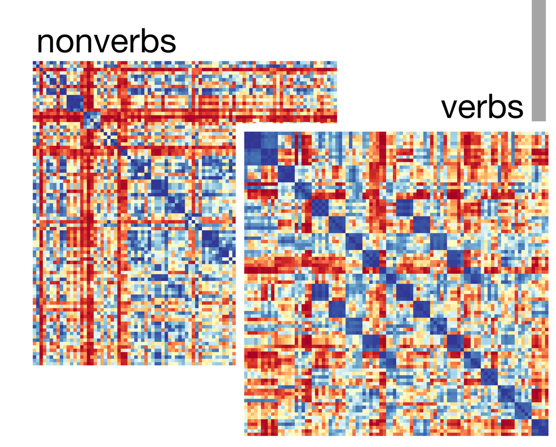


Motion energy was computed using spatiotemporal Gabor filters at different positions, orientations, spatial and temporal frequencies in quadrature (6,555 channels per frame)[8]. Correlations between vectorized channel weights per stimulus were used to construct a motion RDM.

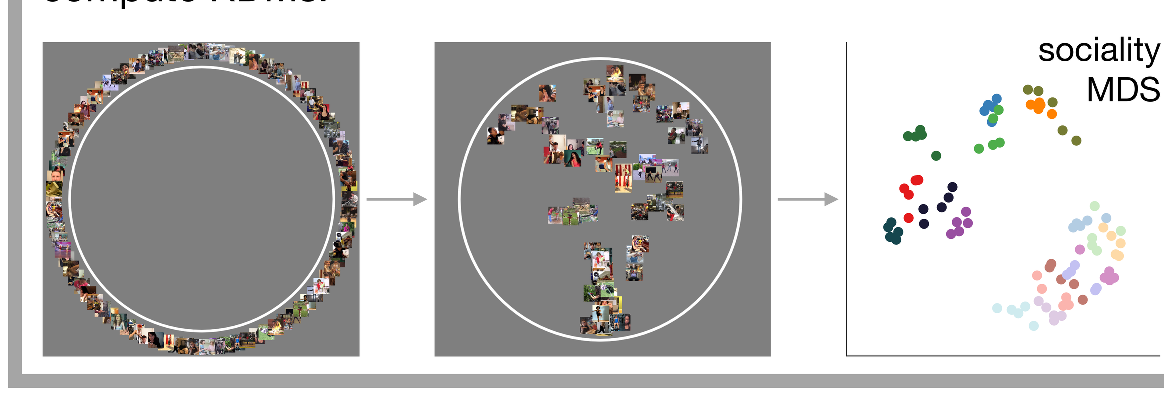
Gaze trajectories were measured in a separate cohort, median-filtered, interpolated across blinks, and downsampled. Euclidean distances between trajectories were used to compute a gaze RDM.

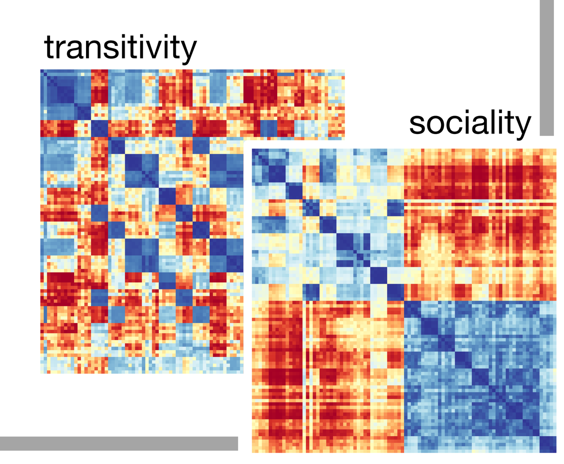


Two annotators manually assigned **nonverb** and **verb** labels to the 90 clip stimuli. Pre-trained 300-dimensional word embeddings from word2vec were assigned to each stimulus. Semantic embeddings were averaged per stimulus and cosine distances between embeddings was used to construct RDMs.



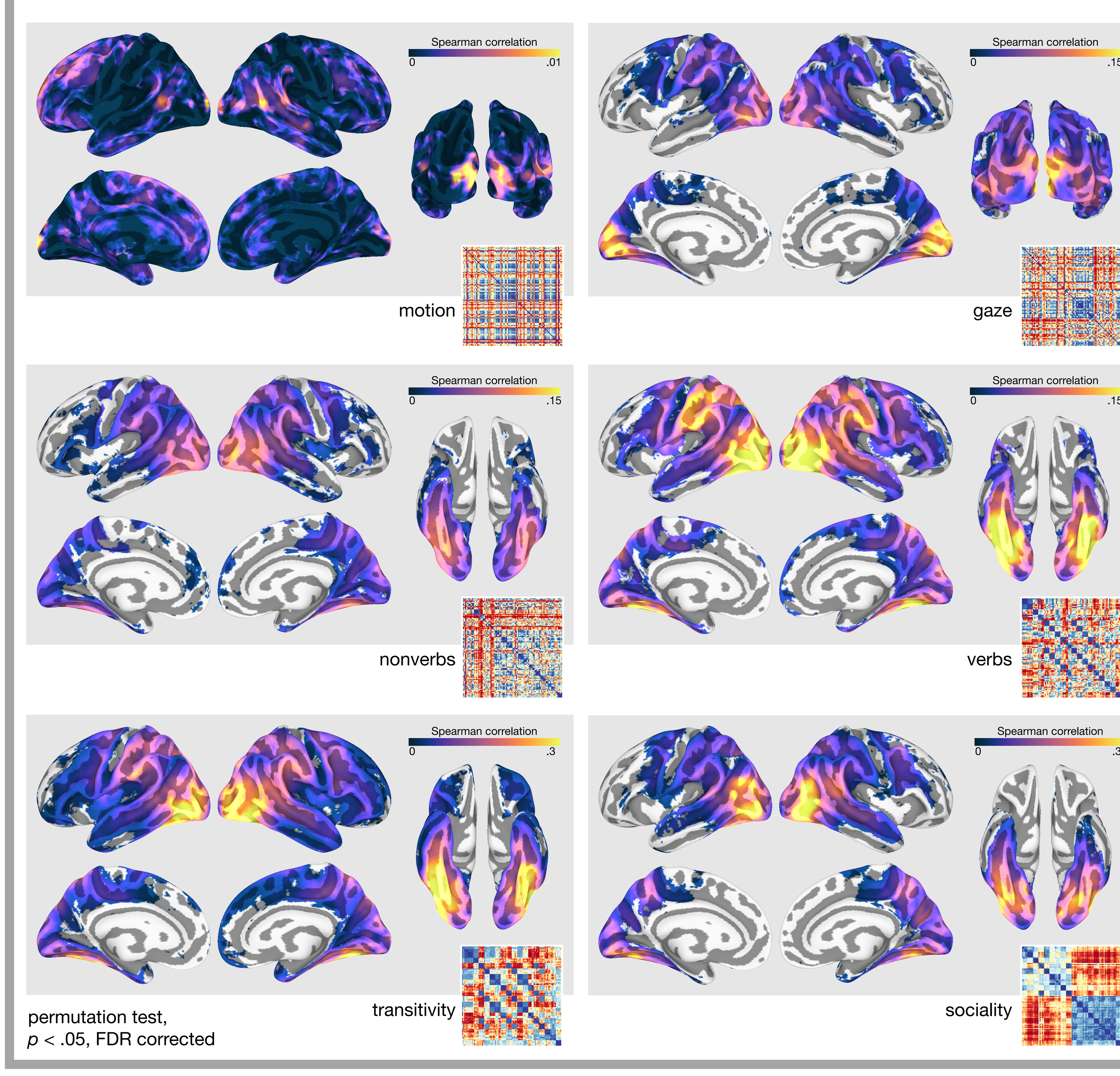
Each participant performed two multiple item arrangement tasks where they organized the stimuli according to **transitivity** or **sociality**. Euclidean distances across subsets were used to compute RDMs.





Searchlight analysis

Each representational model was tested using Spearman correlation in 10 mm radius surface-based searchlights.



Conclusions

The geometry of observed action representation can be disentangled using representational models of visual, semantic, and social content.

Transitivity, sociality, and verb semantics emerged as key dimensions of neural representation in downstream areas, such as LO and VT. These models captured a surprisingly large portion of variance in VT.

Static image stimuli and non-naturalistic tasks provide a limited view onto internal representational spaces. Dynamic, naturalistic stimuli provide complementary insights.

Using a rich variety of naturalistic stimuli, we can replicate several findings from the literature in a single data set.

The best-performing models (e.g., transitivity in LO) still only accounted for ~14% of variance in neural representation and only reached halfway to the noise ceiling—we can do better!

References:
1. Marr, D., & Vaina, L. (1982). Representation and recognition of the movements of shapes. *Proceedings of the Royal Society of London B: Biological Sciences, 214*(1197), 501–524.
2. Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2*, 4.
3. Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Aliaga, A. E., ... Gorgolewski, K. J. (2018). FMRIPrep: a robust preprocessing pipeline for functional MRI. *bioRxiv*, 306951.
4. Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., & Haxby, J. V. (2016). A model of representational spaces in human cortex. *Cerebral cortex, 26*(6), 2919–2934.
5. Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics, 7*(1), 37–53.
6. Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2013). Crossmodal and action-specific: neuroimaging the human mirror neuron system. *Trends in Cognitive Sciences, 17*(7), 311–318.
7. Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLOS Computational Biology, 10*(4), e1003553.
8. Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology, 21*(19), 1641–1646.
9. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems* (pp. 3111–3119).
10. Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers, 26*(4), 381–386.