Representational geometries supporting social action understanding in naturalistic vision

Jane Han (jane.han.gr@dartmouth.edu)
Department of Psychological and Brain Sciences, Dartmouth College
Hanover, NH 03755 USA

Maria Ida Gobbini (mariaida.gobbini@unibo.it)
Department of Medical and Surgical Sciences, University of Bologna
Bologna, BO 40138 Italy

James V. Haxby (james.v.haxby@dartmouth.edu)
Department of Psychological and Brain Sciences, Dartmouth College
Hanover, NH 03755 USA

Samuel A. Nastase (snastase@princeton.edu)
Princeton Neuroscience Institute, Princeton University
Princeton, NJ 08540 USA

Abstract

To understand the actions of others, humans effortlessly extract rich, behaviorally relevant information about actions and intentions from dynamic visual input. How are our internal neural representations of observed actions structured? Here we measured brain activity using fMRI while participants viewed video clips of a wide range of naturalistic action clips. We measured the perceived similarity of the clips based on different criteria by asking participants to arrange them on a screen based on the purpose of the actions (transitive goals and type of social interaction) and visual content (person, object, scene). In a searchlight analysis we tested the correlation of neural representational geometries derived from the arrangement tasks. Results revealed an extensive system for action representation that encompassed most visual cortex in occipital, temporal, parietal, and prefrontal cortices. Representational geometries based on sociality and transitivity judgments better predicted neural representational geometry than did geometries based on similarities of people, scenes, and objects. Transitivity better predicted neural representational geometry than sociality with notable exceptions in the precuneus, posterior superior temporal sulcus, and lateral occipitotemporal cortex bilaterally. Our findings indicate that visual representation is dominated by agentic action and is best accounted for by behavioral measures of the purpose of actions.

Keywords: action perception; fMRI; representational geometry; hyperalignment

Introduction

Humans have a remarkable ability to recognize actions and intentions from complex visual scenes. Previous studies investigated this ability with static, controlled images (Lingnau & Downing, 2015) or controlled videos (e.g., Tucciarella, et al., 2019). Static images and controlled video stimuli, however, may not capture the richness of action representation (Haxby et al., 2020). Recent work has begun investigating the structure of observed action recognition during naturalistic vision (Tarhan & Konkle, 2020; McMahon, Bonner, & Isik, 2023).

To investigate how the cortical processing hierarchy supports action understanding, we used functional magnetic resonance imaging (fMRI) to measure brain activity while participants watched a variety of video clips displaying social and non-social actions. We used several behavioral arrangement tasks to interpret the structure of neural representational spaces supporting action understanding.

Methods

Twenty-three adults participated in the fMRI experiment. Each participant completed two 1-hour fMRI sessions viewing video clips of human actions, a 1-hour movie-viewing session outside the scanner (first half of the movie) immediately followed by a 1.5-hour movie-viewing session in the scanner (second half of the movie). A subset of 17 subjects completed 1-hour behavioral sessions.

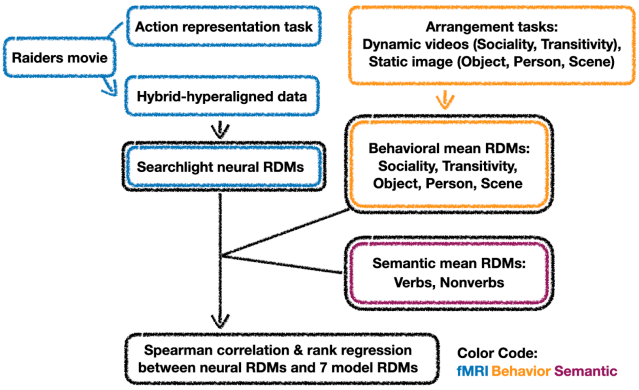

Figure 1. Experimental design. Color fMRI data (blue), behavioral arrangements (orange), and semantic vectors (purple) used sessions to compute the Spearman correlation and rank regression between searchlight neural RDMs and seven model RDMs (black).

For the action observation fMRI task, participants viewed 90 video clips depicting 18 categories of naturalistic social and nonsocial human actions. Each clip was presented for 2.5 seconds with jittered 2-5 seconds inter-trial intervals. Intermittently, participants were asked to identify which action word most accurately described the action depicted in the previous video. Each block contained one video from each category. Each clip was presented 8 times in 40 blocks of 18 videos. Participants watched Raiders of the Lost Ark in a separate movie-viewing session in order to calculate transformation matrices for hyperalignment. All action representation task data were hyperaligned using hybrid-hyperalignment (Busch et al. 2021) to increase between-subject correlations of representational geometry (Kriegeskorte, Mur, & Bandettini, 2008; Fig. 1, blue). We used a surface-based searchlight to construct 90 × 90 neural representational dissimilarity matrices (RDMs) across the cortex.

We collected behavioral judgments using a multiple arrangements task (Goldstone, 1994). Participants completed two types of arrangement tasks: (1) arranging video stimuli according to the similarity of action purpose—transitivity (object- and goal-relatedness) or sociality; and (2) arranging static images of representative video frames according to the similarity of visual content—people, objects, or scenes (Fig. 1, orange). Arrangement distances were averaged across partici-

pants to create group-mean behavioral RDMs. In addition to the five behavioral RDMs based on arrangement tasks, we constructed two more RDMs (Fig. 1, purple) based on semantic (word2vec) embeddings derived from annotations of the videos using verbs and non-verbs. Finally, we used Spearman correlation and rank regression to relate these behavioral and semantic RDMs to the searchlight neural RDMs (Fig. 1, black).

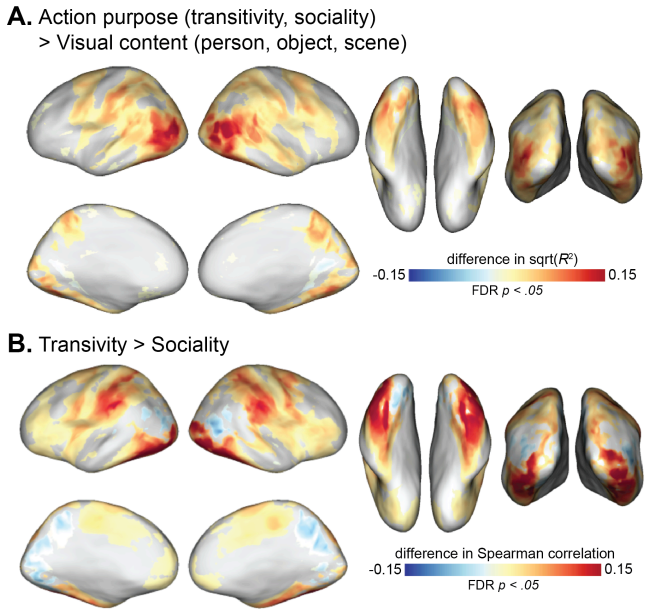

Figure 2. Model comparisons. (A) Action purpose (transitivity, sociality) RDMs better predicted neural RDMs than visual content RDMs (person, object, scene). (B) Contrast of transitivity versus sociality RDMs. Significance was assessed using a *t*-test across subjects, controlling FDR at $p < .05$.

Results

We evaluated the variance in neural RDMs explained by two groups of behavioral RDMs: (1) transitivity and sociality RDMs capturing action purpose, and (2) person, object, and scene RDMs capturing visual content (Fig. 2A). The difference in R^2 revealed that representational geometries based on dynamic, video stimuli and reflecting the purpose of actions—the transitivity and sociality RDMs—significantly outperform RDMs based on static images reflecting behavioral judgments of visual content (person, object, scene) throughout the visual system, including ventral temporal, lateral occipitotemporal, posterior superior temporal, intraparietal, premotor, and posterior medial regions. Together these two behavioral action-purpose RDMs accounted for a maximum of 22% of variance in searchlight neural RDMs. By contrast, the behavioral RDMs based on static visual content—people, objects, and scenes—accounted for about half as much variance (maximum $R^2 = .12$), while the semantic RDMs accounted for about a third as much variance (maximum $R^2 = .07$).

To estimate the amount of meaningful variance in neural data (the noise ceiling) we calculated intersubject correlations (ISCs) of searchlight neural RDMs (the correlation between each participant's neural RDM and the mean neural RDM of the other participants). The mean ISCs were strongest in the same cortical areas with a maximum of $r = .75$ ($R^2 = .56$), indicating that the action-purpose RDMs accounted for roughly 40% of the reliable variance in neural representational geometry.

Both the transitivity and sociality RDMs were strongly correlated with the neural RDMs throughout the action representation system. The transitivity RDM correlated more strongly with neural RDMs throughout most of the action observation network, with notable exceptions where sociality significantly outperformed transitivity in the bilateral precuneus, lateral occipitotemporal cortex, and posterior superior temporal sulcus (Fig. 2B). Of the behavioral arrangement RDMs based on visual content, the object RDM yielded the strongest performance, possibly due to collinearity between the objects and the transitive nature of actions. The semantic RDMs produced much weaker correlations, but the verb RDM, which better captured actions, generally outperformed the non-verb RDM (nouns and adjectives generally describing the content of the videos).

Conclusion

Overall, the representational geometry of transitivity, corresponding to behavioral judgments of action goals, most strongly correlated with the neural representational patterns in the ventral temporal, inferior lateral occipital, parietal, and premotor cortices. Conversely, the representational geometry associated with sociality was a stronger predictor of neural patterns in the precuneus, superior lateral occipitotemporal cortex, and the posterior superior temporal sulcus. Our results suggest that the neural representational spaces supporting action understanding are predominantly organized according to behaviorally relevant interpretations of action goals and social interaction.

Acknowledgments

This work was supported by NSF grants 1835200 (MIG) and 1607845 (JVH). and by NIMH grant 1R01MH127199 (JVH and MIG).

References

Busch, E. L., Slipski, L., Feilong, M., Guntupalli, J. S., di Oleggio Castello, M. V., Huckins, J. F., ... & Haxby, J. V. (2021). Hybrid hyperalignment: A single high-dimensional model of shared information embedded in cortical patterns of response and functional connectivity. *NeuroImage. 233*, 117975.

Goldstone, R. (1994). An efficient method for obtaining similarity data, *Behavior Research Methods, Instruments, & Computers, 26*, 381-386.

Haxby, J.V., Gobbini, M. I., & Nastase, S. A. (2020) Naturalistic stimuli reveal a dominant role for agentic action in visual representation. *NeuroImage*, *216*, 116561.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience, 2*, 249.

Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in cognitive sciences*, *19*(5), 268-277.

McMahon, E., Bonner, M., & Isik, L. (2023). Hierarchical organization of social action features along the lateral visual pathway. *Current Biology*, *33*(23), 5035-5047.e8.

Tarhan, L., & Konkle, T. (2020). Sociality and interaction envelope organize visual action representations. *Nature Communications*, *11*(1), 3002.

Tucciarelli, R., Wurm, M., Baccolo, E., & Lingnau, A. (2019). The representational space of observed actions. *eLife*, *8*, e47686.