# Hierarchical representations of naturalistic social interactions in the lateral visual pathway

**Emalie McMahon (emaliemcmahon@jhu.edu)**
Johns Hopkins University, Department of Cognitive Science
237 Krieger Hall, 3400 N. Charles St, Baltimore, MD 21218

**Michael F. Bonner (mfbonner@jhu.edu)**
Johns Hopkins University, Department of Cognitive Science
237 Krieger Hall, 3400 N. Charles St, Baltimore, MD 21218

**Leyla Isik (lisik@jhu.edu)**
Johns Hopkins University, Department of Cognitive Science
237 Krieger Hall, 3400 N. Charles St, Baltimore, MD 21218

## Abstract

**In our daily lives, we quickly and effortlessly perceive features of others' interactions. Extracting these social details is crucial for deciding how to act in the social world, but little is understood about how this is solved in the mind and brain. Recent work has identified a visually-selective region for the presence of a social interaction in posterior superior temporal sulcus (pSTS), but whether and how diverse features of a social interaction (ranging from visual to high-level) are presented in the pSTS or elsewhere in the brain is unknown. To answer this question, we showed participants 250 3-second video clips of naturalistic two-person interactions while undergoing fMRI. We used an encoding model approach to model where visual and social features of the videos are represented in the brain. We replicate known preference for scene and object features in visual cortex. We also find preference for facing bodies in EBA and joint action in pSTS, extending prior findings with controlled stimuli to natural settings. Finally, we identify regions along the extent of the STS that show a preference for third-party communication. Together, these results suggest a hierarchy of visual to social feature processing along the lateral surface of the brain.**

**Keywords:** social perception; social interactions; fMRI; naturalistic stimuli; fMRI encoding models

## Introduction

In everyday life, we effortlessly perceive rich details about the relations between individuals, their social goals, and interactions. The ability to detect a social interaction and perceive features of that interaction develops early in infancy (Hamlin, Wynn, & Bloom, 2007) and is shared with non-human primates (Cheney, Seyfarth, & Smuts, 1986). Recent work has also shown that social interaction perception is preferentially processed in both the macaque and human brain (Sliwa & Freiwald, 2017; Isik, Koldewyn, Beeler, & Kanwisher, 2017; Walbrin, Downing, & Koldewyn, 2018; Walbrin & Koldewyn, 2019; Lee Masson & Isik, 2021), but we still don't know how diverse aspects of a social interaction are represented in the brain, particularly under more natural viewing conditions. In the current study, we investigate how the brain processes diverse aspects of a social interaction in natural stimuli.

To address where the features of social interaction are represented in the brain, we curated a rich set of videos of naturalistic two-person interactions (Figure 1). The use of naturalistic stimuli allowed us to investigate the contribution of many co-varying features simultaneously and has been argued to be particularly important to investigating dynamic, social perception (Nastase et al., 2017; Redcay & Moraczewski, 2020; Haxby, Gobbini, & Nastase, 2020). The videos in our set sampled variance along many relevant dimensions of daily life. These dimensions include visual features such as whether the action occurred indoors or outdoors, visual features associated with a social interaction (here referred to as "social primitives") like the distance of the people and whether they were facing one another, and high-level social features like the the extent to which people are communicating. We presented this datasets to subjects in a condition-rich fMRI experiment and used an encoding model approach to investigate social interaction feature coding across the brain.

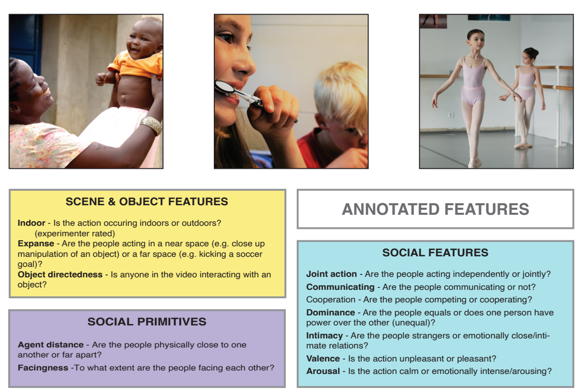


Figure 1: Example still frames from videos in the dataset. The features that were annotated and the questions that were presented are shown.

## Methods

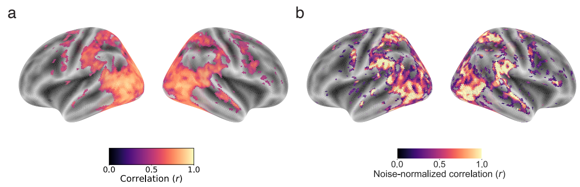


Figure 2: a) The split-half reliability of the data in the test set. b) The noise-normalized prediction accuracy of the encoding model in the reliable voxels.

### Annotations

We collected annotations for a range of scene and object (yellow), social primitive (purple) and high-level social (teal) features (Figure 1). Participants recruited on Prolific rated the videos on each of the dimensions. The question text that was presented is shown in Figure 1. All questions were rated on a five point Likert scale by at least ten independent raters each.

### fMRI experiment

We ran a condition-rich fMRI experiment where a few participants (n=4) watched the 250 videos many times over four two-hour fMRI sessions. We split these videos into a 200 video training set and 50 video test set. Participants viewed the training videos 10 times each and the testing videos 20 times each to increase SNR. This led to fMRI data that was highly reliable within each individual participant (Figure 2A).

### fMRI encoding models

Using the annotated features and low-level visual features (principle components from the second convolutional layer of AlexNet (Karpathy et al., 2014) and a motion energy model (Adelson & Bergen, 1985)), we fit an encoding model in the training videos. We assessed performance of the model in the test set (Figure 2B).

## Results

In order to understand where the features of social interactions are represented in the brain, we visualized the model weights ($\beta$ weights) within the voxels that were significantly predicted by the full model (Figure 2). For brevity, results are only shown for a few dimensions (Figure 3).

For the features of the scene and object model, the model loadings for expanse and object directedness replicate previous work on object and scene selectively. In particular, expanse has high loadings in the scene-selective parahippocampal place area (PPA), and object directedness has high loadings in object-selective lateral occipital cortex (LOC) although this was limited to the left hemisphere.

Facingness loaded highly on EBA in the left hemisphere replicating findings from (Abassi & Papeo, 2020) for the first time in a naturalistic stimulus set.

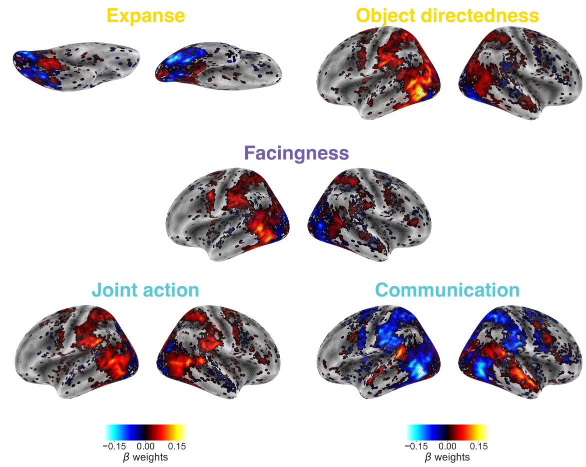


Figure 3: Visualized $\beta$ weights for selected features from the encoding model.

We found that joint action was represented in the pSTS replicating previous work on social interaction detection (Isik et al., 2017) but is also represented in many lateral regions that are motion responsive.

Communication was represented along the STS bilaterally and extending beyond the pSTS into more anterior regions.

## Discussion

Using a large-scale natural neuroimaging approach, we mapped the organization of a range of visual and social dimensions across the brain. Our work replicates known visual selectivity for scene, object, and social primitives, validating our dataset and approach. Further, by using a rich set of annotations, we were able to isolate representations for different types of social interactions. We find that social primitives such as the degree to which the people in the videos are facing one another are represented in the EBA, and the joint action of agents is represented in pSTS, as previously shown with controlled stimuli. Unexpectedly, we find that third-person communicative actions are represented in regions of the STS anterior to what has been previously reported. Together, these results suggest a hierarchy of visual to social feature processing along the lateral surface of the brain.

These findings of social interaction representations, ranging from social primitives to higher-level features like communication, may be key aspects of the recently proposed third visual pathway for social perception along the lateral surface of the brain (Pitcher & Ungerleider, 2021; Wurm & Caramazza, 2022). Our results add to a growing body of literature finding that social perception is an important organizing principle of human visual system.

## Acknowledgments

This work was supported by NSF GRFP (DGE-1746891) awarded to Emalie McMahon.

## References


Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *Journal of Neuroscience*, *40*(4), 852–863. doi: 10.1523/JNEUROSCI.1378-19.2019

Adelson, E. H., & Bergen, J. R. (1985, Feb). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, *2*(2), 284–299. doi: 10.1364/JOSAA.2.000284

Cheney, D., Seyfarth, R., & Smuts, B. (1986). Social relationships and social cognition in nonhuman primates. *Science*, *234*(4782), 1361–1366. doi: 10.1126/science.3538419

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*, 557–559. doi: 10.1038/nature06288

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. doi: 10.1016/j.neuroimage.2020.116561

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *114*(43), E9145–E9152. doi: 10.1073/pnas.1714471114

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *2014 ieee conference on computer vision and pattern recognition* (p. 1725-1732). doi: 10.1109/CVPR.2014.223

Lee Masson, H., & Isik, L. (2021). Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. *NeuroImage*, *245*, 118741. doi: 10.1016/j.neuroimage.2021.118741

Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Visconti di Oleggio Castello, M., ... Haxby, J. V. (2017, 06). Attention selectively reshapes the geometry of distributed semantic representation. *Cerebral Cortex*, *27*(8), 4277-4291. doi: 10.1093/cercor/bhx138

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Science*, *25*, 100 – 110. doi: 10.1016/j.tics.2020.11.006

Redcay, E., & Moraczewski, D. (2020). Social cognition in context: A naturalistic imaging approach. *NeuroImage*, *216*, 116392. doi: 10.1016/j.neuroimage.2019.116392

Sliwa, J., & Freiwald, W. A. (2017). A dedicated network for social interaction processing in the primate brain. *Science*, *356*(6339), 745–749. doi: 10.1126/science.aam6383

Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia*, *112*, 31-39. doi: 10.1016/j.neuropsychologia.2018.02.023

Walbrin, J., & Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *NeuroImage*, *198*, 296-302. doi: 10.1016/j.neuroimage.2019.05.027

Wurm, M. F., & Caramazza, A. (2022). Two 'what' pathways for action and object recognition. *Trends in Cognitive Sciences*, *26*(2), 103-116. doi: 10.1016/j.tics.2021.10.003