Neural Representations of Social Interactivity: A Perceptual and Language Model Analysis

Sajjad Torabian
Department of Cognitive Sciences
University of California, Irvine

John A. Pyles
Center for Human Neuroscience
University of Washington

Hongjing Lu
Department of Psychology
University of California, Los Angeles

Emily D. Grossman
Department of Cognitive Sciences
University of California, Irvine

Abstract

When given the opportunity, humans naturally engage in anthropomorphism, which may reflect a bias to engage in mentalistic attributions in understanding social interactions. In this experiment, we evaluate whether neural activity in social perceptual brain regions can be explained by perceptual cues of agency and interactivity, or by semantic models of written descriptions of Heider-Simmel style animations. Models were compared in representational similarity space using variance partitioning of the neural response from the STS, TPJ, and PCC. The right STS and TPJ were best explained by perceptual models of distance between the agents, an indicator of interactivity, and separately by the similarity structure of the free responses, which captured both action and interaction terms. Together, these results implicate the importance of contextual framing, either through perceptual features of interactivity or social context as implied by the nature of interactions, as defining features in neural representations of interactivity.

Keywords: social cognition; theory of mind; language models; fMRI

Introduction

Humans may spontaneously attribute social characteristics to moving objects around them, even when the objects lack explicit features that convey animacy such as eyes or body limbs. In the original Heider & Simmel (1944) report, participants described animated shapes as autonomous agents when instructed to simply "write down what happened in the picture." Since that discovery numerous studies have followed, investigating various types of attributions made when viewing these animations and the extent to which this may reveal tendencies to engage the cognitive process of theory of mind (see Schultz & Frith (2022) and Torabian & Grossman (2023) for a comprehensive review). Abell et al. (2000) showed that typical observers commonly use mentalistic terms when viewing Heider-Simmel type animations that make reference to the mental states of the actors, whereas simple interactions garner descriptions weighted with terms describing directly observable action states. Likewise, observers are also able to draw inferences as to the emotional state of the actors as conveyed by the animated shapes (Moessnang et al., 2017). Thus, the nature of language used by the observer when describing the events and mental states of the geometric actors reflects, in part, the depth of mentalistic processes engaged by the viewer.

In Heider-Simmel animations, object trajectories serve as a cue for goal-directed action while relative position cues may signal interactivity, and emotional states must be inferred through a complex interplay of the two. Thus, a number of studies have evaluated the kinds of perceptual cues that convey interactivity, and in what cases humans take a deep mentalistic stance versus using functional descriptions more consistent with a teleological stance (Gergely & Csibra, 2003). As evidence in favor of the potency of perceptual cues, studies have noted distinct differences in the movement patterns of animations that depict intentional vs. simple goal-directed movements (Roux et al., 2013). Eye-tracking studies have further shown that animations conveying mentalistic interactions require deeper processing as evident by longer fixations as compared to goal-directed animations or random movement (Klein et al., 2009). Moreover, computational work successfully predicted human performance for a variety of actions (e.g., push, tickle, argue, flirt) with models solely based on features derived from motion trajectories (i.e., rotation and relative distance; Roemmele et al. (2016)). Likewise, Shu et al. (2018) created a model that was able to discriminate interactive vs. independent actions from de-contextualized aerial videos of human movements in an open space (with the humans replaced by shapes) based on the characters' motion trajectories alone and without the need for high-level reasoning. Movement trajectories may also combine with relational visual representations (the distance between shapes that appear as agents), which recent modeling work shows could better explain human judgments on the nature of Heider-Simmel interactions (adversarial, neutral, or friendly) as compared to a model with explicit information about the social and physical world (Malik & Isik, 2023).

In contrast, at least one other study has demonstrated the benefit of cognitive contextual information in modeling human judgments. Ullman et al. (2009) showed observers animations with one agent that moved towards one of two object goals and another agent that either helped the first agent or hindered it. The authors found a model that inferred social goals to be a better fit to human judgments as compared to a perceptual distance-based model. Importantly, the goals of the agents were not directly observable by humans or the models, suggesting that attributions were made by drawing inferences at the mentalistic level. These results indicate a potential advantage of engaging theory of mind processes, against the potential cost of additional cognitive load. The extent to which observers are willing to engage that load is not entirely clear, but recent findings from Tarhan et al. (2021) showed that observers intuitively arrange videos of interaction as more similar when they convey similar goals, rather than based on visual or action-based similarities. Taken together, the above studies illustrate the challenges in evaluating how and at what level of complexity humans recognize actions and interactions.

Brain imaging studies have revealed a network of regions in the social cognitive brain network and the lateral temporal pathway (Pitcher & Ungerleider, 2021) to be engaged when making goal-directed and mentalistic attributions to these simple animations (Castelli et al., 2000; Martin & Weisberg, 2003). In a recent fMRI study with naturalistic viewing of everyday actions of two humans, McMahon et al. (2023) found a hierarchy of increasing complexity in the representations of action recognition along the lateral visual pathway. Whereas early visual cortex (EVC) and the middle temporal area (MT) have neural representations well described by low-level visual features (for example, as captured by a motion energy model), the extrastriate body area (EBA) and the lateral occipital cortex (LOC) are better described by features such as scenes and objects, and the superior temporal sulcus (STS) is best explained by features consistent with social interactions and communication. As shown by Tavares et al. (2008), it is these higher-level areas that are more strongly modulated by the goals of the observers, illustrated by changing the attentional cue. When observers were cued to attend to the social nature of the interactions between the shapes, neural activity in the STS and adjacent temporoparietal junction (TPJ), posterior cingulate (PCC), and other social cognitive regions all increased relative to when the observers viewed the same animations but attended to the specific visual properties of the scene (i.e., speed and motion trajectories, relative position of the shapes).

Here, we evaluate the kinds of attributions observers make when viewing a large set of Heider-Simmel style vignettes (100 unique animations in total) intended to convey stories through the movements and interactions of simple shapes ("Theatrical" animations, Roemmele et al. (2016)). Using both free-response and constrained-selection approaches, we ask observers to describe the "gist" of the sequences, then using a large language model we reconstruct the similarity structure of the verbal descriptions of animations, targeting key words/phrases for actions, interactions, emotional states, and higher-order attributes about the mental states of the characters. By integrating these semantic models with models capturing low-level visual features, we can evaluate the relative contributions of intentional and perceptual levels of analysis in the representational structure of the social brain network.

Methods

Annotation Experiment

Participants Annotation data was collected from 79 undergraduate students following experimental procedures as approved by the Institutional Review Board at the University of California, Los Angeles.

Experimental Design and Procedure One hundred Heider-Simmel style animations were obtained from a public dataset (https://github.com/asgordon/TriangleCOPA) originally created using the "Heider-Simmel Interactive Theater" tool (Gordon & Roemmele, 2014). Each animation is a depiction of a social narrative involving three characters depicted as distinct shapes: a big triangle, a little triangle, and a circle. The length of the animations varies from 3.88s to 28.71s, with a mean of 11.13s (standard deviation of 4.97s). We will refer to these animations as "Tri-COPA" animations.

Each annotator observed all 100 animations in two stages through a graphical interface designed in MATLAB App Designer. In the first stage, the participant viewed half of the stimuli and then generated one or more keywords that they felt best described the gist of the animation ("free response"). In the second stage, the participant annotated the other half of the animations by selecting from a predefined set of terms that included action, interaction, emotion, and mentalistic features ("predefined responses"). Participants were instructed to select as many labels as they felt were suited to each animation.

Semantic Models From the annotations we created six unique semantic models of similarity. The free response model was constructed by concatenating all the terms identified by all the participants for each animation. The word list of each animation was then projected into a semantic space using Google's Universal Sentence Encoder (USE) language model (Cer et al., 2018), which takes text input, and outputs a 512 dimensional embedding that best captures the semantic structure of the term list. The free response model of similarity space was then calculated as the Pearson correlation distance between the embeddings, for all pairs of animations.

Annotation selections from the predefined terms were combined into five distinct models as shown in Table 1. Action features included terms for observable events that could be completed by a single agent alone; interaction terms described observable events that required the action of two agents; emotion terms described the emotional valence conveyed in the animation; mental cognitive labels described inferences that could be made without reference to emotional valence; and mentalistic emotional labels described inferences that also involved the attribution of emotional coloring. Finally, an additional model included all predefined terms selected for each animation. The selected predefined terms were concatenated across all participants for each animation and projected into semantic space using the USE model. The unique models of similarity space were computed as the Pearson correlation distance between all pairs of animation embeddings.

Table 1: Linguistic Models

Model	Labels
USEact	approach, chase, touch, separation, vibrate, attack, stalk, flee, avoid
USEint	help, cooperation, competition, hug, conversation, argue
USEemo	positive emotion, negative emotion
USEmentcog	unexpected, surprising
USEmentemo	deceive, unwelcome, friendly, protective, scary

Perceptual Models In addition to the language models described in the previous section, we built models of perceptual properties in each animation, specifically targeting features that signal agency (speed variations as a metric for "hidden energy", Scholl & Tremoulet (2000)) and interactivity (distance between agents). Speed models were calculated for each shape as the Euclidean distance traveled on each frame, which was then averaged across all the frames for each animation, resulting in three average speed scores (one for each character). Subsequently, these scores formed the basis of three models of animation dissimilarity based on the Euclidean distance between the average speed of pairs of animations, which were included with three additional models computed from the standard deviation of speed for each agent calculated across the frames per each animation.

Distance between agents describes the frequency with which any two characters are close enough to depict social interactions. This was calculated as the Euclidean distance between all pairs of shapes in each frame of each animation and averaged into a single score for that animation, which were then used to construct dissimilarity models for pairwise potential interactions. Three similar models were built from the standard deviation of distance between characters within each animation.

The average and standard deviation of speed and distance models were further combined in the variance partitioning stage to form a single model group for speed (spd), and another model for distance (dist).

fMRI Experiment

Participants Thirty participants (mean age 22.46, range 18 to 31, 15 male) with normal or correct to normal vision participated in the fMRI experiment which was approved by the Institutional Review Board at the University of California, Irvine.

Procedure All 100 TriCOPA animations were observed in the MRI scanner by each participant over the course of 8 runs in a single scan session (12 or 13 animations per run, with a fixed inter-trial interval of 4.5s). The animations were horizontally flipped on half of the trials to counterbalance the position of the room (left or right) around which the characters interacted. Participants were instructed to view the animations attentively so as to be able to answer questions after the scan session, which asked "What kinds of interactions did you view in the movies? List as many as you can recall."

Data Acquisition Participants were scanned on a 3T Siemens Prisma MRI scanner (Siemens Medical Solutions) equipped with a 32-channel receive-only phase array head coil. T1-weighted images were collected with magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence. Functional images were acquired with an in-plane resolution of 2x2x2mm (multiband accel. factor = 4) and using interleaved slice acquisition (transversal slice orientation, 68 slices, TR = 1500 ms).

fMRI Preprocessing Results included in this manuscript come from preprocessing performed using *fMRIPrep* 21.0.1 (Esteban, Markiewicz, et al. (2018); Esteban, Blair, et al. (2018); RRID:SCR_016216) using the default pipeline (https://fmriprep.org/en/stable/workflows.html), including reconstructions into the fsaverage space using FreeSurfer's reconall. All further analyses were performed under the BIDS standard (Poldrack et al., 2024) using the PyMVPA BIDS-App (Torabian et al., 2023).

Regions of Interest Subjects in the fMRI experiment also participated in two independent scans used for localizing regions of interest. In these scans observers viewed eight 21s Heider-Simmel animations depicting social vignettes, alternated with eight unique animations of mechanical events (Martin & Weisberg, 2003). Regions of interest were identified from a group level random-effects analysis computed on the contrast *social* > *mechanical*, with t-scores thresholded at two standard deviations above the mean (right hemisphere: $t > 6.82$ and left hemisphere: $t > 6.28$). These vertices were then grouped into parcels derived from the Glasser atlas (Glasser et al., 2016), selecting all parcels with at least 10% activated vertices to be included for further analysis. Further parcel exclusion was performed through significance testing as part of variance partitioning, discussed below.

Model fitting and variance partitioning Trial-wise estimates of BOLD activity for each TriCOPA animation was estimated by first de-trending and z-scoring each timeseries for every voxel within each run. The BOLD response for each trial was estimated using the LSS approach (Mumford et al., 2014) that combined a boxcar model specifying the onset and duration for each animation, convolved with a canonical HRF. The design matrix included framewise displacement as a nuisance regressor. The best-fit beta estimates for all trials were computed for each vertex using a fixed-effect general linear model (GLM), which were then grouped by parcel and averaged across participants. Beta estimates were z-scored within each vertex before inclusion in the multivariate analyses.

A representational dissimilarity matrix (RDM) was constructed for each parcel using Pearson correlation and then z-scored, as a metric for distance between trial-wise patterns of activation for each parcel. A regression model was built with the parcel RDM as the dependent variable and a total of 7 language model RDMs and 12 perceptual model RDMs as the independent predictors. The R^2 from this full regression was used to restrict further analysis to the parcels with total variance explained that exceeded levels expected by chance (alpha level of 5%). This criterion was determined by creating null models in which the trial scores of each feature model (semantic, perceptual) were randomly permuted to generate a set of random RDMs from which a null full-model R^2 value can be calculated. This process was repeated 1000 times to generate the distribution of R^2 expected from chance alone, from which significance levels were computed.

The parcels with significant variance explained by the full model were further subjected to variance partitioning using the approach of commonality analysis (Nimon & Reio, 2011). The unique contribution of each class of feature model was assessed through the change in variance explained with that group of models removed:

$$UniqueVar(model) = R^2_{full} - R^2_{model} \quad (1)$$

For each parcel, we calculated the percent contribution of each model group as its unique variance relative to the full model R^2.

Results

Model Comparison Figure 1 shows the pairwise correlations between the perceptual models and the language models obtained from USE embeddings. There was little evidence of shared similarity between the perceptual models and the semantic models. Within the predefined semantic models, correlations were highest between the general USE model (all terms included) and the models of action terms (USEact: $r = .53$) and the models of interaction terms (USEint: $r = .55$). We take this as evidence that people's annotations of the TriCOPA animations slightly favored observable events (action and interaction terms) over the cognitive mentalistic and emotional attributes.

The free response model (USEfree) captured unique dimensions of the similarity structure between the animations, as shown by modest correlations with all predefined models and no meaningful correlation with the model of emotional mentalistic attributions. These results suggest that the intuitive free-response semantic model has the potential to reveal new insights into neural structure that may otherwise not be apparent by the predefined terms.

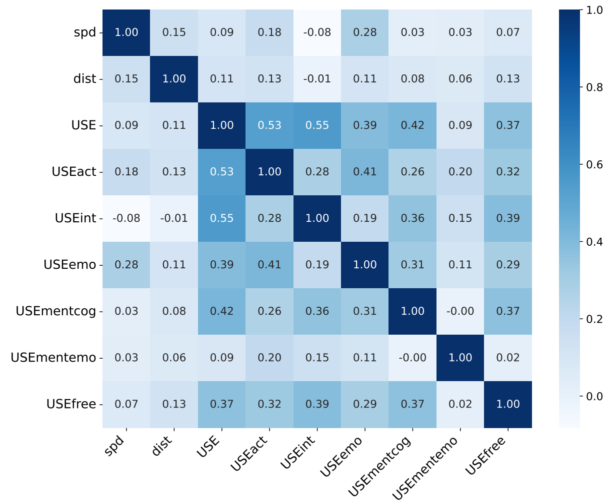

Figure 1: Pairwise correlations between perceptual and language models.

We further analyzed the free response model by clustering the animations into two groups based on their similarities in high-dimensional embedding space. The most repeated labels within each cluster were retained and visualized in Figure 2. The two clusters shared the terms "triangle", "circle", and "together", while one cluster was dominantly represented by action terms (i.e. "leaving", "running", and "knocking") and the other cluster consisted of terms depicting communication ("talking", "fighting", and "arguing").

Regions of Interest Table 2 shows atlas parcels with at least 10% of vertices with stronger BOLD response when observers viewed shape animations depicting social vignettes vs. mechanical events, as assessed in an independent localizer. This analysis identified the right pSTS (parcels STSdp, STSda, STSvp) and bilateral ventral anterior STS (STSva), the bilateral TPJ (PGi, TPOJ1, TPOJ2, TPOJ3 and STV), and two medial parcels at the posterior cingulate and precuneus (7m, 31pd, and PCV). Of these parcels, the bilateral STSva, the right STSda and 7m, and the left STV, pd31, and PCV were excluded due to insufficient variance explained by our full model of perceptual and semantic dissimilarity spaces (alpha = 5%, as determined by estimating null models). The p-value associated with each parcel is reported in the table.

Variance partitioning Figure 3 shows the results of the feature models and the associated variance partitioning for each parcel across both hemispheres. Across all parcels, with one exception, the perceptual models and the free response semantic model captured the most variance in the similarity structure of the BOLD responses. The one exception was the PGi bilaterally, for which the models

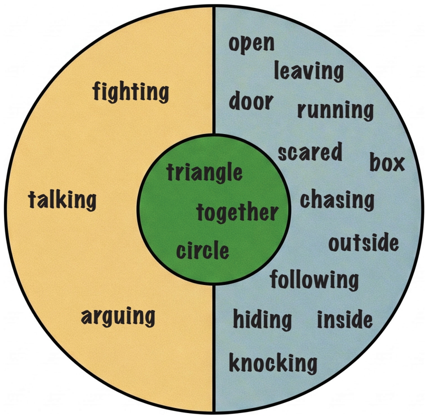

Figure 2: Schematic illustration showing the most frequent labels (*count* > 45) used in the free-response annotations, with the embeddings split into two clusters in similarity space. The green center captures the three most common terms shared by both groups. The left (yellow) cluster shows a dominance of labels signifying communicative interactions and the right (blue) cluster includes a diverse array of action features. Note that repetitions of the same label with minor changes such as "fight" and "fighting" are visualized as a single term.

constructed from the cognitive mentalizing terms (USE-mentcog) best captured the variance among the semantic models.

The perceptual model of interactivity captured the highest proportion of variance in the right TPJ and posterior STS parcels (PGi, STV, TPOJ1, and STSdp). The semantic model of interactive terms (USEint), however, did not well explain the variance in the neural dissimilarity matrices, which suggests that the processing of social interactions in posterior and mid-STS may reflect the perceptual building blocks for perceived interactivity (i.e. distance between the actors) rather than discern the nature of the different types of interactions (such as cooperation and conversation).

In the left hemisphere and the PCC, the perceptual models for speed of the characters and distance between them performed equivalently, with the exception of PGi which was better explained by speed.

Discussion

In this study, we investigated humans' attributions when viewing a large set of Heider-Simmel style animations designed to depict a wide range of simple and mentalistic interactions. We constructed semantic models derived from human annotations that reflected the similarity structure of the descriptions, along with models of visual features relevant to perceived interactivity. Targeting social cognitive brain regions, we performed variance partitioning to evaluate which of the models best explained the BOLD response in parcels in the TPJ, pSTS, and posterior cingulate.

Table 2: Intersection of parcels with functional localizer. Only parcels with at least 10% of vertices with strong BOLD response to social interactions are shown. Further exclusion of parcels was performed if the full-model within each parcel did not exceed chance performance.

Parcel	Left	p-value	Right	p-value
PGi	35.03%	$< .001$	51.81%	$< .001$
TPOJ1	26.98%	$< .001$	55.52%	$< .001$
TPOJ2	35.00%	$< .001$	15.11%	$< .001$
TPOJ3	14.50%	$< .001$	24.43%	$< .001$
STV	65.88%	.102	54.47%	$< .001$
STSdp	–	–	20.20%	$< .01$
STSda	–	–	30.92%	.252
STSvp	–	–	23.60%	$< .001$
STSva	14.54%	.948	72.75%	.102
7m	–	–	15.66%	.134
31pd	15.04%	.393	12.35%	$< .001$
PCV	16.71%	.292	–	–

Pairwise correlations between a language model based on free responses and models formed using predefined features each captured a somewhat different organization of the semantic space. Cluster analysis of the free-response embeddings revealed that when given the opportunity to freely describe the TriCOPA animations, observers relied on a diverse array of terms that conveyed observable actions of individuals and communication between individuals. These results reveal the diversity and multi-level structure of attributions made while intuitively describing social narratives.

The free response model, in particular, was very effective in describing the similarity structure of the neural responses in many of the parcels evaluated as shown by variance partitioning. To the extent that this was driven by the communicative aspect of the annotations, this finding is consistent with at least one previous study which identified the presence of communication in interactions to uniquely capture variance in the STS (McMahon et al., 2023). The diverse and multi-level responses in the posterior and mid-STS are also consistent with studies implicating the bilateral STS as a general lexical conceptual site (Hickok & Poeppel, 2007). Furthermore, regions on and around the angular gyrus (AG) have been suggested as one of two conceptual knowledge zones that are specifically involved in knowledge about events and relations between entities (Matchin & Hickok, 2020). Together, these previous studies and our findings may suggest that the more posterior aspects of the lateral stream may reflect multi-modal representations of social interactions. Alternatively, our results may be taken as evidence that even simple linguistic descriptions that capture narrative structures ("gists") may better model neural responses in the lateral stream than perceptual features alone.

Variance partitioning also showed variance in the similar-

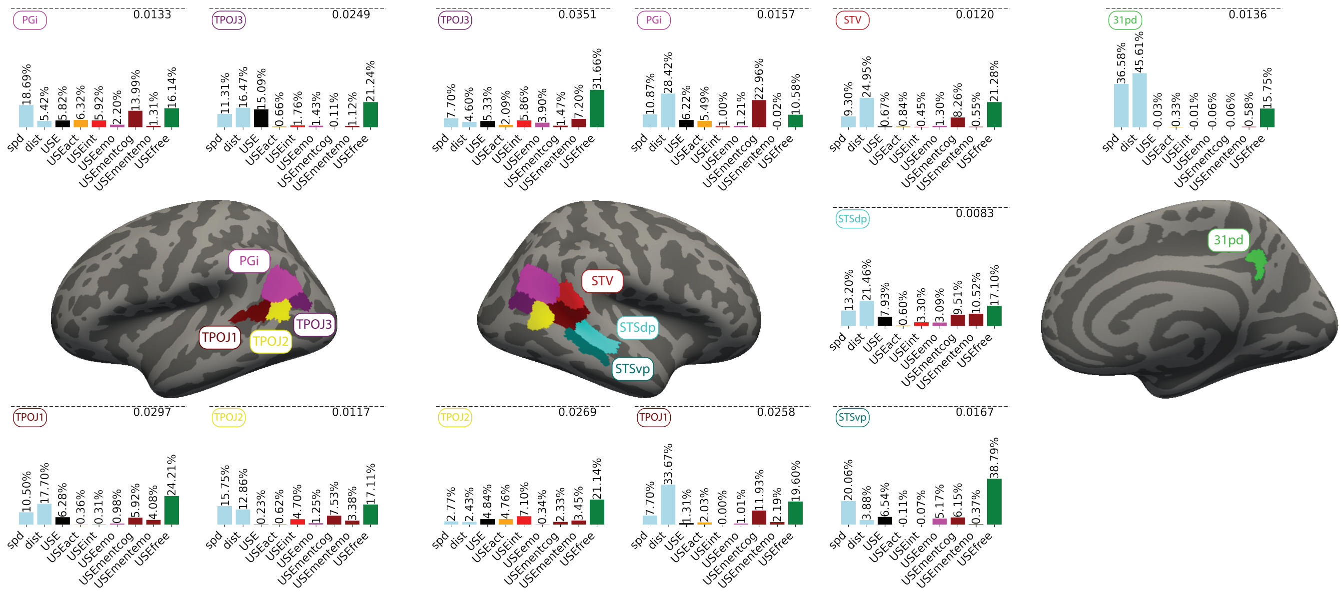

Figure 3: Variance partitioning results. For each parcel, the dotted line denotes the R^2 obtained from the full model that included all perceptual and linguistic models. Percentages show the proportion of unique variance captured by each model.

ity structure of neural responses to be well explained by the visual features in the animations. In the right hemisphere, the TPOJ1 and adjacent STSdp, STV, and PGi were best explained by a perceptual model of interactivity, consistent with previous reports that representations of interactivity are a key feature represented within these regions (Isik et al., 2017; Lahnakoski et al., 2012; Tarhan & Konkle, 2020). Models of the perceptual features also best explained variance in the right posterior cingulate, a region that is both linked to cognitive load and to introspective processes (Foster et al., 2023) and to social interaction (Schilbach et al., 2008). The study by Schilbach et al. (2008) defined social interaction through facial expressions as they appear when initiating communication. Our findings, however, suggest that each of these brain regions have neural signals that can be sufficiently explained by perceptual models of implied interactivity without explicit cues as to animacy (i.e. faces or bodies).

It is worth noting one specific semantic model that stood out among the others, which was the cognitive mentalistic key terms that captured violations of expectation ("unexpected" and "surprising"). Among all the models derived from predefined terms, the RDM constructed from this model best described the similarity structure in the PGi, a large parcel that is centered on the dorsal aspects of the TPJ. Our finding that violations of expectation are linked to the TPJ is consistent with prior literature linking this region to prediction error (Abrahamse & Silvetti, 2016), re-orienting (Patel et al., 2019), and false beliefs (Saxe & Kanwisher, 2003). It is worth noting that the PGi is a relatively large parcel that extends into the pSTS, and its involvement in both interactivity and mentalizing might be due to the activity of distinct subregions within the PGi. Further breaking of the PGi into smaller parcels or a fine-grained searchlight analysis at this region is needed for more conclusive model comparisons.

Acknowledgments

This work was supported by the NSF grants BCS-1658560 to EG, BCS-2142269 to HL, and BCS-1658078 to JP.

References

Abell, F., Happé, F., & Frith, U. (2000, January). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, *15*(1), 1–16. Retrieved 2022-11-09, from `https://www.sciencedirect.com/science/article/pii/S0885201400000149` doi: 10.1016/S0885-2014(00)00014-9

Abrahamse, E. L., & Silvetti, M. (2016, May). Commentary: The Role of the Parietal Cortex in the Representation of Task–Reward Associations. *Frontiers in Human Neuroscience*, *10*, 192. Retrieved 2022-09-22, from `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4861704/` doi: 10.3389/fnhum.2016.00192

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000, September). Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns. *NeuroImage*, *12*(3), 314–325. Retrieved 2020-09-02, from https://linkinghub.elsevier.com/retrieve/pii/S1053811900906128 doi: 10.1006/nimg.2000.0612

Cer, D., Yang, Y., Kong, S.-y., Hua, N., Limtiaco, N., John, R. S., … Kurzweil, R. (2018). Universal Sentence Encoder.

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., … Gorgolewski, K. J. (2018). fmriprep. *Software*. doi: 10.5281/zenodo.852659

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., … Gorgolewski, K. J. (2018). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*. doi: 10.1038/s41592-018-0235-4

Foster, B. L., Koslov, S. R., Aponik-Gremillion, L., Monko, M. E., Hayden, B. Y., & Heilbronner, S. R. (2023, March). A tripartite view of the posterior cingulate cortex. *Nature Reviews Neuroscience*, *24*(3), 173–189. Retrieved 2025-02-02, from https://www.nature.com/articles/s41583-022-00661-x (Publisher: Nature Publishing Group) doi: 10.1038/s41583-022-00661-x

Gergely, G., & Csibra, G. (2003, July). Teleological reasoning in infancy: the naıve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. Retrieved 2020-04-25, from https://linkinghub.elsevier.com/retrieve/pii/S1364661303001281 doi: 10.1016/S1364-6613(03)00128-1

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., … Van Essen, D. C. (2016, August). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. Retrieved 2025-02-03, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4990127/ doi: 10.1038/nature18933

Gordon, A. S., & Roemmele, M. (2014). An Authoring Tool for Movies in the Style of Heider and Simmel. In A. Mitchell, C. Fernández-Vara, & D. Thue (Eds.), *Interactive Storytelling* (Vol. 8832, pp. 49–60). Cham: Springer International Publishing. Retrieved 2025-02-03, from http://link.springer.com/10.1007/978-3-319-12337-0_5 (Series Title: Lecture Notes in Computer Science) doi: 10.1007/978-3-319-12337-0_5

Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, *57*(2), 243–259. Retrieved 2022-03-14, from https://www.jstor.org/stable/1416950 (Publisher: University of Illinois Press) doi: 10.2307/1416950

Hickok, G., & Poeppel, D. (2007, May). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. Retrieved 2023-01-08, from http://www.nature.com/articles/nrn2113 doi: 10.1038/nrn2113

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017, October). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *114*(43), E9145–E9152. Retrieved 2025-02-03, from https://www.pnas.org/doi/10.1073/pnas.1714471114 (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.1714471114

Klein, A. M., Zwickel, J., Prinz, W., & Frith, U. (2009, June). Animated triangles: An eye tracking investigation. *Quarterly Journal of Experimental Psychology*, *62*(6), 1189–1197. Retrieved 2025-01-30, from https://doi.org/10.1080/17470210802384214 (Publisher: SAGE Publications) doi: 10.1080/17470210802384214

Lahnakoski, J. M., Glerean, E., Salmi, J., Jääskeläinen, I. P., Sams, M., Hari, R., & Nummenmaa, L. (2012, August). Naturalistic fMRI Mapping Reveals Superior Temporal Sulcus as the Hub for the Distributed Brain Network for Social Perception. *Frontiers in Human Neuroscience*, *6*, 233. Retrieved 2025-02-03, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3417167/ doi: 10.3389/fnhum.2012.00233

Malik, M., & Isik, L. (2023, November). Relational visual representations underlie human social interaction recognition. *Nature Communications*, *14*(1), 7317. Retrieved 2024-07-04, from https://www.nature.com/articles/s41467-023-43156-8 (Publisher: Nature Publishing Group) doi: 10.1038/s41467-023-43156-8

Martin, A., & Weisberg, J. (2003). Neural Foundations For Understanding Social And Mechanical Concepts. *Cognitive neuropsychology*, *20*(3-6), 575–587. Retrieved 2022-05-19, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1450338/ doi: 10.1080/02643290342000005

Matchin, W., & Hickok, G. (2020, March). The Cortical Organization of Syntax. *Cerebral Cortex (New York, N.Y.: 1991)*, *30*(3), 1481–1498. doi: 10.1093/cercor/bhz180

McMahon, E., Bonner, M. F., & Isik, L. (2023, December). Hierarchical organization of social action features along the lateral visual pathway. *Current biology : CB*, *33*(23), 5035–5047.e8. Retrieved 2025-01-30, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10841461/ doi: 10.1016/j.cub.2023.10.015

Moessnang, C., Otto, K., Bilek, E., Schäfer, A., Baumeister, S., Hohmann, S., … Meyer-Lindenberg, A. (2017). Differential responses of the dorsomedial prefrontal cortex and right posterior superior temporal sulcus to spontaneous mentalizing. *Human Brain Mapping*, *38*(8), 3791–3803. Retrieved 2021-09-09, from https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.23626 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.23626) doi: 10.1002/hbm.23626

Mumford, J. A., Davis, T., & Poldrack, R. A. (2014, December). The impact of study design on pattern estimation for single-trial multivariate pattern analysis. *NeuroImage*, *103*, 130–138. Retrieved 2022-03-04, from https://linkinghub.elsevier.com/retrieve/pii/S105381191400768X doi: 10.1016/j.neuroimage.2014.09.026

Nimon, K., & Reio, T. G. (2011, September). Regression Commonality Analysis: A Technique for Quantitative Theory Building. *Human Resource Development Review*, *10*(3), 329–340. Retrieved 2025-02-03, from https://doi.org/10.1177/1534484311411077 (Publisher: SAGE Publications) doi: 10.1177/1534484311411077

Patel, G. H., Sestieri, C., & Corbetta, M. (2019, September). The evolution of the temporoparietal junction and posterior superior temporal sulcus. *Cortex*, *118*, 38–50. Retrieved 2020-04-06, from https://linkinghub.elsevier.com/retrieve/pii/S0010945219300462 doi: 10.1016/j.cortex.2019.01.026

Pitcher, D., & Ungerleider, L. G. (2021, February). Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences*, *25*(2), 100–110. doi: 10.1016/j.tics.2020.11.006

Poldrack, R. A., Markiewicz, C. J., Appelhoff, S., Ashar, Y. K., Auer, T., Baillet, S., … Gorgolewski, K. J. (2024, March). The past, present, and future of the brain imaging data structure (BIDS). *Imaging Neuroscience*, *2*, 1–19. Retrieved 2025-02-03, from https://doi.org/10.1162/imag_a_00103 doi: 10.1162/imag_a_00103

Roemmele, M., Morgens, S.-M., Gordon, A. S., & Morency, L.-P. (2016, March). Recognizing Human Actions in the Motion Trajectories of Shapes. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (pp. 271–281). Sonoma California USA: ACM. Retrieved 2021-09-16, from https://dl.acm.org/doi/10.1145/2856767.2856793 doi: 10.1145/2856767.2856793

Roux, P., Passerieux, C., & Ramus, F. (2013, February). Kinematics matters: A new eye-tracking investigation of animated triangles. *Quarterly Journal of Experimental Psychology*, *66*(2), 229–244. Retrieved 2025-02-03, from https://doi.org/10.1080/17470218.2012.704052 (Publisher: SAGE Publications) doi: 10.1080/17470218.2012.704052

Saxe, R., & Kanwisher, N. (2003, August). People thinking about thinking peopleThe role of the temporo-parietal junction in "theory of mind". *NeuroImage*, *19*(4), 1835–1842. Retrieved 2022-03-14, from https://linkinghub.elsevier.com/retrieve/pii/S1053811903002301 doi: 10.1016/S1053-8119(03)00230-1

Schilbach, L., Eickhoff, S. B., Mojzisch, A., & Vogeley, K. (2008, March). What's in a smile? Neural correlates of facial embodiment during social interaction. *Social Neuroscience*, *3*(1), 37–50. Retrieved 2025-02-02, from https://doi.org/10.1080/17470910701563228 (Publisher: Routledge _eprint: https://doi.org/10.1080/17470910701563228) doi: 10.1080/17470910701563228

Scholl, B. J., & Tremoulet, P. D. (2000, August). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299–309. Retrieved 2020-05-07, from https://linkinghub.elsevier.com/retrieve/pii/S1364661300015060 doi: 10.1016/S1364-6613(00)01506-0

Schultz, J., & Frith, C. D. (2022, September). Animacy and the prediction of behaviour. *Neuroscience & Biobehavioral Reviews*, *140*, 104766. Retrieved 2025-02-04, from https://www.sciencedirect.com/science/article/pii/S014976342200255X doi: 10.1016/j.neubiorev.2022.104766

Shu, T., Peng, Y., Fan, L., Lu, H., & Zhu, S.-C. (2018). Perception of Human Interaction Based on Motion Trajectories: From Aerial Videos to Decontextualized Animations. *Topics in Cognitive Science*, *10*(1), 225–241. Retrieved 2022-11-08, from https://onlinelibrary.wiley.com/doi/abs/10.1111/tops.12313 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tops.12313) doi: 10.1111/tops.12313

Tarhan, L., De Freitas, J., & Konkle, T. (2021, December). Behavioral and Neural Representations en route to Intuitive Action Understanding. *Neuropsychologia*, *163*, 108048. Retrieved 2025-01-30, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8649031/ doi: 10.1016/j.neuropsychologia.2021.108048

Tarhan, L., & Konkle, T. (2020, June). Sociality and interaction envelope organize visual action representations. *Nature Communications*, *11*(1), 3002. Retrieved 2025-02-03, from https://www.nature.com/articles/s41467-020-16846-w (Publisher: Nature Publishing Group) doi: 10.1038/s41467-020-16846-w

Tavares, P., Lawrence, A. D., & Barnard, P. J. (2008, August). Paying Attention to Social Meaning: An fMRI Study. *Cerebral Cortex*, *18*(8), 1876–1885. Retrieved 2020-04-17, from https://academic.oup.com/cercor/article-lookup/doi/10.1093/cercor/bhm212 doi: 10.1093/cercor/bhm212

Torabian, S., & Grossman, E. D. (2023, September). When shapes are more than shapes: perceptual, developmental, and neurophysiological basis for attributions of animacy and theory of mind. *Frontiers in Psychology*, *14*. Retrieved 2025-02-01, from https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1168739/full (Publisher: Frontiers) doi: 10.3389/fpsyg.2023.1168739

Torabian, S., Vélez, N., Sochat, V., Halchenko, Y. O., & Grossman, E. D. (2023, August). The PyMVPA BIDS-App: a robust multivariate pattern analysis pipeline for fMRI data. *Frontiers in Neuroscience*, *17*. Retrieved 2025-02-03, from https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2023.1233416/full (Publisher: Frontiers) doi: 10.3389/fnins.2023.1233416

Ullman, T. D., Baker, C. L., Macindoe, O., Evans, O., Goodman, N. D., & Tenenbaum, J. B. (2009). Help or Hinder: Bayesian Models of Social Goal Inference.