

Multidimensional neural representations of social features during movie viewing

Haemy Lee Masson,¹ Lucy Chang,² and Leyla Isik²

¹Department of Psychology, Durham University, Durham DH1 3LE, UK

²Department of Cognitive Science, Johns Hopkins University, Baltimore 21218, USA

Correspondence should be addressed to Haemy Lee Masson, Department of Psychology, Durham University, South Road, Durham DH1 3LE, UK.

E-mail: haemy.lee-masson@durham.ac.uk.

[†]These authors share first authorship.

Abstract

The social world is dynamic and contextually embedded. Yet, most studies utilize simple stimuli that do not capture the complexity of everyday social episodes. To address this, we implemented a movie viewing paradigm and investigated how everyday social episodes are processed in the brain. Participants watched one of two movies during an MRI scan. Neural patterns from brain regions involved in social perception, mentalization, action observation and sensory processing were extracted. Representational similarity analysis results revealed that several labeled social features (including social interaction, mentalization, the actions of others, characters talking about themselves, talking about others and talking about objects) were represented in the superior temporal gyrus (STG) and middle temporal gyrus (MTG). The mentalization feature was also represented throughout the theory of mind network, and characters talking about others engaged the temporoparietal junction (TPJ), suggesting that listeners may spontaneously infer the mental state of those being talked about. In contrast, we did not observe the action representations in the frontoparietal regions of the action observation network. The current findings indicate that STG and MTG serve as key regions for social processing, and that listening to characters talk about others elicits spontaneous mental state inference in TPJ during natural movie viewing.

Keywords: theory of mind; social communication; social interaction perception; action observation; naturalistic fMRI; representational similarity analysis

Introduction

Humans form impressions about others, such as their personality traits and social status, based on observable social cues manifesting through actions, communication and interactions (Mehl *et al.*, 2006; Ames *et al.*, 2011; Quadflieg and Koldewyn, 2017). An early sensitivity to social cues found in infancy (Hamlin *et al.*, 2007; Hamlin and Wynn, 2011) may be a precursor to functional neural selectivity in adulthood (Grossmann and Johnson, 2007). Neuroimaging studies have identified two widely recognized brain systems related to distinctive social functions: the action observation network and the mentalizing network (Van Overwalle and Baetens, 2009). Observing others' actions activates the action observation network, including the inferior frontal gyrus (IFG), intraparietal sulcus (IPS) and superior temporal sulcus (STS) (Caspers *et al.*, 2010; Kilner, 2011). Inferring the mental states of others (i.e. theory of mind (Premack and Woodruff, 1978)) activates the mentalizing network, including the medial prefrontal cortex (mPFC), temporoparietal junction (TPJ), precuneus and temporal pole (TP) (Gallagher and Frith, 2003; Jacoby *et al.*,

2016). Beyond those systems, prior work has identified regions in the STS that show functionally selective to social interaction in both controlled experiments using simple stimuli (Isik *et al.*, 2017; Walbrin *et al.*, 2018) and in more ecologically valid studies that involve natural viewing (Lee Masson and Isik, 2021; McMahon *et al.*, 2023). Furthermore, the STS has shown selective neural responses to visually presented social communication (McMahon *et al.*, 2023) and speech-based social communication (Landsiedel and Koldewyn, 2023). Recent data-driven work has further suggested that communication and antisocial behavior elicit responses in the superior temporal gyrus (STG) and middle temporal gyrus (MTG), respectively (Santavirta *et al.*, 2023).

In the real social world, social communication and interaction co-occur frequently with theory of mind and action observation. However, previous research has predominantly examined these facets in isolation, resulting in gaps in our understanding of complex social processes in extended, real-world contexts. The goal of the current study is to provide an in-depth understanding of the neural mechanisms underlying complex social pro-

Received: 21 November 2023; Revised: 5 March 2024; Accepted: 3 May 2024

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

cesses in real-world contexts by adopting a natural movie viewing approach. Specifically, by densely labeling social features of movies and employing representational similarity analysis (RSA) to the two movie functional magnetic resonance imaging (fMRI) datasets, we identified how social interaction, action observation, mentalization and three contents of spoken communication (characters talking about themselves, others or things) are represented in the three neural systems implicated in social perception, action observation and mentalization. This study builds off our prior work with the same datasets (Lee Masson and Isik, 2021) in two important ways. First, by using an RSA approach vs voxel-wise modeling, we can examine representational match within key hypothesized regions of interest (ROIs). Second, we expanded the set of social features labeled to include richer speech labels. While different types of social interaction (e.g. helping vs hindering; arguing vs celebrating; social touch) have been investigated in cognitive neuroscience (Isik et al., 2017; Lee Masson et al., 2018; Walbrin et al., 2018; Walbrin and Koldewyn, 2019), no previous studies have explored the neural representations of different targets of speech that vary based on the spoken social content.

We find that STG and MTG are responsible for processing various social features in both movies, including three contents of spoken communication (regardless of content—self, others and things), social interactions (including touch), mentalization and others' action. The mentalizing network, excluding the precuneus, processes mentalization and social interaction features. Listening to characters talking about others is processed in TPJ, whereas listening to conversations revolving around objects or inanimate items is processed in TP within the mentalizing network. The frontoparietal regions of the action observation network did not represent others' action.

Methods

By performing RSA (Kriegeskorte et al., 2008) on two fMRI movie datasets, we evaluated how various social features are represented in the brain areas implicated in social perception, action observation and theory of mind. To this end, using two publicly available fMRI movie datasets, where participants watched the movie *Sherlock* and *500 Days of Summer* (Figure 1A), we analyzed the neural responses of 11 brain regions to 10 different sensory and social features. Brain regions and features were selected based on prior hypotheses on social processing.

Movie feature annotations

The movies were split into three segments. For the *Summer* movie, the opening and ending credits were cropped. This resulted in a total of 988 segments for *Sherlock* and 1722 segments for *500 Days of Summer*. Prior to fMRI analysis, all movie segments were labeled with six social features by two raters—social interactions, mentalization, characters talking about themselves, talking about someone else, talking about something else and actions. For the *Summer* movie, social touch was also labeled. The *Sherlock* movie rarely contained social touch scenes, so this feature was not included in *Sherlock*. Features were labeled 1 if the feature was present in a scene and 0 if it was absent.

Precisely, for the social interaction feature, we labeled scenes that involved any human–human interaction either through conversation (e.g. speaking) or action (e.g. hugging). For the spoken communication feature, we created separate vectors that represented the various talking scenes—characters talking about themselves participating in communication (lines with I/you—'You must be an army doctor'), characters talking about other

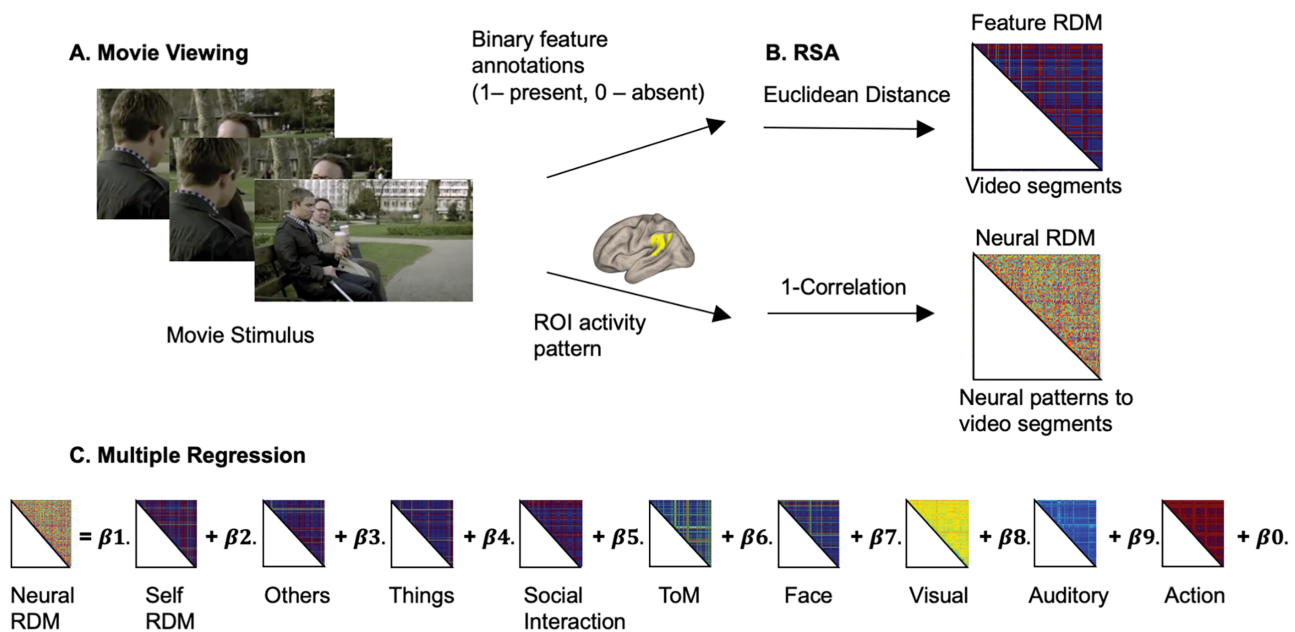


Fig. 1. Overview of methods. (A) fMRI data from participants viewing two movies—*Sherlock* and *500 Days of Summer*—were used for analyses. Each video segment was annotated for its social content, including the presence of a social interaction, theory of mind (ToM), characters speaking about themselves (self), characters speaking about others and characters speaking about things. Sensory features such as the presence of a face, audio amplitude and visual features extracted from the fifth convolutional layer of AlexNet were also included. (B) The representational dissimilarity matrices were created from the neural patterns for each ROI and feature. Each element in the matrix represents the distance or correlation between the features or neural response patterns of every possible pair combination of 3 s video segments. (C) A multiple regression model was fit to each ROI to predict neural activity based on the different feature variables. Beta-weights for each feature served as a measure of how strongly that feature explains the neural data.

Table 1. Inter-rater reliability.

	Social interaction	Self	Others	Things	Mentalization	Touch
Sherlock	97%	94%	95%	95%	79%	
Summer	94%	89%	93%	89%	85%	84%

characters who are not part of communication (lines with he/she/they/other character's names—'Yeah, he's always like that') and characters talking about things (lines with it/object name—'On the desk there is a number.')

In cases where multiple communication features were present in a scene (e.g. 'You got all that because you realized the case would be pink'; this scene has both characters talking about themselves and about an object—the case'), this scene was labeled 1 for characters talking about themselves and 1 for characters talking about objects.

The presence of mentalization was labeled in each scene whenever the character was inferring another character's emotions and thoughts (e.g. Sherlock says to Dr Watson 'Dear God, what is it like in that funny little brain of yours. It must be so boring'. In Summer, the narrator infers that '[Tom will] never be truly happy, until the day he met the one'). As labeling mentalization features based on the non-verbal expression of a character can be highly subjective, this feature is labeled solely based on the contents of the conversation. Descriptions of a character's appearance (e.g. 'She is tall and thin') or bodily sensation (e.g. 'She has been sick for 3 days') were not labeled as mentalization. This type of second-order mentalization activates the theory of mind network (Tholen et al., 2020), as a listener may consider the speaker's perspective and what is said about another person's mental state. This annotation method was chosen for objectivity and to prevent raters from making subjective assumptions about a movie viewer's mentalizing activity.

The percentage of agreement ((the total number of agreements in ratings/the total number of video segments) × 100) between two raters is very high for all social features (Table 1). Thus, scores were averaged across two raters.

For the action feature, only one rater labeled the actions present as there is no straightforward way to average action names across raters. The rater remained consistent with wording throughout the movie (e.g. using 'speaking' instead of 'talking' in all the scenes). Only the Summer movie was labeled for social touch, scenes where characters are engaging in physical contact (e.g. hugging and kissing).

To account for co-varying sensory information, we included sensory features—audio amplitude, the presence of a face and other high-level visual features quantified as the activations extracted from the final convolutional layer of a deep neural network. We used sensory features extracted from previous studies (Aliko et al., 2020; Lee Masson and Isik, 2021). For more details see [Supplementary Material \(SM\)](#).

The feature annotations were then turned into representational dissimilarity matrices (RDMs), which were used as predictors in a multiple regression model to explain the neural patterns. The feature RDM was created by calculating the pairwise Euclidean distance of each feature between all pairs of movie segments (Figure 1B). For the action RDM, identical actions were given a value of 0, while different actions (speaking vs hugging) were given a value of 1.

Feature correlations

To determine the correlation between each feature, we conducted a Pearson correlation analysis on all pairs of feature RDMs. We chose to use RDMs instead of raw feature annotations as this approach enabled us to establish correlations between high-dimensional visual features (256 × 13 × 13 for each scene) and other features. The results were visualized using the `corrplot` function in R version 3.6.3 (R Core Team, 2020). Several features were correlated across the movies. In particular, the feature capturing characters speaking about others was positively correlated with the mentalization feature in both movies. The presence of social interaction was correlated with the presence of face and action feature (Figure 2), likely reflecting the fact that 'talking' is the most prevalent action in both movies ([Supplemental Table S1](#)).

fMRI data sources

fMRI data were collected while two sets of participants watched the first episode of Sherlock BBC TV series (N = 17) and 500 Days of Summer (N = 20). Datasets are publicly available from two different studies (Chen et al., 2017; Aliko et al., 2020). A brief description of the scanning parameters can be found in SM. For the Sherlock movie, we removed one participant (subject 5) from analysis due to missing data at the end of the scan, resulting in a total of 16 participants. For the Summer movie, we removed two participants (IDs 14 and 16) as one was scanned with a different head coil and the other was given glasses only after the first run, resulting in a total of 18 participants. The studies were approved by the Princeton University Institutional Review Board and the Ethics Committee of University College London, respectively. All subjects provided their written informed consent before the experiment.

Definition of brain ROIs

We conducted ROI-based analyses on fMRI data that underwent multiple preprocessing steps performed by the authors of the original study. For more details about the preprocessing steps, see SM. We measured neural representations of various social features in three well-defined networks (action observation, mentalizing and social perception). First, anatomical ROI masks were created by using various templates. The TPJ (anterior and posterior parcels), anterior portion of mPFC (clusters 3 and 4) and posterior portion of mPFC (clusters 1 and 2) templates were taken from the connectivity-based parcellation atlas (Mars et al., 2012; Sallet et al., 2013). Following a previous study on social norm processing (Pegado et al., 2018), we separated the mPFC into two distinct ROIs. The original templates only included the right hemisphere, despite finding similar parcellation in the left hemisphere. To have bilateral ROIs, we created a mirror ROI on the left side and merged it with the original template. The STG, MTG, precuneus, opercular part of IFG and TP templates were taken from the automatic anatomical labeling atlas (Tzourio-Mazoyer et al., 2002)

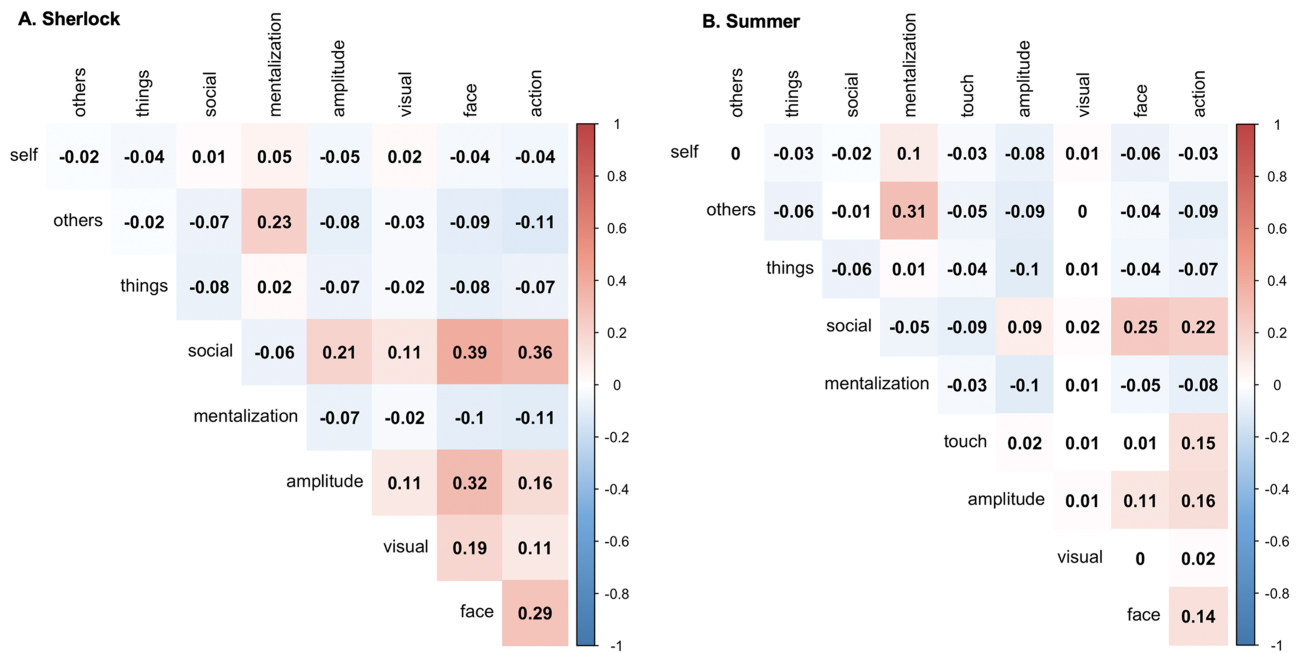


Fig. 2. Pairwise Pearson's correlation coefficient between feature RDMs in Sherlock (A) and Summer movies (B). A positive number (red) indicates a positive correlation, while a negative number (blue) indicates a negative correlation. Self = characters talking about themselves, Others = characters talking about other characters, Things = characters talking about things, Social = social interaction, and Visual = the fifth layer of Alexnet.

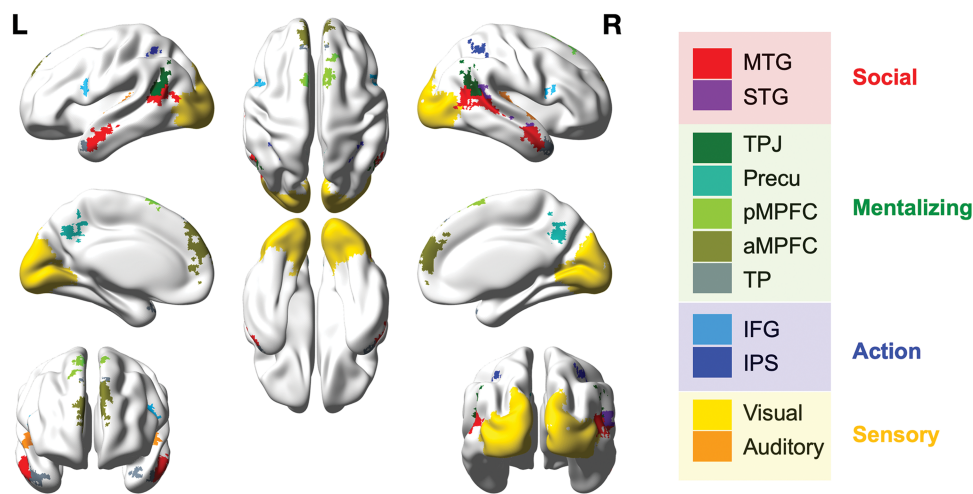


Fig. 3. An illustration of selected ROIs visualized through BrainNetViewer (Xia et al., 2013).

using PickAtlas software version 3.0.5b (Maldjian et al., 2003). We opted to use STG and MTG templates as the STS map is less standardized and encompasses TPJ, another ROI in the current study. The IPS mask consists of hIP1, 2 and 3 templates taken from the SPM Anatomy Toolbox Version 2.2b (Eickhoff et al., 2005). We also included two control sites for visual and auditory processing. The visual brain mask consists of V1, V2, V3, V4, V5 and lateral occipital cortex extracted from the SPM Anatomy toolbox. The auditory brain mask covers the primary auditory cortex extracted from the same toolbox.

Second, to select the functionally relevant voxels within each anatomical mask, we obtained brain activation maps (thresholded at z -score >3) from Neurosynth (<https://neurosynth.org/>) by searching the keywords—social, action observation and mentalizing. Using FSL from the FMRIB Software Library (Jenkinson et al., 2012), these activation maps were binarized. To define STG

and MTG, we selected all voxels from the social map that were restricted to each anatomical mask (Figure 3). Anterior mPFC, posterior mPFC, TPJ, precuneus and TP were defined by selecting all voxels from the mentalizing map within the corresponding anatomical mask. IFG and IPS were defined using the action observation map. The visual and auditory ROIs were defined with anatomical templates as they only serve as reference sites. We removed any overlapping voxels between pairs of ROIs to ensure that all ROIs were anatomically independent of each other (Table 2).

Neural RDMs

Prior to creating neural RDMs, blood-oxygen-level-dependent signals were averaged for every two TRs (e.g. first and second TR, third and fourth TR) for Sherlock fMRI data and three TRs for Sum-

Table 2. Overlapping voxels from ROIs pair and final ROI size.

	Overlapping voxels	Final voxel size
STG	11 voxels overlapping with TPJ	77
MTG	19 with visual, 80 with TPJ	649
TPJ	11 with STG, 80 with MTG	182
TP	No overlap	150
Precuneus	No overlap	221
aMPFC	8 with pMPFC	289
pMPFC	8 with aMPFC	98
IFG	No overlap	97
IPS	No overlap	49
Auditory	No overlap	266
Visual	19 with MTG	4424

mer fMRI data to make fMRI data have a resolution of 3 s matching to feature annotations. This resulted in 988 neural patterns for Sherlock and 1722 neural pattern for Summer. To create the neural RDMs for each participant and ROI, we used the CoSMoMVA toolbox in MATLAB (Oosterhof et al., 2016). This involved calculating the pairwise correlation distance (1—Pearson correlation across all voxels within an ROI) of the neural patterns in response to each scene of the movie (Figure 1B). Finally, feature and neural RDMs were normalized to have the mean of 0 and an s.d. of 1. For the subsequent analyses, the entries from the upper diagonal of RDMs were used as variables as all matrices are symmetric (Ritchie et al., 2017).

Statistical analysis

For the group-level statistical inference, we conducted a one-tailed sign permutation test with 5000 iterations. *P*-values were corrected for multiple comparisons using a maximum correlation threshold across all ROIs (Nichols and Holmes, 2002).

Inter-subject correlation as a measure of reliability

To determine the reliability of the neural data, we performed a leave-one-subject-out correlation. Specifically, for each ROI, the neural RDM of one participant was correlated with the averaged neural RDMs of the other participants. A permutation test revealed that the averaged correlation value across participants is above chance for all ROIs, indicating that neural data are reliable (all *P*_{corrected} values <0.05 and *r*-values listed in Supplemental Table S2). The inter-subject correlation value can also be interpreted as the noise ceiling, which is the expected highest correlation between neural data and other predictors, taking into account the noise present in the neural data (Nili et al., 2014).

Multiple regression analysis

All features were assigned as predictors in the multiple regression model to explain the neural patterns in each ROI for each participant (Figure 1C). A fitlm function in MATLAB was used for this analysis. Prior to this analysis, we first checked for multicollinearity using variance inflation factors (VIFs) (Marquardt, 1970). Our analysis showed that each predictor had a VIF value of 1.3 or lower for both movies. Typically, a VIF value >5 indicates moderate multicollinearity (Belsley, 1991). Since VIF values were well below that threshold, we assumed that critical levels of multicollinearity would not be present in our model. After the regression analysis, we performed a sign permutation test on beta-values from each predictor. *P*-values were adjusted for multiple comparisons to account for the number of ROIs tested.

Neural pattern similarity between ROIs

Lastly, we performed a correlation analysis between the neural RDMs to identify the representational relationship between ROIs (Kriegeskorte et al., 2008). This approach differs from comparing averaged neural responses between ROIs as RSA on neural RDMs enables us to compare the representational structure across the pattern of voxels in each ROI (Pillet et al., 2020). The representational similarity between ROIs was calculated through pairwise Pearson correlation of the neural RDMs averaged across all participants for all ROI pairs. We visualized the results through multidimensional scaling (MDS) reconstruction with the mdscale function on MATLAB.

Results

Neural patterns from 11 brain regions were extracted from two fMRI datasets recorded while subjects viewed different movies, Sherlock and 500 Days of Summer. We first computed a neural RDM for each ROI based on the pairwise similarity of each region's response to different movie segments. To determine how brain regions involved in social perception, mentalizing and action observation represent various social features while watching natural movies, we fit a multiple regression model to each ROI using the different feature RDMs as predictors (Figure 1). We determined that a feature was represented in the brain when it showed statistical significance in both movie fMRI datasets.

High-level social features are represented in STG and MTG

All high-level social features—social interaction (including social touch interaction), mentalization, three contents of speaking features (talking about themselves, others and things) and action—were significantly represented in both STG and MTG (Figure 4A). In contrast, visual (including the face feature) and auditory features were not consistently represented across two movies in these regions (beta-values and statistics are listed in Supplemental Table S3).

The mentalizing network represents the mentalization feature and characters speaking about others

As expected, we found that the mentalization feature was represented across the mentalizing network for both movies, except precuneus, where the results were only significant for the Sherlock data (Figure 4B). All of them represented the social interaction feature during both movies and the touch feature for Summer, with the exception of precuneus.

We next examined whether the mentalizing network represented the feature capturing characters speaking about others. We hypothesized that listening to a protagonist speaking about another character would engage the mentalizing network as it may invite a viewer to make social judgments about others. However, a key difference between the mentalizing feature and this feature is that the mentalization feature only refers to the character speaking about another character's thoughts and feelings and does not include when talking about the character's appearance or bodily sensations, unlike the speaking about other features, which include all of these. We found that out of all the mentalizing brain regions, only TPJ consistently represented the feature capturing speaking about others in both movies (*P*_{corrected} = 0.01 for Summer and *P*_{corrected} < 0.05 for Sherlock). Surprisingly, TP represented the feature capturing the speaking about things

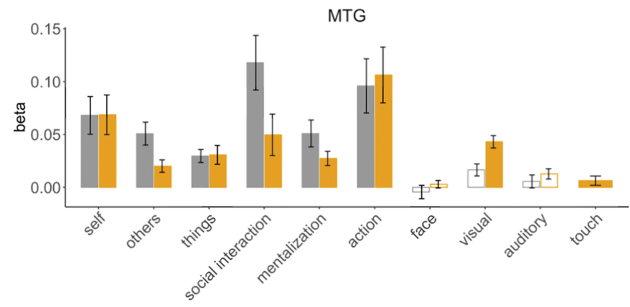
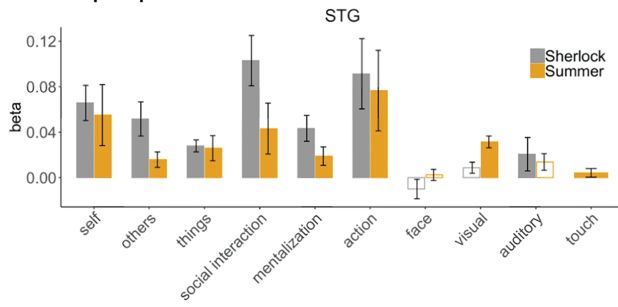
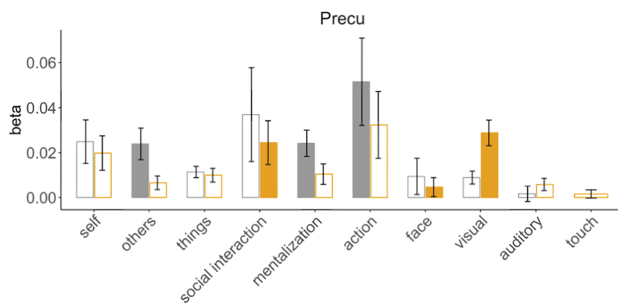
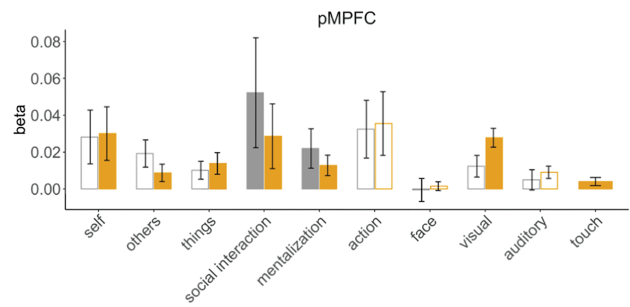
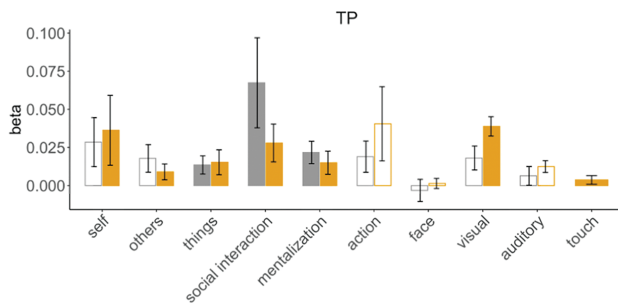
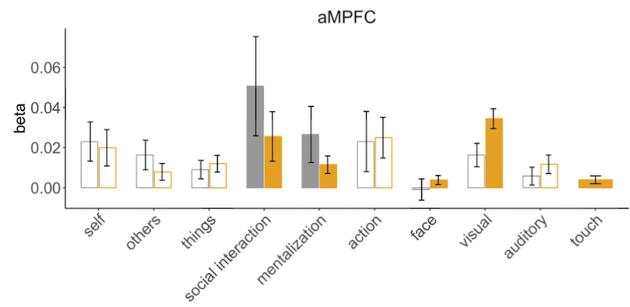
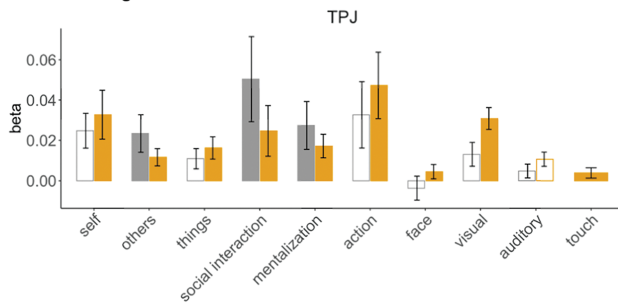
A. Social perception network**B. Mentalizing network**

Fig. 4. (A) Feature representations in ROIs from the social perception network STG and MTG. (B) Feature representations in the mentalizing network. The Y-axis displays the average beta-value across all participants from the multiple regression model for each feature. The bar graphs include error bars to demonstrate the standard deviation. Solid bars represent statistical significance ($P_{\text{corrected}} < 0.05$), unlike empty ones.

across both movies ($P_{\text{corrected}} < 0.05$ for both movies). No other regions in this network represented any speaking feature in a manner that was consistent across both movies.

The action feature is not represented in the frontoparietal action observation network

Surprisingly, the beta-values for observed action were not significantly above chance in either IPS or IFG (Figure 5). This is in contrast to STG and MTG which both significantly represent action information (Figure 4A). The social touch feature was represented in IPS for Summer ($P_{\text{corrected}} < 0.05$). Other social features were represented in IFG (e.g. social interaction for Sherlock), but the results were not consistent across two movies.

Visual and auditory features are represented in sensory regions

As described earlier, we included two sensory regions—the visual cortex encompassing V1 to V5 and the auditory cortex. Mentalization, action, face, and visual features extracted from the fifth layer of Alexnet were represented in the visual cortex across two movies (Figure 6). All speaking features and the amplitude of the audio were represented in the auditory cortex across two movies (Figure 6).

ROIs within the same network have similar representational structure

Lastly, to understand the representational relationship between ROIs, we performed a correlation analysis on all pairs of neural

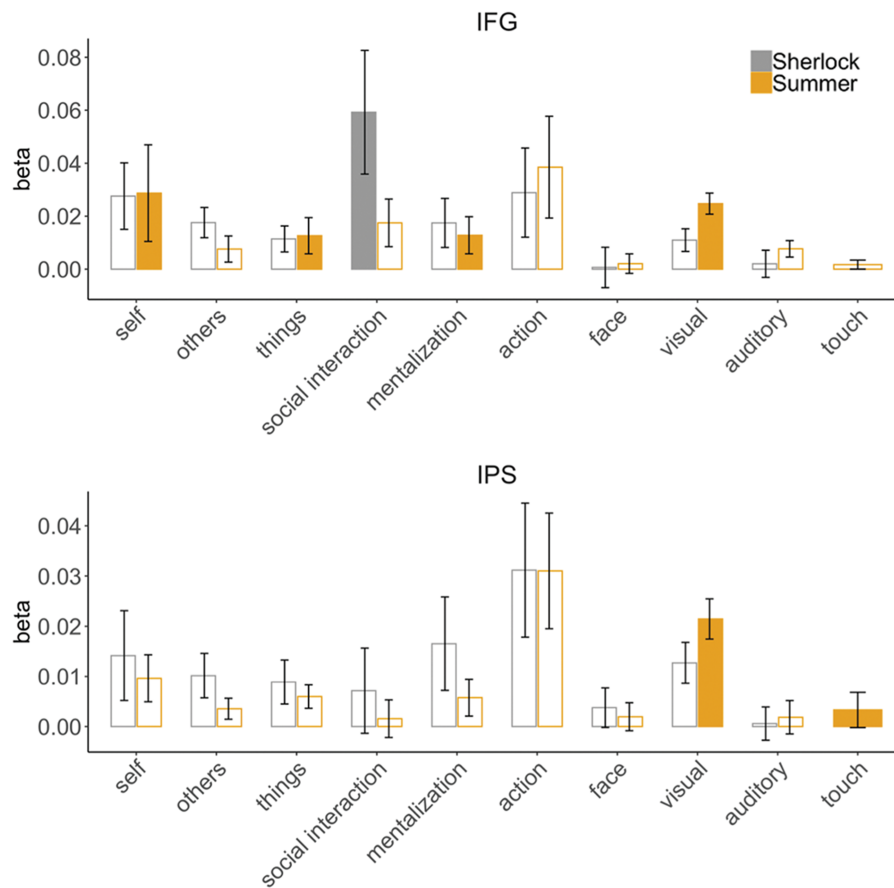


Fig. 5. Feature representations in the action observation network. Plotting conventions are the same as in Figure 4.

RDMs. Results are visualized with MDS plots (Figure 7). We found that ROIs within each brain network, except for the action observation network, showed similar representational structure. For example, r -value between STG and MTG within the social perception network was 0.83 and 0.87 for Sherlock and Summer, respectively (Supplemental Figure S1 including matrices with r -values). However, IFG and IPS within the action observation network did not show strong neural pattern similarity. Instead, IFG neural patterns were more strongly correlated with those of STG ($r = 0.44$ and 0.59) and MTG ($r = 0.54$ and 0.65) rather than the IPS ($r = 0.36$ and 0.45).

Discussion

We investigated the brain mechanisms underlying naturalistic social processing in a context resembling real-life situations by using publicly available two movie fMRI datasets. By densely annotating features and performing RSA (Figure 1), we identified the brain regions representing a broad range of social features and those representing mentalization during movie viewing (Figure 4). We, for the first time, showed how the brain represents three contents of speech, varying depending on the object being spoken about—self, others or things (Figure 4B). Lastly, comparing the representational structure of different brain regions, we identified that those in the same network had similar representations, with the exception of the action observation network (Figure 7). By analyzing two distinct sets of movie fMRI data, obtained from different participants and labs, we cross-confirmed

our results, thereby enhancing the validity of the current findings.

The STG and MTG serves as key regions for social processing during natural movie viewing

RSA and multiple regression methods revealed that the neural patterns in both the MTG and STG represented all of the selected social features in both movies (Figure 4A). This finding emphasizes the critical role of the human temporal cortex in the perception of social interaction, understanding others' actions and mental states, as well as processing different contents of communication varying depending on the objects being spoken about. We did not observe the face representation in these regions, which include both anterior and posterior parts. It may be that face selectivity in STS occurs at a posterior portion (Deen et al., 2015) and voxels in the anterior part may not be sensitive to the presence of face in the scene. Social interaction and observed action were the most prominent social features driving the configuration of neural patterns in both regions. Additionally, these two regions share highly similar representational structures with a neural similarity of >0.8 , though the strong correlation observed between the two regions may have been slightly overestimated due to their close anatomical proximity (Pillet et al., 2020). It is worth noting that both STG and MTG clusters in this study encompassed clusters in the STS. Despite using the anatomical template to define the STG, we found that the functionally selected voxels were mostly situated in the right STS (Supplemental Figure S2). A similar trend was

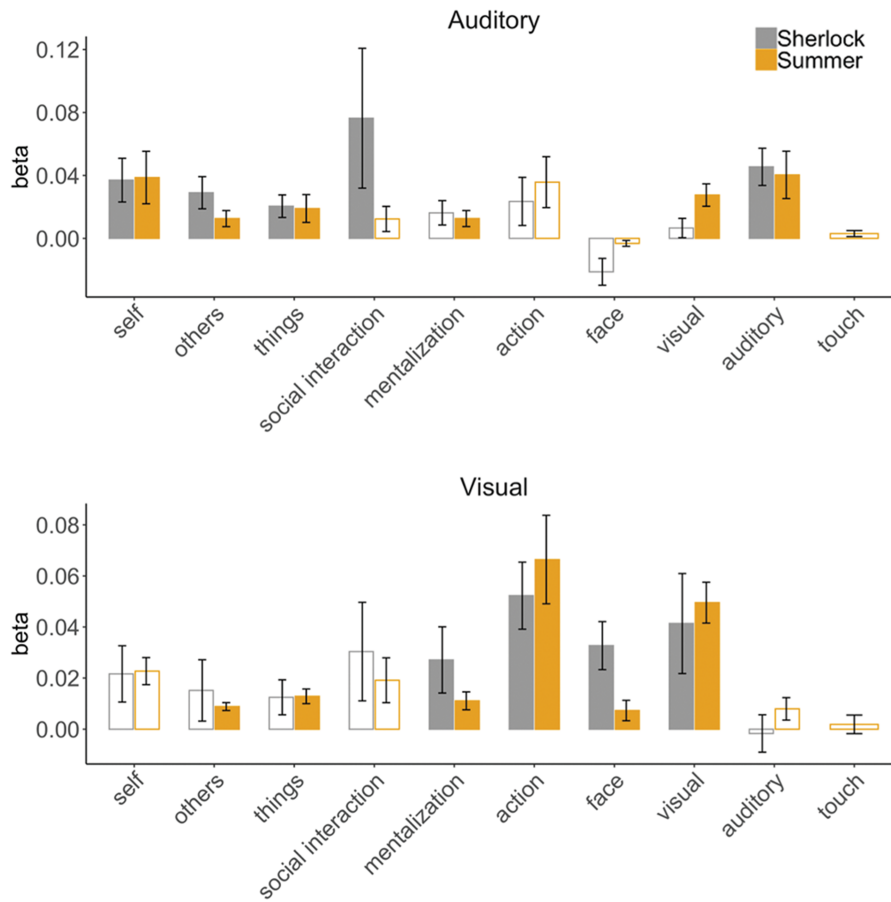
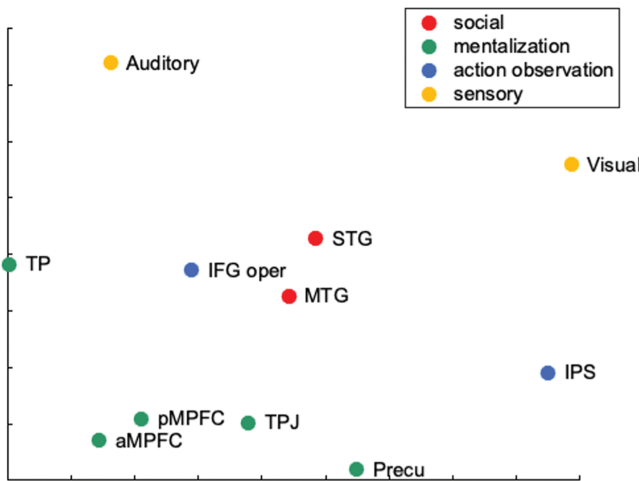


Fig. 6. Feature representations in the visual and auditory cortex. Plotting conventions are the same as in Figure 4.

A. Sherlock



B. Summer

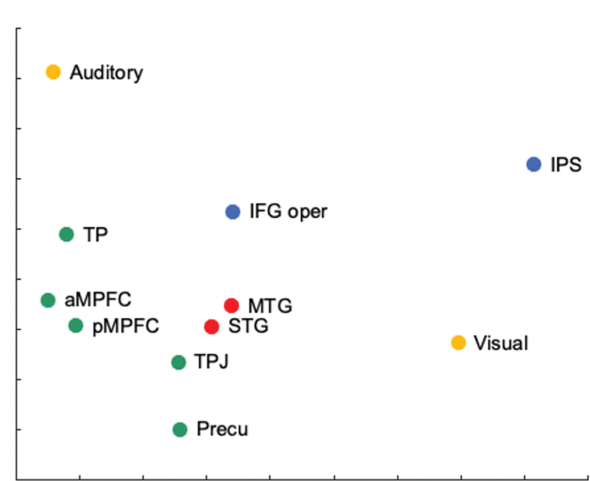


Fig. 7. A 2D MDS plot showing the neural pattern similarity between different ROIs for Sherlock (A) and Summer (B). The distance between them is based on the 2D projection of the correlation coefficient of their neural patterns. An ROI pair with stronger correlation are closer in proximity as compared to a pair with weaker correlation. The dots on the figure are color-coded, as illustrated on the plot legend, according to which network an ROI belongs to—social perception (STG and MTG in red), mentalization (TP, TPJ, aMPFC, pMPFC, Precu in green), action observation (IFG and IPS in blue) or sensory network (Visual and Auditory in yellow).

observed in the MTG, except that the clusters were considerably larger and located in both hemispheres, with distinct separation between anterior and posterior regions (Supplemental Figure S3). Our findings extend previous neuroimaging studies addressing

social functions within the temporal cortex. Importantly, the current study investigated novel social features that have not been previously examined, specifically, three contents of speech that vary based on the object being spoken about.

Increased STS responses have been linked to a broad range of social cognitive processes, including perceiving biological motion, goal-directed action, social interaction and social communication, extracting meaning from speech, mental state inference and social norm processing (Pelphrey et al., 2004; Herrington et al., 2011; Deen et al., 2015; de Heer et al., 2017; Isik et al., 2017; Pegado et al., 2018; Lee Masson and Isik, 2021; Landsiedel and Koldewyn, 2023; McMahon et al., 2023). Previous work has also found increased inter-subject brain synchronization in STS during movie viewing (Lahnakoski et al., 2014). Our findings suggest that the synchronized STS voxel responses may be explained by similar social feature representations across individuals. The STG and MTG have also been implicated in social cognitive processes, including social signal detection, the integration of verbal and non-verbal social cues, extracting social-affective meaning from observed touch, extracting meaning from speech and perceiving social communication and antisocial behavior (Price, 2012; Sugiura et al., 2014; Holler et al., 2015; Lee Masson et al., 2018; Santavirta et al., 2023). These regions show atypical neural responses to social stimuli in neurodiverse conditions, such as autism and schizophrenia (Zilbovicius et al., 2006; Redcay, 2008; Brent et al., 2014; Köchel et al., 2015; He et al., 2021). Functionally disrupting these regions results in short-term atypical social perception (Grossman et al., 2005; Akiyama et al., 2006; Saitovitch et al., 2016). While subparts of these regions have been shown to be domain-specific (Deen et al., 2015), our findings provide evidence that they respond to multiple social features. These regions may serve as key regions for social processing during natural movie viewing. Though the selected social features are somewhat correlated (Figure 2), it might be the case that STG/MTG process their shared variance rather than each individual social feature.

The mentalizing network represents the social interaction and mentalization feature

We observed that the neural patterns in TPJ, TP and mPFC within the mentalizing network represented the mentalization and social interaction feature (Figure 4B). Intriguingly, the social interaction feature drives the configuration of neural patterns in these regions the most strongly in both movies. Previous work with a movie viewing paradigm has also found increased activity in the mPFC when social interaction is present in the scene (Wagner et al., 2016). Our results expand on previous work by demonstrating that not only the mPFC but also the TPJ and TP show distinct neural patterns in response to scenes with social interaction and those without. Given that the presence of social interaction did not explain the unique variance in voxel-wise neural activity in these regions (Lee Masson and Isik, 2021), it is mostly likely that mentalization during the perception of social interaction might have influenced those findings. Previous behavioral work has suggested that when observing social interaction, individuals may naturally consider the mental states of others (Dziobek et al., 2006; Baksh et al., 2018; Grainger et al., 2019). Regarding mentalization, we replicated previous findings showing the involvement of TPJ, TP and mPFC during mentalization (Dufour et al., 2013; Schurz et al., 2014; Yang et al., 2015; Jacoby et al., 2016; Moessnang et al., 2020). Previous studies have mainly investigated the first-order mentalization of a movie viewer, based on the assumption that a certain scene prompts the viewer to infer a character's mental state (Jacoby et al., 2016; Richardson et al., 2018) or by using a retrospective behavioral sampling method to assess mentalizing components in the movie (Brandman et al., 2021). Future work may compare our operationalized

definition of theory of mind to the viewer's first-order mentalization as they view the movie to determine how the neural underpinnings of these two processes differ.

Our current findings do not entirely confirm our previous work on the same dataset, employing a different methodology, where we predicted the magnitude of voxel-wise neural responses (Lee Masson and Isik, 2021). In contrast to our prior work showing unique selectivity to social interaction in the precuneus across two movies, we did not observe its neural patterns representing social interaction in the Sherlock movie. This slight discrepancy may be due to the relatively small cluster identified in our previous work (the number of voxels = 25 in the Sherlock fMRI) compared to the precuneus region defined in the present study (the number of voxels = 221) or discrepancies between univariate and those from multivoxel pattern analyses (Pillet et al., 2020). This feature may solely account for the voxel-wise activity without explaining the underlying patterns of those voxels.

Lastly, while we replicated our findings regarding social interaction and mentalization features across two movies, we noticed a discrepancy in the remaining features. Specifically, we observed more social features that were significantly represented in these regions in the Summer fMRI data. Due to the vast differences between the two movies in terms of genre (crime vs romance), duration (40 vs 120 min) and the relationship between characters (colleagues vs romantic partners), it is challenging to pinpoint the exact factors that contributed to these discrepancies.

Processing different contents of spoken communication

The current findings on the representations of all three contents of speech in STG, MTG and the auditory cortex align well with the previous literature. The temporal cortex has been long implicated in speech comprehension (Crinion et al., 2003; Lindenberg and Scheef, 2007; Leonard and Chang, 2014), and listening to dialogs between people activates those areas (Landsiedel and Koldewyn, 2023; Santavirta et al., 2023). Intriguingly, we observed subtle differences in how each speaking feature was represented in the mentalization network. The neural patterns in TPJ exhibited sensitivity to the information about others, whereas the neural patterns in TP showed sensitivity to the information about things. This finding may suggest a differentiation in the neural processing of social vs non-social information when listening to others' conversations. The current study did not evaluate whether one region is statistically better at representing a specific content of speech compared to another region. Moreover, in the Summer movie, contents about others and things were represented in TPJ and TP without significant statistical differences. This might be related to many scenes with voice-over monologues where a narrator describes the characters and plots, potentially leading to changes in how speech contents are processed. Future work may explore the distinct neural processing of speech content in voice-over monologues vs social communicative dialogues.

Our finding shows that the TPJ is not only activated when we directly engage in inferring others' mental states during social interaction but also when we passively listen to someone else talking about others during movie viewing. As discussed earlier, the TPJ exhibits functional selectivity to the mentalizing feature, and its activation decreases during the processing of others' bodily states and physical pain (Jacoby et al., 2016; Richardson et al., 2018). The current study demonstrates that the TPJ responses are explained by various social features, including listening to speech that includes both mental and bodily states of others. Similar

to the representation of social interaction in the TPJ, this finding suggests spontaneous mentalizing that occurs when people hear others talk about someone else. When someone talks about another person, a listener may spontaneously make inferences about the person being discussed, even when the spoken content is not about mental states. For example, if someone says that person B got a promotion, the listener may infer that person B is feeling proud, and the speaker may be feeling jealous. In contrast, the anterior temporal lobe, which includes TP, plays a key role in semantic processing (Patterson et al., 2007; Gesierich et al., 2012; Visser et al., 2012). These studies have predominantly investigated the ATL responses to controlled visual stimuli. Our result extends the previous findings by demonstrating the neural representations of others' mental states and things in TP when listening to other people's natural conversations.

Action observation network

The frontoparietal regions in the action observation network do not appear to play much role in representing social features in a manner that is generalized across movies. Moreover, even the action feature was not consistently represented in this network. Prior work with controlled stimuli has suggested that representations in these regions may not generalize across scenes with different kinematic patterns or with variable visual information (Wurm and Lingnau, 2015), which may explain their lack of consistent response in natural movies that involve highly variable visual information. The lack of action representation within the action network may also be linked to the high number of speaking scenes present in the movie (~40%), which makes the action category rather simplistic, reducing it to essentially speaking vs other action categories. Unlike other social features, we chose detailed action labels because almost every scene has action components (>95%). Thus, annotating it as present vs absent would not be informative. One possibility for future work may be categorizing action into broader categories, such as personal care, eating, speaking, exercising and traveling. The frontoparietal part of the action observation network may not differentiate fine-grained action categories, such as brushing hair vs teeth, as it may only be sensitive to broader action categories, such as personal care vs speech.

Future research may use movies with fewer speaking scenes and more action variability to verify whether fine-detailed action categories are represented in this network during natural movie viewing.

Conclusion

Our study investigated the neural representations of various social features in natural movies, a setting closer to real-life social scenarios than typical experiments, including social interaction perception, listening to others' conversations, action observation and mentalization. Our findings highlight the temporal cortex as a key region for naturalistic social processing and suggest that different cognitive processes may come into play depending on whether a conversation concerns the speaker themselves, others or inanimate objects. Moreover, we found high similarities in the activity patterns of brain regions responsible for social perception and mentalization, suggesting a collaborative effort among these regions in natural settings to combine various social cues. Our study draws generalized conclusions from two distinct fMRI datasets, improving the reliability of the current findings. Future research may explore the temporal dimension of these social

processes, which would shed light on the order of computational steps unfolding within these brain networks.

Supplementary data

Supplementary data is available at SCAN online.

Data availability

fMRI data are available from the original authors for Sherlock (<https://dataspace.princeton.edu/handle/88435/dsp01nz8062179>) and Summer (<https://openneuro.org/datasets/ds002837/>, <https://www.naturalistic-neuroimaging-database.org/index.html>). Annotations are available at <https://osf.io/98rfv/>. All code for RSA is available at <https://github.com/lchang31/MultipleRegression>.

Conflict of interest

The authors declared that they had no conflict of interest with respect to their authorship or the publication of this article.

Acknowledgements

This work was supported with funds from The Clare Boothe Luce Program for Women in STEM.

References

- Akiyama, T., Kato, M., Muramatsu, T., Saito, F., Umeda, S., Kashima, H. (2006) Gaze but not arrows: a dissociative impairment after right superior temporal gyrus damage. *Neuropsychologia*, **44**, 1804–10.
- Aliko, S., Huang, J., Gheorghiu, F., Meliss, S., Skipper, J.I. (2020) A naturalistic neuroimaging database for understanding the brain using ecological stimuli. *Scientific Data*, **7**, 1–21.
- Ames, D.L., Fiske, S.T., Todorov, A.T. (2011) Impression formation: a focus on others' intents. *The Oxford handbook of social neuroscience*, 419.
- Baksh, R.A., Abrahams, S., Auyeung, B., MacPherson, S.E., van den Bos, R. (2018) The Edinburgh Social Cognition Test (ESCoT): examining the effects of age on a new measure of theory of mind and social norm understanding. *PLoS One*, **13**, e0195818.
- Belsley, D.A. (1991) A Guide to using the collinearity diagnostics. *Computer Science in Economics and Management*, **4**, 33–50.
- Brandman, T., Malach, R., Simony, E. (2021) The surprising role of the default mode network in naturalistic perception. *Communications Biology*, **4**, 79.
- Brent, B.K., Seidman, L.J., Coombs, G., Keshavan, M.S., Moran, J.M., Holt, D.J. (2014) Neural responses during social reflection in relatives of schizophrenia patients: relationship to subclinical delusions. *Schizophrenia Research*, **157**, 292–8.
- Caspers, S., Zilles, K., Laird, A.R., Eickhoff, S.B. (2010) ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, **50**, 1148–67.
- Chen, J., Leong, Y.C., Honey, C.J., Yong, C.H., Norman, K.A., Hasson, U. (2017) Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, **20**, 115–25.
- Crinion, J.T., Lambon-Ralph, M.A., Warburton, E.A., Howard, D., Wise, R.J.S. (2003) Temporal lobe regions engaged during normal speech comprehension. *Brain*, **126**, 1193–201.
- de Heer, W.A., Huth, A.G., Griffiths, T.L., Gallant, J.L., Theunissen, F.E. (2017) The hierarchical cortical organization of human speech processing. *Journal of Neuroscience*, **37**, 6539–57.

- Deen, B., Koldewyn, K., Kanwisher, N., Saxe, R. (2015) Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex*, **25**, 4596–609.
- Dufour, N., Redcay, E., Young, L., et al. (2013) Similar brain activation during false belief tasks in a large sample of adults with and without autism. *PLoS One*, **8**, e75468.
- Dziobek, I., Fleck, S., Kalbe, E., et al. (2006) Introducing MASC: a movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, **36**, 623–36.
- Eickhoff, S.B., Stephan, K.E., Mohlberg, H., et al. (2005) A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, **25**, 1325–35.
- Gallagher, H.L., Frith, C.D. (2003) Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, **7**, 77–83.
- Gesierich, B., Jovicich, J., Riello, M., et al. (2012) Distinct neural substrates for semantic knowledge and naming in the temporoparietal network. *Cerebral Cortex*, **22**, 2217–26.
- Grainger, S.A., Steinvik, H.R., Henry, J.D., Phillips, L.H. (2019) The role of social attention in older adults' ability to interpret naturalistic social scenes. *Quarterly Journal of Experimental Psychology*, **72**, 1328–43.
- Grossman, E.D., Battelli, L., Pascual-Leone, A. (2005) Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research*, **45**, 2847–53.
- Grossmann, T., Johnson, M.H. (2007) The development of the social brain in human infancy. *European Journal of Neuroscience*, **25**, 909–19.
- Hamlin, J.K., Wynn, K. (2011) Young infants prefer prosocial to antisocial others. *Cognitive Development*, **26**, 30–9.
- Hamlin, J.K., Wynn, K., Bloom, P. (2007) Social evaluation by preverbal infants. *Nature*, **450**, 557–9.
- Herrington, J.D., Nymberg, C., Schultz, R.T. (2011) Biological motion task performance predicts superior temporal sulcus activity. *Brain and Cognition*, **77**, 372–81.
- He, Y., Steines, M., Sammer, G., Nagels, A., Kircher, T., Straube, B. (2021) Modality-specific dysfunctional neural processing of social-abstract and non-social-concrete information in schizophrenia. *NeuroImage: Clinical*, **29**, 102568.
- Holler, J., Kokal, I., Toni, I., Hagoort, P., Kelly, S.D., Özyürek, A. (2015) Eye'm talking to you: speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Social Cognitive & Affective Neuroscience*, **10**, 255–61.
- 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 of the United States of America*, **114**, E9145–52.
- Jacoby, N., Bruneau, E., Koster-Hale, J., Saxe, R. (2016) Localizing pain matrix and theory of mind networks with both verbal and non-verbal stimuli. *NeuroImage*, **126**, 39–48.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M. (2012) FSL. *NeuroImage*, **62**, 782–90.
- Kilner, J.M. (2011) More than one pathway to action understanding. *Trends in Cognitive Sciences*, **15**, 352–7.
- Köchel, A., Schöngaßner, F., Feierl-Gsodam, S., Schienle, A. (2015) Processing of affective prosody in boys suffering from attention deficit hyperactivity disorder: a near-infrared spectroscopy study. *Social Neuroscience*, **10**, 583–91.
- Kriegeskorte, N., Mur, M., Bandettini, P.A. (2008) Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, **2**, 249.
- Lahnakoski, J.M., Glerean, E., Jääskeläinen, I.P., et al. (2014) Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage*, **100**, 316–24.
- Landsiedel, J., Koldewyn, K. (2023) Auditory dyadic interactions through the “eye” of the social brain: How visual is the posterior STS interaction region?. *Imaging Neuroscience*, **1**, 1–20.
- Lee Masson, H., Isik, L. (2021) Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. *NeuroImage*, **245**, 118741.
- Lee Masson, H., Van De Plas, S., Daniels, N., de Beeck H, O. (2018) The multidimensional representational space of observed socio-affective touch experiences. *NeuroImage*, **175**, 297–314.
- Leonard, M.K., Chang, E.F. (2014) Dynamic speech representations in the human temporal lobe. *Trends in Cognitive Sciences*, **18**, 472–9.
- Lindenberg, R., Scheef, L. (2007) Supramodal language comprehension: role of the left temporal lobe for listening and reading. *Neuropsychologia*, **45**, 2407–15.
- Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H. (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, **19**, 1233–9.
- Marquardt, D.W. (1970) Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, **12**, 591–612.
- Mars, R.B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., Rushworth, M.F.S. (2012) Connectivity-based subdivisions of the human right 'temporoparietal junction area': evidence for different areas participating in different cortical networks. *Cerebral Cortex*, **22**, 1894–903.
- McMahon, E., Bonner, M.F., Isik, L. (2023) Hierarchical organization of social action features along the lateral visual pathway. *Current Biology*, **33**(23): 5035–47.
- Mehl, M.R., Gosling, S.D., Pennebaker, J.W. (2006) Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, **90**, 862–77.
- Moessnang, C., Baumeister, S., Tillmann, J., et al. (2020) Social brain activation during mentalizing in a large autism cohort: The Longitudinal European Autism Project. *Molecular Autism*, **11**, 17.
- Nichols, T.E., Holmes, A.P. (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, **15**, 1–25.
- Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N. (2014) A toolbox for representational similarity analysis. *PLoS Computational Biology*, **10**, e1003553.
- Oosterhof, N.N., Connolly, A.C., Haxby, J.V. (2016) CoSMoMVA: multimodal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, **10**, 27.
- Patterson, K., Nestor, P.J., Rogers, T.T. (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews, Neuroscience*, **8**, 976–87.
- Pegado, F., Hendriks, M.H.A., Amelynck, S., et al. (2018) A multitude of neural representations behind multisensory 'Social Norm' processing. *Frontiers in Human Neuroscience*, **12**, 153.
- Pelphrey, K.A., Morris, J.P., McCarthy, G. (2004) Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, **16**, 1706–16.
- Pillet, I., Op de Beeck, H., Lee Masson, H. (2020) A comparison of functional networks derived from representational similarity, functional connectivity, and univariate analyses. *Frontiers in Neuroscience*, **13**, 464352.
- Premack, D., Woodruff, G. (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, **1**, 515–26.
- Price, C.J. (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, **62**, 816–47.

- Quadflieg, S., Koldewyn, K. (2017) The neuroscience of people watching: how the human brain makes sense of other people's encounters. *Annals of the New York Academy of Sciences*, **1396**, 166–82.
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Redcay, E. (2008) The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neuroscience and Biobehavioral Reviews*, **32**, 123–42.
- Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., Saxe, R. (2018) Development of the social brain from age three to twelve years. *Nature Communications*, **9**, 1027.
- Ritchie, J.B., Bracci, S., de Beeck H, O. (2017) Avoiding illusory effects in representational similarity analysis: what (not) to do with the diagonal. *NeuroImage*, **148**, 197–200.
- Saitovitch, A., Popa, T., Lemaitre, H., et al. (2016) Tuning eye-gaze perception by transitory STS inhibition. *Cerebral Cortex*, **26**, 2823–31.
- Sallet, J., Mars, R.B., Noonan, M.P., et al. (2013) The organization of dorsal frontal cortex in humans and macaques. *Journal of Neuroscience*, **33**, 12255–74.
- Santavirta, S., Karjalainen, T., Nazari-Farsani, S., et al. (2023) Functional organization of social perception networks in the human brain. *NeuroImage*, **272**, 120025.
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J. (2014) Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, **42**, 9–34.
- Sugiura, M., Yomogida, Y., Mano, Y., et al. (2014) From social-signal detection to higher social cognition: an fMRI approach. *Social Cognitive & Affective Neuroscience*, **9**, 1303–9.
- Tholen, M.G., Trautwein, F.M., Böckler, A., Singer, T., Kanske, P. (2020) Functional magnetic resonance imaging (fMRI) item analysis of empathy and theory of mind. *Human Brain Mapping*, **41**, 2611–28.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., et al. (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, **15**, 273–89.
- Van Overwalle, F., Baetens, K. (2009) Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage*, **48**, 564–84.
- Visser, M., Jefferies, E., Embleton, K.V., Lambon Ralph, M.A. (2012) Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *Journal of Cognitive Neuroscience*, **24**, 1766–78.
- Wagner, D.D., Kelley, W.M., Haxby, J.V., Heatherton, T.F. (2016) The dorsal medial prefrontal cortex responds preferentially to social interactions during natural viewing. *Journal of Neuroscience*, **36**, 6917–25.
- Walbrin, J., Downing, P., Koldewyn, K. (2018) Neural responses to visually observed social interactions. *Neuropsychologia*, **112**, 31–9.
- Walbrin, J., Koldewyn, K. (2019) Dyadic interaction processing in the posterior temporal cortex. *NeuroImage*, **198**, 296–302.
- Wurm, M.F., Lingnau, A. (2015) Decoding actions at different levels of abstraction. *Journal of Neuroscience*, **35**, 7727–35.
- Xia, M., Wang, J., He, Y., Csermely, P. (2013) BrainNet Viewer: a network visualization tool for human brain connectomics. *PLoS One*, **8**, e68910.
- Yang, D.Y.-J., Rosenblau, G., Keifer, C., Pelphrey, K.A. (2015) An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience and Biobehavioral Reviews*, **51**, 263–75.
- Zilbovicius, M., Meresse, I., Chabane, N., Brunelle, F., Samson, Y., Boddaert, N. (2006) Autism, the superior temporal sulcus and social perception. *Trends in Neurosciences*, **29**, 359–66.