



# Functional organization of social perception networks in the human brain

Severi Santavirta<sup>a,b,\*</sup>, Tomi Karjalainen<sup>a,b</sup>, Sanaz Nazari-Farsani<sup>a,b</sup>, Matthew Hudson<sup>a,b,c</sup>, Vesa Putkinen<sup>a,b</sup>, Kerttu Seppälä<sup>a,b,d</sup>, Lihua Sun<sup>a,b,e</sup>, Enrico Glerean<sup>f</sup>, Jussi Hirvonen<sup>g,h</sup>, Henry K. Karlsson<sup>a,b</sup>, Lauri Nummenmaa<sup>a,b,i</sup>

<sup>a</sup> Turku PET Centre, University of Turku, Turku, Finland

<sup>b</sup> Turku PET Centre, Turku University Hospital, Turku, Finland

<sup>c</sup> School of Psychology, University of Plymouth, Plymouth, United Kingdom

<sup>d</sup> Department of Medical Physics, Turku University Hospital, Turku, Finland

<sup>e</sup> Department of Nuclear Medicine, Pudong Hospital, Fudan University, Shanghai, China

<sup>f</sup> Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo, Finland

<sup>g</sup> Department of Radiology, University of Turku and Turku University Hospital, Turku, Finland

<sup>h</sup> Medical Imaging Center, Department of Radiology, Tampere University and Tampere University Hospital, Tampere, Finland

<sup>i</sup> Department of Psychology, University of Turku, Turku, Finland

## ARTICLE INFO

### Keywords:

Social perception

fMRI

Multivariate pattern analysis

Intersubject correlation

## ABSTRACT

Humans rapidly extract diverse and complex information from ongoing social interactions, but the perceptual and neural organization of the different aspects of social perception remains unresolved. We showed short movie clips with rich social content to 97 healthy participants while their haemodynamic brain activity was measured with fMRI. The clips were annotated moment-to-moment for a large set of social features and 45 of the features were evaluated reliably between annotators. Cluster analysis of the social features revealed that 13 dimensions were sufficient for describing the social perceptual space. Three different analysis methods were used to map the social perceptual processes in the human brain. Regression analysis mapped regional neural response profiles for different social dimensions. Multivariate pattern analysis then established the spatial specificity of the responses and intersubject correlation analysis connected social perceptual processing with neural synchronization. The results revealed a gradient in the processing of social information in the brain. Posterior temporal and occipital regions were broadly tuned to most social dimensions and the classifier revealed that these responses showed spatial specificity for social dimensions; in contrast Heschl gyri and parietal areas were also broadly associated with different social signals, yet the spatial patterns of responses did not differentiate social dimensions. Frontal and subcortical regions responded only to a limited number of social dimensions and the spatial response patterns did not differentiate social dimension. Altogether these results highlight the distributed nature of social processing in the brain.

## 1. Introduction

Humans live in a complex and ever-changing social world, but how do we make sense of the high-dimensional and time-variable information constantly conveyed by our conspecifics? Prior functional imaging studies have localized specific aspects of social perception into different brain regions (Brooks et al., 2020). Fusiform gyrus (FG) is consistently involved in the perception of faces (Haxby et al., 2000) and lateral occipitotemporal cortex (LOT) in the perception of bodies (Downing et al., 2001). Temporoparietal junction (TPJ) is in turn involved in reflecting the mental states of others (Saxe and Kanwisher, 2003) as well as in processing of social context and in focusing attention (Carter and Huettel, 2013). Polysensory areas in the superior temporal sulcus (STS)

have been associated with multiple higher-order aspects of social perception (Deen et al., 2015; Isik et al., 2017; Lahnakoski et al., 2012; Nummenmaa and Calder, 2009), while medial frontal cortex (MFC) has been extensively studied in the context of self-representation and theory of mind (Amodio and Frith, 2006). Finally, speech-based social communication is accomplished by a network consisting of superior temporal gyrus (STG) and its proximal areas STS (Wernicke area in left pSTS), TPJ, angular gyrus, middle temporal gyrus (MTG) and inferior frontal gyrus (Broca's area in the left IFG) (Price, 2012).

Humans can however reliably process numerous simultaneously occurring features of the social world ranging from others' facial identities and emotions to their intentions and mental contents to the fine-grained affective qualities of the social interaction. Given the computational

\* Corresponding author: Severi Santavirta, Turku PET Centre c/o Turku University, Kiinamyllynkatu 4-6, 20520 Turku, Finland.

E-mail address: [svtsan@utu.fi](mailto:svtsan@utu.fi) (S. Santavirta).

limits of the human brain, it is unlikely that all features and dimensions of the social domain are processed by distinct areas and systems (Huth et al., 2012). Although the brain basis of perceiving specific isolated social features has been successfully delineated, the phenomenological as well as neural organization of the different social perceptual processes have remained poorly understood and neural responses to complex stimuli cannot necessarily be predicted on statistical combination of responses to simple stimuli (Felsen and Dan, 2005). Therefore studies based on neural responses to isolated social features may not generalize to real-world social perception (Adolphs et al., 2016) where social features such as facial identities, body movements, and nonverbal communication often overlap with distinct temporal occurrence patterns.

In psychological domains including actions (Huth et al., 2012), language (Huth et al., 2016), and emotions (Koide-Majima et al., 2020), neuroimaging studies have tackled this issue by first generating a comprehensive set of modelled dimensions for the complex dynamic stimulus. Then, using dimension reduction techniques, they have assessed the representational similarities of the modelled dimensions, or the representational similarities of the brain activation patterns associated with each dimension. For example, a recent study found that linguistic and visual semantic representations converge so that visual representations locate on the border of occipital cortex while linguistic representations are located anterior to the visual representations (Popham et al., 2021). However, a detailed representational space for social features at both perceptual and neural level is currently lacking.

Commonly applied univariate analyses modelling the BOLD response in each voxel or region separately cannot reveal the specificity of spatial brain activation patterns resulting from the perception of different social features. Consequently, they do not allow testing whether different perceptual features can be reliably discriminated based on their spatial brain activation patterns. Multivariate pattern analysis (MVPA) allows the analysis of information carried by fine-grained spatial patterns of brain activation (Tong and Pratte, 2012). Pattern recognition studies have established that regional multivariate patterns allow distinguishing brain activation related to multiple high-level social features such as faces (Haxby et al., 2001) and their racial group (Brosch et al., 2013) in FG and facial expressions in FG and STS (Harry et al., 2013; Said et al., 2010; Wegryn et al., 2015). Perception of different goal-orientated motor actions with different levels of abstraction can be decoded in LOTC and in inferior parietal lobe, suggesting that these regions process the abstract concepts of the goal-orientated actions, not just their low-level visual properties (Wurm and Lingnau, 2015). Furthermore, decoding of goal-orientated actions was successful in LOTC when subjects observed the actions in both first and third person perspectives (Oosterhof et al., 2012). It however remains unresolved how specific these regional response profiles are across different social perceptual features. The perception and interpretation of sensory social information is vital for planning social interaction in everyday life of people, and neuroimaging studies have also highlighted the centrality of social information in the brain function (Hari et al., 2015). Accordingly, it is important to establish how sensory social information is organized at the phenomenological and neural levels.

We define social perception as perception of all possible information relevant to interpret social interaction. To our knowledge, there is currently no consensus on a combined taxonomy for this broad definition. In social psychology, social situation has been described as a triad of person, situation and consequent behaviour (Lewin, 1936) where these elements have close interact between each other (Funder, 2006). However, data-driven taxonomies have only been proposed for the elements separately. Person perception has been extensively studied and person characteristics can be categorised as a limited set of trait dimensions, such as Big Five (Goldberg, 1990) or Big Six (Lee and Ashton, 2004). For psychological situations, data-driven lexical studies have proposed limited dimensionality (Parrigon et al., 2017; Rauthmann et al., 2014). Recently, in behavioural domain categorization of human actions have

also been proposed (Thornton and Tamir, 2022). For two reasons, these established taxonomies are suboptimal for studying social perception as whole. First, these taxonomical studies base their results on questionnaires regarding social situations or rated similarities of different words describing social situations instead of the actual perception of social situations in real-life dynamic environment. Second, since the three elements (person, situation, and behaviour) are intimately linked, it would be sensible to study them together. Therefore, we selected a large set of features from the person, situation, and behaviour domains, collected perceptual ratings for these features from the stimulus used in this neuroimaging study and then limited the social perceptual space of the stimulus with clustering analysis.

### 1.1. The current study

In this fMRI study, we mapped the perceptual and neural representations of naturalistic social episodes using both univariate and multivariate analyses (Fig. 1). We used short movie clips as stimuli because cinema contains rich and complex social scenarios and as it also elicits strong and consistent neural responses in functional imaging studies (Hasson et al., 2010; Lahnakoski et al., 2012). We first aimed at establishing a perception-based taxonomy of the social dimensions that human observers use for describing social scenarios, and then mapped the brain basis of this social perceptual space. We mapped the perceptual space of social processes based on subjective annotations of a large array of social features ( $n = 112$ ) in the movies ( $n = 96$ ). We then used dimension reduction techniques to establish the representational space of social perception, and to reduce the multidimensional space into a limited set of reliable perceptual dimensions of social features. Using a combination of univariate regression analysis and multivariate pattern analysis we established that posterior temporal and occipital regions are the main hubs for social perception and that brain shows a gradient in social perceptual processing from broadly tuned but spatially dimension-specific responses in posterior temporal and occipital regions towards more selective responses in frontal and subcortical areas.

## 2. Materials and methods

### 2.1. Participants

Altogether 102 volunteers participated in the study. The exclusion criteria included a history of neurological or psychiatric disorders, alcohol or substance abuse, BMI under 20 or over 30, current use of medication affecting the central nervous system and the standard MRI exclusion criteria. Two additional subjects were scanned but excluded from further analyses because unusable MRI data due to gradient coil malfunction. Two subjects were excluded because of anatomical abnormalities in structural MRI and additional three subjects were excluded due to visible motion artefacts in preprocessed functional neuroimaging data. This yielded a final sample of 97 subjects (50 females, mean age of 31 years, range 20 – 57 years). All subjects gave an informed, written consent and were compensated for their participation. The ethics board of the Hospital District of Southwest Finland had approved the protocol and the study was conducted in accordance with the Declaration of Helsinki.

### 2.2. Neuroimaging data acquisition and preprocessing

MR imaging was conducted at Turku PET Centre. The MRI data were acquired using a Phillips Ingenuity TF PET/MR 3-T whole-body scanner. High-resolution structural images were obtained with a T1-weighted (T1w) sequence (1 mm<sup>3</sup> resolution, TR 9.8 ms, TE 4.6 ms, flip angle 7°, 250 mm FOV, 256 × 256 reconstruction matrix). A total of 467 functional volumes were acquired for the experiment with a T2\*-weighted echo-planar imaging sequence sensitive to the blood-oxygen-level-dependent (BOLD) signal contrast (TR 2600 ms, TE 30 ms, 75° flip angle, 240 mm FOV, 80 × 80 reconstruction matrix, 62.5 kHz

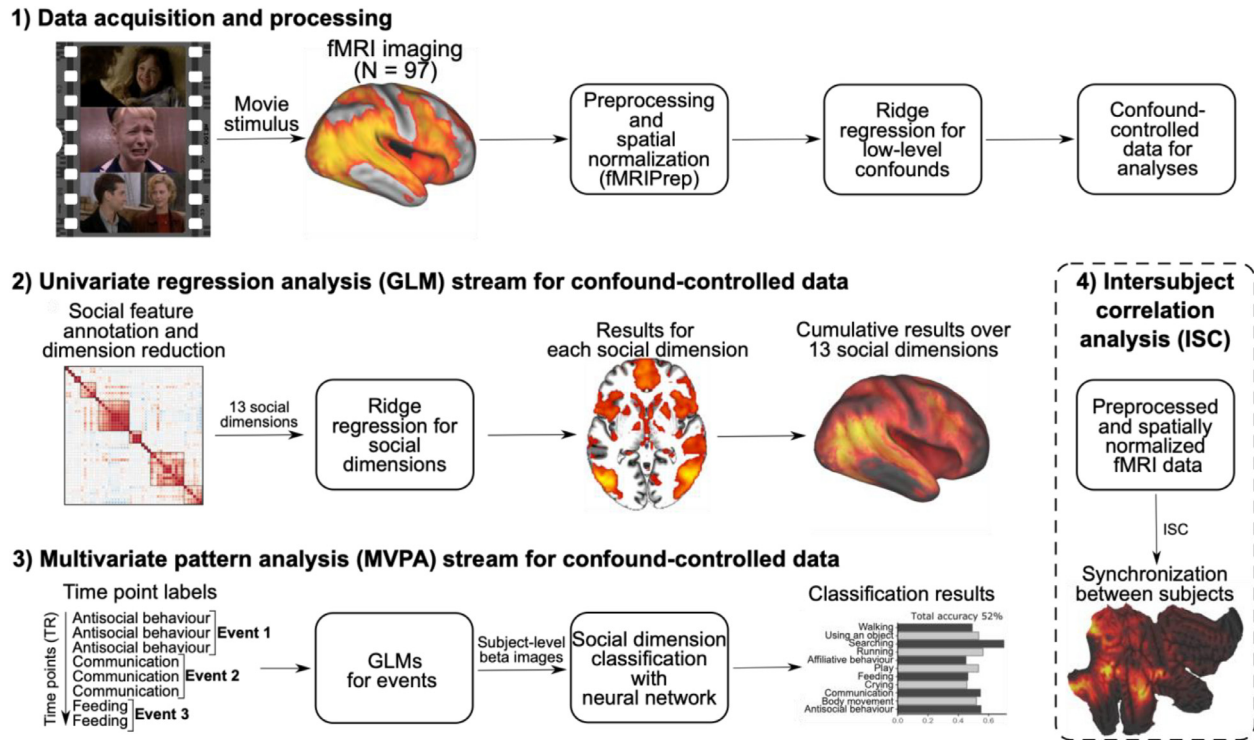


Fig. 1. Analysis stream from data acquisition and processing to the univariate regression analysis, multivariate pattern analysis and intersubject correlation analysis.

bandwidth, 3.0 mm slice thickness, 45 interleaved axial slices acquired in ascending order without gaps).

The functional imaging data were preprocessed with fMRIprep (Esteban et al., 2019) (v1.3.0), a Nipype (Gorgolewski et al., 2011) based tool that internally uses Nilearn (Abraham et al., 2014). During the preprocessing, each T1w volume was corrected for intensity non-uniformity using N4BiasFieldCorrection (v2.1.0) (Tustison et al., 2010) and skull-stripped using antsBrainExtraction.sh (v2.1.0) using the OASIS template. Brain surfaces were reconstructed using recon-all from FreeSurfer (v6.0.1) (Dale et al., 1999), and the brain masque estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical grey-matter of Mindboggle (Klein et al., 2017). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009) was performed through nonlinear registration with the antsRegistration (ANTs v2.1.0) (Avants et al., 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid, white-matter and grey-matter was performed on the brain-extracted T1w image using FAST (Zhang et al., 2001) (FSL v5.0.9).

Functional data were slice-time-corrected using 3dTshift from AFNI (Cox, 1996) (v16.2.07) and motion-corrected using MCFLIRT (Jenkinson et al., 2002) (FSL v5.0.9). These steps were followed by co-registration to the T1w image using boundary-based registration (Greve and Fischl, 2009) with six degrees of freedom, using bbrgister (FreeSurfer v6.0.1). The transformations from motion-correction, coregistration, and spatial normalization were concatenated and applied in a single step using antsApplyTransforms (ANTs v2.1.0) using Lanczos interpolation. Independent-component-analysis-based Automatic Removal Of Motion Artifacts (ICA-AROMA) was used to denoise the data nonaggressively after spatial smoothing with 6-mm Gaussian kernel (Pruim et al., 2015). The data were then detrended using 240-s-Savitzky-Golay filtering to remove the scanner drift (Cukur et al., 2013), and finally downsampled to original 3 mm isotropic voxel size. The BOLD signals were demeaned to make the regression coefficients

comparable across different individuals (Chen et al., 2017). First and last two functional volumes were discarded to ensure equilibrium effects and to exclude the time points before and after the stimulus.

### 2.3. Stimulus

To map brain responses to different social features, we used our previously validated socioemotional “localizer” paradigm that allows reliable mapping of various social and emotional functions (Karjalainen et al., 2017, 2019; Lahnakoski et al., 2012; Nummenmaa et al., 2021). The experimental design and stimulus selection has been described in detail in the original study with this setup (Lahnakoski et al., 2012). Briefly, the subjects viewed a medley of 96 movie clips (median duration 11.2 s, range 5.3 – 28.2 s, total duration 19 min 44 s) that have been curated to contain large variability of social and emotional content. The videos were extracted from mainstream Hollywood movies with audio track in English. To limit experiment duration, a subset of 87 of the previously validated 137 clips were selected. 71 of these clips contained people in various social situations and contexts (one person: 15, two people: 22, more than two people: 34). To distinguish person perception from other audiovisual perception, the stimulus contained four clips with animals and 12 control clips without people (showing e.g. scenery and objects). Additionally, nine erotic scenes showing heterosexual intercourse were added to broaden the emotional content of the original stimuli. Short descriptions about movie clips can be found from Table SI-1. Because this task was designed to map neural processing of naturalistic socioemotional events, the clips were not deliberately matched with respect to, for example, human motion or optic flow. The videos were presented in fixed order across the subjects without breaks to allow the brain synchronization analyses between subjects (see section 2.11). Subjects were instructed to view the movies similarly as if they were viewing a movie at a cinema or at home and no specific task was assigned. Visual stimuli were presented with NordicNeuroLab VisualSystem binocular display. Sound was conveyed with Sensimetrics S14 insert earphones. Stimulation was

controlled with Presentation software. Before the functional run, sound intensity was adjusted for each subject so that it could be heard over the gradient noise.

#### 2.4. Stimulus features

We collected ratings for 112 predefined social features (see **Table SI-2**) from the movie clips. We selected a broad range of socioemotional features describing persons, social situations and behaviours from following categories: sensory input (e.g. smelling, tasting), basic bodily functions (e.g. facial expressions, walking, eating), person characteristics (e.g. pleasantness, trustworthiness) and person's inner states (e.g. pleasant feelings, arousal), social interaction signals (e.g. talking, communicating with gestures) and social interaction characteristics (e.g. hostility, sexuality). Collecting perceptual ratings from a large set of individual social features enables reliable mapping of the whole social perceptual space that can be derived from the stimulus movie clips and ensures that the data-driven dimensionality arises from the used stimulus. It was stressed to the observers that they should rate the perceived features of the social interaction rather than the observer's own inner states (such as emotions evoked by the movies). The ratings were collected separately for each video clip in short time intervals (median 4.0 s, range: 3.1 – 7.3 s). Features were annotated in a continuous and abstract scale from “absent” to “extremely much”. For analyses the ratings were transformed to continuous scale from 0 (absent) to 100 (extremely much). Annotators watched the video clips altogether 12 times, rating an average of 10 features on each viewing to reduce the cognitive load. The ratings were done using an online rating platform Onni (<http://onni.utu.fi>) developed at Turku PET Centre (Heikkilä et al., 2020).

#### 2.5. Feature reliability

We first evaluated whether the *a priori* features were frequently and consistently perceived in the stimulus movies. Features with low occurrence rate and/or inter-rater reliability were excluded, because i) high occurrence rate is needed to reliably estimate the stimulus-dependant variation in BOLD signal, and ii) high inter-rater reliability is necessary to study brain activity in a sample of participants independent from the raters. The occurrence rate was defined as the number of time points where the mean rating (minus standard error of the mean) over the annotators exceeded 5 (on a scale ranging from 0 to 100, 5% of the slider length). Features were included in the analyses if they occurred at least five times throughout the experiment; this was estimated to yield sufficient statistical power in the BOLD-fMRI GLM analyses. Inter-rater reliability of the features was assessed using intra-class correlation coefficient (ICC) as calculated in the R package psych (<https://cran.r-project.org/package=psych>). ICC(A,1) was selected as appropriate model for ICC since it treats both video clips and raters as random effects and measures the absolute agreement between raters (McGraw and Wong, 1996). ICCs below 0.5 are considered poor (Koo and Li, 2016), thus we only included features with ICC over 0.5. A total of 45 features satisfied both criteria. The occurrence rate and inter-rater reliability of each feature are shown in **Figure SI-1**.

#### 2.6. Reliability check for the social feature annotations

Dynamically rating 112 social perceptual features from 96 movie clips is extremely laborious (28 672 individual rating decisions). The reliability of online data collecting platforms have also been questioned (Webb and Tangney, 2022), thus we chose to recruit local subjects to the laboratory for doing the annotations. The decision ensured that the neuroimaging subjects and annotators belong to the same population and that we could ensure that the annotators adhered to the instructions. Moreover, the fMRI analysis only focused on the social features that were perceived most coherently (based on ICC) between the raters. Prior

high-dimensional annotation studies have also used similar annotator pools (Huth et al., 2012).

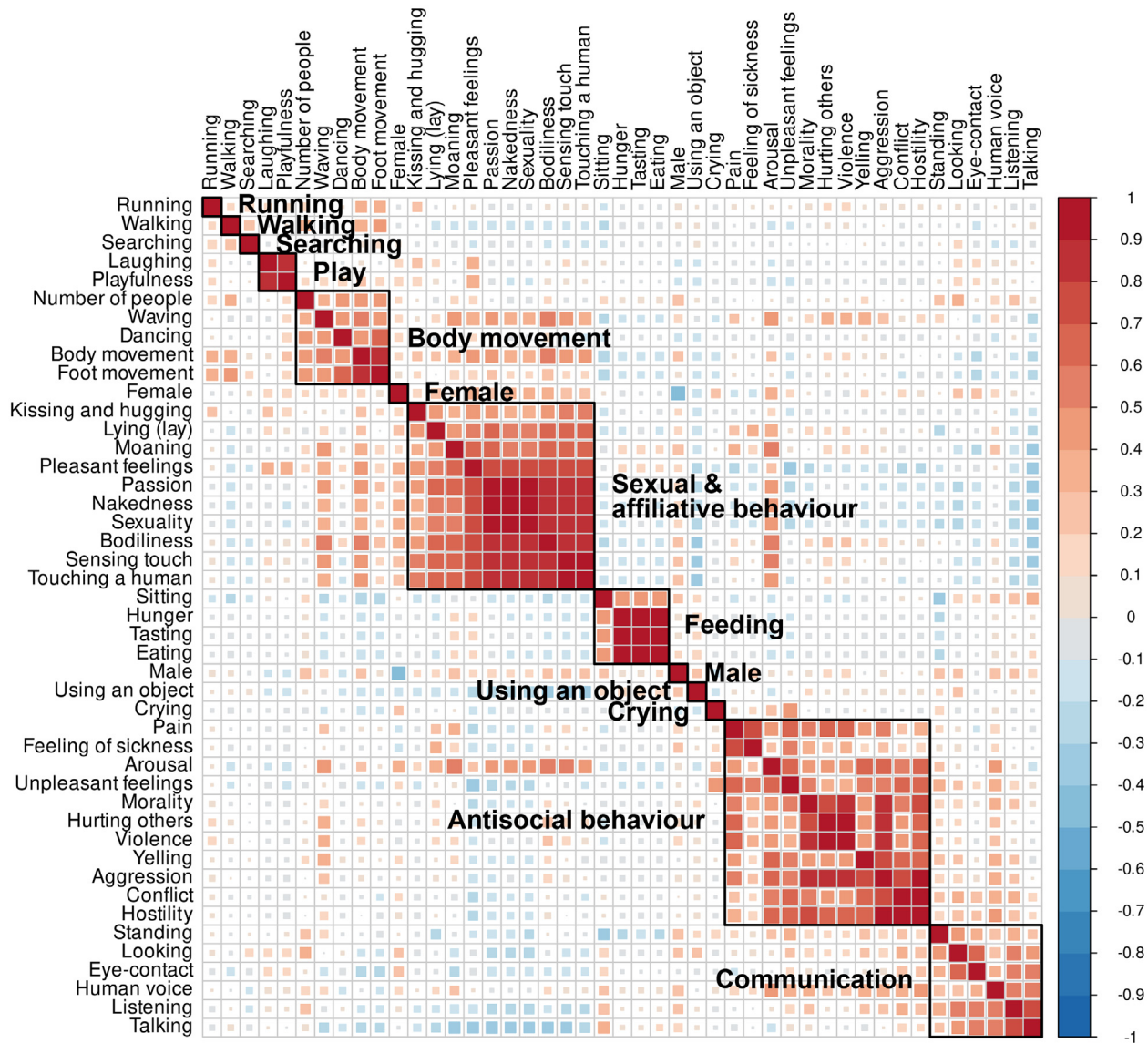
To further ensure the reliability of our ratings from a small subject pool, we also compared the average perceptual ratings of our five annotators with an independent, significantly larger dataset collected for another study using the same stimulus deck. In this dataset we collected ratings for a subset of the reliable social features in the current study (33 out of 45) from the majority of the movie clips (87 out of 96). These annotations were collected only once for each video clip instead of shorter intervals used for generating the fMRI stimulus model. The validation data were collected using online platform Prolific (<https://www.prolific.co/>) and included ten ratings for each social feature in each video clip. This dataset contained ratings from 1096 participants (one participant annotated only few social features and few movie clips). The correlation of the population average ratings over all 33 social features between local and online data collection was high ( $r = 0.78$ ,  $p < 0.05$ , range: 0.60 – 0.97) confirming that social features with high ICC are perceived similarly between people. Hence, only five annotations were sufficient, and these independent annotations could be used to model the brain activity of the neuroimaging subjects.

#### 2.7. Dimension reduction of the social perceptual space

The reliable 45 features were linearly correlated (**Fig. 2**) and it is unlikely that each social feature is processed in different brain regions or networks. We performed dimension reduction with hierarchical clustering on the correlation matrix of selected features to define the perceptual dimensions that characterize different aspects of social interaction. Clustering was chosen over principal component analysis for easier interpretation of the dimensions, because it is likely more sensitive than principal component analysis (PCA) in finding perceptually important social features, or their combinations, that do not share a large proportion of variance with other social features. Initially we chose Pearson correlation as the similarity measure because the co-occurrence of features measured in abstract and possibly not strictly continuous scale is more interesting than the absolute distance between them (considered in PCA). Unweighted pair group method with arithmetic mean (UPGMA), as implemented in R, was used as the clustering algorithm (<https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/hclust>). Other average linkage clustering methods implemented in the R package (WPGMA, WPGMC and UPGMC) yielded highly similar clustering hierarchy. Hierarchical clustering requires a desired number of resulting clusters as an input for automatic definition of cluster boundaries from hierarchical tree (**Figure SI-2**). To estimate the optimal number of clusters we chose three criteria that the clustering result should satisfy. These were cluster stability, theoretically meaningful clustering, and sufficient reduction in collinearity between the clusters. To assess the stability of clusters we conducted a consensus clustering analysis with ConsensusClusterPlus R package (Wilkerson and Hayes, 2010). Theoretically meaningful clustering was then assessed, and collinearity was measured using Pearson correlation and variance inflating factor (VIF). Detailed information of the cluster analysis and consensus clustering results can be found in Supplementary Materials (see also **Figure SI-3**). Cluster analysis grouped social features into six clusters and seven independent features not belonging to any cluster (**Fig. 2**) and these social dimensions formed the final model for social perception. The cluster regressors were created by averaging across the individual feature values in each cluster (**Figure SI-4**).

#### 2.8. Modelling low-level sensory features

Our goal was to map perceived social dimensions in the human brain. The stimulus clips were not balanced with respect to their low-level audiovisual properties, thus these were controlled statistically when estimating the unique contribution of social dimensions to the BOLD signal. We extracted 14 different dynamic audiovisual properties



**Fig. 2.** The results of the hierarchical clustering of reliably rated social features. The correlation matrix is ordered based on hierarchical clustering, and clustering results ( $k = 13$ ) are shown. The analysis suggested that the social perceptual space of the stimulus can be reduced to six clusters and seven individual features.

from the stimulus movie clips including six visual features (luminance, first derivative of luminance, optic flow, differential energy, and spatial energy with two different frequency filters) and eight auditory features (RMS energy, first derivative of RMS energy, zero crossing, spectral centroid, spectral entropy, high frequency energy and roughness). Optic flow was estimated with opticalFlowLK -function with basic options (<https://www.mathworks.com/help/vision/ref/opticalflowlk.html>). Custom functions were used for estimating the other visual features (see Code availability). Auditory features were extracted using MIRTtoolbox1.8.1 (Lartillot et al., 2008). First eight principal components (PCs) explaining over 90% of the total variance were selected as regressors for low-level audiovisual features. As the stimulus movie clips included control clips with no human interaction, we created a “nonsocial” block regressor by assigning a value of 1 to the time points where the stimulus did not contain people, human voice, or animals. A low-level model was formed by combining the eight audiovisual PCs, the nonsocial regressor and subjectwise mean signals from cerebrospinal fluid (CSF) and white matter (WM). See **Figure SI-5** for correlations between low-level features and social dimensions.

## 2.9. Univariate regression analysis of social perceptual dimensions

### 2.9.1. Overview of the regression analysis

Ridge regression (Hoerl and Kennard, 1970) was used to estimate the contributions of the low-level features and cluster-based composite social dimensions to the BOLD signals for each subject. Ridge regression was preferred over ordinary least squares (OLS) regression because even after dimension reduction, the social regressors were moderately correlated (range:  $-0.38 - 0.32$ ) and we wanted to include all perceptual dimensions in the same model to estimate their unique contributions to the BOLD signals. We also wanted to avoid overfitting while retaining generalizability of the results. To conservatively control for low-level features, the demeaned BOLD signals were first predicted with the low-level model and the residual BOLD signals were then used as input in the following regression analysis with the social stimulus model (see Fig. 1). The low-level regressors were still included as nuisance covariates in the analysis of social dimensions for the possible interaction between the social dimensions and low-level features. In both consecutive analyses ridge parameter was optimized using leave-one-subject-out

cross-validation. Prior to statistical modelling the regressor time series were convolved with the canonical HRF and the columns of the design matrices were standardised ( $\mu = 0$ ,  $\sigma = 1$ ).

### 2.9.2. Ridge regression optimization

Optimization of ridge penalty for each voxel separately could have yielded in large differences in the penalty parameters values throughout the brain thus making it more difficult to interpret the regional differences in the results. Thus, we selected an unbiased sample of grey matter voxels for the optimization by randomly sampling 20% (~5000) of grey matter voxels uniformly throughout the brain. Only voxels within population level EPI mask where the population level probability of grey matter was over 0.5 were available for sampling. Detailed description of ridge regression modelling is included in supplementary materials (Figure SI-6).

### 2.9.3. Statistical inference in the regression analysis

For the social perceptual model, the regression analysis was run both at voxel-level and at region-of-interest (ROI) level. The population level EPI mask was used in all analyses to include only voxels with reliable BOLD signal from each subject and thus brain areas including parts of orbitofrontal, inferior temporal and occipital pole areas were not included in the analyses. In voxel-level analysis, subject-level  $\beta$ -coefficient-maps were subjected to group-level analysis to identify the brain regions where the association between intensity of each social dimension and haemodynamic activity was consistent across the subjects. Voxels outside the population level EPI mask were excluded from the analysis. Statistical significance was identified using the randomise function of FSL (Winkler et al., 2014). Voxel-level FDR with q-value of 0.05 was used to correct for multiple comparisons (Benjamini and Hochberg, 1995). Anatomical ROIs were extracted from AAL2 atlas (Rolls et al., 2015). ROIs with at least 50% of voxels outside the population level EPI mask, were excluded from the analysis and only voxels within population level EPI mask were considered for the included ROIs. This resulted in inclusion of 41 bilateral ROIs into the analysis. A parametric t-test on the  $\beta$ -weights of a ROI was used to assess statistical inference across subjects. ROI-analysis results were considered significant with p-value threshold of 0.05 Bonferroni corrected for multiple comparisons. The results for ROI analyses are reported as union of bilateral ROIs.

## 2.10. Multivariate pattern analysis of social perceptual dimensions

### 2.10.1. Overview of the multivariate pattern analysis

To reveal the regional specialization in processing of different social features, between-subject classification of 11 perceptual dimensions<sup>1</sup> was performed in Python using the PyMVPA toolbox (Hanke et al., 2009). The aim of the classification analysis was to complement univariate regression analysis by testing whether the human brain expressed regional specificity for distinct social dimensions. This approach was based on classification of discrete social dimensions from brain activity, rather than computationally more complex approach to predict actual values of multiple social predictors simultaneously based on brain activity. The classification was performed by i) first labelling each time point with only one social label, ii) then splitting the stimulus into time windows and fitting general linear models separately for different social labels within each time window and iii) finally running the between-subject classification on the subject level beta images labelled with social dimensions.

<sup>1</sup> Dimensions “Male” and “Female” were excluded from classification, because unlike the rest of the dimensions, they are genuinely binary features and thus not comparable with the other dimensions in the implemented classification framework (see Figure SI-4).

### 2.10.2. Discrete social labelling for each stimulus time point

For discrete classification from mixed signal, only one dimension label for each time point (each TR) could be given. For the continuous signal, the currently most salient category is not always unambiguous, because more than one social feature could be present simultaneously and salient changes in social information may attract more attention (“Somebody starts crying”) than those occurring frequently (“People are talking”). To resolve this issue, we first normalized the dimension rating time series ( $\mu = 0$ ,  $\sigma = 1$ ) and then, for each time point, chose the feature with the highest Z-score as the category label for that time point (Figure SI-7). To ensure that the included time points would be representative of the assigned categories, we chose only time points where Z-scores for the chosen dimension were positive. This procedure ensured that each time is labelled with representative category and that infrequently occurring social information is weighted more than constantly present categories.

### 2.10.3. Time window selection and general linear modelling before classification

Classifying every time point separately would not be adequate since single EPI scans are noisy and it cannot be assumed that adjacent time points assigned with the same label would be independent from each other. Accordingly, we split the data into 29 time windows and all time points with the same label within a time window were considered as a single event of that class. The number of time windows was selected based on the response length of canonical HRF (~30 s). Over 30 second time windows would be less dependant from each other than shorter time windows while the data would contain enough events for classification. The time window boundaries were adjusted so that adjacent time points with the same label would not be interspersed to different time windows because temporal autocorrelation of adjacent time points may yield in artificial increase in the classification accuracy. After adjustment, the average time window length was 39 s (range: 34 s – 49 s). The time windows were longer than the movie clips and therefore time points from different clips with similar social context could be judged as one event if they belong to the same time window. Altogether the data consisted of 87 events (Events: using an object: 16, communication: 15, antisocial behaviour: 11, feeding: 10, walking: 9, sexual & affiliative behaviour: 8, body movement: 5, crying: 4, play: 4, running: 3 and searching: 2). 75 out of the total 87 events (86%) included only adjacent time points and the mean length of the events was ~12 s. For generating input for between-subject classification an ordinary least squares (OLS) GLM without covariates was fit to the normalized ( $\mu = 0$ ,  $\sigma = 1$ ) residual BOLD time series (confound-controlled data) for each subject and each event. These subject-level beta-images and their social labels were used as input for the classifier (see Fig. 1).

### 2.10.4. Classifier algorithm and cross-validation method

A neural network (NN) model ([https://scikit-learn.org/stable/modules/generated/sklearn.neural\\_network.MLPClassifier.html](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html)) was trained to classify the perceptual dimensions using leave-one-subject-out cross-validation, where the model was trained on the data from all except one subject and tested on the hold-out subject's data; this procedure was repeated N times so that each subject was used once as the hold-out subject. Such leave-one-subject-out cross-validation tests the generalizability of the results across the sample of the subjects. The analysis was performed using whole brain data (with non-brain voxels masked out) and regional data using anatomical ROIs. In the whole-brain analysis, an ANOVA feature selection was applied to the training set within each cross-validation and 3000 voxels with the highest F-score were selected. The regional MVPA was first performed using data from all voxels within a region. To control for the effect of ROI size to the classification accuracy the regional MVPA was also performed with an ANOVA feature selection where the size of the smallest ROI (lateral orbitofrontal cortex, 119 voxels) was selected as the number of features for the feature selection.

### 2.10.5. Classifier hyperparameter tuning

Hyperparameters of the NN algorithm were optimized within a limited set of predefined hyperparameter values in the whole brain analysis. Hyperparameter values reflecting the best prediction accuracy with acceptable runtime were used in both full brain and ROI analyses (see **Table SI-3** for hyperparameter tuning). The optimized NN included two hidden layers with 100 nodes in each ( $\alpha = 1.00$ , max\_iter 500, other hyperparameters set to default). In the model learning process, the order of events was shuffled in each training iteration which minimized the model's ability to learn the order of the events in the stimulus. A support vector machine (SVM) classifier had similar classification accuracy in the whole brain analysis, but NN model was chosen because the computation time was shorter and the variance of classification accuracies between subjects were lower with NN model compared to SVM classifier.

### 2.10.6. Outcome measures of the classification analysis and statistical significance testing

Classification accuracy was quantified by computing the proportion of correctly classified events relative to the total number of events (i.e., recall). To estimate the null distribution, the following procedure was repeated 500 times: we 1) randomly shuffled social class labels; 2) ran the whole-brain MVPA with 97 leave-one-subject-out cross-validations, where the classifier was trained on the data with shuffled labels from N-1 subjects and tested on data with correct labels from the remaining subject; and 3) calculated the classification accuracies on each of the 500 iterations. The null distribution estimation was computationally prohibitive as one iteration took approximately one hour, and we decided that 500 iterations would be sufficient to assess the statistical significance of our findings. If the true accuracy was larger than 99% of the accuracies obtained with the randomly shuffled labels, the true accuracy was considered significant with an  $\alpha$  of 0.01. We cannot assume that the null distribution of classification accuracies for each class is equal and centre around the naïve chance level because the number of events is unbalanced between classes. For this reason, we only report if the total accuracy of the classification is statistically significant. In the whole-brain classification we also report the precision of the classifications which is the number of correct predictions for a class divided by the total number of predictions into that class. In ROI analyses, the statistical differences between regional classification accuracies were tested using paired t-tests between subjectwise classification accuracies between each pair of regions.

### 2.11. Intersubject correlation analysis

Watching movies synchronizes brain activity between different individuals particularly in the occipital, temporal, and parietal regions of the brain and the synchronization of brain activity can be quantified with intersubject correlation (ISC) analysis (Hasson et al., 2004). As the only variable factor in the experiment is the time-varying audiovisual stimulus, ISC analysis captures the shared stimulus-dependant activation in the brain. It is well known that ISC is greatest on the sensory cortices, but an important yet unresolved question is which variables drive the degree of synchronization of BOLD response. Some prior studies suggest that emotions and top-down perspectives play a role (Lahnakoski et al., 2014; Nummenmaa et al., 2014), but the role of social features remains unknown. As a *post hoc* analysis, we assessed whether the regional differences in brain response profiles for social dimensions relate to the intersubject response reliability of BOLD response. To this end, we calculated the ISC across subjects over the whole experiment and compared the regional ISC with the results from regression and MVPA analyses. ISC-toolbox with default settings was used for ISC calculations (Kauppi et al., 2014).

## 3. Results

### 3.1. How people perceive the social world?

A total of 45 out of the 112 social features had sufficient inter-rater reliability and occurrence rate (see **Figure SI-1**). Hierarchical clustering identified six clusters that were labelled as “antisocial behaviour”, “sexual & affiliative behaviour”, “communication”, “body movement”, “feeding” and “play”. Seven perceptual dimensions did not link with any cluster and were analysed separately. These dimensions were “using an object”, “crying”, “male”, “female”, “running”, “walking” and “searching”. **Fig. 2** shows the clustering of the dimensions. Median pairwise correlation between any two of the 13 dimensions was 0.02 (range:  $-0.38 - 0.32$ ) and the maximum variance inflation factor (VIF) in the design matrix excluding nuisance covariates was 3.3 (male regressor) while the median VIF value was 1.3. These diagnostics indicate that regression coefficients for the dimensions will be stable in linear model estimations, and they could thus be included in the same model. See **Figure SI-4** for visualized time series of social dimensions and **Figure SI-5** for correlations matrices for low-level features and social dimensions.

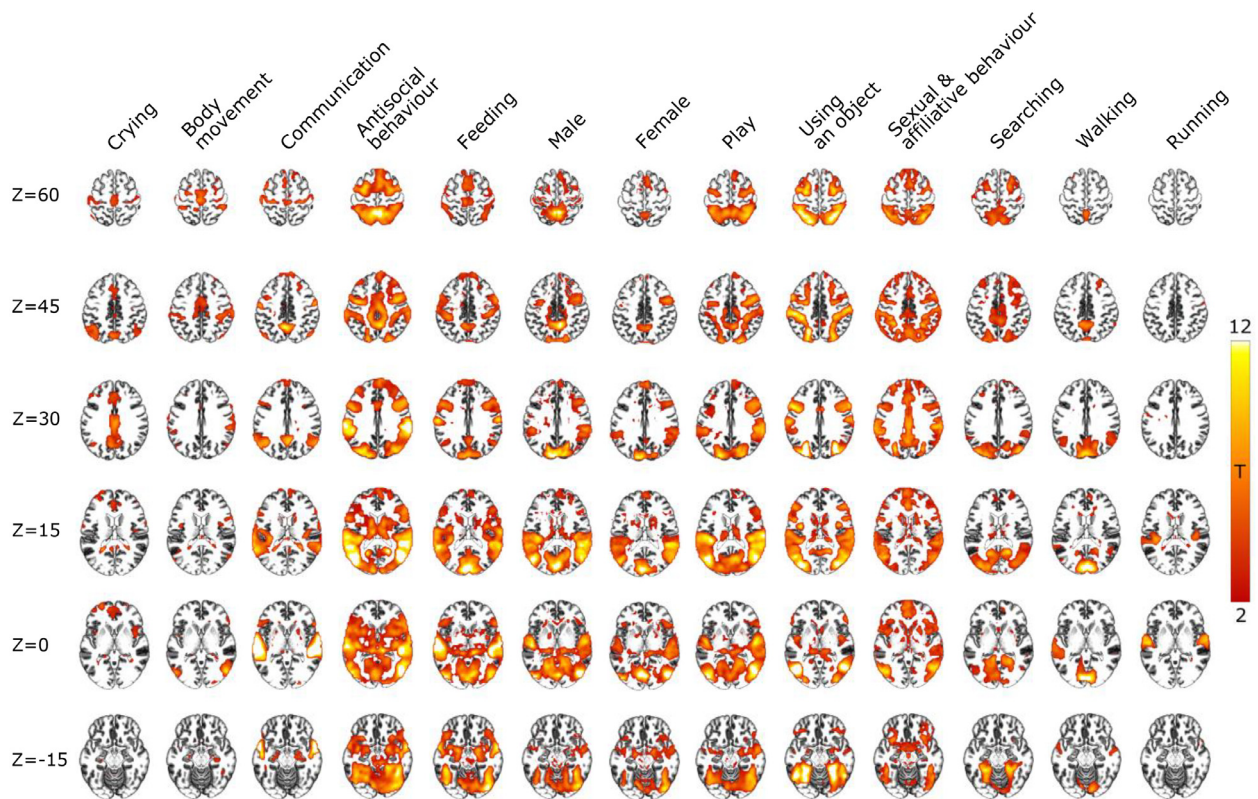
### 3.2. Cerebral topography of social perception

Regularized ridge regression was used to establish the full-volume activation patterns for the 13 perceptual social dimensions (**Fig. 3**). Social information processing engaged all brain lobes and both cortical and subcortical regions. Robust responses were observed in occipital, temporal, and parietal cortices (**Fig. 4**). There was a clear gradient in the responses, such that posterior temporal, occipital and parietal regions showed the strongest positive association with most of the social dimensions, with significantly less consistent activations in the frontal lobes and subcortical regions. Yet, frontal, and subcortical activations were also observed for some dimensions such as sexual & affiliative behaviour, antisocial behaviour, and feeding.

In ROI analysis, broadly tuned responses for social dimensions were observed in STG and MTG with strongest responses for communication and antisocial behaviour, respectively. In parietal lobe, all regions except angular gyrus and paracentral lobule responded to a wide range of perceptual dimensions. In frontal regions the associations between social dimensions and haemodynamic activity were less consistent than in more posterior regions, yet still statistically significant in some of the regions including IFG, cingulate cortex and precentral gyrus. Most consistent frontal effects were found for sexual & affiliative and antisocial behaviour. For subcortical regions the observed associations were generally weak. Most notable subcortical associations with perceptual dimensions were seen in amygdala and thalamus. Consistent negative associations were restricted to occipital lobe and were observed for communication, crying, body movement and running.

### 3.3. Cerebral gradients in social perception

**Fig. 5a** shows the cumulative brain activation maps for all 13 perceptual dimensions. There was a gradient in the regional selectivity for social dimensions. Posterior temporal and occipital cortices as well as parietal cortices responded to most social dimensions, while responses become more selective in the frontal cortex although IFG, precentral gyrus and the frontal part of the medial superior frontal gyrus (SFG) had some consistency in their response profiles. Because the same stimulus was used across the subjects, we hypothesized that the brain activation in the areas with the broadest response profiles would be temporally most synchronized across subjects. We thus calculated the ISC of brain activation over the whole experiment (**Fig. 5b**) and correlated the regional ISC values with corresponding response selectivity values (i.e. number of social features resulting in significant activations in each region). Scatterplot in **Fig. 5c** shows the association between ISC and cor-



**Fig. 3.** Brain regions showing increased BOLD activity for the social dimensions. Results show the voxelwise T-values (FDR-corrected,  $q = 0.05$ ) of increased BOLD activity for each social dimension from the multiple regression analysis. The results are also visualized on inflated cortical surfaces in **Figure SI-8**.

responding brain response selectivity for perceptual dimensions (Pearson  $r = 0.86$ ).

### 3.4. Multivariate pattern analysis

Finally, we trained a between-subject neural network model to decode presence of perceptual social dimensions from the spatial haemodynamic activation patterns to reveal which social dimensions are consistently represented in each cerebral region. Whole brain classification was performed in 3000 voxels that passed through the ANOVA feature selection. Most of the selected voxels (**Fig. 6a**) localized into temporal (STG, MTG, Heschl gyrus and superior temporal pole), occipital (calcarine and lingual gyri, cuneus, FG, superior occipital gyrus (SOccG), middle occipital gyrus (MOccG) and inferior occipital gyrus (IOccG)) and parietal cortices (supramarginal, superior parietal gyrus (SPG) and inferior parietal gyrus (IPG)). The permuted chance level for the total classification accuracy in the whole brain analysis was 0.128 which is above naïve chance level ( $\frac{1}{11} \cong 0.09$ ). At the whole brain level, the NN model was able to classify all 11 social dimensions significantly above chance level with the total classification accuracy of 0.52 ( $p < 0.01$ ). Classification accuracies/precisions for each social dimension were: walking: 0.49/0.51, using an object: 0.53/0.50, searching: 0.70/0.69, running 0.56/0.62, sexual & affiliative behaviour 0.45/0.48, play 0.53/0.51, feeding 0.46/0.48, crying 0.46/0.51, communication 0.55/0.55, body movement 0.52/0.50 and antisocial behaviour 0.55/0.53 (**Fig. 6a**).

The classification was also performed within anatomical ROIs (**Fig. 6b**). Most accurate classifier performance was observed in lingual gyrus (0.34,  $p < 0.01$ ), calcarine gyrus (0.33,  $p < 0.01$ ), cuneus (0.29,  $p < 0.01$ ), SOccG (0.29,  $p < 0.01$ ), MOccG (0.27,  $p < 0.01$ ), STG (0.27,  $p < 0.01$ ) and MTG (0.25,  $p < 0.01$ ). Although the prediction accuracies were statistically significantly above permuted chance level for each ROI, the

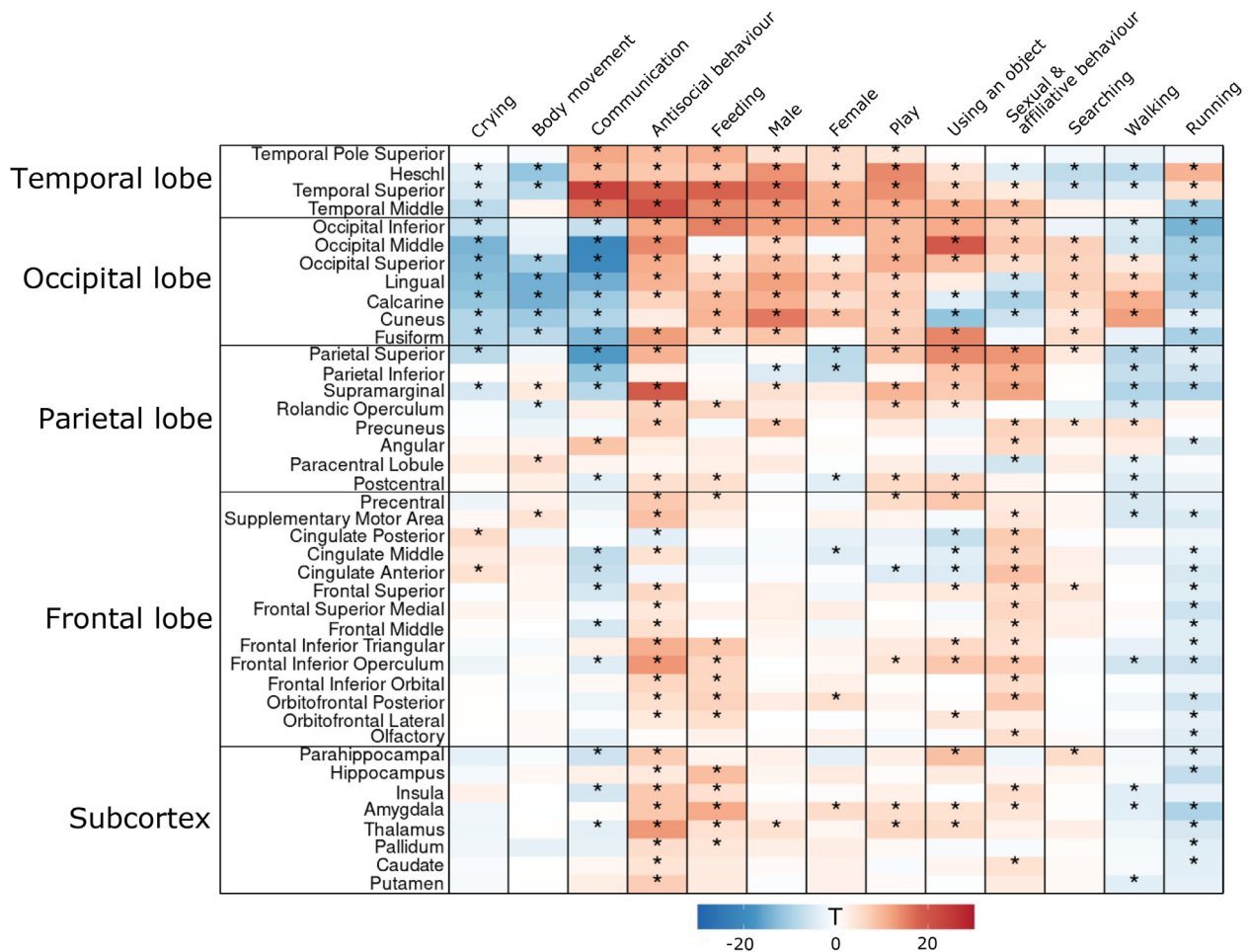
gradient in brain responses for social perception was also observed in the classification accuracies so that highest accuracy was observed in occipital and temporal areas, followed by parietal cortices and frontal and cingulate cortices. Lowest accuracies were found in the subcortical regions. (**Fig. 6b**). We also validated that this gradient was not an artefact stemming from the sizes of the ROIs, as similar gradient was observed in the regional classification with ANOVA feature selection limited to the number of voxels in the smallest ROI (**Fig. 6c**). **Figure SI-9** shows the statistical differences of the classification accuracies between all pairs of ROIs confirming the observed gradient in classification accuracies. Occipital and temporal areas (excluding temporal pole) showed significantly higher classification accuracy than frontal and subcortical regions.

### 3.5. Relationship between classification accuracy and ISC

Regional classification accuracy and ISC were positively correlated (Pearson  $r = 0.85$ , **Fig. 7a**). Most occipital regions, STG and MTG showed high synchrony ( $ISC > 0.1$ ) and high classification accuracy ( $acc > 0.25$ ). Most parietal regions showed average ISC and average classification accuracy while frontal and subcortical regions showed low ISC and low classification accuracies. The most notable exception to this pattern was Heschl gyrus which had high ISC (0.28) yet average classification accuracy ( $acc = 0.22$ ). **Fig. 7b** summarizes the results from separate regression, ISC and classification analyses where the findings overlap most in temporal and occipital cortices.

## 4. Discussion

Our findings provide the currently most detailed map of the social perceptual mechanisms in the human brain using naturalistic stimuli.



**Fig. 4.** Regional results from the multiple regression analysis. The heatmap shows T-values for regression coefficients for each ROI and social dimension. Statistically significant ( $p < 0.05$ , Bonferroni-corrected for each dimension independently) ROIs are marked with an asterisk.

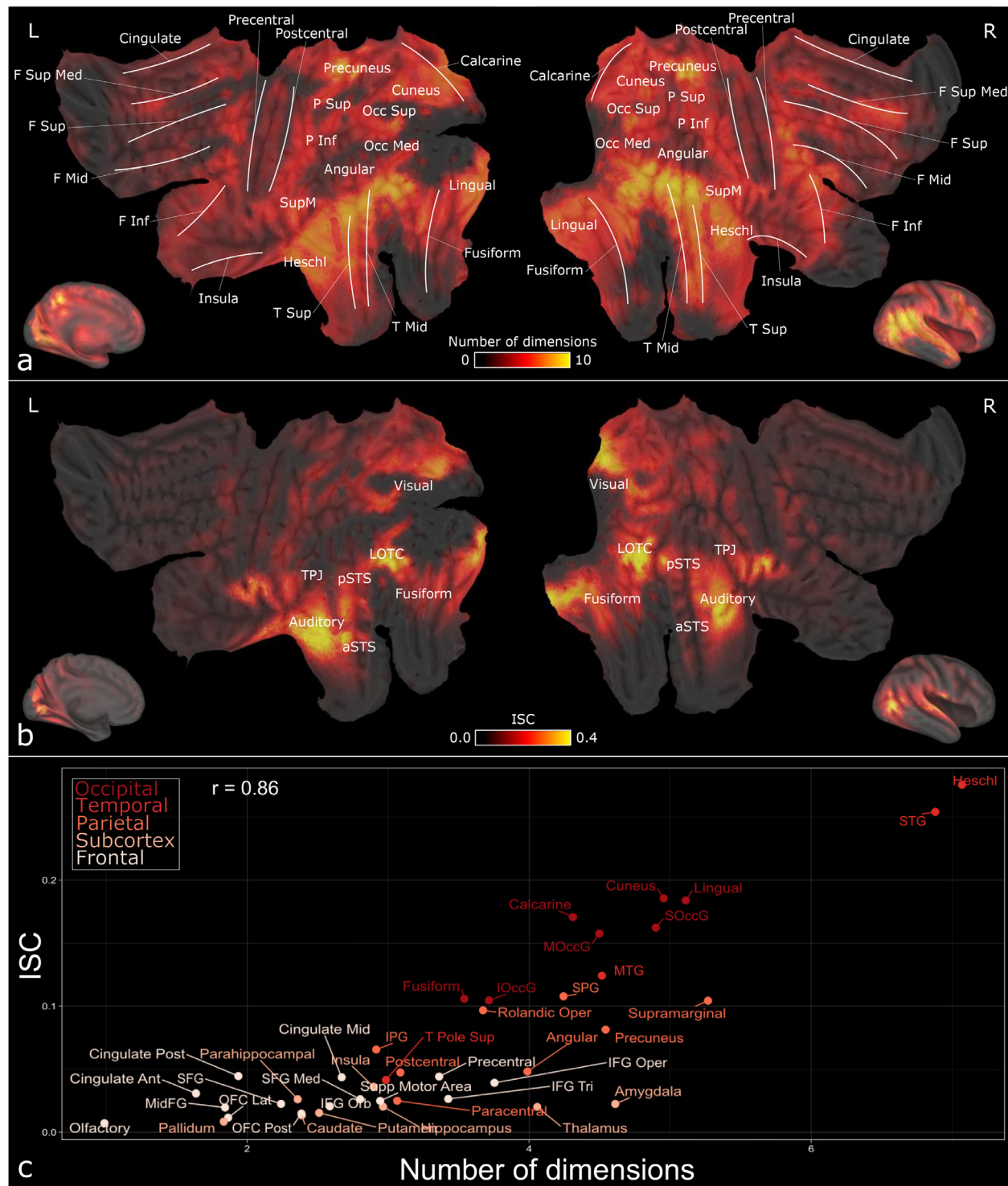
The behavioural data established that 13 social dimensions reliably capture the social perceptual space contained in the video stimulus. The cerebral topography for social perception was organized along an axis, where posterior temporal and cortical regions served a central general-purpose role in social perception, while the regional selectivity for social dimensions increased towards frontal and subcortical regions. Multivariate pattern recognition established that particularly occipito-temporal and parietal regions carry detailed and spatially dimension-specific information regarding the social world, as evidenced by the highest classification accuracies in the multi-class classification approach. Both classification accuracy and consistency of the responses for specific social dimensions were the highest in the brain regions having most reliable (indicated by ISC) activation patterns throughout the experiment. These effects were observed although low-level sensory features were statistically controlled for. Altogether these results show that multiple brain regions are jointly involved in representing the social world and that different brain regions have variable specificity in their spatial response profiles towards social dimensions.

#### 4.1. Dimensions of social perception

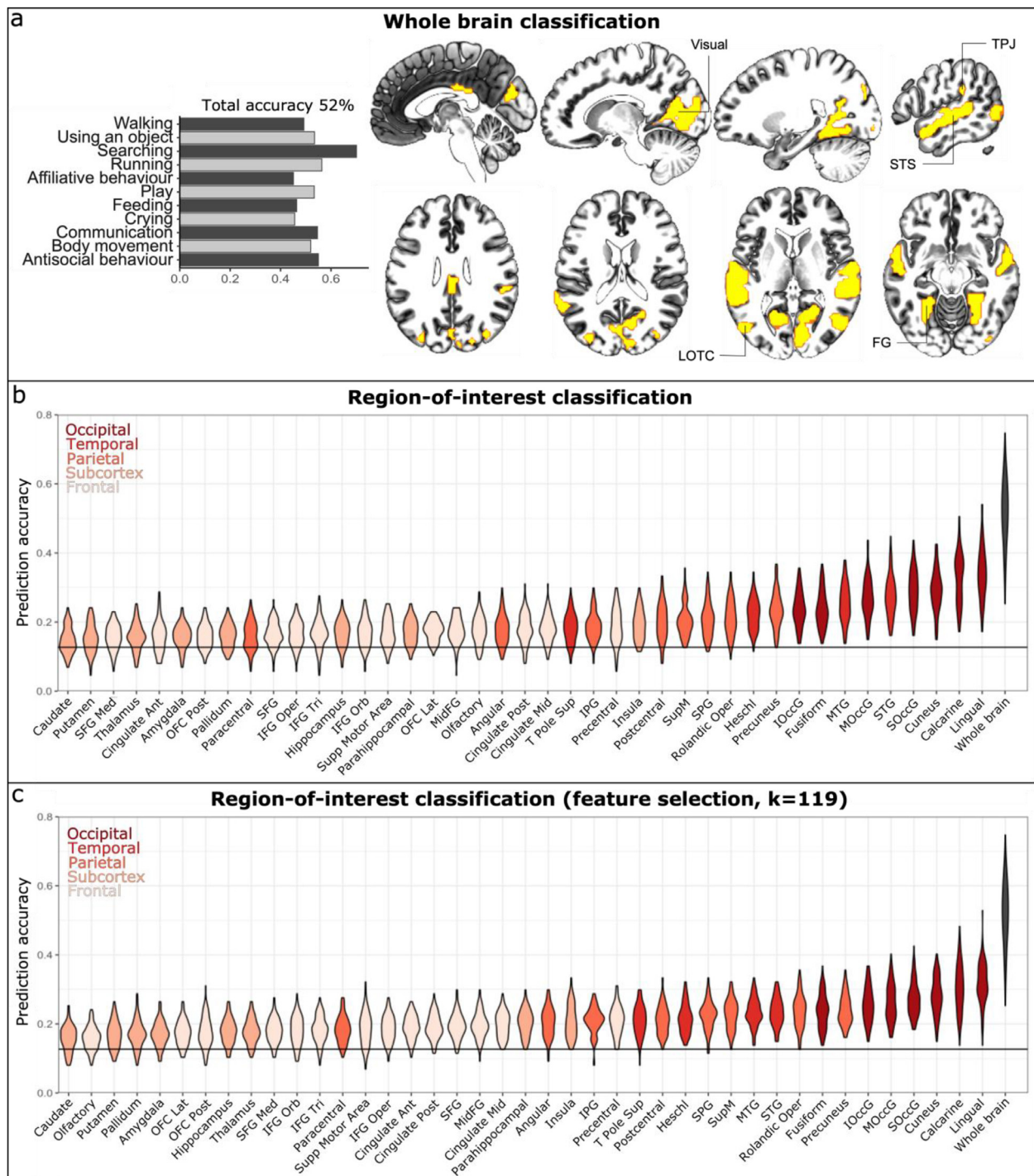
The behavioural experiment established that the observers used consistently a set of 45 descriptors when evaluating the social contents of the movies. Dimension reduction techniques further revealed that these 45 features could be adequately summarized in 13 social dimensions. The largest clusters were organized along the valence dimension of the social interaction containing sexual & affiliative (e.g., kissing, touching,

sexuality) versus antisocial (hurting others, yelling) behaviors indicating a close link between emotion and social interaction. Social communicative behaviors (e.g., eye contact, talking) and body movements (e.g., waving, moving a foot) also formed large clusters. Play-related behaviors (laughing, playfulness) as well as feeding-related actions (e.g., tasting, eating) were also represented into smaller clusters. Notably, some features such as presence of males versus females, walking, and using objects remained independent of any of the clusters. Average hierarchical clustering algorithm was used because it yields clearly interpretable clusters and because feature similarity could be measured with correlation instead of absolute distance. Further research could establish how behavioural clusters found with hierarchical clustering relate to, for example, principal components off the same data and how the clusters generalize to other naturalistic stimuli.

The stimulus movie clips cannot portray all possible social scenarios and Hollywood movies are only a proxy of real-life social interaction. Still, 99 of the predefined 112 social features had sufficient occurrence rate in the stimulus video clips (**Figure SI-1**) which indicate that the stimulus contains a broad range of social information. The average duration of movie clips was ~10 s and we acknowledge that this timescale does not allow examination of social processes occurring at slower temporal frequencies such as pair bonding and long-term impression formation. However, social perception can be astonishingly fast. Semantic, social, and affective categorization may happen in few hundred milliseconds (Nummenmaa et al., 2010) and the judgements do not significantly change from the initial judgments after longer consideration (Willis and Todorov, 2006). Electroencephalography (EEG) has also confirmed reli-



**Fig. 5.** (a) Cumulative activation map for social dimensions. Voxel intensities indicate how many social dimensions (out of 13) activated the voxel statistically significantly (FDR-corrected,  $q = 0.05$ ). White lines indicate the localizations of major gyri. (b) Significant ISC (FDR-corrected,  $q = 0.05$ ) across subjects over the whole experiment (c) Scatterplot showing the association between regional ISC values and tuning for social perceptual dimensions. ISC is plotted in Y-axis and the X-axis shows the number of social dimensions (out of 13) associated significantly with the BOLD response. Regional values are calculated as the average over all regional voxels. CARET software (Van Essen, 2012) was used for mapping results from ICBM 152 Nonlinear Asymmetrical template version 2009c space to the flatmap surface.

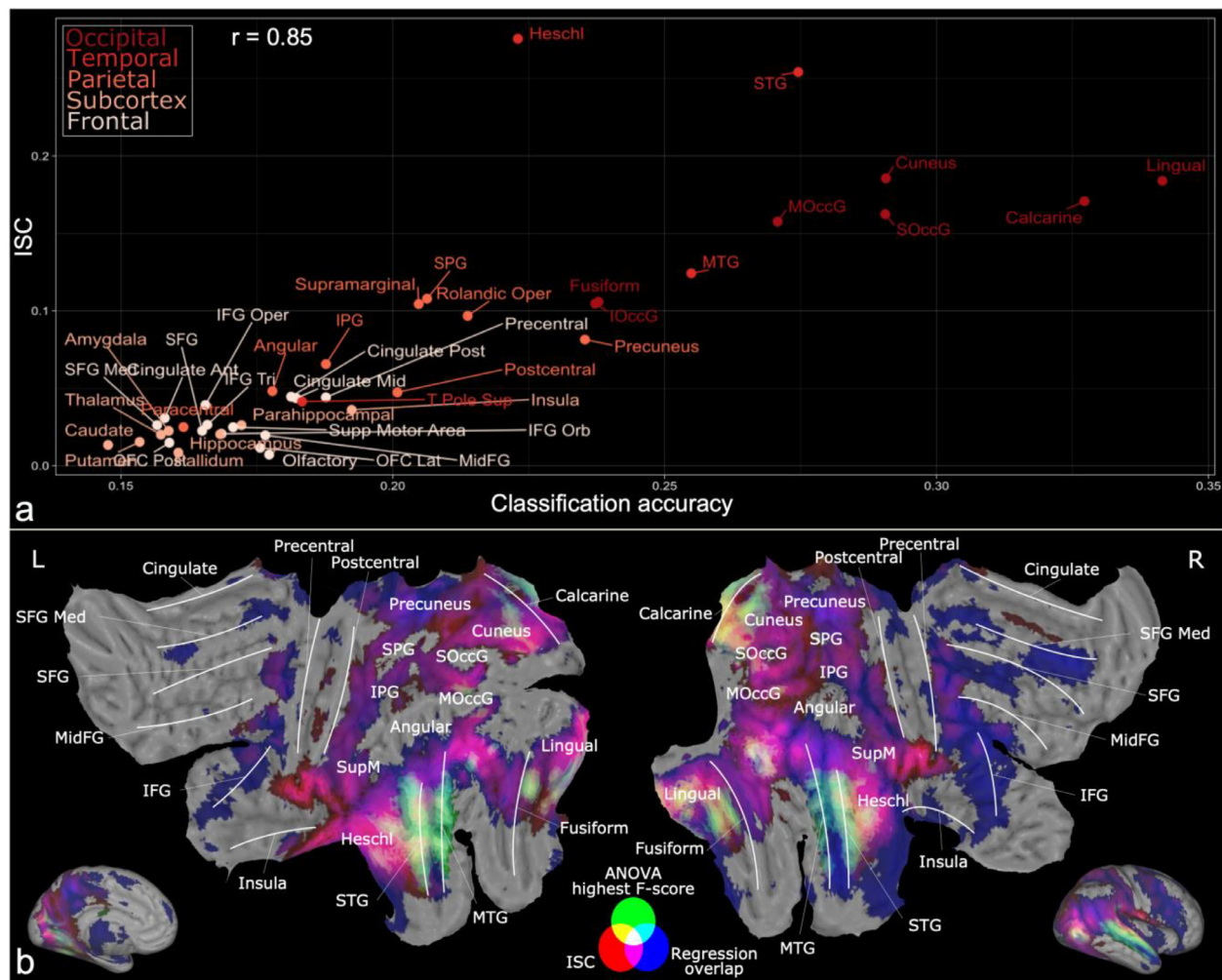


**Fig. 6.** Results from the multivariate pattern analysis of social dimensions. (a) Whole brain classification accuracies and the voxels used in the whole brain classification analysis (based on ANOVA feature selection). (b) Regional classification accuracies compared with the whole brain classification accuracy (the rightmost black bar). The permuted chance level accuracy ( $\text{acc} = 0.128$ ) is shown as a horizontal line. The mean prediction accuracy was significantly ( $p < 0.01$ ) above the chance level accuracy in the whole brain analysis and for each region-of-interest. (c) Regional classification accuracies using only 119 voxels (the size of the smallest region) with the highest F-scores as input for the classifier.

able associations between social perceptual features and brain response already 400 ms after the stimulus (Dima et al., 2022) concluding that short video clips can capture some temporal scales of social perception. However, the temporal resolution of fMRI is limited to the TR of the scanner (2.6 s in this study) and the social features were rated in approximately 4 second time intervals. Therefore, our study do not measure instantaneous brain responses for perceptual social features. Additionally,

the haemodynamic response ( $\sim 30$  s) is longer than the average length of the movie clips, but the convolution of the predictors accounts for the delayed response.

Data-driven models for characterising social perception (Adolphs et al., 2016) constitute an important and complementary alternative for the theory-based models for separate taxonomies of person, situation, and action perception since i) the clusters observed in



**Fig. 7.** (a) Scatterplot showing the relationship between regional ISC and classification accuracy. (b) Additive RGB map summarizing the main findings. The overlap between activation patterns for perceptual dimensions in regression analysis is shown as blue (areas where at least 3 dimensions expressed FDR-corrected brain activation). Significant (FDR-corrected,  $q = 0.05$ ) ISC across subjects is shown as red and the ANOVA selected voxels for the whole brain classification are shown as green.

our data are based on the actual perception of the social context, ii) the data-driven model does not separate persons, situations, and actions and is based on the subjects' net percept of the stimulus and iii) only dimensions actually present in the stimulus are considered. Importantly, this data-driven model for social perception has many similarities with previously proposed taxonomies. The largest observed clusters (sexual & affiliative and antisocial behaviour) closely relate to the emotional valence which is at the core of human emotions (Russell, 1980) and is also considered in taxonomies describing persons (Simms, 2007) and situations (Parrigon et al., 2017; Rauthmann et al., 2014). Clusters pertaining to play and feeding closely relate to dimensions "humour" from situation taxonomy (Parrigon et al., 2017) and "food" from action domain (Thornton and Tamir, 2022), respectively. Mapping of the neural space for social perception requires the social features to be consistently rated amongst the independent set of annotators. 61 of the total 112 rated social features showed low between-rater agreement ( $ICC < 0.5$ , Figure SI-1) which is an important finding in itself regarding the consistency of the perceptual taxonomy individuals use for describing social events. The exclusion of these features had the effect that more abstract, or idiosyncratically judged dimensions cannot be addressed in this experiment and pushed the studied perceptual processes towards action and situation domains. Further research should nevertheless investigate the shared versus idiosyncratic social

evaluations across individuals, as this would be informative regarding what are the core building blocks of the social environment that are shared across most observers.

#### 4.2. Cerebral gradient in social perception

The univariate BOLD-fMRI analysis based on social dimensions revealed that a widely distributed cortical and subcortical networks encode the social contents of the video stimuli. Most dimensions activated LOTC, STS, TPJ, as well as other occipitotemporal and parietal regions. There was a gradual change from these unselective social responses in occipitotemporal and parietal regions towards more selective responses in frontal and subcortical regions, suggesting that social perception is mainly processed in lateral and caudal parts of the brain. This effect was also confirmed by the ROI analysis. Most consistent responses were observed in all occipital regions and in temporal regions STG and MTG (which outline STS) and Heschl gyrus. In parietal cortex, most consistent responses were observed in supramarginal gyrus (a part of TPJ), SPG and precuneus. The responses were less consistent in frontal cortex, although brain activity in IFG, precentral gyrus and frontal part of medial SFG associated with a limited number of dimensions including sexual & affiliative behaviour, antisocial behaviour, feeding and using an object. These data are consistent with previous univariate studies addressing social

functions for LOTC (Downing et al., 2001; Lingnau and Downing, 2015; Wurm and Caramazza, 2019; Wurm et al., 2017), STS (Deen et al., 2015; Isik et al., 2017; Lahnakoski et al., 2012; Walbrin et al., 2018), TPJ (Carter and Huettel, 2013; Saxe and Kanwisher, 2003), and MFC (de la Vega et al., 2016). The results were controlled with an extensive set of PCA rotated audiovisual features. A non-social regressor was also built from the stimulus time points where no social interaction was present, and this feature was added to the low-level model. The fMRI data were collected in one scan, hence ruling out the possibility to control for low-level features by cross-validation across scans. Therefore the separation of social perceptual features from all possible low and mid-level features is not possible. However, higher-level information such as body parts and actions have already been shown to associate with BOLD response better than low-level visual features in occipital cortex outside V1 (Tarhan and Konkle, 2020). Additionally, it has been shown that social features of actions explain more variance of EEG responses to videos than low-level visual features (Dima et al., 2022) further supporting the conclusion that the results reflect social information processing rather than low-level audiovisual perception.

#### 4.3. Decoding of perceptual social dimensions from brain activation patterns

The univariate analysis revealed the overall topography and regional brevity of the tuning for different social signals. However, this analysis cannot determine whether a single anatomical region activated by multiple social dimensions reflects responses to shared features across all the dimensions (such as biological motion perception or intentionality detection; Allison et al., 2000; Nummenmaa and Calder, 2009), or spatially overlapping yet dimension-specific processing. Multivariate classification analysis revealed that the answer to this question depends on the region. The ANOVA feature selection for the whole-brain classification retrieved voxels from STS, LOTC, TPJ and FG (Fig. 6 and Fig. 7) yielding classification accuracy exceeding 50% for the multi-class classification. Regional classification confirmed that occipital, temporal and parietal regions showed average to high classification accuracies, whereas the classification accuracies diminished towards chance level in frontal and subcortical regions. These results show that even if the regional univariate responses for social dimensions were overlapping, the specificity of the spatial activation patterns was different between regions. Interestingly, the ANOVA selected voxels found to best discriminate social features closely resemble the network proposed for social aspects of human actions in a recent study (Tarhan and Konkle, 2020) with the exception that our results are more bilateral. Previous multivariate studies have also shown that individual social features are represented in these regions. For example, specific response patterns to pictures of faces versus animals, houses or man-made objects can be found in FG and LOTC (Haxby et al., 2001) and semantic information from different human actions judged from static images are represented in LOTC (Tucciarelli et al., 2019). Subsequent classification studies have shown that, for example, different facial expressions can be classified from activation patterns in FG, and STS (Said et al., 2010; Wegrzyn et al., 2015) and goal-orientated actions in LOTC and interior parietal lobe (Smirnov et al., 2017; Wurm and Lingnau, 2015). Importantly, our results show that BOLD-fMRI can be used for classification of multiple overlapping event categories from continuous naturalistic stimulation. Previous multivariate pattern analyses of social categories have used block designs and categorical stimuli matching the *a priori* category labels. In addition, these studies have only focused on a certain detailed aspect of socioemotional processing. The present results thus underline that even with high-dimensional naturalistic stimulus, the response properties of certain brain areas show high degree of category specificity.

The results from the classification analysis complement the results from the regression analysis with some limitations. To minimize the autocorrelation between successive events while preserving sufficient

number of events for classification the data were split into 29 time windows and the time points with the same social label were combined as one social event within each time window. The time window approach in combination with the Z-score method for labelling each time point split the data into clearly consecutive events (86% events included only adjacent time points) rather than dispersed the social labels within the time window. The mean duration of the events was ~12 s. These findings indicate that this data-driven approach split the data into representative social events that could be used as input for classification. Nevertheless, the capability of this event generating method is dependent on the temporal dynamics of the stimulus and the time window length should be adjusted accordingly at the same time limiting the temporal closeness between two events with the same label. The results are an important first proof of concept for this type of category classification during naturalistic and uninterrupted perception.

The video clips were shown in the same order for all subjects, which may artificially boost classification accuracy, although the model should not learn the actual order of the events since the data were shuffled in each learning iteration. Regardless, the observed differences in classification accuracies in different brain regions should not be due to the order of the stimulus which is more interesting than the actual classification accuracies. It is likely that people focus attention in the most salient social details in the stimulus movies instead of continuously monitoring for multiple sources of information with possibly low importance. Hence, we chose a classification approach where each time point was labelled with the social dimension of the highest relative intensity instead of trying to predict the values of all social features simultaneously. Future studies could aim to predict multiple intensities from multiple categories in the stimulus set. Due to naturalistic and uncontrolled stimuli the classification dataset was unbalanced. Even in regions with near chance level total accuracy, some classes with large number of events were classified with relatively high accuracy (Figure SI-10) which may reflect the differences in the number of events in these classes and might not reflect the actual social information processing in the brain. Consequently, regional differences in the prediction accuracies to individual classes cannot be addressed.

#### 4.4. Reliability versus specificity of responses to social perceptual dimensions

We observed robust intersubject correlation of brain activity in temporal and occipital regions while subjects viewed the video clips. Previous studies have found that the ISC is in general the strongest in sensory regions, and it progressively becomes weaker towards the polysensory and associative cortices (Hasson et al., 2010). Regional ISC has been shown to depend on features such as emotions conveyed by the film (Nummenmaa et al., 2012, 2014). The spatial ISC patterns also depend on the structure of the narrative presented in the stimulus. Structured films with clear plot result in significantly larger ISC than unstructured videos that merely contain social signals (Hasson et al., 2010). Our data revealed that the strength of the ISC was contingent on the number of social features each region responded to in the univariate analysis ( $r = 0.86$ , Fig. 5c) and regional ISC was also associated with the corresponding regional classification accuracy ( $r = 0.85$ , Fig. 7a). These data highlight the relevance of the social domain to the cortical information processing, as the consistency of the regional neural responses was associated with the brevity of the tuning for social signals in each region. In other words, regions responding to multiple social signals also do so in a time-locked fashion across subjects. Importantly, this effect was not just an artefact of the consistency of sensory cortical responses to social signals but was also observed in higher-order associative areas including LOTC and STS. The results indicate that social perception is a key factor in synchronizing brain responses across individuals, supporting the idea that “mental resonance” underlies mutual understanding of social environment and supports the centrality of social interaction in human brain function (Hari et al., 2015).

#### 4.5. Functional organization of social perception networks in the human brain

The regional response profiles towards social signals can be summarized based on the combination of the regional response consistency (univariate regression analysis), the spatial response pattern specificity (MVPA) and the reliability of the BOLD signal across subjects (ISC). First, posterior temporal and occipital regions responded consistently to most social dimensions, while the presence of specific social dimensions could also be classified accurately from these regions. High classification accuracy suggests that these regions already hold dimension-specific and integrated information regarding the social world. Additionally, these regions responded consistently to the social stimuli (as indicated by high ISC) across subjects. LOTC, STS, TPJ, FG and occipital regions thus constitute the most fundamental hubs for social perception in the human brain and are likely involved in integration of the multisensory information and semantic representations regarding the social events (Allison et al., 2000; Lahnakoski et al., 2012).

Second, Heschl gyrus, the site of the auditory cortex (Da Costa et al., 2011), responded consistently to social dimensions but the classification accuracy was only moderate in that region while the ISC of the response was the highest of all regions. This suggests that Heschl gyrus processes domain-general social (most likely auditory) information but does not carry detailed information about the distinct social dimensions, as evidenced by the relatively low classification accuracy in the region. Third, parietal regions especially precuneus, supramarginal gyrus and SPG showed consistent responses with numerous social dimensions and yet their ISC and classification accuracies were only moderate. Previously, precuneus has been linked with attention and memory retrieval (Cavanna and Trimble, 2006), supramarginal gyrus with phonological (Hartwigsen et al., 2010) and visual (Stoeckel et al., 2009) processing of words, and SPG in visuospatial processing and working memory (Koenigs et al., 2009). These parietal regions thus likely respond to some general features of the social signals or idiosyncratic brain states associated with social dimensions.

Frontal and subcortical regions responded only to a limited number of social dimensions, and classification accuracy and ISC remained low. The regression analysis showed some consistency in IFG, precentral gyrus, the frontal part of the medial SFG, amygdala and thalamus, yet the classification accuracies remained low. MFC has previously been associated with higher-level social and affective inference such as linking social processing with decision making, affective processing and theory of mind (Amodio and Frith, 2006; de la Vega et al., 2016). However, previous classification studies have not found specificity for responses to social perceptual dimensions in the frontal cortex (Haxby et al., 2001; Oosterhof et al., 2012; Wegrzyn et al., 2015; Wurm and Lingnau, 2015). Thus, frontal areas may subserve higher-order social process by linking low-level social perception into more complex and abstract cognitive processes such as making predictions of the next actions or linking perception with the brain's affective system. Indeed, there is evidence that MFC could be responsible in giving affective meaning for the ongoing experiences and that MFC processing is highly idiosyncratic (Chang et al., 2021). Limbic regions such as amygdala and thalamus in turn have been linked with processing of (negative) emotions (Karjalainen et al., 2019) and accordingly, they showed reliable responses primarily to perception of antisocial behaviours. Finally, frontal regions and thalamus have been associated with felt, but not perceived emotions while general social hubs TPJ, STS, and LOTC associated with both perceived and felt emotions establishing distinction between emotion perception and emotional experience in the human brain (Saarimäki et al., 2023).

The present study focused on the functional organization of social perception. Still, passive observation is different from active engagement in social interaction. Recently, this difference between spectator accounts versus truly interactive models of neural basis in social cognition has been highlighted and researchers are increasingly interested in measuring real interactive social processes in the brain (Redcay and

Schilbach, 2019). An important yet challenging future question would thus be mapping the organization of the building blocks of active social interaction in the human brain.

## 5. Conclusions

Using a combination of data-driven approaches and multivariate pattern recognition we established the perceptual space for social features and mapped the cerebral topography of social perception that can be adequately described with 13 perceptual dimensions. Social perceptual space of the video stimuli included clusters of social features describing sexual & affiliative and antisocial behaviour, feeding, body movement, communication, and playfulness, as well as individual dimensions male, female, running, walking, searching, crying, and using an object. Clear gradient in response selectivity was observed from broad response profiles in temporal, occipital and parietal regions towards narrow and selective responses in frontal and subcortical regions. Perceptual social dimensions could be reliably decoded from regional activation patterns using multivariate pattern analysis. Both regression analysis and multivariate pattern analysis highlighted the importance of LOTC, STS, TPJ and FG and other occipitotemporal regions as dimension-specific social information processors, while parietal areas and Heschl's gyrus process domain-general information from the social scenes. Additionally, regional response profiles for social perception closely related to the overall reliability of the BOLD responses. Altogether these results highlight the distributed nature of social processing in the brain as well as the spatial specificity of brain regions to social dimensions.

## Data availability

The stimulus movie clips can be made available for researchers upon request, copyrights preclude public redistribution of the stimulus set. Short descriptions of each movie clip can be found in the supplementary material. According to Finnish legislation, the original (even anonymized) neuroimaging data used in the experiment cannot be released for public use. The voxelwise (unthresholded) result maps are available in NeuroVault (<https://neurovault.org/collections/IZWVFEYL/>).

## Code availability

We developed in-house scripts for low-level audiovisual feature extraction and ridge regression optimization and used previously validated packages for the clustering analysis, multivariate pattern analysis and intersubject correlation analysis. The analysis scripts are freely available in GitHub (<https://github.com/santavirta/functional-organization-of-social-perception>).

## Declaration of Competing Interest

The authors declare no competing financial or non-financial interests.

## Credit authorship contribution statement

**Severi Santavirta:** Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Tomi Karjalainen:** Conceptualization, Methodology, Software, Formal analysis, Resources, Data curation, Writing – original draft, Writing – review & editing. **Sanaz Nazari-Farsani:** Methodology, Software, Resources, Data curation, Writing – review & editing. **Matthew Hudson:** Conceptualization, Investigation, Resources, Writing – review & editing. **Vesa Putkinen:** Investigation, Resources, Writing – review & editing. **Kerttu Seppälä:** Investigation, Resources, Writing – review & editing. **Lihua Sun:** Investigation, Resources, Writing – review & editing.

**Enrico Glerean:** Methodology, Software, Writing – review & editing.  
**Jussi Hirvonen:** Investigation, Resources, Writing – review & editing.  
**Henry K. Karlsson:** Resources, Writing – review & editing. **Lauri Nummenmaa:** Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

## Acknowledgements

The study was supported by grants to LN from European Research Council (#313000) and the Academy of Finland (#332225, #294897). SNF thanks Turku University Hospital for the financial support of her postdoctoral fellowship at Turku PET Centre. We thank Tuulia Malen for the fruitful conversations around the involved analysis methods and Juha Lahnakoski for the help with low-level feature extraction.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.neuroimage.2023.120025](https://doi.org/10.1016/j.neuroimage.2023.120025).

## References

- Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn [Methods]. *Front. Neuroinform.* 8 (14). doi:[10.3389/fninf.2014.00014](https://doi.org/10.3389/fninf.2014.00014).
- Adolphs, R., Nummenmaa, L., Todorov, A., Haxby, J.V., 2016. Data-driven approaches in the investigation of social perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371 (1693). doi:[10.1098/rstb.2015.0367](https://doi.org/10.1098/rstb.2015.0367).
- Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4 (7), 267–278. doi:[10.1016/s1364-6613\(00\)01501-1](https://doi.org/10.1016/s1364-6613(00)01501-1).
- Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7 (4), 268–277. doi:[10.1038/nrn1884](https://doi.org/10.1038/nrn1884).
- Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12 (1), 26–41. doi:[10.1016/j.media.2007.06.004](https://doi.org/10.1016/j.media.2007.06.004).
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57 (1), 289–300. doi:[10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x).
- Brooks, J.A., Stoller, R.M., Freeman, J.B., 2020. Computational approaches to the neuroscience of social perception. *Soc. Cogn. Affect. Neurosci.* doi:[10.1093/scan/nsaa127](https://doi.org/10.1093/scan/nsaa127).
- Brosch, T., Bar-David, E., Phelps, E.A., 2013. Implicit race bias decreases the similarity of neural representations of black and white faces. *Psychol. Sci.* 24 (2), 160–166. doi:[10.1177/0956797612451465](https://doi.org/10.1177/0956797612451465).
- Carter, R.M., Huettel, S.A., 2013. A nexus model of the temporal-parietal junction. *Trends Cogn. Sci.* 17 (7), 328–336. doi:[10.1016/j.tics.2013.05.007](https://doi.org/10.1016/j.tics.2013.05.007).
- Cavanna, A.E., Trimble, M.R., 2006. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129 (Pt 3), 564–583. doi:[10.1093/brain/awl004](https://doi.org/10.1093/brain/awl004).
- Chang, L.J., Jolly, E., Cheong, J.H., Rapuano, K.M., Greenstein, N., Chen, P.A., Manning, J.R., 2021. Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. *Sci. Adv.* 7 (17). doi:[10.1126/sciadv.abf7129](https://doi.org/10.1126/sciadv.abf7129).
- Chen, G., Taylor, P.A., Cox, R.W., 2017. Is the statistic value all we should care about in neuroimaging? *Neuroimage* 147, 952–959. doi:[10.1016/j.neuroimage.2016.09.066](https://doi.org/10.1016/j.neuroimage.2016.09.066).
- Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29 (3), 162–173. doi:[10.1006/cbmr.1996.0014](https://doi.org/10.1006/cbmr.1996.0014).
- Cukur, T., Nishimoto, S., Huth, A.G., Gallant, J.L., 2013. Attention during natural vision warps semantic representation across the human brain. *Nat. Neurosci.* 16 (6), 763–770. doi:[10.1038/nn.3381](https://doi.org/10.1038/nn.3381).
- Da Costa, S., van der Zwaag, W., Marques, J.P., Frackowiak, R.S., Clarke, S., Saenz, M., 2011. Human primary auditory cortex follows the shape of Heschl's gyrus. *J. Neurosci.* 31 (40), 14067–14075. doi:[10.1523/JNEUROSCI.2000-11.2011](https://doi.org/10.1523/JNEUROSCI.2000-11.2011).
- Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage* 9 (2), 179–194. doi:[10.1006/nimg.1998.0395](https://doi.org/10.1006/nimg.1998.0395).
- de la Vega, A., Chang, L.J., Banich, M.T., Wager, T.D., Yarkoni, T., 2016. Large-scale meta-analysis of human medial frontal cortex reveals tripartite functional organization. *J. Neurosci.* 36 (24), 6553–6562. doi:[10.1523/JNEUROSCI.4402-15.2016](https://doi.org/10.1523/JNEUROSCI.4402-15.2016).
- Deen, B., Koldewyn, K., Kanwisher, N., Saxe, R., 2015. Functional organization of social perception and cognition in the superior temporal sulcus. *Cereb. Cortex* 25 (11), 4596–4609. doi:[10.1093/cercor/bhv111](https://doi.org/10.1093/cercor/bhv111).
- Dima, D.C., Tomita, T.M., Honey, C.J., Isik, L., 2022. Social-affective features drive human representations of observed actions. *Elife* 11. doi:[10.7554/eLife.75027](https://doi.org/10.7554/eLife.75027).
- Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. *Science* 293 (5539), 2470–2473. doi:[10.1126/science.1063414](https://doi.org/10.1126/science.1063414).
- Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S.S., Wright, J.P., Duriez, J., Poldrack, R.A., Gorgolewski, K.J., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* 16 (1), 111–116. doi:[10.1038/s41592-018-0235-4](https://doi.org/10.1038/s41592-018-0235-4).
- Felsen, G., Dan, Y., 2005. A natural approach to studying vision [Article]. *Nat. Neurosci.* 8 (12), 1643–1646. doi:[10.1038/nn1608](https://doi.org/10.1038/nn1608).
- Fonov, V.S., Evans, A.C., McKinsty, R.C., Alml, C.R., Collins, D.L., 2009. Unbiased non-linear average age-appropriate brain templates from birth to adulthood. *Neuroimage* 47, S102. doi:[10.1016/S1053-8119\(09\)70884-5](https://doi.org/10.1016/S1053-8119(09)70884-5).
- Funder, D.C., 2006. Towards a resolution of the personality triad: persons, situations, and behaviors. *J. Res. Pers.* 40 (1), 21–34. doi:[10.1016/j.jrp.2005.08.003](https://doi.org/10.1016/j.jrp.2005.08.003).
- Goldberg, L.R., 1990. An alternative "description of personality": the Big-Five factor structure. *J. Pers. Soc. Psychol.* 59 (6), 1216–1229. doi:[10.1037/0022-3514.59.6.1216](https://doi.org/10.1037/0022-3514.59.6.1216).
- Gorgolewski, K., Burns, C., Madison, C., Clark, D., Halchenko, Y., Waskom, M., Ghosh, S., 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python [original research]. *Front. Neuroinform.* 5 (13). doi:[10.3389/fninf.2011.00013](https://doi.org/10.3389/fninf.2011.00013).
- Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48 (1), 63–72. doi:[10.1016/j.neuroimage.2009.06.060](https://doi.org/10.1016/j.neuroimage.2009.06.060).
- Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., Pollmann, S., 2009. PyMVP: a python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7 (1), 37–53. doi:[10.1007/s12021-008-9041-y](https://doi.org/10.1007/s12021-008-9041-y).
- Hari, R., Henriksson, L., Malinen, S., Parkkonen, L., 2015. Centrality of social interaction in human brain function. *Neuron* 88 (1), 181–193. doi:[10.1016/j.neuron.2015.09.022](https://doi.org/10.1016/j.neuron.2015.09.022).
- Harry, B., Williams, M.A., Davis, C., Kim, J., 2013. Emotional expressions evoke a differential response in the fusiform face area. *Front. Hum. Neurosci.* 7, 692. doi:[10.3389/fnhum.2013.00692](https://doi.org/10.3389/fnhum.2013.00692).
- Hartwigsen, G., Baumgaertner, A., Price, C.J., Koehnke, M., Ulmer, S., Siebner, H.R., 2010. Phonological decisions require both the left and right supramarginal gyri. *Proc. Natl. Acad. Sci. U. S. A.* 107 (38), 16494–16499. doi:[10.1073/pnas.1008121107](https://doi.org/10.1073/pnas.1008121107).
- Hasson, U., Malach, R., Heeger, D.J., 2010. Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* 14 (1), 40–48. doi:[10.1016/j.tics.2009.10.011](https://doi.org/10.1016/j.tics.2009.10.011).
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. *Science* 303 (5664), 1634–1640. doi:[10.1126/science.1089506](https://doi.org/10.1126/science.1089506).
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293 (5539), 2425–2430. doi:[10.1126/science.1063736](https://doi.org/10.1126/science.1063736).
- Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. *Trends Cogn. Sci. (Regul. Ed.)* 4 (6), 223–233. doi:[10.1016/s1364-6613\(00\)01482-0](https://doi.org/10.1016/s1364-6613(00)01482-0).
- Heikkilä, T.T., Laine, O., Savela, J., & Nummenmaa, L., December 4). (2020). *Onni: an online experiment platform for research*. In (Version 1.0) Zenodo.
- Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12 (1), 55–67. doi:[10.1080/00401706.1970.10488634](https://doi.org/10.1080/00401706.1970.10488634).
- Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532 (7600), 453–458. doi:[10.1038/nature17637](https://doi.org/10.1038/nature17637).
- Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76 (6), 1210–1224. doi:[10.1016/j.neuron.2012.10.014](https://doi.org/10.1016/j.neuron.2012.10.014).
- Isik, L., Koldewyn, K., Beeler, D., Kanwisher, N., 2017. Perceiving social interactions in the posterior superior temporal sulcus. *Proc. Natl. Acad. Sci. U. S. A.* 114 (43), E9145–E9152. doi:[10.1073/pnas.1714471114](https://doi.org/10.1073/pnas.1714471114).
- Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17 (2), 825–841. doi:[10.1006/nimg.2002.1132](https://doi.org/10.1006/nimg.2002.1132).
- Karjalainen, T., Karlsson, H.K., Lahnakoski, J.M., Glerean, E., Nuutila, P., Jaaskelainen, I.P., Hari, R., Sams, M., Nummenmaa, L., 2017. Dissociable roles of cerebral mu-opioid and type 2 dopamine receptors in vicarious pain: a combined PET-fMRI study. *Cereb. Cortex* 27 (8), 4257–4266. doi:[10.1093/cercor/bhx129](https://doi.org/10.1093/cercor/bhx129).
- Karjalainen, T., Seppälä, K., Glerean, E., Karlsson, H.K., Lahnakoski, J.M., Nuutila, P., Jaaskelainen, I.P., Hari, R., Sams, M., Nummenmaa, L., 2019. Opioidergic regulation of emotional arousal: a combined PET-fMRI study. *Cereb. Cortex* 29 (9), 4006–4016. doi:[10.1093/cercor/bhy281](https://doi.org/10.1093/cercor/bhy281).
- Kauppi, J.P., Pajula, J., Tohka, J., 2014. A versatile software package for inter-subject correlation based analyses of fMRI. *Front. Neuroinform.* 8, 2. doi:[10.3389/fninf.2014.00002](https://doi.org/10.3389/fninf.2014.00002).
- Klein, A., Ghosh, S.S., Bao, F.S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., Keshavan, A., 2017. Mindboggling morphometry of human brains. *PLoS Comput. Biol.* 13 (2), e1005350. doi:[10.1371/journal.pcbi.1005350](https://doi.org/10.1371/journal.pcbi.1005350).
- Koenigs, M., Barbey, A.K., Postle, B.R., Grafman, J., 2009. Superior parietal cortex is critical for the manipulation of information in working memory. *J. Neurosci.* 29 (47), 14980–14986. doi:[10.1523/JNEUROSCI.3706-09.2009](https://doi.org/10.1523/JNEUROSCI.3706-09.2009).
- Koide-Majima, N., Nakai, T., Nishimoto, S., 2020. Distinct dimensions of emotion in the human brain and their representation on the cortical surface. *Neuroimage* 222, 117258. doi:[10.1016/j.neuroimage.2020.117258](https://doi.org/10.1016/j.neuroimage.2020.117258).
- Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15 (2), 155–163. doi:[10.1016/j.jcm.2016.02.012](https://doi.org/10.1016/j.jcm.2016.02.012).
- Lahnakoski, J.M., Glerean, E., Jaaskelainen, I.P., Hyona, J., Hari, R., Sams, M., Nummenmaa, L., 2014. Synchronous brain activity across individuals underlies shared psychological perspectives. *Neuroimage* 100, 316–324. doi:[10.1016/j.neuroimage.2014.06.022](https://doi.org/10.1016/j.neuroimage.2014.06.022).

- Lahnakoski, J.M., Glerean, E., Salmi, J., Jaaskelainen, I.P., Sams, M., Hari, R., Nummenmaa, L., 2012. Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Front. Hum. Neurosci.* 6, 233. doi:[10.3389/fnhum.2012.00233](https://doi.org/10.3389/fnhum.2012.00233).
- Lartillot, O., Toivianen, P., Eerola, T., 2008. A Matlab Toolbox for Music Information Retrieval. In: Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R. (Eds.), *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg.
- Lee, K., Ashton, M.C., 2004. Psychometric properties of the HEXACO personality inventory. *Multivariate Behav. Res.* 39 (2), 329–358. doi:[10.1207/s15327906mbr3902\\_8](https://doi.org/10.1207/s15327906mbr3902_8).
- Lewin, K., 1936. *Principles of Topological Psychology*. McGraw-Hill.
- Lingnau, A., Downing, P.E., 2015. The lateral occipitotemporal cortex in action. *Trends Cogn. Sci.* 19 (5), 268–277. doi:[10.1016/j.tics.2015.03.006](https://doi.org/10.1016/j.tics.2015.03.006).
- McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1 (1), 30–46. doi:[10.1037/1082-989x.1.1.30](https://doi.org/10.1037/1082-989x.1.1.30).
- Nummenmaa, L., Calder, A.J., 2009. Neural mechanisms of social attention [Review]. *Trends Cogn. Sci. (Regul. Ed.)* 13 (3), 135–143. doi:[10.1016/j.tics.2008.12.006](https://doi.org/10.1016/j.tics.2008.12.006).
- Nummenmaa, L., Glerean, E., Viinikainen, M., Jaaskelainen, I.P., Hari, R., Sams, M., 2012. Emotions promote social interaction by synchronizing brain activity across individuals. *Proc. Natl. Acad. Sci. U. S. A.* 109 (24), 9599–9604. doi:[10.1073/pnas.1206095109](https://doi.org/10.1073/pnas.1206095109).
- Nummenmaa, L., Hyönä, J., Calvo, M.G., 2010. Semantic categorization precedes affective evaluation of visual scenes. *J. Exp. Psychol. Gen.* 139 (2), 222–246. doi:[10.1037/a0018858](https://doi.org/10.1037/a0018858).
- Nummenmaa, L., Lukkarinen, L., Sun, L., Putkinen, V., Seppälä, K., Karjalainen, T., Karlsson, H.K., Hudson, M., Venetjoki, N., Salomaa, M., Rautio, P., Hirvonen, J., Lauerma, H., Tiihonen, J., 2021. Brain Basis of Psychopathy in Criminal Offenders and General Population. *Cereb. Cortex* 31 (9), 4104–4114. doi:[10.1093/cercor/bhab072](https://doi.org/10.1093/cercor/bhab072).
- Nummenmaa, L., Saarimäki, H., Glerean, E., Gotsopoulos, A., Jaaskelainen, I.P., Hari, R., Sams, M., 2014. Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. *Neuroimage* 102 (Pt 2), 498–509. doi:[10.1016/j.neuroimage.2014.07.063](https://doi.org/10.1016/j.neuroimage.2014.07.063).
- Oosterhof, N.N., Tipper, S.P., Downing, P.E., 2012. Viewpoint (in)dependence of action representations: an MVPA study. *J. Cogn. Neurosci.* 24 (4), 975–989. doi:[10.1162/jocn\\_a.00195](https://doi.org/10.1162/jocn_a.00195).
- Parrigon, S., Woo, S.E., Tay, L., Wang, T., 2017. CAPTION-ing the situation: a lexically-derived taxonomy of psychophysical situation characteristics. *J. Pers. Soc. Psychol.* 112 (4), 642–681. doi:[10.1037/pspp0000111](https://doi.org/10.1037/pspp0000111).
- Popham, S.F., Huth, A.G., Bilenko, N.Y., Deniz, F., Gao, J.S., Nunez-Elizalde, A.O., Gallant, J.L., 2021. Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nat. Neurosci.* 24 (11), 1628–1636. doi:[10.1038/s41593-021-00921-6](https://doi.org/10.1038/s41593-021-00921-6).
- 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 (2), 816–847. doi:[10.1016/j.neuroimage.2012.04.062](https://doi.org/10.1016/j.neuroimage.2012.04.062).
- Pruim, R.H.R., Mennes, M., van Rooij, D., Llera, A., Buitelaar, J.K., Beckmann, C.F., 2015. ICA-AROMA: a robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* 112, 267–277. doi:[10.1016/j.neuroimage.2015.02.064](https://doi.org/10.1016/j.neuroimage.2015.02.064).
- Rauthmann, J.F., Gallardo-Pujol, D., Guillaume, E.M., Todd, E., Nave, C.S., Sherman, R.A., Ziegler, M., Jones, A.B., Funder, D.C., 2014. The Situational Eight DIAMONDS: a taxonomy of major dimensions of situation characteristics. *J. Pers. Soc. Psychol.* 107 (4), 677–718. doi:[10.1037/a0037250](https://doi.org/10.1037/a0037250).
- Redcay, E., Schilbach, L., 2019. Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nat. Rev. Neurosci.* 20 (8), 495–505. doi:[10.1038/s41583-019-0179-4](https://doi.org/10.1038/s41583-019-0179-4).
- Rolls, E.T., Joliot, M., Tzourio-Mazoyer, N., 2015. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *Neuroimage* 122, 1–5. doi:[10.1016/j.neuroimage.2015.07.075](https://doi.org/10.1016/j.neuroimage.2015.07.075).
- Russell, J.A., 1980. A Circumplex Model of Affect. *J. Pers. Soc. Psychol.* 39 (6), 1161–1178. doi:[10.1037/h0077714](https://doi.org/10.1037/h0077714).
- Saarimäki, H., Nummenmaa, L., Volynets, S., Santavirta, S., Aksiuto, A., Sams, M., Jaaskelainen, I.P., & Lahnakoski, J. (2023). Cerebral Topographies of Perceived and Felt Emotions. *bioRxiv preprint*. <https://doi.org/10.1101/2023.02.08.521183>
- Said, C.P., Moore, C.D., Engell, A.D., Todorov, A., Haxby, J.V., 2010. Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J. Vis.* 10 (5), 11. doi:[10.1167/10.5.11](https://doi.org/10.1167/10.5.11).
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind. *Neuroimage* 19 (4), 1835–1842. doi:[10.1016/s1053-8119\(03\)00230-1](https://doi.org/10.1016/s1053-8119(03)00230-1).
- Simms, L.J., 2007. The big seven model of personality and its relevance to personality pathology. *J. Pers.* 75 (1), 65–94. doi:[10.1111/j.1467-6494.2006.00433.x](https://doi.org/10.1111/j.1467-6494.2006.00433.x).
- Smirnov, D., Lachat, F., Peltola, T., Lahnakoski, J.M., Koistinen, O.-P., Glerean, E., Vehtari, A., Hari, R., Sams, M., Nummenmaa, L., 2017. Brain-to-brain hyperclassification reveals action-specific motor mapping of observed actions in humans. *PLoS One* 12 (12), e0189508. doi:[10.1371/journal.pone.0189508](https://doi.org/10.1371/journal.pone.0189508).
- Stoeckel, C., Gough, P.M., Watkins, K.E., Devlin, J.T., 2009. Supramarginal gyrus involvement in visual word recognition. *Cortex* 45 (9), 1091–1096. doi:[10.1016/j.cortex.2008.12.004](https://doi.org/10.1016/j.cortex.2008.12.004).
- Tarhan, L., Konkle, T., 2020. Sociality and interaction envelope organize visual action representations. *Nat. Commun.* 11 (1), 3002. doi:[10.1038/s41467-020-16846-w](https://doi.org/10.1038/s41467-020-16846-w).
- Thornton, M.A., Tamir, D.I., 2022. Six dimensions describe action understanding: the ACT-FAST taxonomy. *J. Pers. Soc. Psychol.* 122 (4), 577–605. doi:[10.1037/pspa0000286](https://doi.org/10.1037/pspa0000286).
- Tong, F., Pratte, M.S., 2012. Decoding patterns of human brain activity. *Annu. Rev. Psychol.* 63, 483–509. doi:[10.1146/annurev-psych-120710-100412](https://doi.org/10.1146/annurev-psych-120710-100412).
- Tucciarelli, R., Wurm, M., Baccolo, E., Lingnau, A., 2019. The representational space of observed actions. *Elife* 8. doi:[10.7554/eLife.47686](https://doi.org/10.7554/eLife.47686).
- Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 Bias Correction. *IEEE Trans. Med. Imaging* 29 (6), 1310–1320. doi:[10.1109/TMI.2010.2046908](https://doi.org/10.1109/TMI.2010.2046908).
- Van Essen, D.C., 2012. Cortical cartography and Caret software. *Neuroimage* 62 (2), 757–764. doi:[10.1016/j.neuroimage.2011.10.077](https://doi.org/10.1016/j.neuroimage.2011.10.077).
- Walbrin, J., Downing, P., Koldewyn, K., 2018. Neural responses to visually observed social interactions. *Neuropsychologia* 112, 31–39. doi:[10.1016/j.neuropsychologia.2018.02.023](https://doi.org/10.1016/j.neuropsychologia.2018.02.023).
- Webb, M.A., Tangney, J.P., 2022. Too good to be true: bots and bad data from mechanical turk. *Perspect. Psychol. Sci.* 17456916221120027. doi:[10.1177/17456916221120027](https://doi.org/10.1177/17456916221120027).
- Wegrzyn, M., Riehle, M., Labudda, K., Woermann, F., Baumgartner, F., Pollmann, S., Bien, C.G., Kissler, J., 2015. Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex* 69, 131–140. doi:[10.1016/j.cortex.2015.05.003](https://doi.org/10.1016/j.cortex.2015.05.003).
- Wilkerson, M.D., Hayes, D.N., 2010. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26 (12), 1572–1573. doi:[10.1093/bioinformatics/btq170](https://doi.org/10.1093/bioinformatics/btq170).
- Willis, J., Todorov, A., 2006. First impressions: making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* 17 (7), 592–598. doi:[10.1111/j.1467-9280.2006.01750.x](https://doi.org/10.1111/j.1467-9280.2006.01750.x).
- Winkler, A.M., Ridgway, G.R., Webster, M.A., Smith, S.M., Nichols, T.E., 2014. Permutation inference for the general linear model. *Neuroimage* 92, 381–397. doi:[10.1016/j.neuroimage.2014.01.060](https://doi.org/10.1016/j.neuroimage.2014.01.060).
- Wurm, M.F., Caramazza, A., 2019. Lateral occipitotemporal cortex encodes perceptual components of social actions rather than abstract representations of sociality. *Neuroimage* 202, 116153. doi:[10.1016/j.neuroimage.2019.116153](https://doi.org/10.1016/j.neuroimage.2019.116153).
- Wurm, M.F., Caramazza, A., Lingnau, A., 2017. Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *J. Neurosci.* 37 (3), 562–575. doi:[10.1523/JNEUROSCI.1717-16.2016](https://doi.org/10.1523/JNEUROSCI.1717-16.2016).
- Wurm, M.F., Lingnau, A., 2015. Decoding actions at different levels of abstraction. *J. Neurosci.* 35 (20), 7727–7735. doi:[10.1523/JNEUROSCI.0188-15.2015](https://doi.org/10.1523/JNEUROSCI.0188-15.2015).
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20 (1), 45–57. doi:[10.1109/42.906424](https://doi.org/10.1109/42.906424).