Functional organization of social perception in the human brain

Severi Santavirta[1], Tomi Karjalainen[1], Sanaz Nazari-Farsani[1], Matthew Hudson[1,2], Vesa Putkinen[1], Kerttu Seppälä[1,3], Lihua Sun[1], Enrico Glerean[6,7], Jussi Hirvonen[4,5], Henry K. Karlsson[1], Lauri Nummenmaa[1,8]

[1]Turku PET Centre, University of Turku and Turku University Hospital, Turku, Finland
[2]School of Psychology, University of Plymouth, Plymouth, United Kingdom
[3]Department of Medical Physics, Turku University Hospital, Turku, Finland
[4]Department of Radiology, University of Turku and Turku University Hospital, Turku, Finland
[5]Medical Imaging Center, Department of Radiology, Tampere University and Tampere University Hospital, Tampere, Finland
[6]Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo, Finland
[7]International Laboratory of Social Neurobiology, Institute of Cognitive Neuroscience, National Research University Higher School of Economics, Moscow, Russian Federation
[8]Department of Psychology, University of Turku, Turku, Finland

Address correspondence to
Severi Santavirta
Turku PET Centre c/o Turku University
Kiinamyllynkatu 4-6, 20520 Turku, Finland
Fax +358 2 231 8191
Email: svtsan@utu.fi

Abstract

Humans can readily perceive a multitude of features from social interactions, but the phenomenological and neural basis of social perception has yet to be solved. Short film clips with rich social content were shown to 97 healthy participants while their haemodynamic brain activity was measured with fMRI. The stimulus clips were annotated for 112 social features yielding the initial stimulus model. Cluster analysis revealed that 13 dimensions were sufficient for describing the social perceptual space. Univariate GLM using these dimensions as predictors was used to map regional neural response profiles to different social features. Multivariate pattern analysis was then utilized to establish the regional specificity of the responses. The results revealed a posterior-anterior gradient in the processing of social information in the brain. Occipital and temporal regions responded to most social dimensions and the classifier revealed that these responses were dimension specific; in contrast Heschl gyri and parietal areas were also broadly tuned to different social signals yet the responses were domain-general and did not differentiate between dimensions. Altogether these results highlight the distributed nature of social processing in the brain as well as the specific contributions of feature-specific versus domain-general social perceptual processes.

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? Functional neuroimaging studies have successfully mapped the brain circuits supporting various dimensions of the social world [1]. Numerous studies have localized specific aspects of social perception into different brain regions. Fusiform gyrus (FG) is involved in the perception of faces (FG) [2] and lateral occipitotemporal cortex (LOTC) in the perception of bodies [3]. Temporoparietal junction (TPJ) in turn is involved in reflecting the mental states of others [4] as well as processing social context and focusing attention [5]. Polysensory areas in the superior temporal sulcus (STS) have been associated with multiple higher-order aspects of social perception [6-9], while medial frontal cortex (MFC) has been extensively studied in the context of self-representation and theory of mind [10]. 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) [11].

Humans can reliably recognize hundreds of types of social events and features 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 limits of the human brain, it is unlikely that all features and dimensions of the social domain are processed by distinct areas and systems [12]. Even though the brain basis of perceiving isolated social features has been previously studied, the phenomenological as well as neural organization of the different social perceptual processes have remained poorly understood. Social features such as facial identities, body movements, and verbal and nonverbal communication often overlap with distinct temporal occurrence patterns. Because neural responses to social features cannot necessarily be predicted on statistical combination of responses to simple stimuli [13], studies based on neural responses to isolated social features may not directly generalize to real-world social perception [14]. In other perceptual domains, 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 assess the representational similarities of the modelled dimensions, or the representational similarities of the brain activation patterns associated with each dimension. Such semantic brain representations have been mapped for visual objects and actions [15], language [16], and experienced emotions [17]. A recent study found that language based representations of social features converge with the visual representations on the border of occipital cortex and that language based representations are located anterior to the visual

representations [18]. Still, a detailed representational space for social features at both phenomenological and neural level is currently lacking.

Commonly applied univariate analyses fail to address the specificity of resultant brain activation patterns. In other words, univariate models do not allow testing whether different perceptual dimensions can be reliably discriminated based on their brain activation patterns. Multivariate pattern analysis (MVPA) allows the analysis of information carried by fine-grained spatial patterns of brain activation [19]. Pattern recognition studies have established that regional multivariate patterns allow distinguishing brain activation related to multiple high-level social features such as faces [20] and their racial group [21] in FG and facial expressions in FG and STS [22-24]. Perception of different goal-oriented 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-oriented actions, not just their low-level visual properties [25]. Furthermore, decoding of goal-oriented actions was successful in LOTC when subjects observed the actions in both first and third person perspectives [26]. It however remains unresolved how specific these regional response profiles are across different social perceptual features.

The current study

In this fMRI study, we mapped the phenomenological and neural representations for social perception using both univariate and multivariate pattern recognition analyses. We aimed at first establishing a taxonomy of the social dimensions that human observers use for describing social scenarios, and then mapping the brain basis of this perceptual space. We modelled real world social perception using short socioemotionally rich film clips as stimuli, as films contain rich and complex social scenarios and powerful emotional episodes. Furthermore, cinema is well-known for eliciting strong and consistent neural responses in functional imaging studies [27,28]. We first mapped the phenomenological representational space of social processes based on subjective annotations of a large array (N=112) of social features in the films. 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 univariate regression analysis, we established an anterior-posterior gradient in social perception in the brain, while multivariate pattern analysis revealed regional sensitivity for different social dimensions.

Materials and methods

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. After preprocessing of functional neuroimaging data, visual quality control identified three subjects whose fMRI data contained artefacts also visible in subject-level regression coefficient estimates; these subjects were also excluded. 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.

Experimental design for fMRI

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 [28-31]. Briefly, the subjects viewed a medley of 96 movie clips (median duration 11.2 seconds, range 5.3 – 28.2 seconds, total duration 19 min 44 seconds) that have been curated to contain large variability of social and emotional content. All videos were extracted from mainstream Hollywood movies with audio track in English. Because this task is designed to map neural processing of naturalistic socioemotional events, the clips are 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. Subjects were instructed to view the movies similarly as if they were viewing a movie at a cinema or at home, i.e. 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.

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^3 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 bandwidth, 3.0 mm slice thickness, 45 interleaved axial slices acquired in ascending order without gaps).

The functional imaging data were preprocessed with FMRIPREP [32] (v1.3.0), a Nipype [33] based tool that internally uses Nilearn [34]. During the preprocessing, each T1w volume was corrected for intensity non-uniformity using N4BiasFieldCorrection (v2.1.0) [35] 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) [36], and the brain mask 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 [37]. Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c [38] was performed through nonlinear registration with the antsRegistration (ANTs v2.1.0) [39], 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 [40] (FSL v5.0.9).

Functional data were slice-time-corrected using 3dTshift from AFNI [41] (v16.2.07) and motion-corrected using MCFLIRT [42] (FSL v5.0.9). These steps were followed by co-registration to the T1w image using boundary-based registration [43] with six degrees of freedom, using bbregister (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 [44]. The data were then detrended using 240-s-Savitzky–Golay filtering to remove the scanner drift [45], and finally downsampled to original 3 mm isotropic voxel size. The BOLD signals were demeaned to make the regression coefficients comparable across different individuals [46] which also downscales the absolute coefficient values. First and last two functional volumes were discarded to ensure equilibrium effects and to exclude the timepoint after the stimulus had ended.

Stimulus features

Five individuals not participating in fMRI rated 112 predefined features (see **Table SI-1**) from the film clips. We selected a broad range of socioemotional features ranging from sensory input (e.g. smelling, tasting), basic body functions (e.g. facial expressions, walking, eating), person characteristics (e.g. pleasantness, trustworthiness) and person's inner states (e.g. pleasant feeling, arousal) to social interaction signals (e.g. talking, communicating with gestures) and social interaction characteristics (e.g. hostility, sexuality). Observers were stressed 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 films). The ratings were collected separately for every video clip in short time-intervals (median 4.0 sec, range: 3.1 – 7.3 seconds). Features were annotated on a continuous scale from 0 (absent) to 100 (present). Annotators watched the video clips altogether 12 times, rating an average of 10 dimensions 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 [47].

Feature selection

We first evaluated whether the *a priori* features were frequently and consistently perceived in the stimulus films. We first excluded features with low occurrence rate and/or inter-rater reliability, because i) high occurrence rate is needed to reliably estimate the stimulus-dependent variation in BOLD signal, and ii) high inter-rater reliability is necessary to ensure that the features were perceived uniformly across the raters. The occurrence rate was defined as the number of datapoints where the mean (minus standard error of the mean) of the rating exceeded 5 (on a scale ranging from 0 to 100). 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). ICCs below 0.5 are considered poor [48], and we thus only included features with ICC > 0.5. 45 features satisfied both criteria. The occurrence rate and inter-rater reliability of each feature are shown in **Figure SI-2**.

Dimension reduction

The selected 45 features were linearly correlated (**Figure 1**), which would affect the modelling of the BOLD-fMRI data [28]. We thus 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. Weighted pair group method with arithmetic mean (WPGMA), as implemented in R, was used as the clustering algorithm [49]. Other average linkage clustering methods

implemented in the R package (UPGMA, WPGMC and UPGMC) yielded highly similar clustering results. After clustering, feature ratings were averaged across the features in each cluster and used as regressors in the BOLD-GLM analysis. The features not assigned to any cluster were modelled independently. Hierarchical clustering automatically defines cluster boundaries from hierarchical tree (**Figure SI-3**) but requires a desired number of resulting clusters as an input. The goal was to find a minimum number of clusters which satisfied two criteria: theoretically meaningful clustering yielding sufficient reduction of collinearity between the clusters. Based on these criteria, we decided to reduce the number of clusters until maximum pairwise correlation between any two regressors of the model was below 0.4. This resulted in 6 clusters and 7 independent features not belonging to any cluster (**Figure 1**). The clustering result was not theoretically meaningful if the maximum correlation limit was set lower than 0.4.

General linear model analysis of the BOLD-fMRI data

Linear modelling of BOLD-fMRI was first conducted separately for each selected feature (N=45) to assess whether perceived similarity (as determined by clustering) of the features would be associated with similarity of BOLD-fMRI activation patterns. BOLD data were modelled by each feature by convolving the feature ratings with canonical HRF and adding nuisance covariates (brightness of the video frames, the intensity of the audio tracks, their time-derivatives as well as detrended mean signals from cerebrospinal fluid and white matter) into the design matrix. Modelling was implemented with linear model (GLM) using SPM12 (Wellcome Trust Center for Imaging, London, UK, https://www.fil.ion.ucl.ac.uk/spm/). Linear models were fit to every subject's fMRI data and group level mean contrast images were calculated (unthresholded β-images). Finally, the representational similarities of social feature ratings were compared to the similarities in the corresponding feature-specific spatial brain activation patterns.

Ridge regression analysis of the main dimensions of social perception

Ridge regression [50] was used to estimate the contributions of the cluster-based composite social dimensions to the BOLD–fMRI signals for each subject. Ridge regression was preferred over ordinary least squares (OLS) regression because even after dimension reduction, the 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–fMRI signals. The ridge parameter was optimised for the whole model (13 perceptual dimensions + 6 nuisance covariates) using leave-one-subject-out cross-validation. Prior to statistical modelling the perceptual dimension time series were convolved with the canonical HRF. The columns of the design matrix were

standardised ($\mu = 0$, $\sigma = 1$) and constant term was also included. Detailed description of ridge regression modelling is included in supplementary materials (**Figure SI-1**).

Ridge regularization parameter optimisation is computationally prohibitive when considering all the voxels in the brain. We used a uniformly distributed sample of grey matter voxels for the optimisation. For selecting a representative sample of grey matter voxels, we first generated a population level grey matter mask by averaging over individual grey matter probability maps and voxels with over 0.5 population level probability were considered as grey matter. Second, we wanted to make sure that only grey matter voxels where each subject has BOLD signal were included. To accomplish this, we calculated a population level EPI-mask based on FMRIPREP derived individual EPI-masks and finally 20% of grey matter voxels within the EPI-mask were randomly identified. Approximately ~5000 voxels were selected for ridge parameter optimization. We developed in-house scripts for ridge regression optimization, and they are freely available in GitHub (https://github.com/santavis/fmri-ridge-regression). The population level EPI-mask was used in all further analyses to include only voxels with 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.

For each social dimension, the regression analysis was run both at voxel-level and at region-of-interest (ROI) level. In voxel-level analysis, subject-level β-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 [51]. Voxel-level FDR with q-value of 0.05 was used to correct for multiple comparisons [52]. Anatomical ROIs were extracted from AAL2 atlas [53]. ROIs, where at least 50% of voxels were outside the population level EPI mask, were excluded from the analysis. This resulted in inclusion of 41 bilateral ROIs into the analysis. A parametric T-test on the β-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.

Inter-subject 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 measured with inter-subject correlation (ISC) analysis [54]. While the only variable factor in the

experiment is the audiovisual stimulus, ISC analysis captures the shared stimulus-dependent activation in the brain. We wanted to test whether the brain regions showing domain-specific versus domain-general response profiles show typically high versus low inter-subject response reliability, respectively. To this end, we calculated the ISC across subjects over the whole experiment. We used the ISC-toolbox with default settings for ISC calculations [55].

Multivariate pattern analysis of social dimensions

Finally, to reveal regional specialization in processing of different social features, between-subject classification of 11 perceptual dimensions[1] was performed in Python using the PyMVPA toolbox [56]. For classification, only one dimension label for each time point could be given. Although, in continuous signal, the choice of the best label for each time point was not always obvious, because more than one feature could be present simultaneously. 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. 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 above zero.

Classifying every timepoint separately is computationally prohibitive and single EPI scans are noisy. Moreover, it cannot be assumed that adjacent timepoints assigned with the same label would be independent from each other. Accordingly, we split the data into 29 time windows and all timepoints with the same label within a time window were considered as a single event of that class. The time window boundaries were adjusted so that adjacent timepoints with the same label would not be interspersed to different time windows because temporal autocorrelation of adjacent timepoints may yield in artificial increase of the classification accuracy. After adjustment, the average time window length was 39 seconds (range: 34 sec – 49 sec). Altogether the data consisted of 87 events for each subject. These events include Using an object: 16, Communication 15, Antisocial behaviour: 11, Feeding: 10, Walking: 9, Prosocial behaviour: 8, Body movement: 5, Crying: 4, Play: 4, Running: 3 and Searching: 2.

A GLM was fit for each time window and social dimension and the resulting subjectwise β-images were used as input for the multivariate pattern analysis (MVPA). The input BOLD time series were

[1] 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.

normalized ($\mu = 0$, $\sigma = 1$) before application of the GLM. A Neural Network (NN) model 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. This analysis was performed using whole-brain data (with non-brain voxel masked out) and within the ROIs used in linear univariate analysis. In the whole-brain MVPA, an ANOVA feature selection was applied to the training set within each cross-validation where 3000 voxels with the highest F-score were selected.

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 for the classification accuracy, the following procedure was repeated 300 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 300 iterations. The null distribution estimation was computationally prohibitive as one iteration took several hours, and we decided that 300 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 center 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.

Results

How people perceive the social world?

A total of 45 out of the 112 socioemotional features had sufficient inter-rater reliability and occurrence rate (see **Figure SI-3**). Hierarchical clustering identified 6 clusters that were labelled as "Antisocial behaviour", "Prosocial 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”. **Figure 1** (left panel) 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) and median VIF value was 1.3. These diagnostics indicate that regression coefficients for the dimensions will be stable in linear model estimations and, and they could thus be included in the same model.

Cerebral topography of social perception

We then tested whether representational structure of the perceived social features is associated with the representational structure of the feature-specific brain activation patterns. General linear model was fitted separately for each selected feature (N=45) and then the spatial correlations between feature-specific activation patterns (uncorrected β-maps) were calculated. **Figure 1** (right panel) shows the correlations between the activity patterns evoked by each dimension. The correlation structure between the perceptual and neural activation patterns was similar (Spearman r = 0.64).

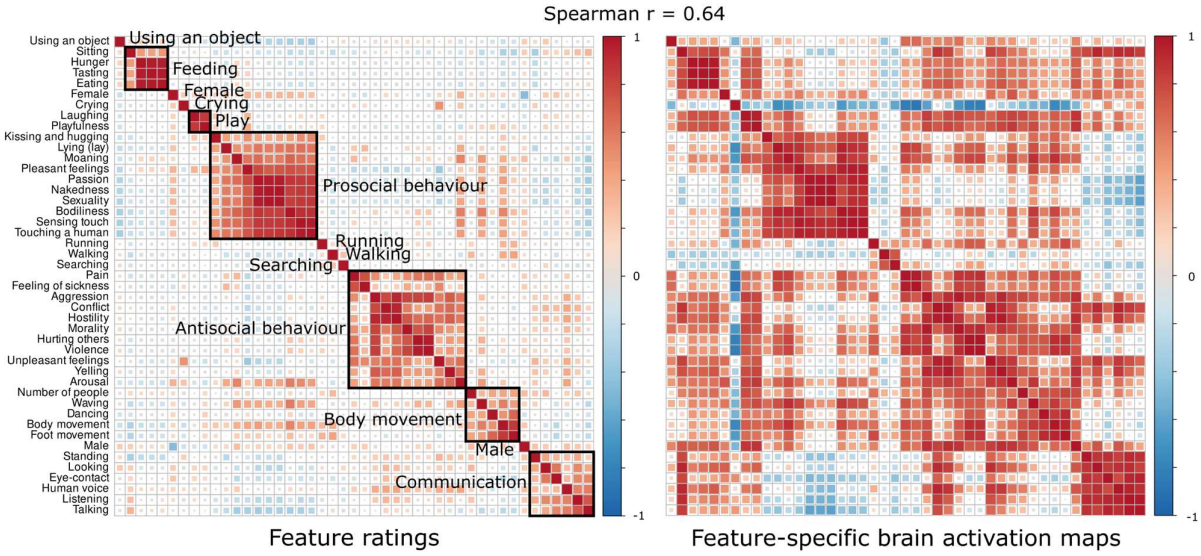

Figure 1. *The results of hierarchical clustering of features and similarities of feature ratings and feature-specific brain activation patterns. Matrices are arranged per clustering solution for the selected perceived features.*

After dimension reduction, we used regularized Ridge regression to establish the dimension-specific full-volume activation patterns (**Figure 2**). ROI analysis (**Figure 3**) revealed that social information processing engaged all brain lobes and both cortical and subcortical regions. Robust responses were

observed in occipital, temporal, and parietal cortices. There was a clear gradient in the responses, such that 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 prosocial behaviour and antisocial behaviour.

In ROI analysis, Heschl gyrus was the only region which showed statistically significant responses for all 13 perceptual dimensions. This region responded most strongly to communication, and this was also the strongest observed positive association throughout the brain. Negative associations between haemodynamic activity and mainly auditory social dimensions communication and crying were observed in occipital lobe. Parietal lobe, superior parietal gyrus, precuneus and supramarginal gyrus showed associations for wide range of perceptual dimensions. In frontal regions the observed associations between social dimensions and haemodynamic activity were weaker than in temporal, occipital and parietal regions while still statistically significant in many regions. Prosocial behaviour was positively associated with brain activity in all frontal regions; similarly antisocial behaviour also elicited consistent activation patterns in most frontal regions. For subcortical regions the observed associations were weak. The largest subcortical associations with perceptual dimensions were seen in amygdala, insula, and thalamus.

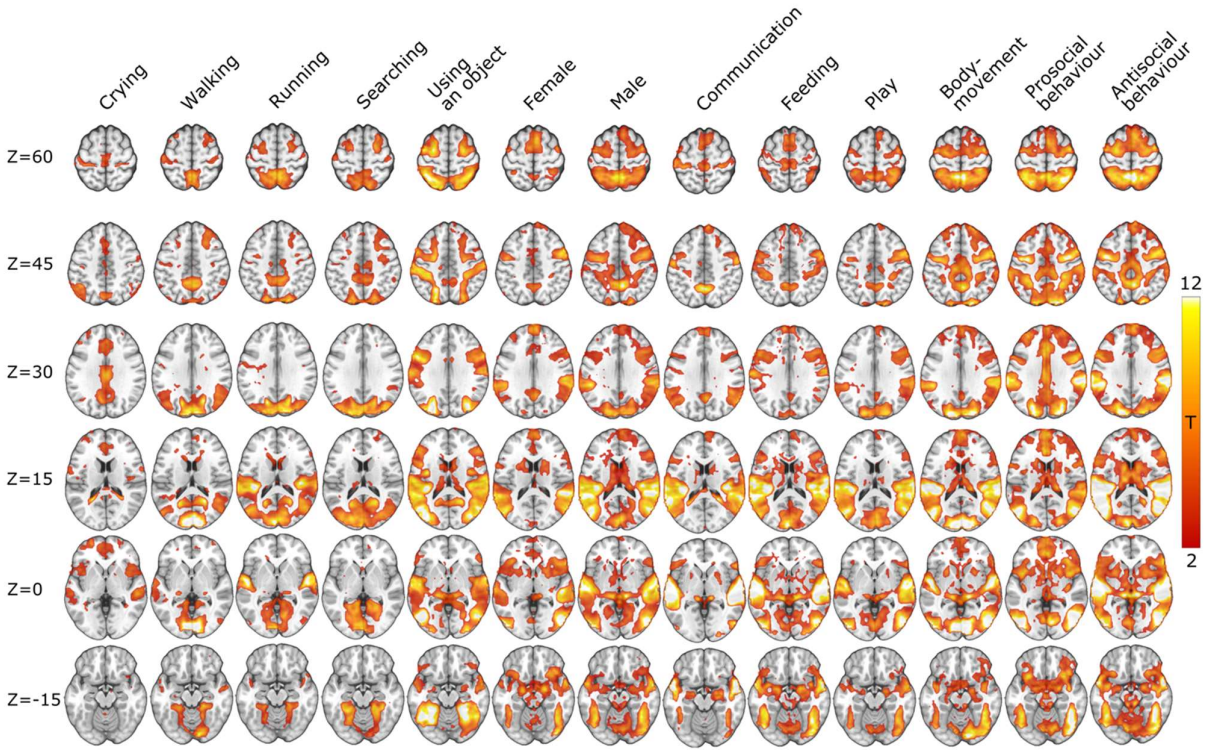

Figure 2. *Brain regions showing increased BOLD activity for the social dimensions. Data are thresholded at p < 0.05, FDR corrected.*

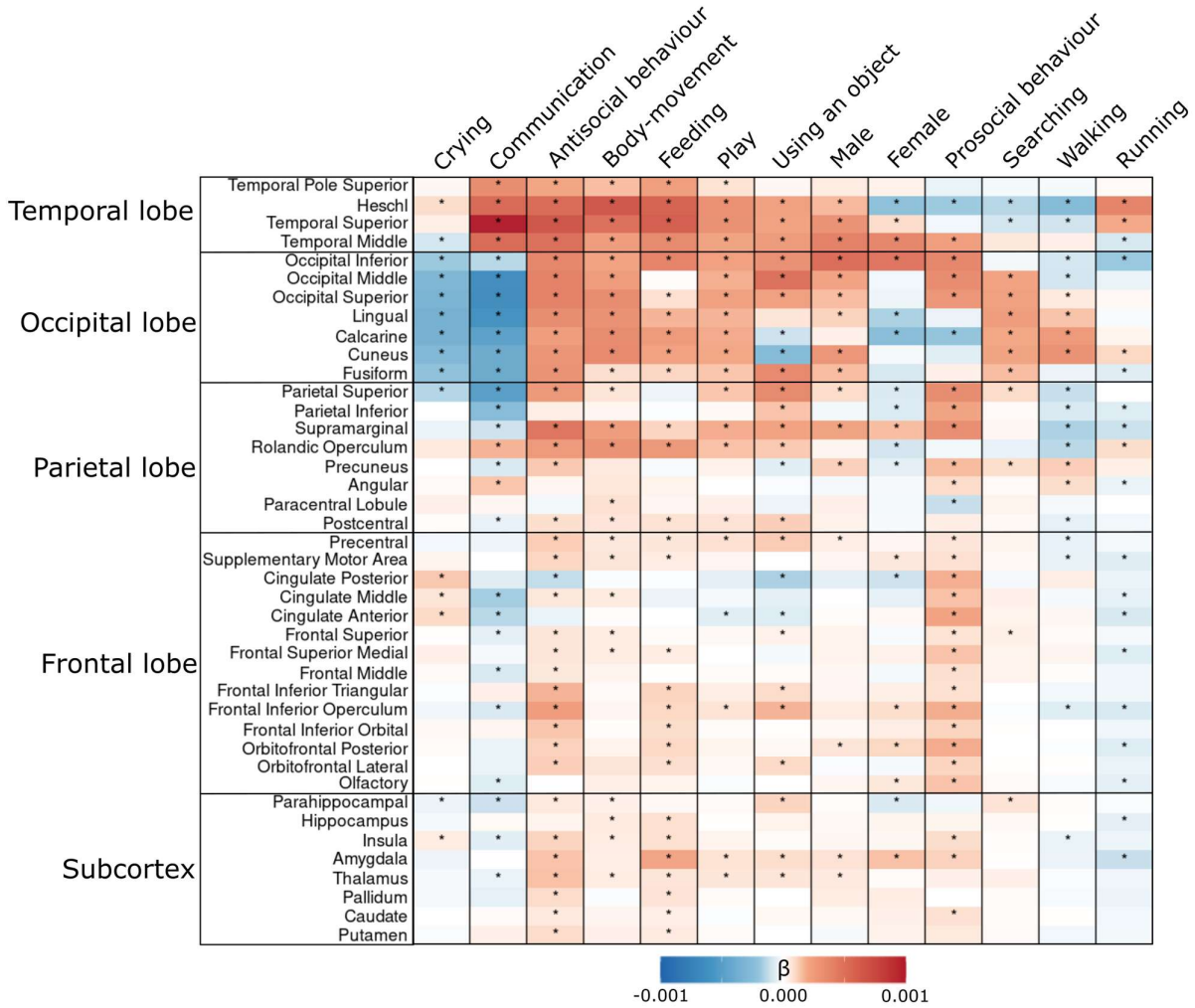

Figure 3. *ROI analysis results. Heatmap and adjacent line plot showing regression coefficients for each ROI and perceptual dimension. Statistically significant (P < 0.05, Bonferroni-corrected for each dimension independently) ROIs are marked with an asterisk.*

Analysis of anterior-posterior gradient of social perception

Figure 4a shows the cumulative brain activation patterns for all 13 perceptual dimensions. High overlap was observed in occipital and temporal cortices as well as in parietal cortices and right inferior frontal gyrus. A decreasing gradient in overlap was observed towards frontal regions. The same stimulus was used across the subjects and therefor it would be reasonable to hypothesize that the brain activation in the areas with high overlap would be temporally synchronized across subjects. We thus calculated the ISC of brain activation over the whole experiment (**Figure 4b**) to test the relationship

between the domain-general versus specific response patterns and overall reliability of the evoked responses during the stimulus. The overlap map of brain responses social dimensions (**Figure 4a**) was cosistent with the ISC map in **Figure 4b** (Pearson $r = 0.81$, $p<0.05$). Scatterplot in **Figure 4c** shows the association between ISC and brain activations overlaps for perceptual dimensions on the level of individual ROIs.

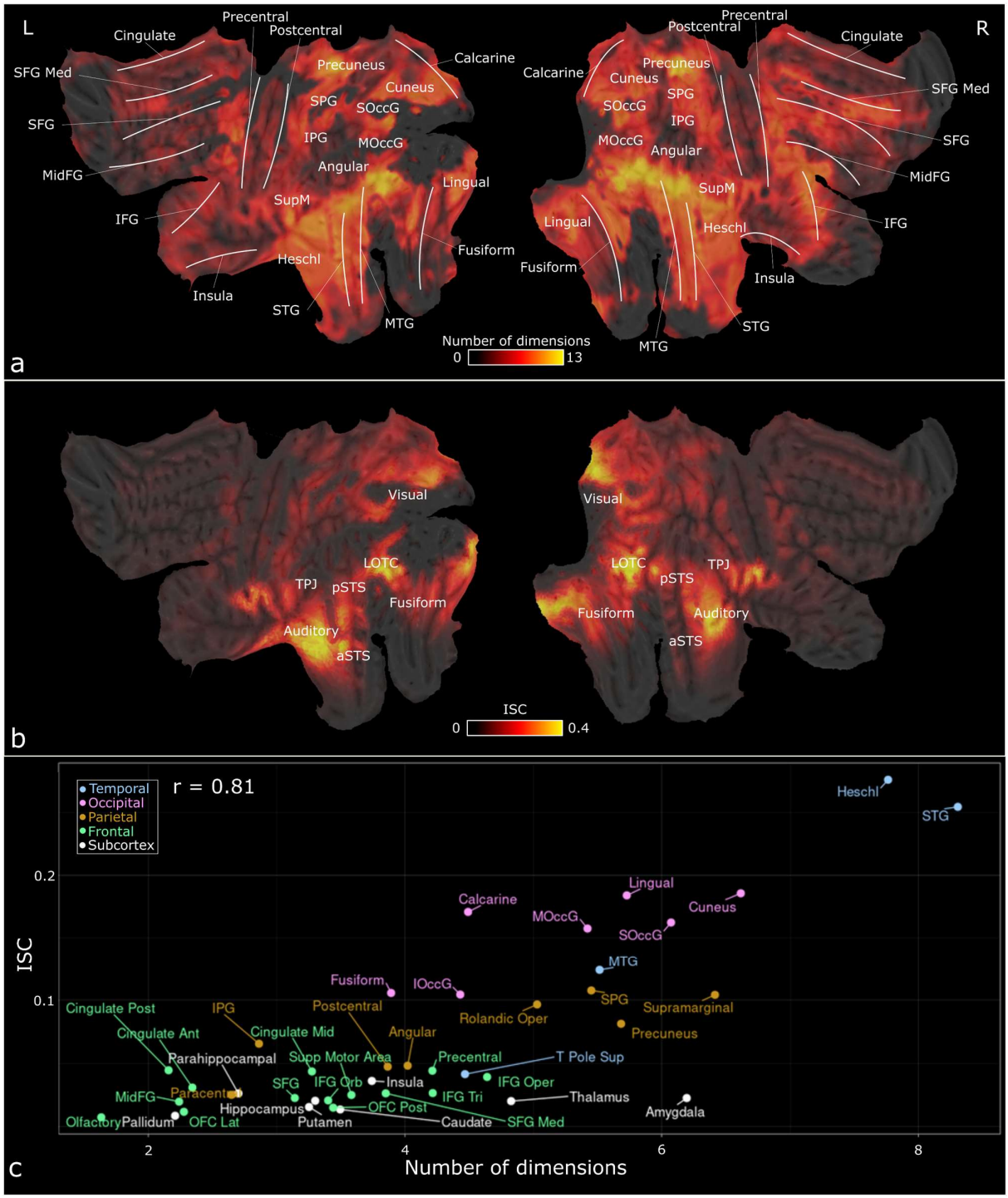

Figure 4. *(a) Cumulative activation map for social dimensions. Voxel intensities indicate how many social dimensions (out of 13) activated the voxel statistically significantly (FDR corrected, q = 0.05). White lines indicate the localizations of major gyri. (b) ISC across subjects over the whole experiment (c) Scatterplot showing the relationship between ISC and the overlap of activation patterns for perceptual dimensions. Regional ISC and overlap of activation patterns reflect the average value of all voxels within a region.*

Multivariate pattern analysis

Finally, we trained a between-subject neural network model to decode perceptual dimensions from the 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 feature selection. The selected voxels (**Figure 5a**) localized in temporal gyrus (Heschl, temporal superior, temporal middle and temporal pole superior), occipital gyrus (lingual, calcarine, cuneus, superior occipital, middle occipital, inferior occipital and fusiform) and parietal cortices (supramarginal, superior parietal and inferior parietal). The permuted chance level for the total classification accuracy in the whole brain analysis was 0.129 which is above naïve chance level ($\frac{1}{11} \cong 0.09$). At whole brain level, the NN model was able to classify all 11 social dimensions significantly above chance level with the total classification accuracy of 0.54 ($p<0.01$). Classification accuracies/precisions for each social dimension were: walking: 0.51/0.52, using an object: 0.52/0.49, searching: 0.58/0.63, running 0.59/0.58, prosocial behaviour 0.51/0.56, play 0.44/0.45, feeding 0.50/0.54, crying 0.46/0.49, communication 0.55/0.56, body movement 0.66/0.57 and antisocial behaviour 0.62/0.60. The classification was also performed within anatomical ROIs without feature selection (**Figure 5b**). Most accurate classifier performance was observed in lingual (0.33, $p<0.01$), calcarine (0.32, $p<0.01$), temporal superior (0.30, $p<0.01$), cuneus (0.29, $p<0.01$), occipital superior (0.29, $p<0.01$), occipital middle (0.28, $p<0.01$), temporal middle (0.28, $p<0.01$) and fusiform (0.26, $p<0.01$) gyri. Although the prediction accuracies were statistically significantly above permuted chance level for each ROI, the anterior-posterior gradient was also observed in the classification accuracies (**Figure 5b**).

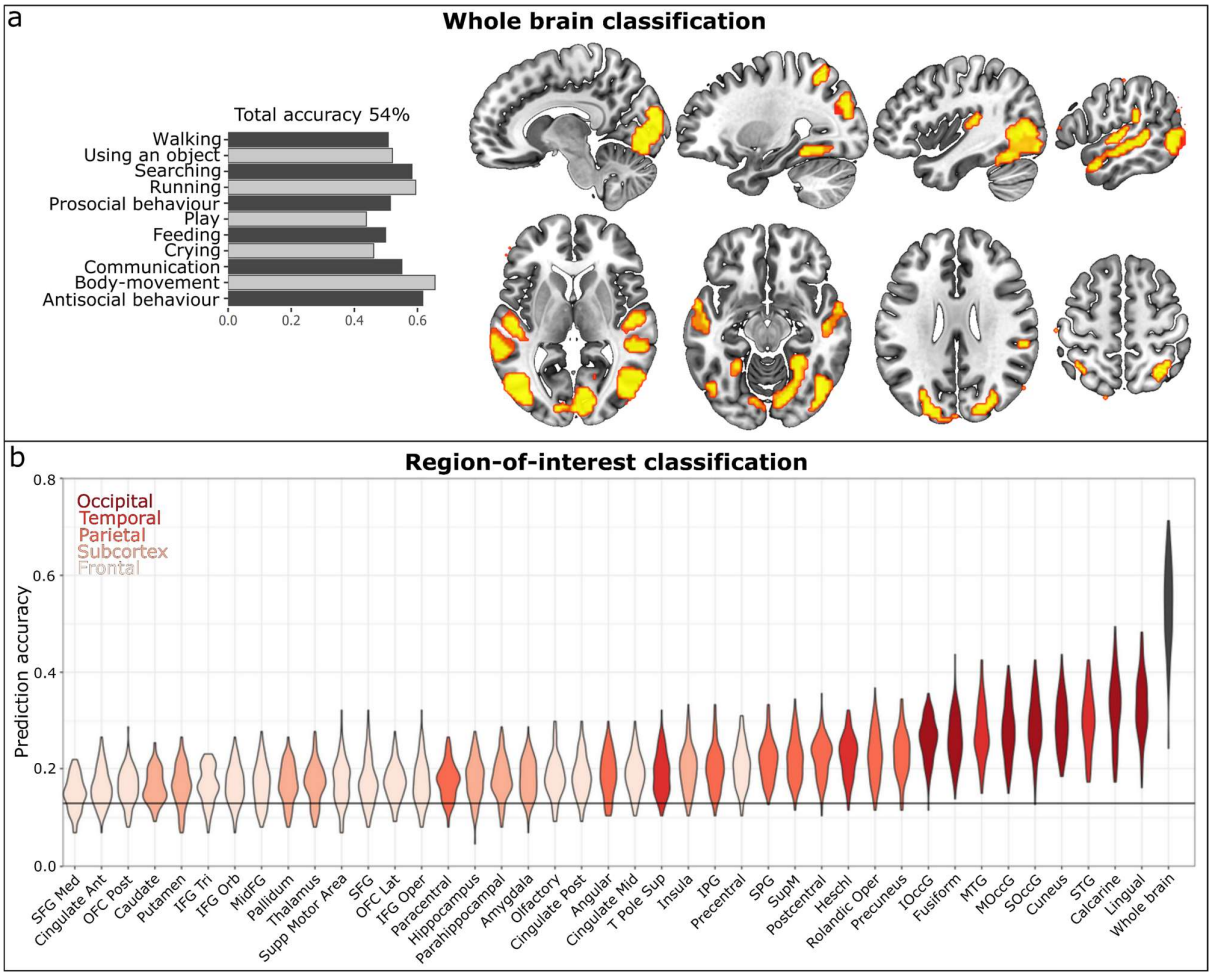

Figure 5*. Results from the multivariate pattern analysis. (a) Whole brain classification accuracies and the localization of ANOVA selected voxels. (b) Regional classification accuracies compared with the whole brain classification accuracy (black). The permuted chance level accuracy (acc = 0.129) is 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 ROI.*

Relationship between classification accuracy and ISC

At level of individual ROIs classification accuracy and ISC were positively correlated (Pearson r = 0.85, p<0.05, **Figure 6a**). Occipital regions, STG and MTG showed high synchrony (ISC > 0.1) and high classification accuracy (acc > 0.25). Although the highest ISC was observed in Heschl gyrus (ISC = 0.28), the classification accuracy for Heschl gyrus was low (acc = 0.23) compared to other high ISC regions. Most parietal region showed average ISC and average classification accuracy while frontal and subcortical regions showed low ISC and low classification accuracies.

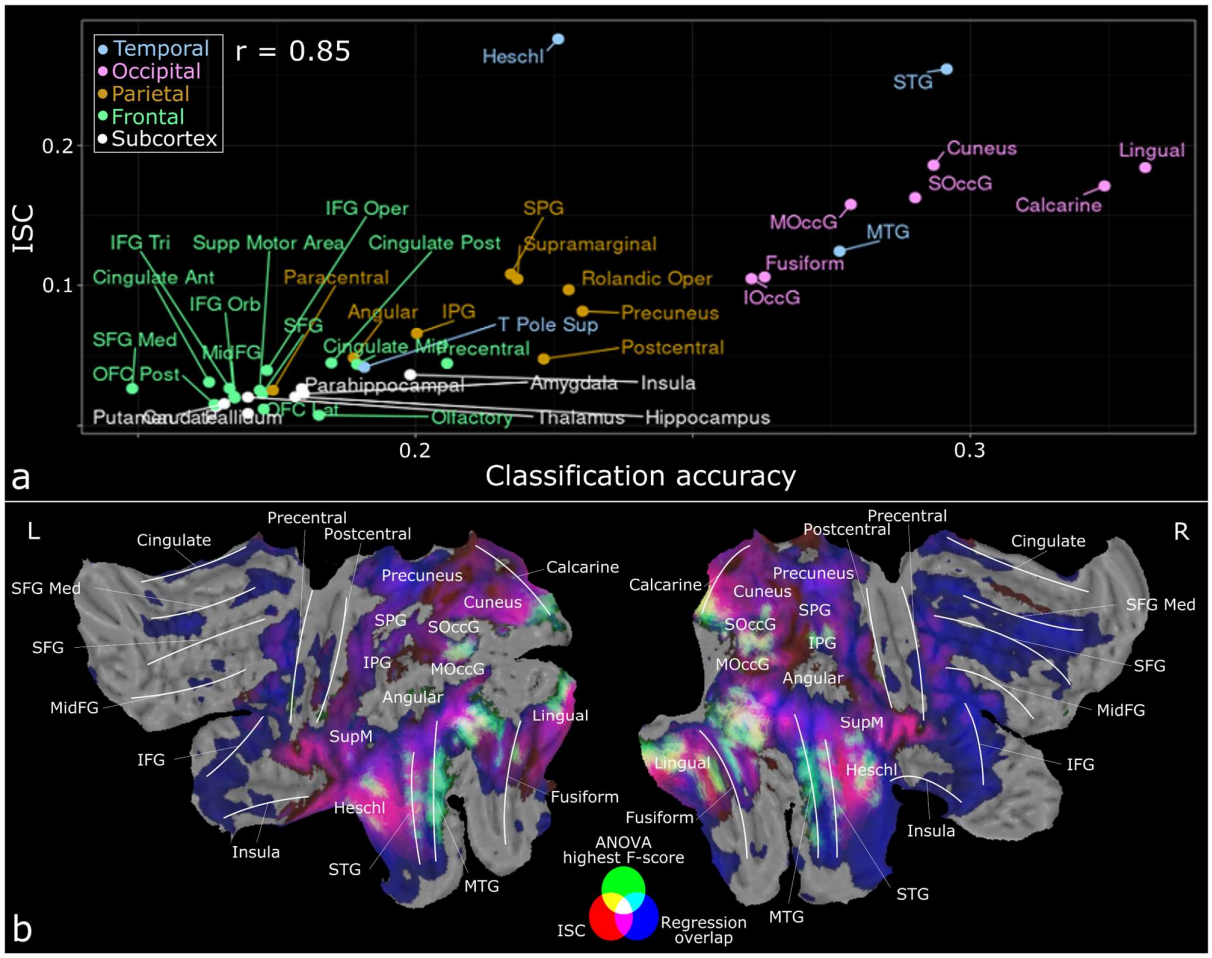

Figure 6. *(a) Scatterplot showing the relationship between ISC and classification accuracy. (b) Additive RGB map of the main findings. The overlap between activation patterns for perceptual dimensions in regression analysis is shown as blue (only areas where at least 3 dimensions expressed FDR corrected brain activation). FDR-corrected ISC across subjects over the whole experiment is shown as red. The selected voxels for the whole brain classification are shown as green.*

Discussion

Our findings provide the currently most detailed map of the social perceptual mechanisms in the human brain. The behavioral data established that 13 main dimensions reliably capture the phenomenological social perceptual space, effectively compressing the original 45 dimensions present in the data. Subjective perceptual space of social features closely corresponds with the neural organization of social processing, as evidenced by the positive association (r = 0.64) between the subjective and neural spaces of social perception. The neural space for social perception revealed a

clear posterior-anterior -axis, with occipital and temporal cortical regions serving general-purpose role in social perception, while the regional specificity increased towards the frontal cortex. Multivariate pattern recognition established that particularly occipito-temporal and parietal regions carry dimension-specific information regarding the social world, as evidenced by the above chance level accuracies in the multi-class classification approach. Finally, both classification accuracy and domain-generality of the social perceptual responses were highest in the brain regions that showed overall the most reliable activation patters throughout the experiment. Altogether these results show that the brain represents the social world in dimensional fashion, and that different brain regions show variable specificity in their response profiles towards these dimensions.

Posterior-anterior gradients in 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 main social dimensions. The univariate BOLD-fMRI analysis based on these dimensions revealed that a widely distributed cortical and subcortical networks encode the social contents of the video stimuli. Most of these dimensions activated LOTC, STS, TPJ and precuneus, as well as other occipito-temporal and parietal regions. There was a gradual decrease in these unselective responses toward more finely tuned 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. Robust effects were observed in all occipital ROIs which serve as visual information processors. Notable associations with perceptual dimensions and brain activity in temporal cortex was observed in Heschl gyrus, which is the site of the primary auditory cortex, and in STG and MTG which outline STS. In parietal cortex, Rolandic operculum (secondary somatosensory cortex, SII), supramarginal gyrus (a part of TPJ), SPG and precuneus showed responses for most social dimensions. These data are consistent with previous univariate studies addressing social functions for LOTC [3,57-59], TPJ [4,5] and STS [6-8,60].

Decoding of perceptual dimensions from brain activation patterns

The univariate analysis revealed the topography and specificity versus generality of the social processing across different brain regions. However, this analysis cannot determine whether a single region activated by multiple social dimensions reflects responses to shared features across all the dimensions such as biological motion perception or intentionality detection; [9,61], or spatially overlapping yet feature-selective processing. Multivariate classification analysis supported that the answer to this question depends on the region. The whole-brain multiclass classification yielded over

50% classification accuracy, suggesting robust differences in the responses across the classes. Whole brain classification accuracy was higher than observed for any individual ROI, suggesting that social perception involves distributed processing across numerous brain regions. The feature selection for the whole brain classification included voxels from regions associated with social perception: STS, LOTC, TPJ and FG (**Figure 5** and **Figure 6**). Previous classification studies have addressed the importance of these regions for distinct social perceptual tasks. For example, faces, animals and objects have been successfully classified by respective brain activation patterns in areas within FG and LOTC [20], and highest classification accuracy for facial expressions is typically observed in FG, STS and anterior temporal cortex [22,62,63]. Additionally, goal-oriented actions can be reliably classified in LOTC and interior parietal lobe [25,64]. 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 specific brain areas show high degree of category specificity.

Reliability versus specificity of responses to social perceptual dimensions

We observed robust inter-subject correlation of brain activity in temporal and occipital regions for visual socioemotional stimulus. Previous studies have found that the ISC is in general the strongest in the sensory regions, and it progressively becomes weaker toward the polysensory and associative cortices [65]. Our data in turn revealed that the strength of the ISC was contingent on the number of social features each region responded to in the univariate analysis and the ISC was also contingent on the classification accuracy in the multivariate analysis (**Figure 4c** and **Figure 6a**). 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. 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.

The regional response profiles can thus be summarized as follows: First, all occipital regions, STG and MTG have with high ISC, overlapping GLM responses across dimensions and high classification accuracy. Most voxels passing the feature selection for the whole brain classification analysis localized in these regions. These occipital and temporal regions thus constitute the most fundamental hubs for social perception in the human brain and are likely involved in integration of the

multisensory information regarding the social events. Second, the overlap between ISC and GLM responses was high in Heschl gyrus near the auditory cortex, but the classification accuracy was only moderate in that region. This suggests that Heschl gyrus processes domain-general social information but may not process carried detailed information about the distinct social dimensions. Third, most frontal and subcortical regions showed low ISC, low regression overlap and low classification accuracy. Frontal areas and more specifically MFC have previously been associated with social and affective processing such as linking social processing with decision making, affective processing and theory of mind [10,66]. Previous classification studies are in line with our results, and they have not found specificity for social perception in frontal cortex [20,22,25,26]. 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 of others based on the perceived information. Amygdala, insula, IFG and precentral gyrus in turn showed specific activation profiles for a limited number of dimensions while ISC and classification accuracy were low in these regions. Additionally, activation in parietal areas supramarginal gyrus and precuneus was associated with numerous social dimensions ant yet their ISC and classification accuracies were only moderate. These fronto-parietal and subcortical regions thus likely respond to some general or low-level features of the social signals. As these features are both general and not of directly sensory nature, both ISC and classification accuracy remained low.

Limitations

We intentionally used naturalistic video clips to mimic real-life social perception. Still, the stimulus film clips were extracted from popular Hollywood films and the social features present may differentiate from real-life social features in their magnitude, occurrence rate and co-occurrence. We did not control for potential low-level differences across the clips. Thus, despite controlling for low-level visual and acoustic differences in the regression analysis, it is possible that such features still confound the results. Over half of the predefined social features had an insufficient inter-rater reliability or insufficient occurrence rate in the stimulus film clips; however, this is itself an important finding regarding the consistency of the perceptual taxonomy individuals use for describing social events. The number of events for classification was unbalanced across categories, which may affect the classifier training. Even in regions with near chance level total accuracy, some classes with large number of events were classified with relatively high accuracy which may be due to the large number of events in these classes and might not reflect the actual social information processing in the brain. Because of this, regional differences in the prediction accuracies in individual classes cannot be addressed.

Conclusion

Using a combination of data-driven approaches and multivariate pattern recognition we established the perceptual space for socioemotional features and mapped the cerebral topography of social perception that can be described with 13 perceptual dimensions. Clear decreasing gradient in brain activation associated with social perception was observed from temporal, occipital and parietal regions to frontal and subcortical regions. Perceptual social dimensions could also 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 among other temporal, occipital regions in social perception, while parietal areas, precuneus and Heschl's gyrus process domain-general information from the social scenes. Altogether these results highlight the distributed nature of social processing in the brain as well as the specific contributions of feature-specific versus domain-general social perceptual processes.

Acknowledgements

The study was supported by grants to LN from European Research Council (#313000) and the Academy of Finland (#332225, #294897). EG was supported by the International Laboratory of Social Neurobiology ICN HSE RF Government grant ag. No. 075-15-2019-1930. The authors declare no competing financial or non-financial interests.

Author contributions

SS developed the analysis methods, analysed the data, and wrote the manuscript.
TK preprocessed the fMRI data, developed the regression analysis methods, and wrote and reviewed the manuscript.
SNF developed the MVPA analysis methods and reviewed the manuscript.
MH conceptualized the study design, collected the data, and reviewed the manuscript.
VP, KS, and LS collected the data and reviewed the manuscript.
EG developed the analysis and preprocessing methods and reviewed the manuscript.
JH evaluated the T1w MR images and reviewed the manuscript and HK supervised the security of MRI data collection and reviewed the manuscript.

LN conceptualized the study design, acquired funding, supervised the project, developed analysis methods, and wrote and reviewed the manuscript.

Data availability

Per Finnish legislation, the original neuroimaging data used in the experiment cannot be released for public use. All necessary files can however be made available for reviewers. The brain activation patterns (unthresholded T-maps) for each perceptual dimension from GLM analysis are available in NeuroVault (https://neurovault.org/collections/IZWVFEYI/).

References

1 Brooks, J. A., Stolier, R. M. & Freeman, J. B. Computational approaches to the neuroscience of social perception. *Soc Cogn Affect Neurosci*, doi:10.1093/scan/nsaa127 (2020).

2 Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. The distributed human neural system for face perception. *Trends in Cognitive Sciences* **4**, 223-233, doi:10.1016/s1364-6613(00)01482-0 (2000).

3 Downing, P. E., Jiang, Y., Shuman, M. & Kanwisher, N. A cortical area selective for visual processing of the human body. *Science* **293**, 2470-2473, doi:10.1126/science.1063414 (2001).

4 Saxe, R. & Kanwisher, N. People thinking about thinking peopleThe role of the temporo-parietal junction in "theory of mind". *NeuroImage* **19**, 1835-1842, doi:10.1016/s1053-8119(03)00230-1 (2003).

5 Carter, R. M. & Huettel, S. A. A nexus model of the temporal-parietal junction. *Trends Cogn Sci* **17**, 328-336, doi:10.1016/j.tics.2013.05.007 (2013).

6 Lahnakoski, J. M. *et al.* 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 (2012).

7 Deen, B., Koldewyn, K., Kanwisher, N. & Saxe, R. Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus. *Cereb Cortex* **25**, 4596-4609, doi:10.1093/cercor/bhv111 (2015).

8 Isik, L., Koldewyn, K., Beeler, D. & Kanwisher, N. Perceiving social interactions in the posterior superior temporal sulcus. *Proc Natl Acad Sci U S A* **114**, E9145-E9152, doi:10.1073/pnas.1714471114 (2017).

9 Nummenmaa, L. & Calder, A. J. Neural mechanisms of social attention. *Trends Cogn. Sci.* **13**, 135-143, doi:10.1016/j.tics.2008.12.006 (2009).

10 Amodio, D. M. & Frith, C. D. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* **7**, 268-277, doi:10.1038/nrn1884 (2006).

11 Price, C. J. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* **62**, 816-847, doi:10.1016/j.neuroimage.2012.04.062 (2012).

12 Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron* **76**, 1210-1224, doi:10.1016/j.neuron.2012.10.014 (2012).

13 Felsen, G. & Dan, Y. A natural approach to studying vision. *Nat. Neurosci.* **8**, 1643-1646, doi:10.1038/nn1608 (2005).

14 Adolphs, R., Nummenmaa, L., Todorov, A. & Haxby, J. V. Data-driven approaches in the investigation of social perception. *Phil Trans B* **371** (2016).

15 Huth, A. G., Nishimoto, S., Vu, A. T. & Gallant, J. L. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* **76**, 1210-1224, doi:10.1016/j.neuron.2012.10.014 (2012).

16 Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453-458, doi:10.1038/nature17637 (2016).

17 Koide-Majima, N., Nakai, T. & Nishimoto, S. 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 (2020).

18 Popham, S. F. *et al.* Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nat Neurosci* **24**, 1628-1636, doi:10.1038/s41593-021-00921-6 (2021).

19 Tong, F. & Pratte, M. S. Decoding patterns of human brain activity. *Annu Rev Psychol* **63**, 483-509, doi:10.1146/annurev-psych-120710-100412 (2012).

20 Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**, 2425-2430, doi:10.1126/science.1063736 (2001).

21 Brosch, T., Bar-David, E. & Phelps, E. A. Implicit race bias decreases the similarity of neural representations of black and white faces. *Psychol Sci* **24**, 160-166, doi:10.1177/0956797612451465 (2013).

22 Wegrzyn, M. *et al.* 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 (2015).

23 Said, C. P., Moore, C. D., Engell, A. D., Todorov, A. & Haxby, J. V. Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J Vis* **10**, 11, doi:10.1167/10.5.11 (2010).

24 Harry, B., Williams, M. A., Davis, C. & Kim, J. Emotional expressions evoke a differential response in the fusiform face area. *Front Hum Neurosci* **7**, 692, doi:10.3389/fnhum.2013.00692 (2013).

25 Wurm, M. F. & Lingnau, A. Decoding actions at different levels of abstraction. *J Neurosci* **35**, 7727-7735, doi:10.1523/JNEUROSCI.0188-15.2015 (2015).

26 Oosterhof, N. N., Tipper, S. P. & Downing, P. E. Viewpoint (in)dependence of action representations: an MVPA study. *J Cogn Neurosci* **24**, 975-989, doi:10.1162/jocn_a_00195 (2012).

27 Hasson, U., Malach, R. & Heeger, D. J. Reliability of cortical activity during natural stimulation. *Trends Cogn Sci* **14**, 40-48, doi:10.1016/j.tics.2009.10.011 (2010).

28 Lahnakoski, J. M. *et al.* Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Front. Hum. Neurosci.* **6**, 14, doi:10.3389/fnhum.2012.00233 (2012).

29 Karjalainen, T. *et al.* Opioidergic Regulation of Emotional Arousal: A Combined PET-fMRI Study. *Cerebral cortex (New York, N.Y. : 1991)*, doi:10.1093/cercor/bhy281 (2018).

30 Nummenmaa, L. *et al.* Brain basis of psychopathy in criminal offenders and general population. *Cereb. Cortex* (2021).

31 Karjalainen, T. *et al.* Dissociable Roles of Cerebral mu-Opioid and Type 2 Dopamine Receptors in Vicarious Pain: A Combined PET-fMRI Study. *Cereb Cortex* **27**, 4257-4266, doi:10.1093/cercor/bhx129 (2017).

32 Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111-116, doi:10.1038/s41592-018-0235-4 (2019).

33 Gorgolewski, K. *et al.* Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics* **5**, doi:10.3389/fninf.2011.00013 (2011).

34 Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics* **8**, doi:10.3389/fninf.2014.00014 (2014).

35 Tustison, N. J. *et al.* N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging* **29**, 1310-1320, doi:10.1109/TMI.2010.2046908 (2010).

36 Dale, A. M., Fischl, B. & Sereno, M. I. Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage* **9**, 179-194, doi:https://doi.org/10.1006/nimg.1998.0395 (1999).

37 Klein, A. *et al.* Mindboggling morphometry of human brains. *PLOS Computational Biology* **13**, e1005350, doi:10.1371/journal.pcbi.1005350 (2017).

38 Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R. & Collins, D. L. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* **47**, S102, doi:https://doi.org/10.1016/S1053-8119(09)70884-5 (2009).

39 Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis* **12**, 26-41, doi:https://doi.org/10.1016/j.media.2007.06.004 (2008).

40 Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging* **20**, 45-57, doi:10.1109/42.906424 (2001).

41 Cox, R. W. AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research* **29**, 162-173, doi:https://doi.org/10.1006/cbmr.1996.0014 (1996).

42 Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage* **17**, 825-841, doi:https://doi.org/10.1006/nimg.2002.1132 (2002).

43 Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* **48**, 63-72, doi:https://doi.org/10.1016/j.neuroimage.2009.06.060 (2009).

44 Pruim, R. H. R. *et al.* ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *NeuroImage* **112**, 267-277, doi:https://doi.org/10.1016/j.neuroimage.2015.02.064 (2015).

45 Cukur, T., Nishimoto, S., Huth, A. G. & Gallant, J. L. Attention during natural vision warps semantic representation across the human brain. *Nat Neurosci* **16**, 763-770, doi:10.1038/nn.3381 (2013).

46 Chen, G., Taylor, P. A. & Cox, R. W. Is the statistic value all we should care about in neuroimaging? *NeuroImage* **147**, 952-959, doi:https://doi.org/10.1016/j.neuroimage.2016.09.066 (2017).

47 Onni: An online experiment platform for research v. 1.0 (Zenodo, 2020).

48 Koo, T. K. & Li, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* **15**, 155-163, doi:https://doi.org/10.1016/j.jcm.2016.02.012 (2016).

49 Mcquitty, L. L. Similarity Analysis by Reciprocal Pairs for Discrete and Continuous Data. *Educ Psychol Meas* **26**, 825-&, doi:Doi 10.1177/001316446602600402 (1966).

50 Hoerl, A. E. & Kennard, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **12**, 55–67, doi:10.1080/00401706.1970.10488634 (1970).

51 Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *NeuroImage* **92**, 381-397, doi:https://doi.org/10.1016/j.neuroimage.2014.01.060 (2014).

52 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B* **57**, 289-300, doi:DOI 10.1111/j.2517-6161.1995.tb02031.x (1995).

53 Rolls, E. T., Joliot, M. & Tzourio-Mazoyer, N. 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 (2015).

54 Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject synchronization of cortical activity during natural vision. *Science* **303**, 1634-1640, doi:10.1126/science.1089506 (2004).

55 Kauppi, J. P., Pajula, J. & Tohka, J. A versatile software package for inter-subject correlation based analyses of fMRI. *Front Neuroinform* **8**, 2, doi:10.3389/fninf.2014.00002 (2014).

56 Hanke, M. *et al.* PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. *Neuroinformatics* **7**, 37-53, doi:10.1007/s12021-008-9041-y (2009).

57 Wurm, M. F., Caramazza, A. & Lingnau, A. Action Categories in Lateral Occipitotemporal Cortex Are Organized Along Sociality and Transitivity. *J Neurosci* **37**, 562-575, doi:10.1523/JNEUROSCI.1717-16.2016 (2017).

58 Wurm, M. F. & Caramazza, A. 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 (2019).

59 Lingnau, A. & Downing, P. E. The lateral occipitotemporal cortex in action. *Trends Cogn Sci* **19**, 268-277, doi:10.1016/j.tics.2015.03.006 (2015).

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

61 Allison, T., Puce, A. & McCarthy, G. Social perception from visual cues: role of the STS region. *Trends Cogn Sci* **4**, 267-278, doi:10.1016/s1364-6613(00)01501-1 (2000).

62 Volynets, S., Smirnov, D., Saarimäki, H. & Nummenmaa, L. Statistical pattern recognition reveals shared neural signatures for displaying and recognizing specific facial expressions. *Soc. Cogn. Affect. Neurosci.* **15**, 803-813, doi:10.1093/scan/nsaa110 (2020).

63 Said, C. P., Moore, C. D., Engell, A. D., Todorov, A. & Haxby, J. V. Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J. Vision* **10**, 11-11, doi:10.1167/10.5.11 (2010).

64 Smirnov, D. *et al.* Brain-to-brain hyperclassification reveals action-specific motor mapping of observed actions in humans. *PLoS One* **12**, e0189508, doi:10.1371/journal.pone.0189508 (2017).

65 Hasson, U., Malach, R. & Heeger, D. J. Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* **14**, 40-48, doi:10.1016/j.tics.2009.10.011 (2010).

66 de la Vega, A., Chang, L. J., Banich, M. T., Wager, T. D. & Yarkoni, T. Large-Scale Meta-Analysis of Human Medial Frontal Cortex Reveals Tripartite Functional Organization. *J Neurosci* **36**, 6553-6562, doi:10.1523/JNEUROSCI.4402-15.2016 (2016).