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[4,5], Jussi Hirvonen[6,7], 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 Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo, Finland

[5]International Laboratory of Social Neurobiology, Institute of Cognitive Neuroscience, National Research University Higher School of Economics, Moscow, Russian Federation

[6]Department of Radiology, University of Turku and Turku University Hospital, Turku, Finland

[7]Medical Imaging Center, Department of Radiology, Tampere University and Tampere University Hospital, Tampere, Finland

[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

Email: svtsan@utu.fi

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 film 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 112 social features. Cluster analysis revealed that 13 dimensions were sufficient for describing the social perceptual space. Univariate GLM was used to map regional neural response profiles to different social features. Multivariate pattern analysis was then utilized to establish the spatial 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 while the responses were domain-general and did not differentiate between dimensions. Frontal and limbic regions responded only to a limited number of social dimensions. Altogether these results highlight the distributed nature of social processing in the brain as well as the posterior-anterior gradient of domain-general versus dimension-specific social perceptual processes in the brain.

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 localized specific aspects of social perception into different brain regions [1]. Fusiform gyrus (FG) is consistently involved in the perception of faces [2] and lateral occipitotemporal cortex (LOTC) in the perception of bodies [3]. Temporoparietal junction (TPJ) is in turn 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 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 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]. 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. Neural responses to complex stimuli cannot necessarily be predicted on statistical combination of responses to simple stimuli [13]. Therefore studies based on neural responses to isolated social features may not directly generalize to real-world social perception [14] where social features such as facial identities, body movements, and nonverbal communication often overlap with distinct temporal occurrence patterns. In other domains including actions [12], language [15], and emotions [16], 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. For example, a recent study found that linguistic and visual semantic representations converge so that visual representations locate on the border of occipital cortex and that linguistic representations are located anterior to the visual representations [17]. However, a detailed representational space for social features at both perceptual and neural level is currently lacking.

Commonly applied univariate analyses cannot reveal the specificity of 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 [18]. Pattern recognition studies have established that regional multivariate patterns allow distinguishing brain activation related to multiple high-level social features such as faces [19] and their racial group [20] in FG and facial expressions in FG and STS [21-23]. 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 [24]. Furthermore, decoding of goal-oriented actions was successful in LOTC when subjects observed the actions in both first and third person perspectives [25]. 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 perceptual and neural representations of naturalistic social episodes using both univariate and multivariate analyses. We used short film clips as stimuli because cinema contains rich and complex social scenarios and as it also elicitis strong and consistent neural responses in functional imaging studies [6,26]. We first aimed at establishing a taxonomy of the social dimensions that human observers use for describing social scenarios, and then mapping 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 films (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 univariate regression analysis, we established a posterior-anterior gradient in social perceptual processes in the brain, while multivariate pattern analysis revealed differences in the specificity of tuning for social dimensions across brain regions.

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 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.

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 [6,27-29]. The original experiment using this stimulus described the experimental design and stimulus selection in detail [6]. 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. The videos were extracted from mainstream Hollywood movies with audio track in English. Nine erotic scenes were added to broaden the social content of the stimulus and a subset of 87 clips from the original set of 137 clips were included to limit the experiment duration. 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. Subjects were instructed to view the movies similarly as if they were viewing a film 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.

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 [30] (v1.3.0), a Nipype [31] based tool that internally uses Nilearn [32]. During the preprocessing, each T1w volume was corrected for intensity non-uniformity using N4BiasFieldCorrection (v2.1.0) [33] 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) [34], 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 [35]. Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c [36] was performed through nonlinear registration with the antsRegistration (ANTs v2.1.0) [37], 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 [38] (FSL v5.0.9).

Functional data were slice-time-corrected using 3dTshift from AFNI [39] (v16.2.07) and motion-corrected using MCFLIRT [40] (FSL v5.0.9). These steps were followed by co-registration to the T1w image using boundary-based registration [41] 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 [42]. The data were then detrended using 240-s-Savitzky–Golay filtering to remove the scanner drift [43], and finally downsampled to original 3 mm isotropic voxel size. The BOLD signals were demeaned to make the regression coefficients comparable across different individuals [44]. First and last two functional volumes were discarded to ensure equilibrium effects and to exclude the time points after the stimulus had ended.

Stimulus features

Five individuals not participating in fMRI rated the 112 predefined features (see **Table SI-1**) from the film clips. We selected a broad range of socioemotional features from the following categories: 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), social interaction signals (e.g. talking, communicating with gestures) and social interaction characteristics (e.g. hostility, sexuality). 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 films). The ratings were collected separately for each 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 [45].

Feature selection

We first evaluated whether the *a priori* features were frequently and consistently perceived in the stimulus films. Features with low occurrence rate and/or inter-rater reliability were excluded, 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 time points 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 [46], and we thus only included features with ICC over 0.5. 45 features satisfied both criteria. The occurrence rate and inter-rater reliability of each feature are shown in **Figure SI-1**.

Dimension reduction

The selected 45 features were linearly correlated (**Figure 1**) and it is unlikely that each social feature is processed in different brain regions or networks. 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. 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 [47]. 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 Methods (see also **Figure SI-3**). Cluster analysis grouped social features into six clusters and seven independent features not belonging to any cluster (**Figure 1**) 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**).

Modelling low-level sensory features

Our goal was to map perceived social dimensions in the human brain. The stimulus film clips were not balanced by 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 from the stimulus film 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 other visual features (see Code availability). Auditory features were extracted using MIRToolbox1.8.1 [48]. First eight principal components (PCs) explaining over 90 % of the total variance were selected as regressors for low-level audiovisual features. As the stimulus film 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.

Modelling brain responses to social perceptual dimensions

Ridge regression [49] 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. 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 ($\mu = 0$, $\sigma = 1$). Detailed description of ridge regression modelling is included in supplementary materials (**Figure SI-6**).

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. 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 the social perceptual model, 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 [50]. Voxel-level FDR with q-value of 0.05 was used to correct for multiple comparisons [51]. Anatomical ROIs were extracted from AAL2 atlas [52]. 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.

As a supplementary analysis we compared the predictive capabilities of separate low-level and social models (**Supplementary Methods, Supplementary Results and Fig-SI8**).

Multivariate pattern analysis of social perceptual dimensions

To reveal the regional specialization in processing of different social features, between-subject classification of 11 perceptual dimensions[1] was performed in Python using the PyMVPA toolbox [53]. 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 time point separately is computationally prohibitive and single EPI scans are noisy. Moreover, 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 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 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 the residual BOLD time series (after regressing out the low-level features) in each time window and social dimension and the resulting subjectwise β-images were used as input for the

[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 (see **Figure SI-4**).

multivariate pattern analysis (MVPA). The input residual BOLD time series were 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. 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 form 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. 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 (**Table SI-2**).

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 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.

Inter-subject correlation analysis

Watching films 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]. As the only variable factor in the experiment is the time-varying audiovisual stimulus, ISC analysis captures the shared stimulus-dependent activation in the brain. This stimulus-dependent activation most likely reflects the population level consistency in perception of low-level audiovisual features but also higher order social perceptual processes. As a *post hoc* analysis, we assessed whether the regional differences in brain response profiles for social dimensions relate to the inter-subject 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 [55].

Results

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", "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** 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. 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.

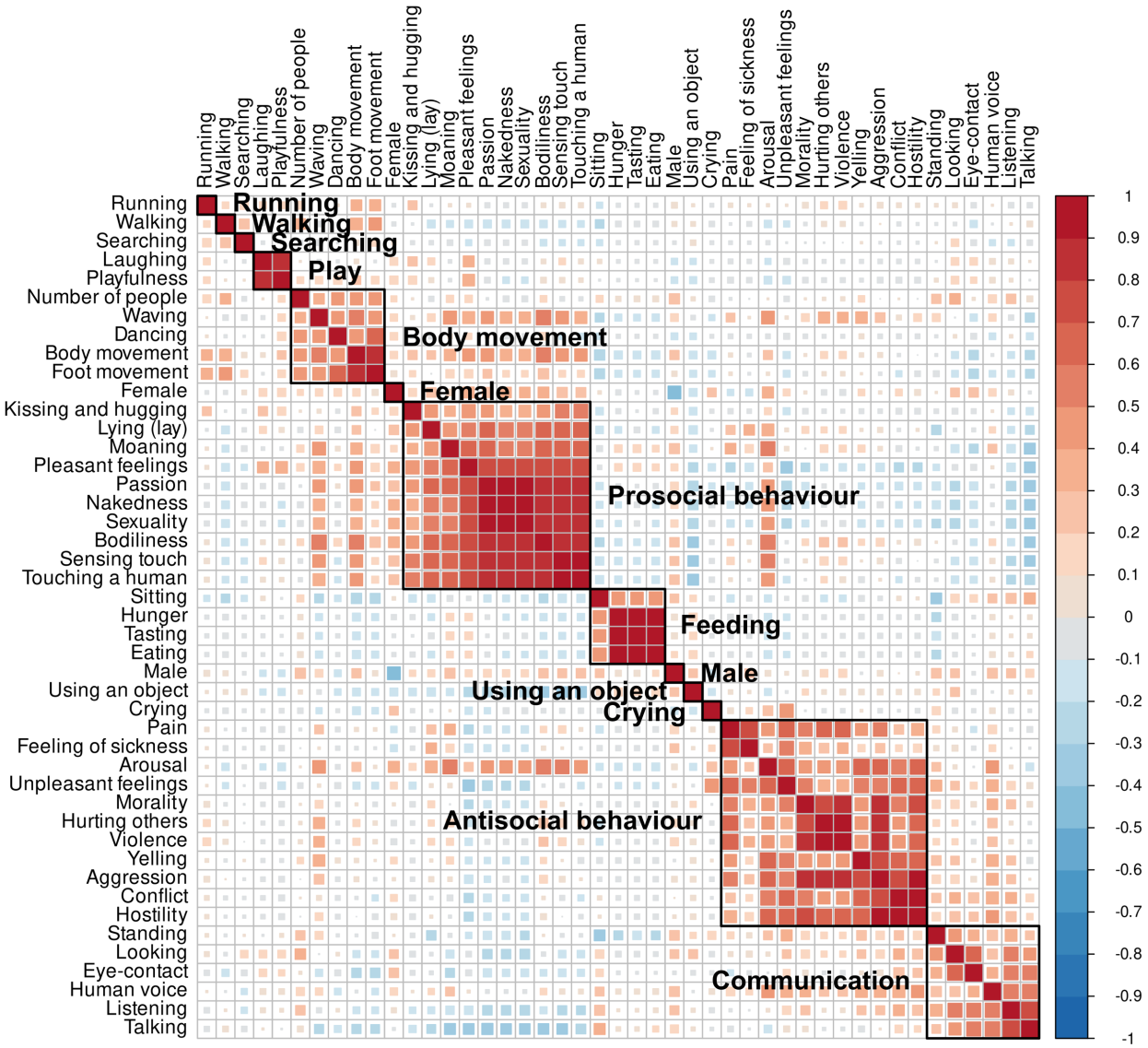

Figure 1. *The results of hierarchical clustering of reliably rated social features. Hierarchical clustering identified six clusters and seven individual features.*

Cerebral topography of social perception

Regularized ridge regression was used to establish the full-volume activation patterns for 13 perceptual social dimensions (**Figure 2**). Social information processing engaged all brain lobes and both cortical and subcortical regions. Robust responses were observed in occipital, temporal, and parietal cortices (**Figure 3**). 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, antisocial behaviour and feeding.

In ROI analysis, broad 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 associated with 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 prosocial 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.

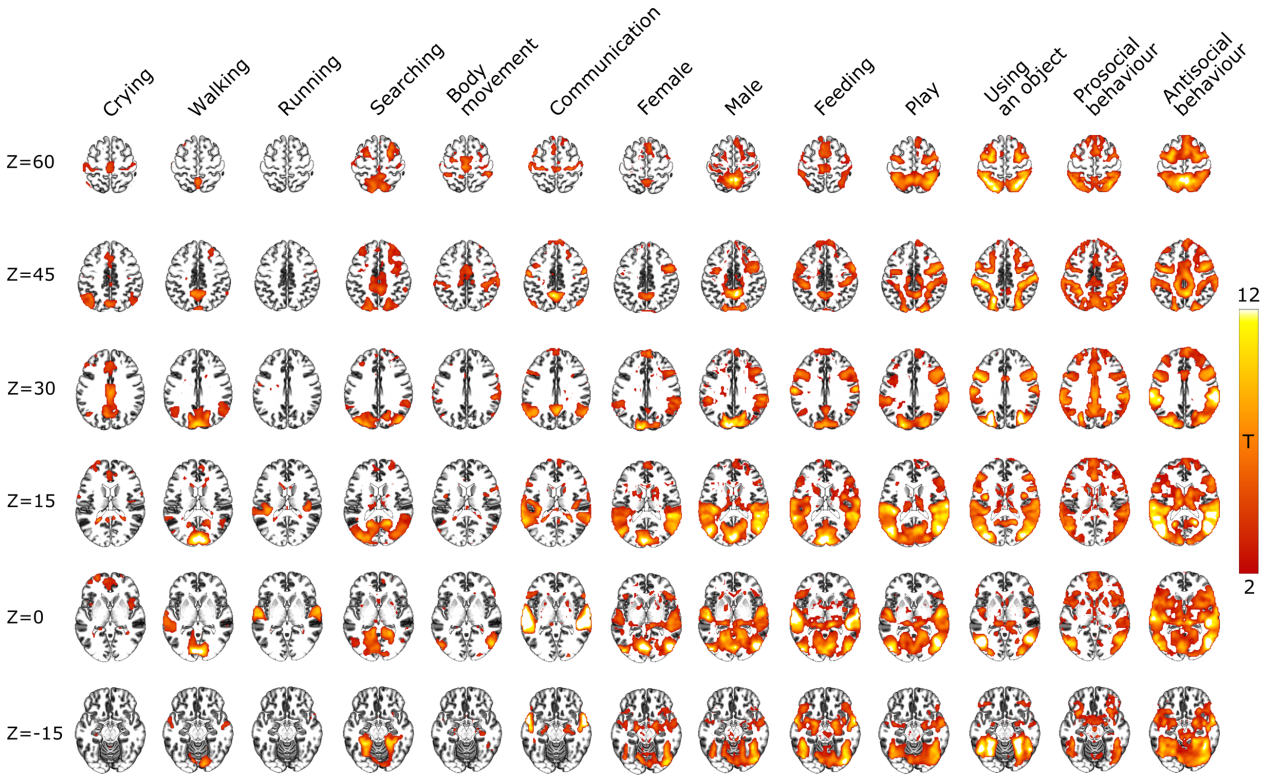

Figure 2. *Brain regions showing increased BOLD activity for the social dimensions. Data are thresholded at $p < 0.05$, FDR corrected.*

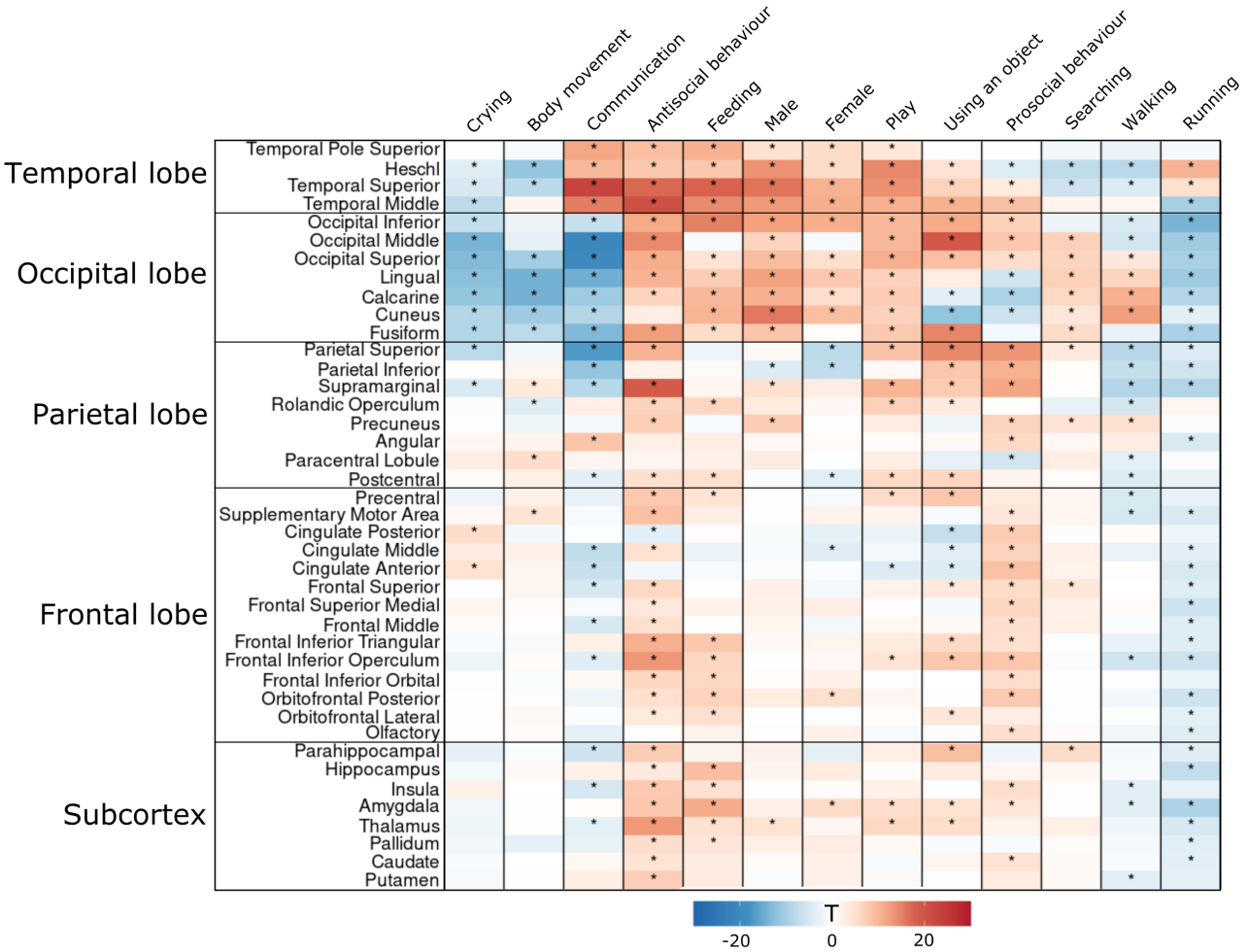

Figure 3. *Results from the ROI analysis. A heatmap showing T-values for regression coefficients in each ROI and perceptual dimension. Statistically significant (P < 0.05, Bonferroni-corrected for each dimension independently) ROIs are marked with an asterisk.*

Analysis of posterior-anterior gradient of social perception

Figure 4a shows the cumulative brain activation maps for all 13 perceptual dimensions. There was a gradient in the regional selectivity for social dimensions. Occipital and temporal 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 consitency in their response profile. 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 (**Figure 4b**) 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 **Figure 4c** shows the association between ISC and corresponding brain response selectivity for perceptual dimensions (Pearson r = 0.86).

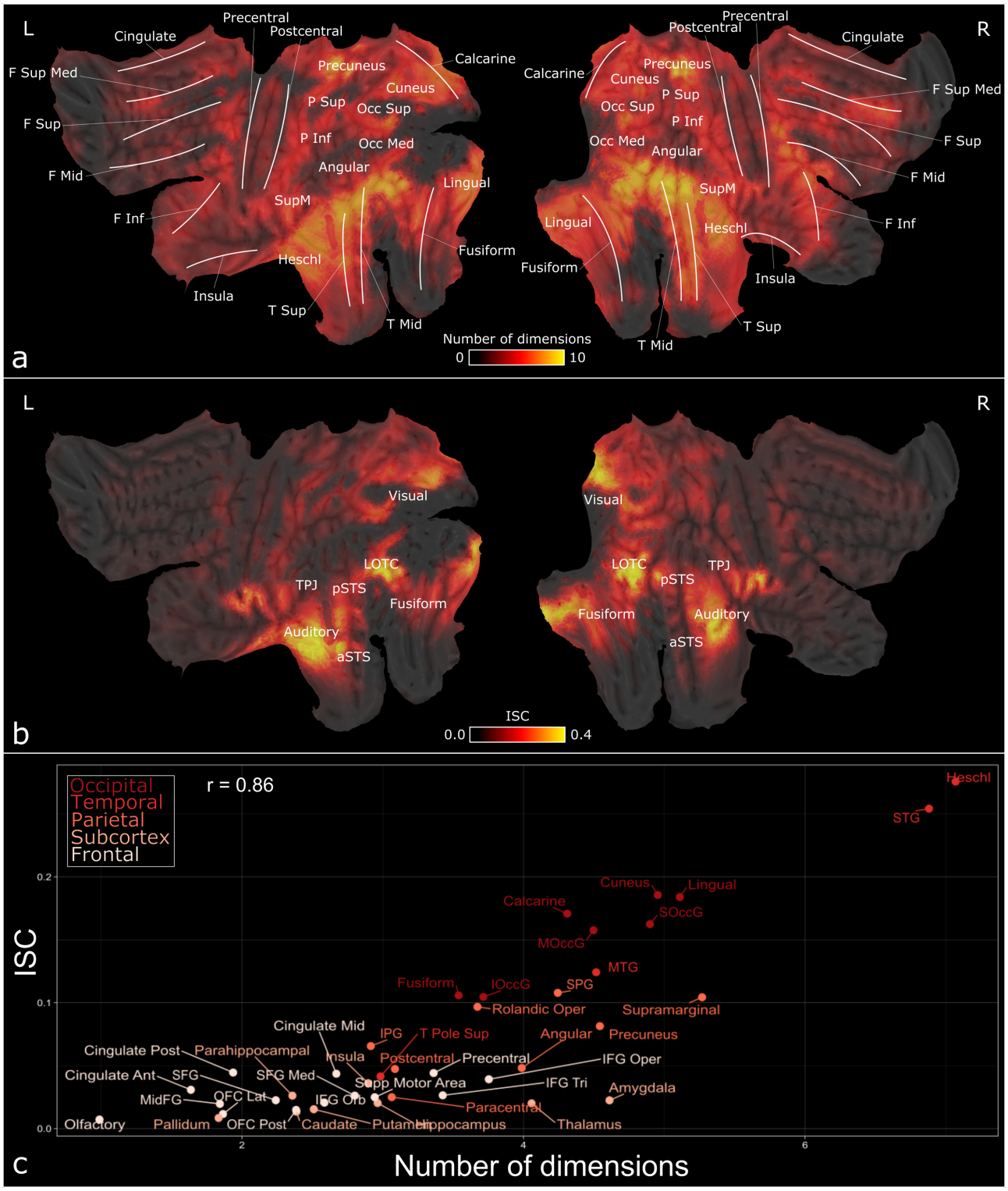

Figure 4. *(a) Cumulative activation map for social dimensions. Voxel intensities indicate how many social dimensions (out of 13) activated the voxel statistically significantly (FDR corrected, q = 0.05). White lines indicate the localizations of major gyri. (b) ISC across subjects over the whole experiment (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. CARET software [56] was used for mapping results from ICBM 152 Nonlinear Asymmetrical template version 2009c space to the flatmap surface.

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 feature selection. The majority of the selected voxels (**Figure 5a**) 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, prosocial 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 (**Figure 5a**).

The classification was also performed within anatomical ROIs (**Figure 5b**). 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 posterior-anterior gradient 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 cubcortical regions. (**Figure 5b**). We also validated that this gradient was not an artefact stemming from the sizes of the ROIs, as similar posterior-anterior gradient was observed in the regional classification with ANOVA feature selection limited to 119 voxels (**Figure 5c**).

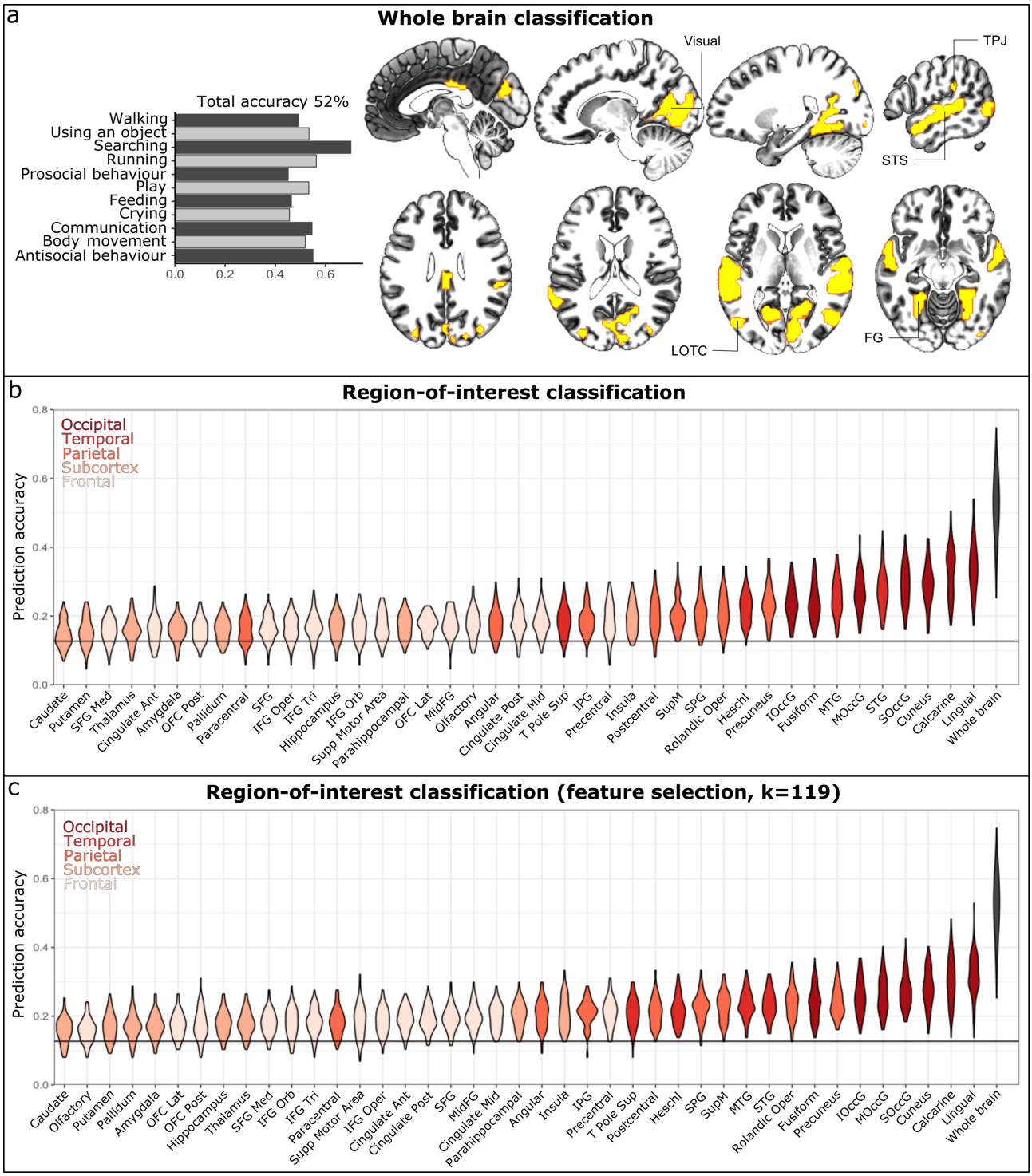

Figure 5. *Results from the multivariate pattern analysis. (a) Whole brain classification accuracies and the localization of ANOVA selected voxels. (b) Regional classification accuracies compared with the whole brain classification accuracy (black). The permuted chance level accuracy (acc = 0.128) is 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 ROI. (c) Regional classification accuracies using only 119 voxels (the size of the smallest ROI) with the highest F-scores as input for the classifier.*

Relationship between classification accuracy and ISC

Regional classification accuracy and ISC were positively correlated (Pearson r = 0.85, **Figure 6a**). 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).

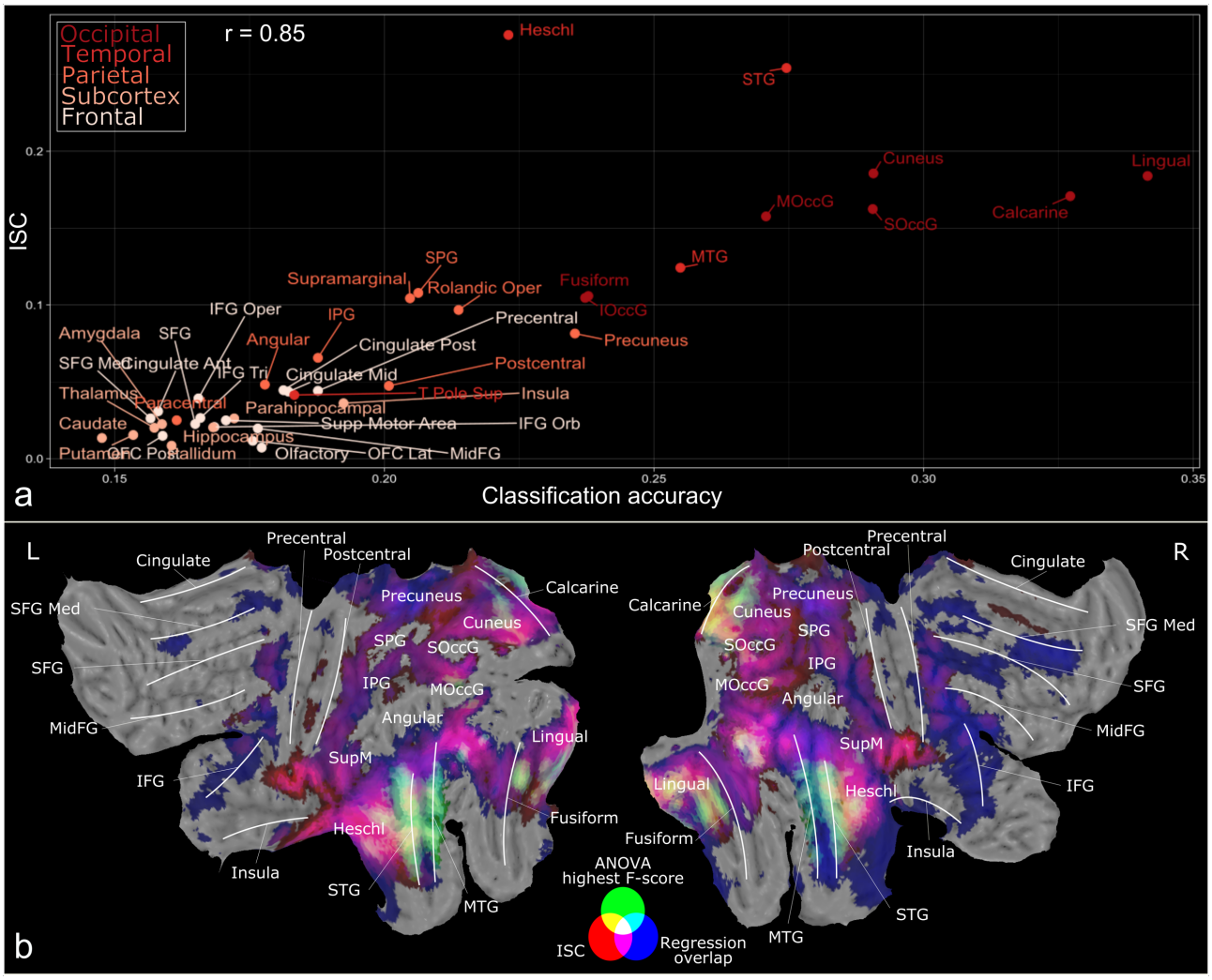

Figure 6. *(a) Scatterplot showing the relationship between ISC and classification accuracy. (b) Additive RGB map of the main findings. The overlap between activation patterns for perceptual dimensions in regression analysis is shown as blue (only areas where at least 3 dimensions expressed FDR corrected brain activation). FDR-corrected ISC across subjects is shown as red and 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 social dimensions reliably capture the social perceptual space. The cerebral topography for social perception was organized along a posterior-anterior -axis, with occipital and temporal cortical regions serving 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 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 patters throughout the experiment. These effects were observed although low-level sensory features were statistically controlled for. Altogether these results show that i) the brain represents the social world in a dimensional fashion, ii) that the pro- versus antisocial nature and social communicative intent and body movements are the primary dimensions extracted from social scenes, and that different brain regions have variable specificity in their response profiles towards social dimensions.

Cerebral posterior-anterior gradient 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 social dimensions. The largest clusters were organized along the valence dimension of the social interaction containing prosocial (e.g., kissing, touching, sexuality) versus antisocial (hurting others, yelling) behaviors. 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. 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 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. Frontally the responses were less consistent although brain activity in IFG, precentral gyrus and frontal part of medial SFG associated with a limited number of dimensions including "Prosocial behaviour", "Antisocial behaviour", "Feeding" and "Using an object". These data are consistent with previous univariate studies addressing social functions for LOTC [3,57-59], TPJ [4,5] and STS [6-8,60], MFC [61].

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 region activated by multiple social dimensions reflects responses to shared features across all the dimensions such as biological motion perception or intentionality detection; [9,62], or spatially overlapping yet dimension-specific processing. Multivariate classification analysis revealed that the answer to this question depends on the region. The feature selection for the whole-brain classification retrieved voxels from STS, LOTC, TPJ and FG (**Figure 5** and **Figure 6**) 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. Previous classification studies have shown how 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 [19]. Subsequent classification studies have shown that, for example, different facial expressions can be classified from activation patterns in FG, and STS [21,22] and goal-oriented actions in LOTC and interior parietal lobe [24,63]. 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.

Reliability versus specificity of responses to social perceptual dimensions

We observed robust inter-subject 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

the sensory regions, and it progressively becomes weaker toward the polysensory and associative cortices [26]. 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. Additionally, regional ISC was also associated with the corresponding regional classification accuracy (**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. In other words, regions responding to multiple social signals also do so in a time-locked fashion across subjects, whereas the responses become more idiosyncratic as they become more selective. 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.

Functional organization of social perception 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, occipitotemporal regions responded consistently to most social dimensions, while 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 [6,62].

Second, Heschl gyrus, the site of the auditory cortex [64] 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 weak ability to classify specific dimensions from this 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 have been linked with attention and memory retrieval [65], supramarginal gyrus with phonological [66] and visual [67] processing of words and SPG in visuospatial processing and working memory [68]. These parietal regions thus likely respond to some general features of the social signals or idiosyncratic brain states associated with social dimensions.

Finally, frontal and subcortical regions responded only to a limited number of social dimensions, and classification accuracy and ISC were 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 have previously been associated with higher-level social and affective inference such as linking social processing with decision making, affective processing and theory of mind [10,61]. Previous classification studies are in line with our results, and they have not found specificity for responses to social perceptual dimensions in frontal cortex [19,21,24,25]. 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. Finally, limbic regions such as amygdala and thalamus in turn have been linked with processing of (negative) emotions [28] and accordingly they showed reliable responses primarily to perception of antisocial behaviours.

Limitations

We used naturalistic video stimuli to model real-life social perception. Yet, because the stimulus film clips were extracted from popular Hollywood films and the social features present may not be completely representative of real-life social encounters. However, the advantage of this approach is the high intensity and frequency of the features, which benefits BOLD response modelling. We did not balance the low-level sensory differences across the clips. Thus, despite carefully controlling for low-level features in the analyses, 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 and these features were excluded from the analysis. 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 (**Figure SI-7**) 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. Consequently, 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 social features and mapped the cerebral topography of social perception that can be adequately described with 13 perceptual dimensions. Social perceptual space included clusters of social features describing prosocial 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 specific contributions of dimension-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. We thank Tuulia Malen for the fruitful conversations around the analysis methods used in this article and Juha Lahnakoski for the help with low-level feature extraction.

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 reviewed the 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 GLM predictor are available in NeuroVault (https://neurovault.org/collections/IZWVFEYI/).

Code availability

We developed in-house scripts for low-level audiovisual feature extraction and ridge regression optimization, and the scripts are freely available in GitHub (https://github.com/santavis/functional-organization-of-social-perception).

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 people. The 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., 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).

16 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).

17 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).

18 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).

19 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).

20 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).

21 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).

22 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).

23 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).

24 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).

25 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).

26 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).

27 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).

28 Karjalainen, T. *et al.* Opioidergic Regulation of Emotional Arousal: A Combined PET-fMRI Study. *Cereb Cortex* **29**, 4006-4016, doi:10.1093/cercor/bhy281 (2019).

29 Nummenmaa, L. *et al.* Brain Basis of Psychopathy in Criminal Offenders and General Population. *Cereb Cortex* **31**, 4104-4114, doi:10.1093/cercor/bhab072 (2021).

30 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).

31 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).

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

33 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).

34 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).

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

36 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).

37 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).

38 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).

39 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).

40 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).

41 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).

42 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).

43 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).

44 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).

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

46 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).

47 Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572-1573, doi:10.1093/bioinformatics/btq170 (2010).

48 Olivier Lartillot & Toiviainen, P. A Matlab Toolbox for Musical Feature Extraction From Audio. *International Conference on Digital Audio Effects* (2007).

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

50 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).

51 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).

52 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).

53 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).

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 Van Essen, D. C. Cortical cartography and Caret software. *Neuroimage* **62**, 757-764, doi:10.1016/j.neuroimage.2011.10.077 (2012).

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 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).

62 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).

63 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).

64 Da Costa, S. *et al.* Human primary auditory cortex follows the shape of Heschl's gyrus. *J Neurosci* **31**, 14067-14075, doi:10.1523/JNEUROSCI.2000-11.2011 (2011).

65 Cavanna, A. E. & Trimble, M. R. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* **129**, 564-583, doi:10.1093/brain/awl004 (2006).

66 Hartwigsen, G. *et al.* Phonological decisions require both the left and right supramarginal gyri. *Proc Natl Acad Sci U S A* **107**, 16494-16499, doi:10.1073/pnas.1008121107 (2010).

67 Stoeckel, C., Gough, P. M., Watkins, K. E. & Devlin, J. T. Supramarginal gyrus involvement in visual word recognition. *Cortex* **45**, 1091-1096, doi:10.1016/j.cortex.2008.12.004 (2009).

68 Koenigs, M., Barbey, A. K., Postle, B. R. & Grafman, J. Superior parietal cortex is critical for the manipulation of information in working memory. *J Neurosci* **29**, 14980-14986, doi:10.1523/JNEUROSCI.3706-09.2009 (2009).