Classification of social anhedonia using temporal and spatial network features from a social cognition fMRI task

Krohne, Lærke Gebser; Wang, Yi; Hinrich, Jesper Løve; Mørup, Morten; Chan, Raymond C K; Madsen, Kristoffer Hougaard

Published in:
Human Brain Mapping

Link to article, DOI:
10.1002/hbm.24751

Publication date:
2019

Document Version
Peer reviewed version

Title:

Classification of social anhedonia using temporal and spatial network features from a social cognition fMRI task

Authors:

Authors: Lærke Gebser Krohne [1,2,3,4], Yi Wang [3,5], Jesper L. Hinrich[1], Morten Mørup[1], Raymond C. K. Chan[3,4,5], Kristoffer H. Madsen[1,2,4]

Correspondence: Kristoffer H. Madsen (kristofferm@drcmr.dk), Kettegard Allé 30, 2650 Hvidovre, Denmark or Raymond C.K. Chan (rckchan@psych.ac.cn), Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing 100101, China.

(1) Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark
(2) Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, Denmark
(3) Neuropsychology and Applied Cognitive Neuroscience Laboratory, CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
(4) Sino-Danish College, University of Chinese Academy of Sciences, Beijing, China
(5) Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

Abbreviations:

(CSAS) Chapman Social Anhedonia Scale, (Emp) Empathy, (EPI) Echo Planar Imaging, (fMRI) functional Magnetic Resonance Imaging, (HSA) High Social Anhedonia, (ICA) Independent Component Analysis, (IPL) Inferior Parietal Lobule, (LSA) Low Social Anhedonia, (MCC) Mathews Correlation Coefficient, (mPFC) medial PreFrontal Cortex, (MSAA) Multi-Subject Archetypal Analysis, (NVR) Nuisance Variable Regressors, (P/ACC) Posterior/Anterior Cingulate Cortex, (PCon) Pooled Condition analysis, (Phy1/2) Physical condition 1 and 2, (ROI) Region Of Interest, (sMSAA) spotlight MSAA, (SVC) Support Vector Classification, (ToM) Theory of Mind, (TPJ) Temporoparietal Junction, (wbMSAA) whole brain MSAA

Published in: Human Brain Mapping, Volume: 40, Issue number: 17
DOI: 10.1002/hbm.24751

Abstract

Previous studies have suggested that the degree of social anhedonia reflects the vulnerability for developing schizophrenia. However, only few studies have investigated how functional network changes are related to social anhedonia. The aim of this fMRI study was to classify subjects according to their degree of social anhedonia using supervised machine learning. More specifically, we extracted both spatial and temporal network features during a social cognition task from 70 subjects, and used support vector machines for classification. Since impairment in social cognition is well established in schizophrenia-spectrum disorders, the subjects performed a comic strip task designed to specifically probe theory of mind (ToM) and empathy processing. Features representing both temporal (time series) and network dynamics were extracted using task activation maps, seed region analysis, independent component analysis (ICA) and a newly developed multi subject archetypal analysis (MSAA), which here aimed to further bridge aspects of both seed region analysis and decomposition by incorporating a spotlight approach.

We found significant classification of subjects with elevated levels of social anhedonia when using the times series extracted using MSAA, indicating that temporal dynamics carry important information for classification of social anhedonia. Interestingly, we found that the same time series yielded the highest classification performance in a task classification of the ToM condition. Finally, the spatial network corresponding to that time series included both prefrontal and temporal-parietal regions as well as insula activity, which previously have been related schizotypy and the development of schizophrenia.

Key words: functional connectivity, social anhedonia, decomposition, archetypical analysis, support vector classification

1. Introduction

In the perspective of schizophrenia as a neurodevelopmental disease, it is very important to study potential early risk groups [Insel, 2010; Lewis and Levitt, 2002; Weinberger, 1987]. Schizotypy refers to a set of positive, negative or disorganized personality traits that are related to schizophrenia [Ettinger et al., 2015]. Individuals with schizotypy are non-clinical subjects, but they have some psychotic-like experiences, ranging from few (low schizotypy) to numerous (high schizotypy), which reflect their vulnerability for developing schizophrenia-spectrum disorders [Blanchard et al., 2011; Kwapil, 1998; Mason, 2015]. The importance of studying schizotypy is twofold. Firstly, it has been suggested that early detection and intervention of schizophrenia might yield substantial improvements in treatment outcome, comparable to what has been reported in preventive approaches to cardiac death [Insel, 2010]. Secondly, schizotypy studies have shown to increase the understanding of the psychopathology of schizophrenia.

Anhedonia, which is the reduced capability to experience pleasure in normal pleasurable situations, is considered as a negative dimension of schizotypy. High levels of anhedonia have consistently been reported in patients with schizophrenia [Blanchard et al., 2011; Bora et al., 2009] and ultra-high risk groups [Bora and Pantelis, 2013]. Furthermore, longitudinal studies have shown that subjects with a high level of *social* anhedonia (reduced pleasure experience in social contexts) are more likely to develop schizophrenia-spectrum disorders later on, compared to control groups or high scorers on positive schizotypy (measured by Perceptual Aberration Scale and Magical Ideation Scale) [Blanchard et al., 2011; Kwapil, 1998] [Gooding et al., 2005; Wang et al., 2014]. For these reasons, social anhedonia will be the focus in this study.

On the other hand, the importance of social cognition research in understanding psychopathology of schizophrenia has been acknowledged [Green et al., 2015; Penn et al., 2007]. Studies have shown that social cognition is substantially impaired in patients with schizophrenia and early risk groups (Bora and Pantelis, 2013; Fett et al., 2011), and changes have even been reported in subjects with schizotypy (Blanchard et al., 2011; Morrison et al., 2013). Theory of mind (ToM) is often defined as the ability to attribute mental states to ourselves and others, and consists of both a cognitive (centered about processing of knowledge and believes) as well as an affective (emotional processing) component [Sebastian et al., 2012; Shamay-Tsoory et al., 2010]. The affective aspect is very similar to what is often defined as cognitive empathy [Sebastian et al., 2012], and will for simplicity, be referred to as empathy (Emp) in the rest of the paper. The abnormalities of ToM or empathy ability has been related to schizotypy [Bora and Pantelis, 2013; Pickup, 2006]. In particular, previous studies consistently suggested an association between high negative schizotypy and poor metalizing ability measured by self-report scales [Bedwell et al., 2014; Henry et al., 2008; Thakkar and Park, 2010; Wang et al., 2013] and behavioral tasks [Pflum and Gooding, 2018; Thakkar and Park, 2010].

Functional imaging studies have correlated the degree of schizotypy and activity in isolated brain regions reviewed in [Ettinger et al., 2015; Nelson et al., 2013], however, until now, only relatively few studies have investigated how functional connectivity

changes in individuals with schizotypy. Lagioia et al. determined six resting state networks and found that functional connectivity in the visual and auditory networks were correlated to the degree of schizotypy [Lagioia et al., 2010]. In terms of social anhedonia, studies found altered connectivity between the striatal seeds and the cingulate cortex as well as the insula during resting state [Wang et al., 2016] and altered functional connectivity of the amygdala during facial emotion processing task [Wang et al., 2018]. Although previous studies have looked at correlations between brain activation or connectivity and the degree of schizotypy, actual classification is of great importance to determine if these changes can be used to categorize or even diagnose subjects already in early stages. Machine learning methods have been used in classification of schizophrenia patients from healthy control using functional imaging data (reviewed in [Madsen et al., 2018]). So far there are a few studies that have investigated the classification performance of individuals with schizotypy based on brain activation during task-based fMRI using machine learning methods [Modinos et al., 2012; Shinkareva et al., 2006], but both studies only focused on the positive dimension of schizotypy instead of negative schizotypy.

The aim of our study was to investigate which features extracted from functional networks during a social cognition task were sufficient to classify subjects according to their degree of social anhedonia using supervised machine learning. To this end, we extracted brain network features using both standard activation maps and traditional seed region analysis [Biswal et al., 1995; Cole et al., 2010], but also decomposition methods based on independent component analysis (ICA) [Beckmann and Smith, 2004; Calhoun et al., 2001] and the multi-subject archetypal analysis (MSAA) described in [Hinrich et al., 2016]. Seed based analysis procedures extract features from defined seed regions, whereas ICA uses unsupervised learning to decompose the data into latent maximally independent spatial components. Each of these components can be thought of as representing a functional brain network. MSAA can be seen as seed region-based analysis where the seeds are automatically defined based on unsupervised learning. The features used for the classification, and the relation between the approaches are illustrated in figure 1.
The use of different methods helped us exploring the separate importance of spatial and temporal network features.

Our second aim was to specifically investigate which features were important for classification. We investigated time series extracted from either specific brain regions or from networks, and hypothesized that the features showing significant classification of subjects with high social anhedonia would entail brain regions previously associated with schizotypy and the development of schizophrenia. Such regions include as prefrontal cortex, temporal-parietal regions and insula [Chan et al., 2011; Kühn et al., 2012; Takahashi et al., 2009].

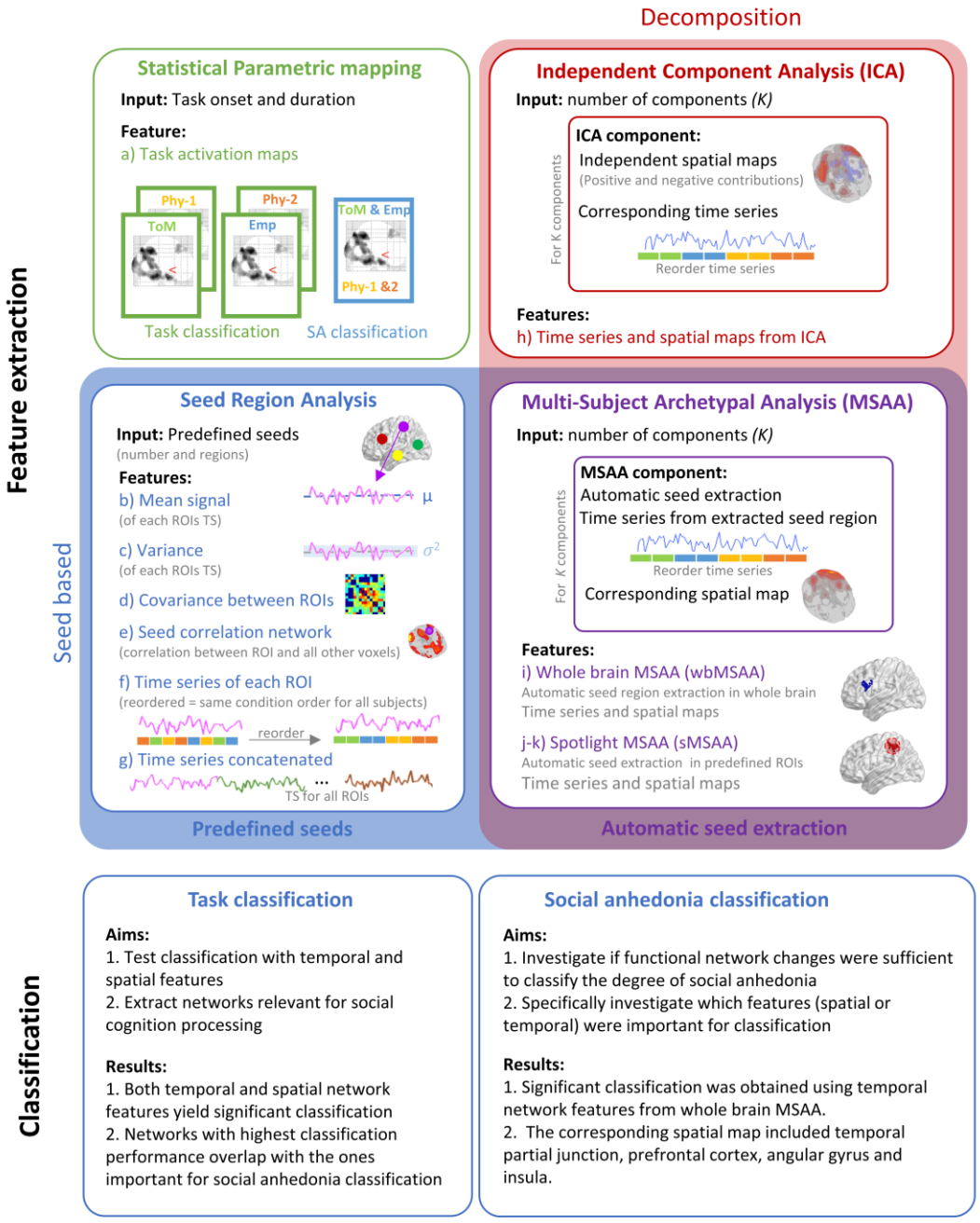


Figure 1: Illustration of the feature extraction methods and aims of classification. We roughly divide the feature extraction methods considered into statistical parametric mapping, unsupervised decomposition, and seed region analysis. Here the letters a–k refers to the results of individual analyses as displayed in Table 1. (a) Refers to spatial maps extracted from statistical parametric mapping and classification approach (b,c) are based on static measures from seed based analysis, (d,e) are expressions of functional connectivity within and between the seeds and (f,g) reflect temporal dynamics of seed based analysis. In analyses (f–k) the time series are rearranged such that the order of the conditions is consistent across subjects, this was necessary as the order of the tasks were randomized across participants. In approach (h) time series and spatial maps obtain from ICA are considered, and approaches (i–k) are based on archetypical analysis which can be seen as seed based analysis with automatical extraction of seeds, merging aspects of ICA and seed region analysis. For approaches (e,f and h–k) classification was performed for each ROI/component separately, and thus multiple comparisons correction was used to assess the significance of the results.

2. Material and Methods

2.1 Participants

This study included 76 college students from Guangzhou Medical University (37/39 male/female) with age between 17–21 years (μ = 19.3 years, σ = 0.9 years). The subjects were chosen such that they covered a continuous range of schizotypy and none had a history of drug abuse, or psychiatric disorders. The Chapman Social Anhedonia Scale (CSAS) was used to assess the inability to experience pleasure from social interactions [Chan et al., 2015; Chapman et al., 1994]. The CSAS consists of 40 items (e.g., "Just being with friends can make me feel really good"; "Making new friends isn't worth the energy it takes") and higher score indicated more severity of anhedonia. The good reliability and validity of the CSAS has been proved in Chinese context [Chan et al., 2015]. The internal consistency coefficient was 0.84 in our sample. The mean and standard deviation of all four Chapman scales and the Becks Depression Inventory can be found in supplementary table 2. All subjects were right-handed and a radiologist screened all scans to exclude subjects with any incidental clinical abnormalities. The study was approved by the Ethics Committee of the Institute of Psychology at the Chinese Academy of Sciences.

In a previous analysis, the same dataset showed specific correlation between the degree of social anhedonia and the mean activity in; the middle temporal gyrus, the temporoparietal junction and the medial prefrontal gyrus. [Wang et al., 2015b]. In contrast, this study investigated if the measured changes were sufficient for actual classification of subject with high and low social anhedonia (HSA/LSA) using support vector machines.

Subjects were defined in the HSA group if their CSAS score was more than one standard deviation above the mean (based on a large independent dataset including 887 subjects [Chan et al., 2012]). This separation threshold was relatively low, but comparable with what previously has been used in the literature [Wang et al., 2015a]. Furthermore, even when using this relatively low separation boundary, the dataset was unbalanced (HSA = 14/LSA = 56 subjects). As it will be discussed more carefully in section 2.9 and 3.4 this complicated the classification procedure.

2.2 Functional imaging task

A Chinese adaption of the visual comic strip task developed by Völlm et al. was presented in a block design [Völlm et al., 2006; Wang et al., 2015b]. The task included four different conditions namely ToM, empathy and two corresponding control conditions; 'Physical causality with one character' (Phy1) and 'Physical causality with two characters' (Phy2). Whereas the ToM and empathy condition were designed to probe the corresponding social cognition processing, the physical conditions were designed to look as similar to the social cognition conditions as possible. Hence, Phy1 included comic strips with only one character, whereas Phy2 included two interacting characters. Each condition was presented twice, resulting in a total of eight blocks, with each block containing five trials of comic strips belonging to the same condition. When the condition was presented the second time, a new set of comic strips were used, hence each comic strip was only seen once by each subject. In each trial, three pictures depicting a short story were displayed in the upper half of the screen for six

seconds. Afterwards, two pictures appeared in the lower half of the screen for another six seconds. During the second six second period, participants were asked to choose one of the two pictures from the lower half of the screen as the appropriate ending to the story by pressing the corresponding button with their right hand. For the ToM trials, the original cartoons from the 'Attribution of intention' [Brunet et al., 2000] condition was used and the question: "What will the main character do next?" was asked. For the empathy condition, scenarios with emotional states attribution was showed and the question "What will make the main character feel better?" was asked. The total duration of the whole task was 8 minutes and 48 seconds. To control for effects of practice and fatigue the blocks were randomized across subjects**.** More details, as well as examples on the comic strip task, can be found in [Völlm et al., 2006], who developed the task.

2.3 Image acquisition and preprocessing

All scans were acquired on a 3T Siemens Verio MR scanner at Guangzhou First People's Hospital in 2012, using a T2* weighted gradient echo based echo planar imaging (EPI) sequence with echo time = 28 ms, repetition time = 2000 ms and flip angle = 90°. 264 whole brain volumes were acquired with a slice thickness of 4 mm, matrix size 64 × 64 (32 slices in coronal plane), field of view = 210 × 210 mm, voxel size = 3.3 × 3.3 × 4 mm and bandwidth = 2232 Hz/px.

The images were preprocessed using Statistical Parameter Mapping (SPM) version 12 revision 6685. The eight first volumes of the scans were removed to ensure T1 equilibrium, and slice timing correction was performed to correct for the descending slice order with the middle slice as reference. The EPI images were normalized to the EPI template (ICBM-152) and the images were re-sliced to 3 × 3 × 3 mm. As this study focused on functional connectivity modeling additional preprocessing steps were included, since artifacts can lead to spurious connections [Power et al., 2012]. First despiking was performed to remove transient phenomena without scrubbing [Patel et al., 2014] using a Daubechies 4 mother wavelet. Then additional nuisance regressors were included in a multiple linear regression and the effect of them was removed from the data. These included; (i) mean signal and second order detrending (ii) nuisance variable regressors (NVRs), (iii) spike percentage from despiking, and (iv) explicit modelling of specific time frames based on the DVARS and frame wise displacement criteria as described in [Patel et al., 2014; Power et al., 2012], using a threshold of 1% and 1mm respectively. NVRs were used to remove both *residual motion* (24-parameter Volterra expansion model [Friston et al., 1996] based on the six head motion parameters estimated during realignment) and *physiological noise* where the mean signals from non-neuronal brain regions was extracted. Non-neuronal tissue included white matter, which was segmented using the SPM12 tissue probability map with a threshold of 0.5, cerebrospinal fluid in the lateral ventricles according to the HarvardOxford atlas [Desikan et al., 2006]. To reduce the influence of partial volume effect with gray matter, the white matter mask was eroded by two voxels. Finally, the images were smoothed using an isotropic Gaussian 8mm full width at half maximum filter.

2.4 Classification using support vector classification

To classify subject into high and low social anhedonia, as well as for the task classification, we used binary support vector machines to perform supervised classification [Cortes and Vapnik, 1995]. The goal of support vector classification (SVC) is to identify a function that discriminates the labels (e.g. high or low social anhedonia) in a training dataset, such that it is possible to use this function to classify the labels of a test dataset. In principle, it is possible to apply SVC directly to the (preprocessed) fMRI images. However, due to the very high dimensionality of fMRI images in relation to the number of subjects, perfect classification in the training dataset is trivial but with poor generalization to the test data due to overfitting (see Madsen et al. for a more thorough description of support vector classification for fMRI data [Madsen et al., 2018]). We therefore applied SVC on 11 spatial and temporal features (analysis a-k listed in table 1), which were extracted from the fMRI data to capture the network changes of interest.

In short, one feature included the task specific activation maps determined by a SPM analysis (section 2.5), six features resulted from a seed region analysis (2.6), and four came from the decomposition methods (2.7). For some of the seed region analysis and decomposition methods, we extracted both time series and spatial maps for each seed region/component respectively (analysis e-f and h-k), and classification was then performed on each extracted feature respectively. Table 1 lists the classification performances of the features yielding the highest classification performance, and maximum permutation statistics was therefore used to correct for multiple comparisons between the components as described in section 2.9.

For classification we used the SVC-C implementation from the LIBSVM [Chang and Lin, 2011] library with a linear kernel. We used nested cross validation to determine the soft margin penalty parameter, and to evaluate the classification performance. For task classification, the cross validation scheme was based on grouped stratified cross validation where each subject was considered a group. In the inner loop the optimal soft margin penalty parameter (C-parameter) was determined in a logarithmic grid containing 11 values $C \in [2^{-5}, 2^{-3}, \dots, 2^{15}]$ by 10-fold cross validation, and an unbiased estimate of the classification accuracy was obtained in another outer 10-fold cross validation loop.

For HSA classification, a similar scheme was followed but without grouping as there was only one sample per subject in this case. Furthermore, the C-parameter was adjusted for each class to counteract the class imbalance [Chang and Lin, 2011]. The inner and outer loops where set to reserve exactly one sample of the least common class (HSA) resulting in 13- and 14-fold cross validation, this ensured that stratification across splits was achievable while preserving sufficient data for training.

2.5 Statistical Parametric mapping

To determine task specific activity maps for all four task conditions (ToM, Emp, Phy1 and Phy2), we ran a standard SPM analysis, performing a parametric statistical test for each voxel separately. The significance level was $\alpha_{RFT} \leq 0.05$, where random field theory was used to correct for multiple comparisons. The activation maps were later used as features (classification approach (a)) for classification. Since the activation maps were constructed based on information about task onset and duration, we

expected that they would obtain a high performance for classifying the tasks conditions. However, for the social anhedonia classification, which was not directly related to the presented task, the static nature of this feature extraction step might not identity information useful for classification. For the task classification, we used one task activation map for each social construct, i.e. the ToM – Phy1 condition, and empathy – Phy2 condition respectively. For the HSA classification, we used one single contrast map, reflecting the pooled effect of ToM and empathy in comparison to the physical control conditions, as illustrated in figure 1.
Furthermore, we used SPM to perform a pooled condition analysis (PCon) identifying the pooled effect of the social cognition tasks (ToM and Emp) compared to the control conditions. This was used as input for the spotlight multi-subject archetypal analysis as described in section 2.7.

2.6 Seed region analysis

Seed region analysis is a very intuitive way to investigate the brain by determining the activity in predefined regions of interest (ROIs). In this study six different methods (approach b-g) were used to investigate the ROI specific activity, which later were used for classification. These included; approach (b): the mean activity and (c) variance within each ROI, (d) the covariance between all *N* ROIs (calculated pairwise), (e) the correlation between the time series of each ROI with all voxels in the brain (classical seed based analysis) resulting in a connectivity map for each seed, (f) the extracted time series of each ROI separately, and (g) the time series of all ROIs concatenated. All of these are illustrated in figure 1, and enabled us to study the importance of temporal dynamics (approach (f) and (g)), network coupling (approach (d) and (e)) and static features separately.
The time series of each ROI were extracted as the first eigenvariate, which reflects the most consistent source across all included voxels. Compared to using the average across the ROI, this can be an advantage if there are multiple sources in the given ROI [Poldrack and Gorgolewski, 2014]. When using the time series as feature for classification, they were rearranged (by simple temporal reordering) such that they reflect the same structure (ToM, Emp, Phy1 and Phy2) for all subjects, despite that the order of the conditions were randomized across subjects. In approach (e) the correlation between the time series of the ROIs, and that of all other voxels in the brain, was determined using Pearson's correlation coefficient, followed by conversion to Z-score through the Fisher Z-transform [Fisher and Fisher, 1915].
In approach (e) and (f) classification was performed independently for each ROI, highlighting the importance of multiple comparisons correction as described more carefully in section 2.9.

2.7 Decomposition methods

One of the most frequently used decomposition methods in neuroscience is the **independent component analysis (ICA),** which determines a predefined numbers of maximally independent sources [McKeown et al., 1998]. For fMRI data, these sources represent spatial networks, where all included regions have similar time series. For multi-subject analysis, common spatial components can be obtained by concatenating subject data in time [Calhoun et al., 2003]. More specifically, ICA seeks to identify latent sources in the data from multiple mixed measurements via the per subject linear mixing model

$$\boldsymbol{X}_i = \boldsymbol{A}_i \boldsymbol{S}_i + \boldsymbol{E}_i,$$

where $\boldsymbol{X}_i \in \mathbb{R}^{T \times V}$ is the data matrix measured at T timepoints and across V voxels for the i'th subject, $\boldsymbol{A}_i \in \mathbb{R}^{T \times K}$ contains K source time series as columns, $\boldsymbol{S}_i \in \mathbb{R}^{K \times V}$ is comprised by the K spatial components as rows, and $\boldsymbol{E}_i \in \mathbb{R}^{T \times V}$ is a residual error term. While the expression above enforces no coupling across subjects, such dependence is usually accomplished by enforcing dependence or equality of $\boldsymbol{S}_i$ across subjects, which we will consider later. Since minimizing the residual leads to rotational ambiguity and thereby non-unique solutions, additional assumptions or constraints are typically imposed on either the time series or spatial components or both. In spatial ICA, this typically amounts to assuming a non-Gaussian source distribution upon the spatial components.

Multi-subject archetypal analysis (MSAA) is another data driven approach, which bridges aspects of seed analysis and decomposition [Hinrich et al., 2016] [Cutler and Breiman, 1994; Mørup and Hansen, 2012]. MSAA is a latent variable model, similar to ICA, but is constrained to have latent factors that reflect representative points in the data, termed "archetypes". For fMRI data, the archetypes are a set of representative time-series which have a corresponding set of spatial networks. Whereas ICA represents the fMRI data by a linear mixture of maximally independent spatial maps, MSAA determines the components through iterative optimization of; (i) a seed region matrix, $\boldsymbol{C}$ (that is identical for all subjects) and (ii) a set of subject specific spatial maps ($\boldsymbol{S}$) corresponding to each archetype. The archetypes for each subject are given as the weighted average of the voxels specified in the seed region matrix, such that

$$\boldsymbol{A}_i = \boldsymbol{X}_i \boldsymbol{C}$$

where $\boldsymbol{X}_i$ is the subject specific data and $\boldsymbol{A}_i \in \mathbb{R}^{T \times K}$ includes all archetypes defining distinct temporal profiles for the i'th subject. Figure 2 illustrates how MSAA represents the fMRI data as archetypes and spatial maps. Each voxel time series is reconstructed by convex combinations as defined in $\boldsymbol{S}_i$ of the archetypes. Thus, both the columns of $\boldsymbol{S}_i$ and $\boldsymbol{C}$ are constrained to be non-negative and to sum to one. The resulting spatial maps can therefore be interpreted as the fractional contribution of all voxels to the archetypal time series as specified in $\boldsymbol{A}_i$.

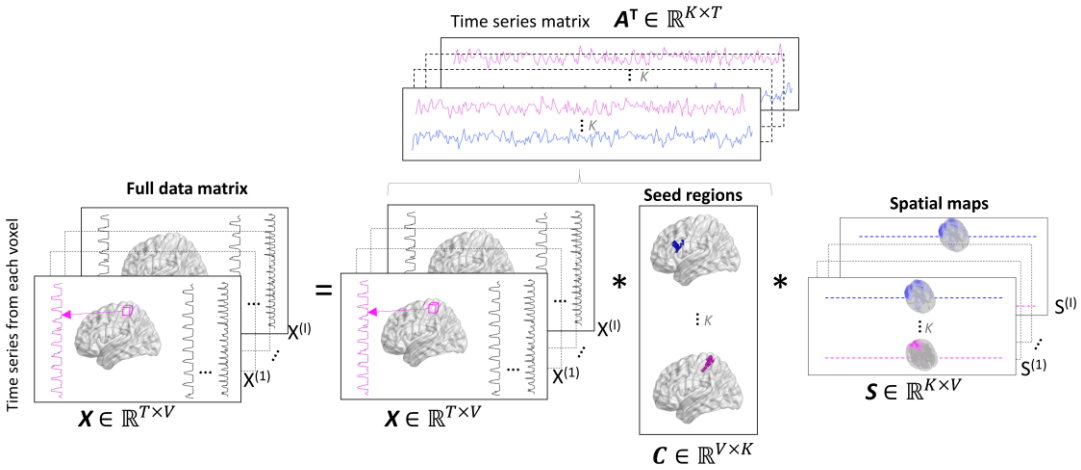


Figure 2: Illustration of whole brain multi-subject archetypal analysis (wbMSAA). The columns data matrix $\boldsymbol{X}$ include the time series for all V voxels. Through iterative optimization, the MSAA algorithm determines a seed region matrix $\boldsymbol{C}$, specifying the optimal choice of K seed regions across subjects, as well as a set of K temporal ($\boldsymbol{X}_i\,\boldsymbol{C}$) and spatial components $\boldsymbol{S}_i$ for all B subjects. The model also includes a subject specific noise map, which is not specified in this figure.

The MSAA decomposition is in general unique [Mørup and Hansen, 2012] and the linear model (per subject) can be formulated as

$$\boldsymbol{X}_i = \boldsymbol{X}_i\boldsymbol{C}\boldsymbol{S}_i + \boldsymbol{E}_i,$$

Under the assumption of independently distributed additive Gaussian noise with heteroscedasticity over voxels we have

$$\boldsymbol{e}_{i,v} \sim \mathbb{N}\left(\mathbf{0}, \sigma_{i,v}^2\right),$$

Where $\boldsymbol{e}_{i,v}$ is a time vector of the residual in voxel v for subject i and $\sigma_{i,v}^2$ is the voxel and subject specific noise variance. This lead to the likelihood

$$\mathcal{L} = \prod_i^B \prod_v^V \frac{1}{\left(2\pi\sigma_{i,v}^2\right)^{T/2}} \exp\left(-\frac{\left\|\boldsymbol{x}_{i,v} - \boldsymbol{X}_i\boldsymbol{C}\boldsymbol{s}_{i,v}\right\|^2}{2\sigma_{i,v}^2}\right).$$

Optimizing this likelihood leads to a sparse seed region matrix $\boldsymbol{C}$, which selects the archetypical voxel time series that best span the entire dataset, and a corresponding set of subject specific spatial maps $\boldsymbol{S}_i$. For explicit derivation of update rules see [Hinrich et al., 2016]. Determining $\boldsymbol{C}$, $\boldsymbol{S}_i$ and $\boldsymbol{\sigma}_i$ is a non-convex optimization problem [Mørup and Hansen, 2012], but a solution can be found by alternating optimization, i.e. optimizing for $\boldsymbol{C}$ while keeping $\boldsymbol{S}_i$ and $\boldsymbol{\sigma}_i$ fixed and vice versa.

Connection between ICA, Seed based analysis and MSAA

In the following, we show how the decomposition scheme of MSAA can be used to bridge spatial group ICA with seed based analysis. The MSAA directly finds subject specific spatial maps ($\boldsymbol{S}_i$) and temporal activations ($\boldsymbol{X}_i\boldsymbol{C}$) which through the common seed matrix ($\boldsymbol{C}^{\boldsymbol{MSAA}}$) express variability across subjects. In contrast, spatial group ICA assumes the spatial sources are fixed across subjects [Calhoun et al., 2003], however, individual subject expressions (spatial maps) can be identified through either back reconstruction or dual regression [Erhardt et al., 2011]. When the spatial sources are known and no additional constraints are imposed upon the time series, solving for $\boldsymbol{A}_i$ reduces to an ordinary least squares regression problem where the solution can be expressed as

$$\boldsymbol{A_i} = \boldsymbol{X_i}\overline{\boldsymbol{S}}^{\top}(\overline{\boldsymbol{S}\boldsymbol{S}}^{\top})^{-\boldsymbol{1}}.$$

Here $\overline{\boldsymbol{S}}$ represents the shared spatial components**.** In back reconstruction, individual subject components are formed through the expression

$$\boldsymbol{X_i} = \boldsymbol{A_i}\tilde{\boldsymbol{S}}_{\boldsymbol{i}},$$

where $\tilde{\boldsymbol{S}}_{\boldsymbol{i}}$ is the individual spatial components and inserting the expression for $\boldsymbol{A_i}$ we obtain

$$\boldsymbol{X_i} = \boldsymbol{X_i}\overline{\boldsymbol{S}}^{\top}(\overline{\boldsymbol{S}\boldsymbol{S}}^{\top})^{-\boldsymbol{1}}\tilde{\boldsymbol{S}}_{\boldsymbol{i}},$$

which again allows the individual spatial maps to be formed by solving an ordinary least squares problem. This establishes an attractive correspondence between MSAA and group ICA, where, in this case, the non-sparse "seed matrix" given by $\boldsymbol{C}^{\boldsymbol{ICA}} = \overline{\boldsymbol{S}}^{\top}(\overline{\boldsymbol{S}\boldsymbol{S}}^{\top})^{-\boldsymbol{1}}$ can take on both positive and negative values whereas the columns are not constrained to sum to one.

Spotlight MSAA

In this study, we considered an expansion to the MSAA algorithm by implementing a spotlight approach that restricted the seed region matrix to pre-specified regions of interest. This allowed specifying a subset of voxels from which the seed regions were then defined,

$$\boldsymbol{X_i} = \widetilde{\boldsymbol{X}}_{\boldsymbol{i}}\boldsymbol{C}\boldsymbol{S_i} + \boldsymbol{E}_{\boldsymbol{i}},$$

where $\widetilde{\boldsymbol{X}}_{\boldsymbol{i}}$ is the subset of voxel time series in the regions of interest as illustrated in figure 3. This approach is useful to investigate "archetypal generating activity" in specific areas, or if only approximate regions of interest are known. The derivation is given in [Hinrich et al., 2016], though they did not investigate the restricted method or considered the stability of its solution.

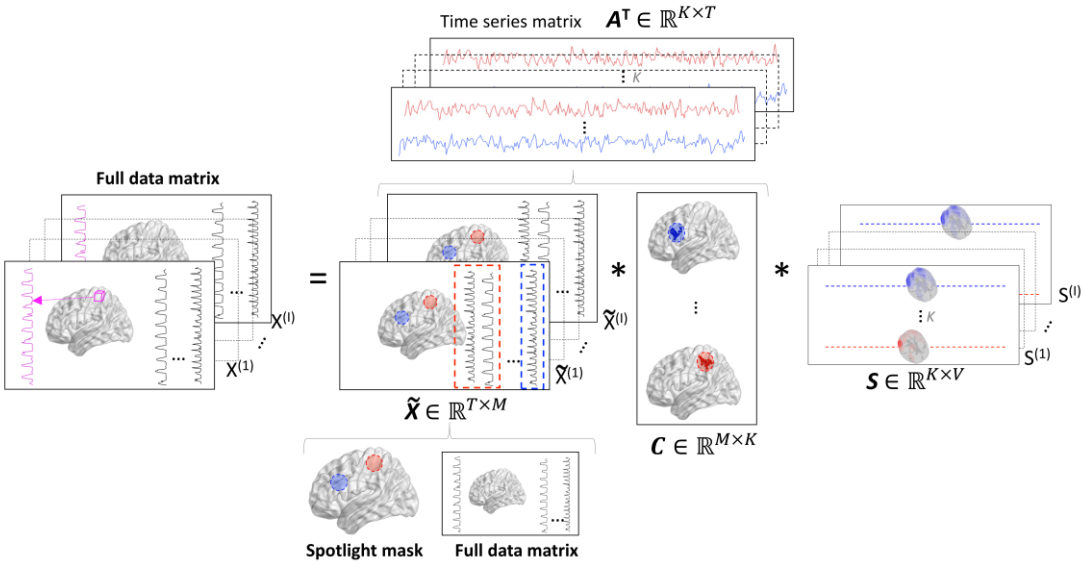


Figure 3: Illustration of the spotlight (sMSAA) approach. For the spotlight MSAA ***C*** and ***X*** are restricted to only include a subset of the voxels corresponding to some predefined regions of interest (for simplicity only two regions are shown here). However, the exact localization and size of the seed regions are still optimized by the algorithm. Apart from the restriction, the model is identical to the wbMSAA shown in Figure 2.

In the remaining manuscript, we will refer to the restricted MSAA as spotlight MSAA (sMSAA) in contrast to the original whole brain MSAA (wbMSAA).
We have run two sMSAA analysis using seed region restriction maps from; (i) a literature study ($sMSAA_{Lit}$) and (ii) from a pooled condition analysis ($sMSAA_{PCon}$) respectively, as described in section 2.8.

Implementation

We applied group ICA through the GroupICATv4.0a GIFT toolbox [Rachakonda et al., 2015], using the Infomax algorithm and the corresponding default settings. The number of components was selected using the minimum description length as proposed in Li et al. [Li et al., 2007], which for our dataset resulted in 25 components. Finally, subject specific spatial and temporal components were determined using the default back reconstruction method implemented in GIFT [Calhoun et al., 2003]. For visualization purposes, the spatial components where z-scored and both positive and negative contributions were shown.
For the MSAA analysis we used the same number of components as for ICA. As the MSAA algorithm is a non-convex optimization problem, there was a risk that the solution would get stuck in a local and not global minimum. As done for other non-convex problems, we therefore repeated the analysis several times with different random initializations for each run, and chose the solution with the lowest final cost at the end of the optimization. Optimization halted after either a maximum of 250 iterations or when the relative decrease in the cost function was less than 10^{-6} as in Hinrich et al. [Hinrich et al., 2016]. Different initializations, such as the FurthestSum initialization [Mørup and Hansen, 2012] have been suggested for archetypal analysis. However, as these resulted in a higher final cost function, random initialization was used in this study.

To increase the stability of MSAA the algorithm was rerun with 10 random initializations choosing the solution that obtained the lowest cost function. To further investigate the stability of the algorithm we repeated the fitting procedure 10 times and compared the spatial maps across runs using spatial correlation, this indicated that components were fairly stable across runs, providing an average correlation of 0.86. Visual inspection revealed that the differences were primarily due to minor changes in network expressions between runs for some components, see the stability of wbMSAA section in the supplementary material for further information. Furthermore, the finding of significant classification of HSA using wbMSAA times series reproduced in all 10 individual runs.

2.8 Predefined regions of interest

For the seed regions analysis and spotlight MSAA predefined ROIs were a prerequisite for the analysis. We defined the ROIs as all voxels in a sphere (8mm radius) around a given center coordinate. These were determined through a literature study of ToM and empathy processing, taking into account both reproducibility of the areas [Abu-Akel and Shamay-Tsoory, 2011; Shamay-Tsoory et al., 2010] and specificity for the comic strip task [Benedetti et al., 2009; Völlm et al., 2006; Wang et al., 2015b]. The center coordinates are illustrated and labeled in figure 4 and the MNI coordinates can be found in supplementary table 3.

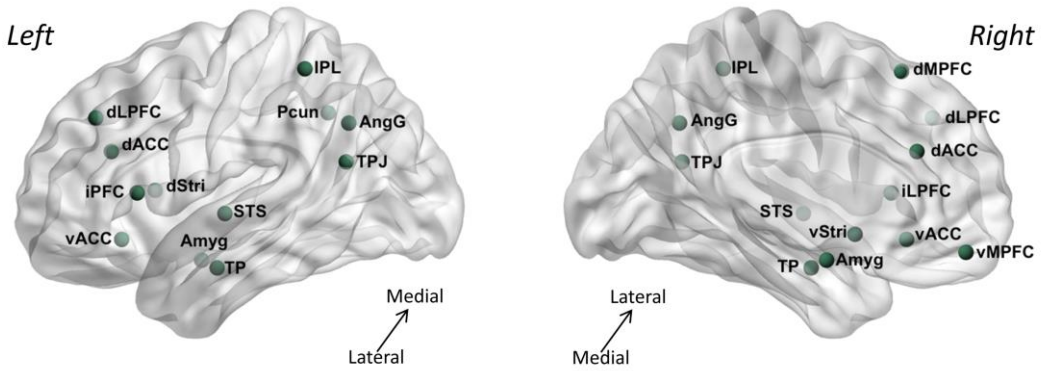


Figure 4: Illustration of center coordinates determined based on the literature. These nodes were used both for the seed region analysis approaches, and for the spotlight MSAA. Abbreviations: Amyg, amygdala; AngG, angular gyrus; d/v ACC, dorsal/ventral anterior cingulate cortex; d/v mPFC, dorsal/ventral medial prefrontal cortex; d/v Stri, dorsal/ventral striatum; IPL, inferior parietal lobule; i/dL PFC, inferior/dorsolateral prefrontal cortex; Pcun, precuneus; STS, superior temporal sulcus; TP, temporal pole; TPJ, temporoparietal junction

Finally, for the classification of social anhedonia using spotlight MSAA, center coordinates were also obtained using the peak coordinates of significant clusters for the pooled condition analysis (PCon) as described in section 2.4. All center coordinates can be found in supplementary material table 3.

2.9 Statistical tests and measures

We used the accuracy as performance measure for the task classification, as it provides a straightforward interpretation for balanced samples. However, for the classification of unbalanced datasets the accuracy measure can be misleading. I.e., even in the case of a trivial classification where all subjects were classified as the dominant class (e.g. in this study: LSA = 56, HSA= 14), the accuracy would be 56/(56+14) = 80%. To mitigate this issue, we used the Matthews correlation coefficient (MCC) for the social anhedonia classification, as it is regarded as being one of the best summary statistic measures for unbalanced datasets [Baldi et al., 2000; Powers, 2011]. MCC returns a value between -1 (worst) and 1 (best) where 0 indicates that the result is no better than random classification.

For all classification procedures statistical inference of the performance was performed using a random permutation testing procedure [Nichols and Holmes, 2003]. For each of 1000 random permutations the entire classification procedure, including the inner and outer nested cross validation loops, were repeated to obtain an empirical null distribution of the performance measure (accuracy and MCC for task and HSA classification respectively).

As mentioned above, for some features the classification was performed for each ROI/network separately, and the significance of these analyses therefore needed to be corrected for multiple comparisons. This was done by the use of maximum permutation statistics, where an empirical null distribution was obtained by considering only the most significant effect over the entire set (here regions or components), which controls the family-wise error over the set.

3. Results and Discussion

This combined results and discussion section is split into five subsections, covering different aspects of the study. The first section includes a general discussion of the networks determined by the decomposition methods (ICA and MSAA), and comments on the stability of these approaches. Section 3.2 and 3.3 cover the results from the task and social anhedonia classification respectively, and discuss how these findings correspond to our hypotheses and previous literature. Since MSAA is a new decomposition method, which previously only has been applied in one neuroimaging study [Hinrich et al., 2016], we comment on the general interpretability and stability of the MSAA networks, and compare it with ICA in section 3.4. Finally, in section 3.5 we discuss general limitations of our study, as well as suggestions for future development and applications.

3.1 Network extraction using decomposition methods.

Visual inspection of the spatial maps from ICA and MSA showed that both methods captured networks which previously have been related to ToM processing [Benedetti et al., 2009; Völlm et al., 2006; Wang et al., 2015b], without any a priori knowledge about the task onset and duration (which was a requirement for the previous studies that used SPM analysis). Furthermore, we observed that both ICA and MSAA successfully captured effect of no interest (such as pulsation and movement artifacts) as well as other specific activity (visual or motor processing) in separate networks. This is an important sanity check, as noise/unrelated activity would otherwise contaminate the task related networks.

Stability: As described in section 2.7, the wbMSAA algorithm was run 10 x 10 times, comparing the stability of the spatial networks, when the best (lowest final cost) solution of 10 runs was compared for 10 repetitions. Using greedy matching a mean correlation of 86% was obtained. Visual inspection showed that the same networks were found in all 10 runs, but with minor differences, resulting in the non-perfect matching. Using the 10 repeated runs to investigate the classification stability, the same feature (discussed later in section 3.3) was found to result in the highest classification performance (MCC varied between 0.49-0.56), which was significant for all 10 repetitions. This stability analysis was only performed for the wbMSAA. For the spotlight approaches the algorithm was repeated 10 times, and the solution with the lowest cost function was chosen.

Cross validation: We used stratified k-fold cross validation as described in section 2.4. For cross validation, it is important that the test and training data sets are independent. For the seed region analysis features this is naturally the case, as the feature extraction was performed for each subject separately. However, in order to limit the computational complexity and to ensure correspondence of components across cross validation splits for ICA and MSAA, the decomposition was run on the entire dataset. Note that this did not lead to biased estimates of the classification performance, as no information about the class labels were used in the decomposition step.

3.2 Classification of task conditions

The aim of the task classification was twofold. Firstly, it was a proof of concept of our classification approach, using either temporal or spatial network features as input to the support vector classification. Secondly, we wanted to investigate if the information captured by the networks was sufficient to actually classify task conditions, and to see how the networks important for classification correspond to previous literature on ToM and empathy processing. The classification performances are listed in table 1, and networks are illustrated in figure 5.

First, we used the activation maps from the **SPM analysis** for classification**.** These activation maps yielded the highest task classification performance (mean accuracy of 83%), which was expected since they were informed about the onset and duration of the task conditions. The center coordinates, cluster size and z-score of the significant clusters can be found in supplementary table 3. This result was mainly used to validate that there was sufficient signal difference between the task conditions.

To investigate our hypothesis about the importance of both temporal and spatial network dynamics, we used six features from the **seed region analysis** as illustrated in figure 1. Firstly, we found that classification was not significant when using static measures such as the mean and variance, indicating that these simple measures do not capture enough signal difference between task blocks for classification in the considered sample. On the contrary, all *spatial networks features* (covariance and seed based analysis) resulted in significant classification with accuracies from 60-73%. As described in section 2.7, classification was performed for each of the 25 networks extracted in the seed based analysis. Table 1 and figure 5 include the networks that obtained the highest classification performance, however more networks yielded significant classification, which can be found in supplementary figure 1. For the empathy condition (Emp-Phy2), the seed of the network yielding the highest classification performance was located in the angular gyrus and the network further

included the inferior parietal lobule (IPL), precuneus, medial temporal gyrus and medial prefrontal cortex (mPFC). For the ToM classification, the seed was located in the dorsal anterior cingulate cortex (ACC), and the network included frontal lobe regions, caudate and the precuneus. Most of these regions were suggested to be involved in the ToM or empathy related processing in previous studies [Abu-Akel and Shamay-Tsoory, 2011; Fan et al., 2011]. In particular, the IPL, precuneus, middle temporal gyrus, mPFC and ACC are key regions of default mode network (DMN), which plays important role in social processing, such as understanding others' beliefs and feelings and self-referencing [Andrews-Hanna et al., 2010; Takeuchi et al., 2014].

Finally, we also tried to classify the conditions using the time series from the 25 seed regions. Here significant classification was only obtained when concatenating the time series from all components (mean accuracy 65%), and not when using the TS from each seed region separately.

For the **decomposition methods,** we used the time series from each component extracted using the three methods: ICA, wbMSAA and $sMSAA_{Lit}$. All decomposition time series yielded a high classification performance with accuracies ranging from 67-79%, which were significant after correcting for multiple comparisons. The reason for the high classification performance when using time series from decomposition methods compared to seed region analysis, might be that the decomposition methods extract components which maximally explain the data. They therefore captured networks (and corresponding time series) which were the most prominent in the data, whereas the seed region analysis relied on seed region points that were manually chosen based on previous literature, and thus were not specific for the given dataset. The corresponding spatial maps of the best components are shown in figure 5, and other significant networks can be found in supplementary figures 2-4. Generally, we found that the best networks across most methods included similar regions. For the **ToM-Phy1** classification (left column), the networks include inferior and medial frontal gyrus, temporoparietal junction (TPJ), posterior cingulate cortex (PCC), and postcentral gyrus activation, which all are known to be involved in ToM processing [Amodio and Frith, 2006; Ettinger et al., 2015; Frith and Frith, 2006; Pickup, 2006]. For the **Emp-Phy2** classification the networks included similar regions as for the ToM-Phy1 classification, but generally there was more activation in posterior parietal regions, such as precuneus and PCC.

To summarize, our findings show that both spatial networks and temporal dynamics capture important information, which enabled significant classification of the ongoing social cognition task. The networks which yielded the highest classification performance, generally included temporoparietal and prefrontal areas, which consistently have been considered core regions for ToM and empathy processing [Frith and Frith, 2006; Schurz et al., 2014].

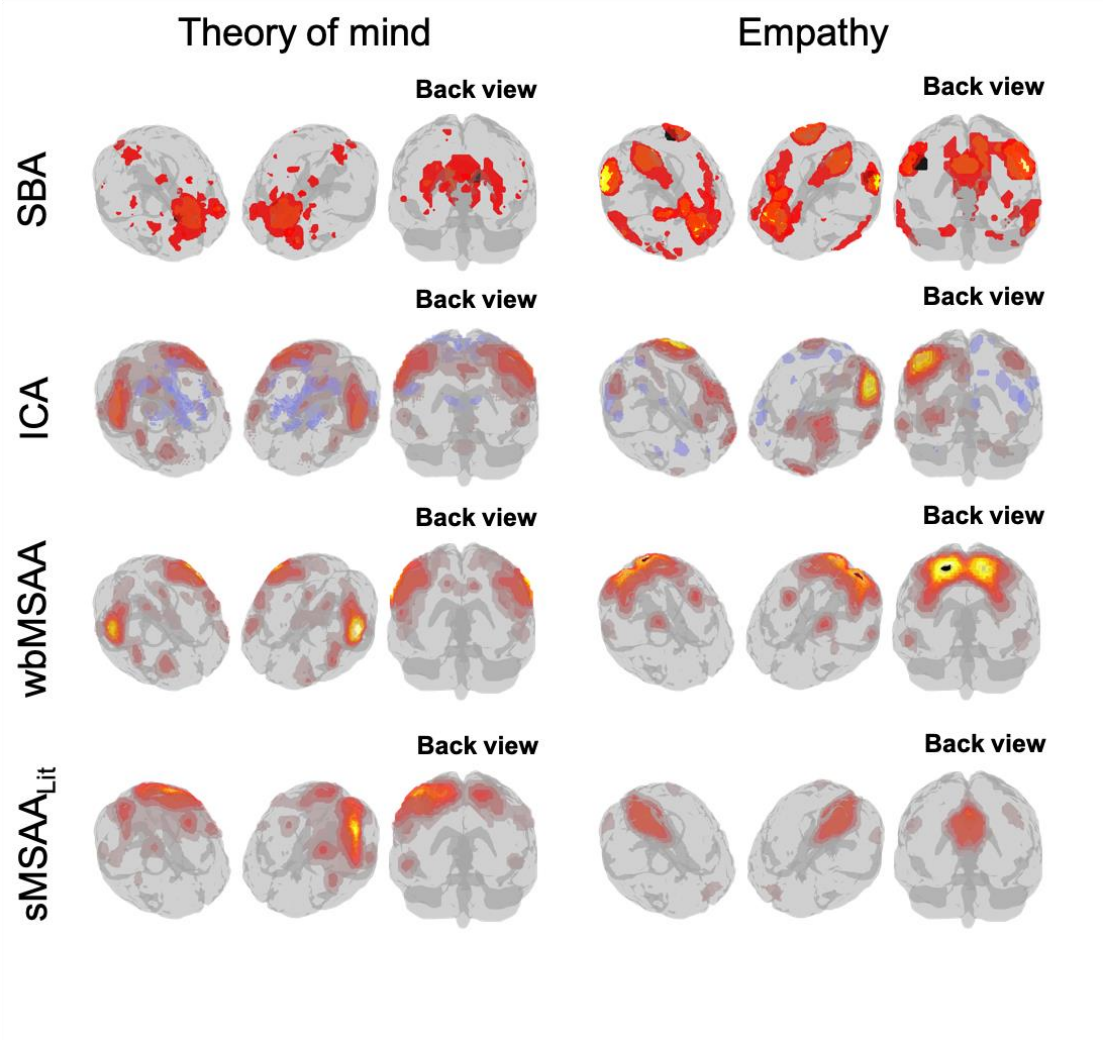


Figure 5: Mean spatial maps across subjects of the networks for ToM-Phy1 classification (left) and Emp-Phy2 classification (right), for SBA, ICA, wbMSAA, and sMSAA$_{lit}$, respectively. More significant networks can be found in Data S1–Figures S2–S4. For all four methods, the ToMPhy1 classifying networks have most activity in the temporoparietal regions, and prefrontal regions. For the Emp-Phy2, processing similar regions are included, but generally more activity is located in posterior parietal regions. For visualization, the SBA networks include the most significant 10% of the network correlations, each ICA map was z-scored and thresholded at $Z = 1$, and the MSAA networks include voxels with 10% or more fractional contribution.

3.3 Classification of social anhedonia

In this section, we show and discuss the results from the social anhedonia classification. The classification performances, measured by the MCC are listed in table 1 and figure 6 shows the spatial maps of the features obtaining the highest classification performance.

Whereas the activation maps from the SPM analysis resulted in the highest task classification performance of all methods, our results show that neither the raw maps, nor the seed based static measures (mean and variance) enabled significant classification of social anhedonia. In fact, for the **seed region analysis** features, only the covariance feature obtained significant classification with a MCC = 0.43 (p = 0.005). This indicates that simple network coupling between regions that are known to be involved in social cognition processing, seems to capture important information to differentiate the high and low social anhedonia group. Additional analysis of which part of the covariance were important for classification, revealed that the only feature surviving correction for multiple comparisons was the variance within the left TPJ. This region has been associated with social cognition and theory of mind in several previous studies [Bodnar et al., 2014; Dodell-Feder et al., 2014; Kronbichler et al., 2017] and was also a prominent region in the decomposition methods to presented below. For more details on this analysis wee the "interpretation of covariance features for HSA classification" section in the supplementary material. The second highest classification performance was obtained when using the time series from the inferior lateral prefrontal cortex seed (MCC = 0.35, $p_{un\text{-}corrected} = 0.007$), however, classification was not significant after multiple comparisons correction, which was necessary since classification was performed for each seed region separately.

On the contrary, several features from the **decomposition methods** yielded significant classification even after multiple comparisons correction. Here we used both the time series and spatial maps for each network as classification feature, and corrected for multiple comparisons using maximum permutation statistics across components. For each decomposition method, only one (or sometimes no) feature yielded significant classification.

The highest classification performance was obtained when using one time series from the wbMSAA approach (MCC = 0.56, p = 0.008). Very interestingly this was the same TS that also obtained the highest task classification performance for the ToM condition, highlighting the coupling between schizotypy and ToM processing [Bora and Pantelis, 2013; Pickup, 2006]. In future studies, such coupling between schizotypy and a relevant task (e.g. theory of mind), could be used to pre-select relevant network features instead of testing the classification for all features extracted by the MSAA.

Furthermore, the spatial map corresponding to this time series also obtained the highest classification performance of all wbMSAA spatial maps, which was borderline significant (MCC = 0.42, p = 0.09, $p_{un\text{-}corrected} = 0.006$). The seed of this network was in the TPJ, and the network further included inferior and medial PFC and insula. Furthermore, the ICA feature (time series) that resulted in the highest classification performance (MCC = 0.45, p = 0.07, $p_{un\text{-}corrected} = 0.005$), had a corresponding spatial map, that was nearly identical to the network from the wbMSAA analysis (see figure 6 and supplementary figure 5).

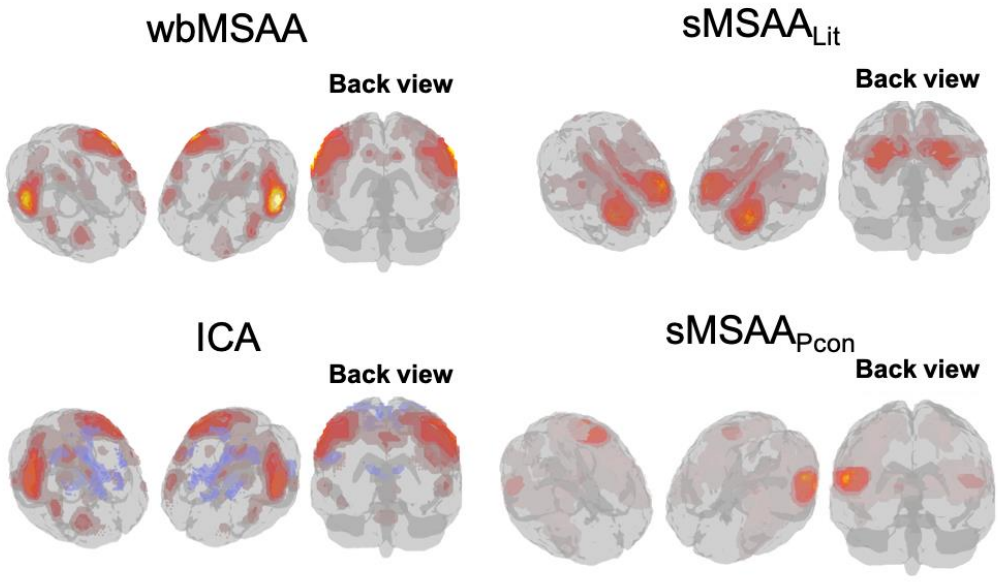


Figure 6: Mean spatial maps of the components yielding significant HSA classification performance. For all features, highest classification performance was obtained when using the times series (TS). This figure shows the corresponding spatial maps. The top row shows the two networks, corresponding to the TS features that obtained significant classification after multiple comparisons correction (wbMSAA and SMSAA$_{Lit}$). Visualization threshold was 10% fractional contribution. The networks in the bottom row are from the ICA and sMSAA$_{PCon}$ analysis, where the un-corrected p-value was below .05. For visualization, the ICA map was thresholded at a Z-score of 1 and the sMSAA$_{PCon}$ network include voxels with 5% fractional contribution.

The second decomposition feature that yielded significant classification, was the time series from the spotlight sMSAA$_{lit}$ approach (MCC = 0.49, p = 0.03). The spatial map corresponding to this TS had its seed region is in the dorsolateral PFC and furthermore the network included cingulate cortex and motor areas. For the sMSAA$_{lit}$ approach, we chose the seed regions which were known to be involved in ToM and empathy processing, since it is well established that social cognition is reduced in patients with schizophrenia [Bora et al., 2009; Brunet et al., 2000], and in subjects with schizotypy [Ettinger et al., 2015; Pickup, 2006].

As described in section 2.8, we also tested another spotlight approach where we used the peak coordinates from a pooled condition analysis (sMSAA$_{PCon}$), because this would be a more data driven way to choose center coordinates. Since the pooled condition analysis was specific for the given task, we hypothesized that the features extracted by this approach would result in a higher classification performance than for the sMSAA$_{Lit}$ approach. However, neither of the time series or spatial maps from the sMSAA$_{PCon}$ analysis resulted in significant classification after multiple comparisons correction. Only one component (time series) obtained a classification performance, with an un-corrected p-value < 0.05. Most activation in this network was in the TPJ and angular gyrus, but also included thalamus, insula and i/m FG.

To summarize, the components from the decomposition methods which obtained the highest classification performance generally included temporoparietal and prefrontal regions, as well as insula and cingulate cortex. These findings are in accordance with earlier studies which have reported lower white matter integrity in the fronto-

temporal tracts (measured by diffusion tensor imaging) in subjects with a high degree of schizotypy [Nelson et al., 2011], and both structural as well as functional studies have related changes in the PFC to schizotypy [Kühn et al., 2012; Raine et al., 1992]. Furthermore, earlier studies have shown a decrease in insula grey matter volume in UHR groups [Chan et al., 2011], and it has even been suggested that structural insular abnormalities might be related to the vulnerability for the development of later psychosis [Takahashi et al., 2009]. In future studies, it could thus be interesting to investigate if functional imaging could support the structural findings of Takahasi et al., and maybe enable identification of schizotypy in even earlier stages than what is possible with the structural changes. As for insula, grey matter volume reductions in thalamus have also been found in both schizophrenia [Ettinger et al., 2001] as well as in schizotypy [Kühn et al., 2012]. Furthermore, fMRI studies have shown correlation between reduced activation in thalamus and the degree of schizotypy [Aichert et al., 2012; Kumari et al., 2008], but it should be noted that the subjects performed another task in these studies.

In summary, the included areas of the two networks which are able to significantly classify HSA, have consistently been related to schizotypy and the schizophrenia development, which highlight the potential importance of these networks.

Finally, we want to comment on the use of **spatial and temporal network features** for the classification. Whereas many spatial network features resulted in significant classification of the task conditions, the time series generally resulted in a higher classification performance for the social anhedonia classification. This finding indicates that the temporal dynamics during the social cognition task captures important information to differentiate between high and low social anhedonia. In comparison, the connectivity measures used to extract spatial network features in this study are regarded static. In future studies, it would thus be interesting to look at dynamic functional connectivity, where the connectivity is estimated repeatedly for different windows of the time series, and thus also reflect the dynamic variations in the time series [Hutchison et al., 2013] [Damaraju et al., 2014; Nielsen et al., 2018].

3.4 Discussion of the MSAA method

This study is one of the first to use the MSAA method on neuroimaging data, and the first to implement the spotlight approach that further bridges aspects of data-driven decomposition methods and seed based analysis. We, therefore, highlight some of the important aspects of MSAA.

Interpretability: Due to the non-negativity and sum-to-one constraints, the spatial maps in MSAA have a clear interpretation, showing the fractional contributions of the components (archetypes) at each voxel. We used a threshold of 0.1 for visualization, meaning that for each voxel shown in a spatial map, this component had a relative contribution of at least 10 % to that given observation. A similar interpretation of the scale in ICA is not immediately possible without additional post processing, and furthermore as the ICA allows both positive and negative contributions, the components can include cancellation effects leading to less straightforward interpretation.

Noise modelling: the MSAA approach enables heteroscedastic noise modeling, i.e. the noise can be estimated for each subject and each voxel separately, instead of assuming it to be constant, which is done in previous decomposition methods such as ICA. Visual inspection of the spatial distribution of these noise levels (supplementary

figure 3) showed that most noise was present around the edges of the brain and close to known major blood vessels, which probably reflects residual movement effects and noise due to blood pulsation respectively. A more elaborate discussion of this noise modeling can be found in [Hinrich et al., 2016].

Spotlight: The spotlight restriction of MSAA showed to successfully enforce the algorithm to reveal functional networks, which otherwise were obscured by other salient signal features. This is somewhat similar to what was done by seed based analysis, but for the spotlight MSAA the optimal seed is determined by data driven optimization instead of manual assignment. Restriction of the seed regions can be especially valuable if a specific hypothesis needs to be tested, e.g. how the connectivity between the whole brain and a particular region changes in relation to disease progression. However, compared to the wbMSAA approach, it requires the user to choose a number of seed regions, which can be difficult to choose. In this study, we have chosen center coordinates based on the social cognition task, either based on previous literature or from a pooled condition analysis. Another approach could have been to choose seeds which have been related to social anhedonia and/or schizotypy progression.

Non-convex optimization and number of components: as for ICA, the MSAA algorithm is a non-convex optimization problem, which means that the optimization might get struck in a local and not global minimum. In practice, this means that repeated runs can result in somewhat different networks. How severe this problem is, depends on the stability of the given dataset (signal to noise ratio, inter-subject differences etc.) as well as on the number of components chosen. In this study, we used 25 components as this was found to be the optimal number using the minimum description length criteria which is the default implemented in GIFT toolbox [Li et al., 2007]. Using 25 components, resulted in relatively stable networks, with a mean spatial correlation of 86% for the wbMSAA when choosing the best (lowest cost) solution between 10 runs as described in section 2.7. Visual inspection of the networks showed that the overall network structures between runs were very stable, and the non-perfect machining resulted in small network differences between runs (networks illustrated in supplementary section “stability of wbMSAA”). Furthermore, we noted that the number of components within one run seemed reasonable, such that known networks were captured by separate spatial maps (mixing of e.g. task and visual processing networks would indicate that the number of components was too low), and did not split networks up into separate components (this would indicate that the number of components was too high). All in all, this indicates that the number of runs and components were appropriate for the given study. However, we want to emphasize the importance of investigating the stability in future studies applying MSAA.

Toolbox: we have implemented the MSAA (both whole brain and spotlight) code into a SPM plugin (compatible with SPM 12), which interested users can download here: http://www.brain-fmri.com/MSAA/. The plugin enables the user to apply the MSAA algorithm on fMRI data, by simply loading the pre-processed images and choose the optimization parameters specified in the toolbox.

		Description of classification feature	Task classification Accuracy (%) (p value)		HSA[1] vs. LSA[2] classification MCC[3] (p-value)	
			ToM[4] – Phy1	Emp[5]- Phy2		
Seed region analysis features						
	(a) Task activation maps	Task specific activity maps determined using mass univariate analysis	84 % p = 0.001 1×V	81% p = 0.001 1×V	0.13 p = 0.199 1×V	
Static measures	(b) Mean activity	Average activity of each ROI	41% p = 0.801 1×25	56 % p = 0.115 1×25	-- (†)	
Static measures	(c) Variance	Variance within each ROI	58 % p = 0.070 1×25	58 % p = 0.091 1×25	-0.02 p = 0.569 1×25	
Network coupling	(d) Covariance (network coupling)	Covariance of the time series of ROIs	60 % p = 0.039 1×325	60 % p = 0.037 1×325	0.43 p = 0.005 1×325	
Network coupling	(e) Seed based network	Correlation between time series of a ROI and all voxels in the brain	73 % p = 0.001 25×V	73 % p = 0.001 25×V	0.19 p = 0.897 p_{UC} = 0.125 25×V	
Time series	(f) Time series (ROI specific)	Time series of each ROI separately	59 % p = 0.666 25×T_1	61 % p = 0.393 25×T_1	0.35 p = 0.189 p_{UC} = 0.007 25×T_2	
Time series	(g) Time series (concatenated)	Time series of each ROI, concatenated	63 % p = 0.010 1×25T_1	68 % p = 0.001 1×25T_1	-0.15 p = 0.937 1×25T_2	
Decomposition features						
					Feature type	
			TS		**TS**	**SM**
	(h) ICA	Time series and spatial maps from ICA	73 % p = 0.001 25×T_1	79 % p = 0.001 25×T_1	0.45 p = 0.072 p_{UC} = 0.005 25×T_2	0.24 p = 0.912 p_{UC} = 0.093 25×V
	(i) wbMSAA	Time series and spatial maps from wbMSAA	74 % p = 0.001 25×T_1	69 % p = 0.002 25×T_1	0.56 p = 0.008 p_{UC} = 0.002 25×T_2	0.42 p = 0.097 p_{UC} = 0.006 25×V
	(j) $sMSAA_{Lit}$	Time series and spatial maps from spotlight MSAA (using literature coordinates)	67 % p = 0.020 25×T_1	73 % p = 0.001 25×T_1	0.49 p = 0.032 p_{UC} = 0.003 25×T_2	0.25 p = 0.744 p_{UC} = 0.059 25×V
	(k) $sMSAA_{PCon}$	Time series and spatial maps from spotlight MSAA (using PCon coordinates)	-- (*)	--(*)	0.31 p = 0.463 p_{UC} = 0.030 25×T_2	0.25 p = 0.732 p_{UC} = 0.066 25×V

Table 1 Classification performance of both task and HSA. For each performed analysis, this table yields a short explanation of the input feature and classification performance measured in accuracy (task classification) or Mathews correlation coefficient, MCC (HSA classification). For the HSA classification both time series (TS) and spatial maps (SM) were used as features for the decomposition methods. For seed region analysis features e-f and decomposition methods (h-j) the table lists the classification performance of the component yielding the highest classification performance. The p-value was non-parametrically estimated with random permutation testing and maximum permutation statistics was used to correct for multiple comparisons when necessary. The number of comparisons × feature dimensionality are stated for each of the classification models, where the size of the voxel dimension is V=60704 , T_1=60 (time points for each condition) and T_2=264 (total number of time points). The uncorrected p-value (p_{UC}) was also based on random permutation and is stated for some HSA classifications. (†) HSA was not classified as the overall mean per subject was subtracted during preprocessing. (*) task classification was not calculated for the $sMSAA_{PCon}$ analysis, since this result would be biased.

[1] HSA: Moderately high social anhedonia
[2] LSA: Low social anhedonia
[3] MCC: Matthews correlation coefficient
[4] ToM: Theory of Mind
[5] Emp: Empathy

3.5 Limitations and future perspectives

As discussed in the previous section there are some challenges for decomposition methods, such as non-convexity and choosing an appropriate number of components. Another large challenge of this study was the relatively small difference between subjects of the high and low social anhedonia respectively. Firstly, classification was challenged by the low separation boundary which was used (mean plus one standard deviation). Though similar boundaries have been used in previous group comparison studies of schizotypy [Wang et al., 2015a], it was challenging for the support vector machine to learn from the data of two relatively similar classes. Secondly, even with this low separation threshold, we had an unbalanced dataset, with 56 (LSA) and 14 (HSA) subjects in each group. This further challenged the supervised classification procedure, and made the classification performance sensitive the classification of few subjects. We tried to mitigate this problem by (i) using weights in the support vector machine to counteract the imbalance and (ii) used the MCC measure to access classification performance. Additionally, it is important to note that while full correction of multiple comparisons was considered within each feature extraction method, this was not done across these different methods. This was motivated by the main aim of comparing a set of, in many aspects, very similar feature extraction methods. With these limitations in mind we consider the present study an explorative investigation of features for classification of social anhedonia rather than a study of the neural correlates of social anhedonia itself. Still, we strongly expect that a larger group, particularly with more subjects with high social anhedonia, would make classification easier and more stable. Furthermore, including subjects with more pronounced social anhedonia, or subjects belonging to other risk groups, would also be very interesting from a clinical perspective.

However, even with these challenges, the whole brain and spotlight MSAA algorithms extracted features that yielded significant classification. Using the same methods on ultra-high-risk groups or patients with schizophrenia would thus be very interesting to investigate how networks alterations are related to the development of schizophrenia. Optimally, this could be investigated through a longitudinal study starting with a large group of subjects with a continuous range of schizotypy and a specific and well-designed experimental set-up.

Conclusion

Using a variety of different feature extraction methods, we found significant classification of social anhedonia for two features, both consisting of times series extracted by the MSAA decomposition methods. The highest classification performance was achieved using the whole brain MSAA. Importantly, the same time series also obtained the highest task classification performance, making a strong coupling between the processing of the theory of mind task and the degree of social anhedonia. This indicates that future studies could focus on components representing task-relevant networks for classification of schizotypy, thereby circumventing the need for correction for multiple comparisons across components. The spatial map corresponding to the time series yielding highest classification performance, included the TPJ, prefrontal cortex, angular gyrus and insula, which all have been consistently related to schizotypy as well as to the development of schizophrenia in earlier studies.

Finally, a nearly identical feature was also identified as the best performing when using features extracted by ICA. The repeated occurrence of the same feature highlights the potential importance of this network for early identification of schizotypy. Thus, in future studies, it would be very interesting to investigate if the same network would also be important for subjects with more pronounced schizotypy and other high-risk groups through the spectrum of schizophrenia development.

Acknowledgements

Raymond Chan was supported by the National Basic Research Programme of China (Precision Psychiatry Programme) (2016YFC0906402), the Beijing Municipal Science & Technology Commission Grant (Z161100000216138), and the Beijing Training Project for the Leading Talents in S & T (Z151100000315020).

Morten Mørup was supported by the Lundbeckfonden (fellowship grant R105-9813).

We gratefully acknowledge the support of NVIDIA Corporation who donated the Titan Xp, which was used while testing GPU support in the MSAA toolbox.

The authors have no professional or financial interests that could be perceived as having biased the presentation.

Data availability statement

The data that support the findings of this study are available on request from the corresponding authors. The data are not publicly available due to privacy or ethical restrictions.

References

Abu-Akel A, Shamay-Tsoory S (2011): Neuroanatomical and neurochemical bases of theory of mind. Neuropsychologia 49:2971–2984.

Aichert DS, Williams SCR, Möller HJ, Kumari V, Ettinger U (2012): Functional neural correlates of psychometric schizotypy: An fMRI study of antisaccades. Psychophysiology 49:345–356.

Amodio DM, Frith CD (2006): Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci 7:268–277.

Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL (2010): Functional-anatomic fractionation of the brain's default network. Neuron 65:550–62.

Baldi P, Brunak S, Chauvin Y, Andersen C a, Nielsen H (2000): Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16:412–424.

Beckmann CF, Smith SM (2004): Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging 23:137–152.

Bedwell JS, Compton MT, Jentsch FG, Deptula AE, Goulding SM, Tone EB (2014): Latent Factor Modeling of Four Schizotypy Dimensions with Theory of Mind and Empathy. Ed. Yinglin Xia. PLoS One 9:e113853.

Benedetti F, Bernasconi A, Bosia M, Smeraldi E (2009): Functional and structural brain correlates of theory of mind and empathy deficits in schizophrenia. Schizophr Res 114:154–160.

Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995): Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med Off J Soc Magn Reson Med 34:537–41.

Blanchard JJ, Collins LM, Aghevli M, Leung WW, Cohen AS (2011): Social anhedonia and schizotypy in a community sample: The Maryland longitudinal study of schizotypy. Schizophr Bull 37:587–602.

Bodnar M, Hovington CL, Buchy L, Malla AK, Joober R, Lepage M (2014): Cortical Thinning in Temporo-Parietal Junction (TPJ) in Non-Affective First-Episode of Psychosis Patients with Persistent Negative Symptoms. Ed. Tianzi Jiang. PLoS One 9:e101372.

Bora E, Pantelis C (2013): Theory of mind impairments in first-episode psychosis, individuals at ultra-high risk for psychosis and in first-degree relatives of schizophrenia: Systematic review and meta-analysis. Schizophr Res 144:31–36.

Bora E, Yucel M, Pantelis C (2009): Theory of mind impairment in schizophrenia: Meta-analysis. Schizophr Res 109:1–9.

Brunet E, Sarfati Y, Hardy-Baylé MC, Decety J (2000): A PET investigation of the attribution of intentions with a nonverbal task. Neuroimage 11:157–166.

Calhoun VD, Adali T, Pearlson GD, Pekar JJ (2001): A Method for Making Group Inferences from Functional MRI Data Using Independent Component Analysis. Hum Brain Mapp 14:140–151.

Calhoun V, Adali T, Hansen L (2003): ICA of functional MRI data: an overview.

Chan RCK, Di X, McAlonan GM, Gong QY (2011): Brain anatomical abnormalities in high-risk individuals, first-episode, and chronic schizophrenia: An activation likelihood estimation meta-analysis of illness progression. Schizophr Bull 37:177–188.

Chan RCK, song Shi H, lei Geng F, hua Liu W, Yan C, Wang Y, Gooding DC (2015): The Chapman psychosis-proneness scales: Consistency across culture and time. Psychiatry Res 228:143–149.

Chan RCK, Wang Y, Yan C, Zhao Q, McGrath J, Hsi X, Stone WS (2012): A study of trait anhedonia in non-clinical chinese samples: Evidence from the chapman scales for physical and social anhedonia. PLoS One 7:3–8.

Chang C-C, Lin C-J (2011): Libsvm A library for support vector machines. ACM Trans Intell Syst Technol 2:1–27.

Chapman L, Chapman J, Kwapil T, Eckblad M, Zinser M (1994): Putatively psychosis-prone subjects 10 years later. J Abnorm Psychol:1689-171–183.
Cole DM, Smith SM, Beckmann CF (2010): Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. Front Syst Neurosci 4:8.
Cortes C, Vapnik V (1995): Support-Vector Networks. Kluwe Acad Publ 297:273–297.
Cutler A, Breiman L (1994): Archetypal Analysis. Technometrics 36:338–347.
Damaraju E, Allen EA, Belger A, Ford JM, McEwen S, Mathalon DH, Mueller BA, Pearlson GD, Potkin SG, Preda A, Turner JA, Vaidya JG, Van Erp TG, Calhoun VD (2014): Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. NeuroImage Clin 5:298–308.
Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006): An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968–980.
Dodell-Feder D, Tully LM, Lincoln SH, Hooker CI (2014): The neural basis of theory of mind and its relationship to social functioning and social anhedonia in individuals with schizophrenia. NeuroImage Clin 4:154–163.
Erhardt EB, Rachakonda S, Bedrick E, Allen E, Adali T, Calhoun VD (2011): Comparison of multi-subject ICA methods for analysis of fMRI data. Brain 32:2075–2095.
Ettinger U, Chitnis XA, Kumari V, Fannon DG, Sumich AL, O'Ceallaigh S, Doku VC, Sharma T (2001): Magnetic resonance imaging of the thalamus in first-episode psychosis. Am J Psychiatry Vol.158:116–118.
Ettinger U, Mohr C, Gooding DC, Cohen AS, Rapp A, Haenschel C, Park S (2015): Cognition and brain function in schizotypy: A selective review. In: . Schizophrenia Bulletin Vol. 41, pp S417--S426.
Fan Y, Duncan NW, de Greck M, Northoff G (2011): Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. Neurosci Biobehav Rev 35:903–911.
Fett AKJ, Viechtbauer W, Dominguez M de G, Krabbendam L (2011): The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: A meta-analysis. Neurosci Biobehav Rev 35:573–588.
Fisher R a., Fisher R a. (1915): Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika.
Friston KJ, Williams S, Howard R, Frackowiak RSJ (1996): Movement Related Effects In FMRI. Magn Reson Med 3:346–355.
Frith CD, Frith U (2006): The Neural Basis of Mentalizing. Neuron 50:531–534.
Gooding DC, Tallent KA, Matts CW (2005): Clinical Status of At-Risk Individuals 5 Years Later: Further Validation of the Psychometric High-Risk Strategy. J Abnorm Psychol 114:170–175.
Green MF, Horan WP, Lee J (2015): Social cognition in schizophrenia. Nat Rev Neurosci 16:620–631.
Henry JD, Bailey PE, Rendell PG (2008): Empathy, social functioning and schizotypy. Psychiatry Res 160:15–22.
Hinrich JL, Bardenfleth S, Roge R, Churchill N, Madsen KH, Morup M (2016): Archetypal Analysis for Modeling Multi-Subject fMRI Data. IEEE J Sel Top Signal Process:1–1.
Hutchison RM, Womelsdorf T, Allen E a., Bandettini P a., Calhoun VD, Corbetta M, Della Penna S, Duyn JH, Glover GH, Gonzalez-Castillo J, Handwerker D a., Keilholz S, Kiviniemi V, Leopold D a., de Pasquale F, Sporns O, Walter M, Chang C (2013): Dynamic functional connectivity: Promise, issues, and interpretations. Neuroimage 80:360–378.
Insel TR (2010): Rethinking schizophrenia. Nature 468:187–193.
Kronbichler L, Tschernegg M, Martin AI, Schurz M, Kronbichler M (2017): Abnormal Brain Activation During Theory of Mind Tasks in Schizophrenia: A Meta-Analysis. Schizophr Bull 43:1240–1250.

Kühn S, Schubert F, Gallinat J (2012): Higher prefrontal cortical thickness in high schizotypal personality trait. J Psychiatr Res 46:960–965.
Kumari V, Antonova E, Geyer MA (2008): Prepulse inhibition and "psychosis-proneness" in healthy individuals: An fMRI study. Eur Psychiatry 23:274–280.
Kwapil TR (1998): Social anhedonia as a predictor of the development of schizophrenia-spectrum disorders. J Abnorm Psychol 107:558–565.
Lagioia A, Van De Ville D, Debbané M, Lazeyras F, Eliez S (2010): Adolescent resting state networks and their associations with schizotypal trait expression. Front Syst Neurosci 4:1–12.
Lewis DA, Levitt P (2002): Schizophrenia as a disorder of neurodevelopment. Annu Rev Neurosci 25:409–432.
Li Y-O, Adali T, Calhoun VD (2007): Estimating the number of independent components for functional magnetic resonance imaging data. Hum Brain Mapp 28:1251–1266.
Madsen KH, Krohne LG, Cai X, Wang Y, Chan RCK (2018): Perspectives on Machine Learning for Classification of Schizotypy Using fMRI Data. Schizophr Bull:1–11.
Mason OJ (2015): The assessment of schizotypy and its clinical relevance. Schizophr Bull 41:S374–S385.
McKeown MJ, Makeig S, Brown GG, Jung T-P, Kindermann SS, Bell AJ, Sejnowski TJ (1998): Analysis of fMRI data by blind separation into independent components. Hum Brain Map 6:160–188.
Modinos G, Pettersson-Yeo W, Allen P, McGuire PK, Aleman A, Mechelli A (2012): Multivariate pattern classification reveals differential brain activation during emotional processing in individuals with psychosis proneness. Neuroimage 59:3033–3041.
Morrison SC, Brown LA, Cohen AS (2013): A multidimensional assessment of social cognition in psychometrically defined schizotypy. Psychiatry Res 210:1014–1019.
Mørup M, Hansen LK (2012): Archetypal analysis for machine learning and data mining. Neurocomputing 80:54–63.
Nelson MT, Seal ML, Pantelis C, Phillips LJ (2013): Evidence of a dimensional relationship between schizotypy and schizophrenia: A systematic review. Neurosci Biobehav Rev 37:317–327.
Nelson MT, Seal ML, Phillips LJ, Merritt AH, Wilson R, Pantelis C (2011): An investigation of the relationship between cortical connectivity and schizotypy in the general population. J Nerv Ment Dis 199:348–53.
Nichols T, Holmes A (2003): Nonparametric Permutation Tests for Functional Neuroimaging. Hum Brain Funct Second Ed 15:887–910.
Nielsen SFV, Levin-Schwartz Y, Vidaurre D, Adali T, Calhoun VD, Madsen KH, Hansen LK, Morup M (2018): Evaluating Models of Dynamic Functional Connectivity Using Predictive Classification Accuracy. ICASSP, IEEE Int Conf Acoust Speech Signal Process - Proc 2018-April:2566–2570.
Patel AX, Kundu P, Rubinov M, Jones PS, Vértes PE, Ersche KD, Suckling J, Bullmore ET (2014): A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. Neuroimage 95:287–304.
Penn DL, Sanna LJ, Roberts DL (2007): Social Cognition in Schizophrenia: An Overview. Schizophr Bull 34:408–411.
Pflum MJ, Gooding DC (2018): Context matters: Social cognition task performance in psychometric schizotypes. Psychiatry Res 264:398–403.
Pickup GJ (2006): Theory of mind and its relation to schizotypy. Cogn Neuropsychiatry 11:177–192.
Poldrack RA, Gorgolewski KJ (2014): Making big data open: data sharing in neuroimaging. Nat Neurosci 17:1510–1517.
Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE (2012): Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59:2142–54.

Powers D (2011): Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. J Mach Learn Technol 2:37–63.
Rachakonda S, Egolf E, Correa N, Calhoun V, Neuropsychiatry O (2015): Group ICA/IVA of fMRI Toolbox (GIFT) Manual.
Raine A, Sheard C, Reynolds GP, Lencz T (1992): Pre-frontal structural and functional deficits associated with individual differences in schizotypal personality. Schizophr Res 7:237–247.
Schurz M, Radua J, Aichhorn M, Richlan F, Perner J (2014): Fractionating theory of mind: A meta-analysis of functional brain imaging studies. Neurosci Biobehav Rev 42:9–34.
Sebastian CL, Fontaine NMG, Bird G, Blakemore S-J, De Brito SA, McCrory EJP, Viding E (2012): Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. Soc Cogn Affect Neurosci 7:53–63.
Shamay-Tsoory SG, Harari H, Aharon-Peretz J, Levkovitz Y (2010): The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. Cortex 46:668–677.
Shinkareva S V, Ombao HC, Sutton BP, Mohanty A, Miller GA (2006): Classification of functional brain images with a spatio-temporal dissimilarity map. Neuroimage 33:63–71.
Takahashi T, Wood SJ, Yung AR, Velakoulis D, Pantelis C (2009): Insular cortex gray matter changes in individuals at ultra-high-risk of developing psychosis. Schizophr Res 111:94–102.
Takeuchi H, Taki Y, Nouchi R, Sekiguchi A, Hashizume H, Sassa Y, Kotozaki Y, Miyauchi CM, Yokoyama R, Iizuka K, Nakagawa S, Nagase T, Kunitoki K, Kawashima R (2014): Association between resting-state functional connectivity and empathizing/systemizing. Neuroimage 99:312–322.
Thakkar KN, Park S (2010): Empathy, schizotypy, and visuospatial transformations. Cogn Neuropsychiatry 15:477–500.
Völlm B a., Taylor ANW, Richardson P, Corcoran R, Stirling J, McKie S, Deakin JFW, Elliott R (2006): Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. Neuroimage 29:90–98.
Wang Y, Li Z, Liu W, Wei X, Jiang X, Lui SSY, Ho-wai So S, Cheung EFC, Debbane M, Chan RCK, Ho-wai So S, Li Z, Jiang X, Wang Y, Chan RCK, Lui SSY, Liu W, Wei X (2018): Negative Schizotypy and Altered Functional Connectivity During Facial Emotion Processing. Schizophr Bull 44:S491–S500.
Wang Y, Liu W-HH, Li Z, Wei X-HH, Jiang X-QQ, Geng F-LL, Zou L-QQ, Lui SSY, Cheung EFC, Pantelis C, Chan RCK (2016): Altered corticostriatal functional connectivity in individuals with high social anhedonia. Psychol Med 46:1–11.
Wang Y, Liu WH, Li Z, Wei XH, Jiang XQ, Geng FL, Zou LQ, Lui SS, Cheung EF, Pantelis C, Chan RC (2015a): Altered corticostriatal functional connectivity in individuals with high social anhedonia. Psychol Med:1–11.
Wang Y, Liu W-H, Li Z, Wei X-H, Jiang X, Neumann DL, Shum DHK, Cheung EFC, Chan RCK (2015b): Dimensional schizotypy and social cognition: an fMRI imaging study. Front Behav Neurosci 9:133.
Wang Y, Lui SSY, Zou L quan, Zhang Q, Zhao Q, Yan C, Hong X hong, Chan RCK (2014): Individuals with psychometric schizotypy show similar social but not physical anhedonia to patients with schizophrenia. Psychiatry Res 216:161–167.
Wang Y, Neumann DL, Shum DHK, Liu W, Shi H, Yan C, Lui SSY, Zhang Q, Li Z, Cheung EFC, Chan RCK (2013): Cognitive empathy partially mediates the association between negative schizotypy traits and social functioning. Psychiatry Res 210:62–68.
Weinberger D (1987): Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry 44:660–669.

Supplementary Material

This supplementary material includes additional figures and tables which are referred to in the main paper, and that are described by their corresponding captions. In short, **supplementary table** 1 lists the literature center coordinates that are used for the seed region analysis and spotlight multi-subject archetypal analysis (MSAA).

Supplementary table 2 gives additional information on the Chapman and the Beck depression inventory scales.

Supplementary figure 1-4 shows all networks that obtained significant task classification for either the "Theory of Mind" or "Empathy" condition, furthermore the figure lists the accuracies obtained by each network.

Supplementary figure 5 shows axial slices of the networks that obtained significant classification for the social anhedonia classification.

Supplementary figure 6 illustrates the average (over subjects) noise map for the whole brain MSAA.

The section **stability of wbMSAA** describes the consistency of the wbMSAA analysis across multiple runs of the algorithm.

The section **Interpretation of covariance features for HSA classification** investigate which ROIs were responsible for the significant classification of HSA based on covariance features.

Finally, **supplementary table 3** includes the center coordinates from the pooled condition analysis that were used for the spotlight MSAA classification.

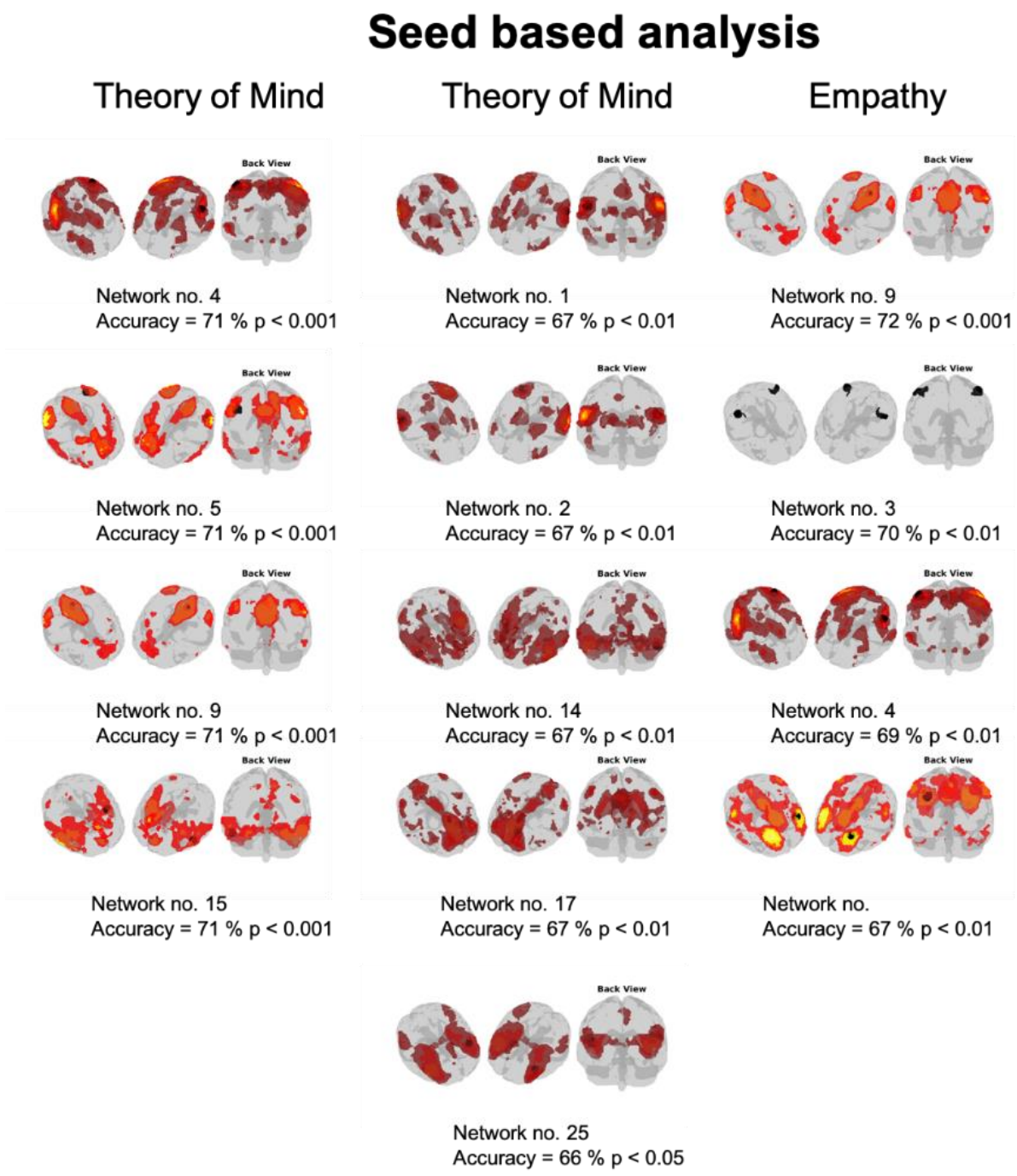
Supplementary table 1: MNI coordinates of the 25 center coordinates used for the seed region analysis and spotlight $sMSAA_{Lit}$. The number in the brackets indicate the network number (right before left), e.g. TPJ (1-2) indicates that rTPJ was seed number 1.

Literature coordinates	Left hemisphere MNI coordinate			Right hemisphere MNI coordinates		
	x	y	z	X	y	Z
Temporoparietal junction (TPJ) (1-2)	-53	-59	20	53	-59	20
Inferior parietal Lobe (IPL) (3-4)	-45	-43	56	45	-43	56
Angular gyrus (AngG) (5-6)	-45	-60	35	45	-60	35
Superior Temporal Sulcus (STS) (7-8)	-54	-12	0	54	-12	0
Precuneus (Pcun) (9)	-2	-52	39			
Amygdala (Amyg) (10-11)	-20	-3	-18	20	-3	-18
Ventral Striatum (12-13)	-10	15	9	10	8	-8
Dorsal Temporal Pole (dTP) (14-15)	-54	-9	-21	54	-9	-21
Dorsal Anterior Cingulate Cortex (dACC) (16-17)	-10	32	24	10	32	24
Ventral Anterior Cingulate Cortex (vACC) (18-19)	-12	28	-10	12	28	-10
Ventral medial Prefrontal Cortex (vmPFC) (20)				3	51	-15
Dorsal medial Prefrontal Cortex (dmPFC) (21)				6	26	55
Dorsolateral Prefrontal Cortex (dLPFC) (22-23)	-33	38	37	33	38	37
Inferiorlateral Prefrontal Cortex (dLPFC) (24-25)	-46	22	8	46	22	8

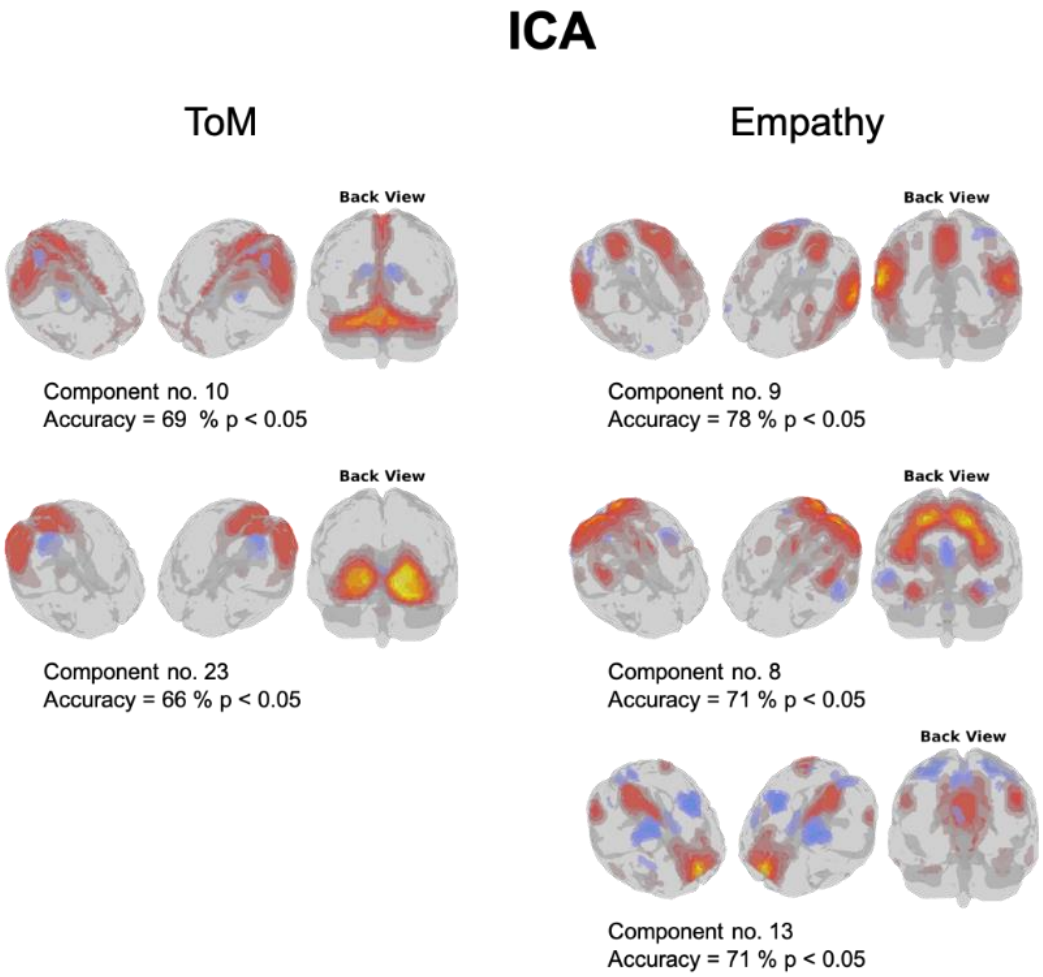
Supplementary table 2: Chapman scale scores and Beck Depression inventory for all 70 included subjects.

	Chapman scale scores				Beck Depression Inventory (BDI)
	Chapman Social Anhedonia (CSAS)	Chapman Physical Anhedonia (CPAS)	Magical ideation (MIS)	Perceptual aberration (PAS)	
Mean	7.97	12.59	6.59	10.44	4,06
Standard deviation	5.65	9.87	7.06	5.64	4.51

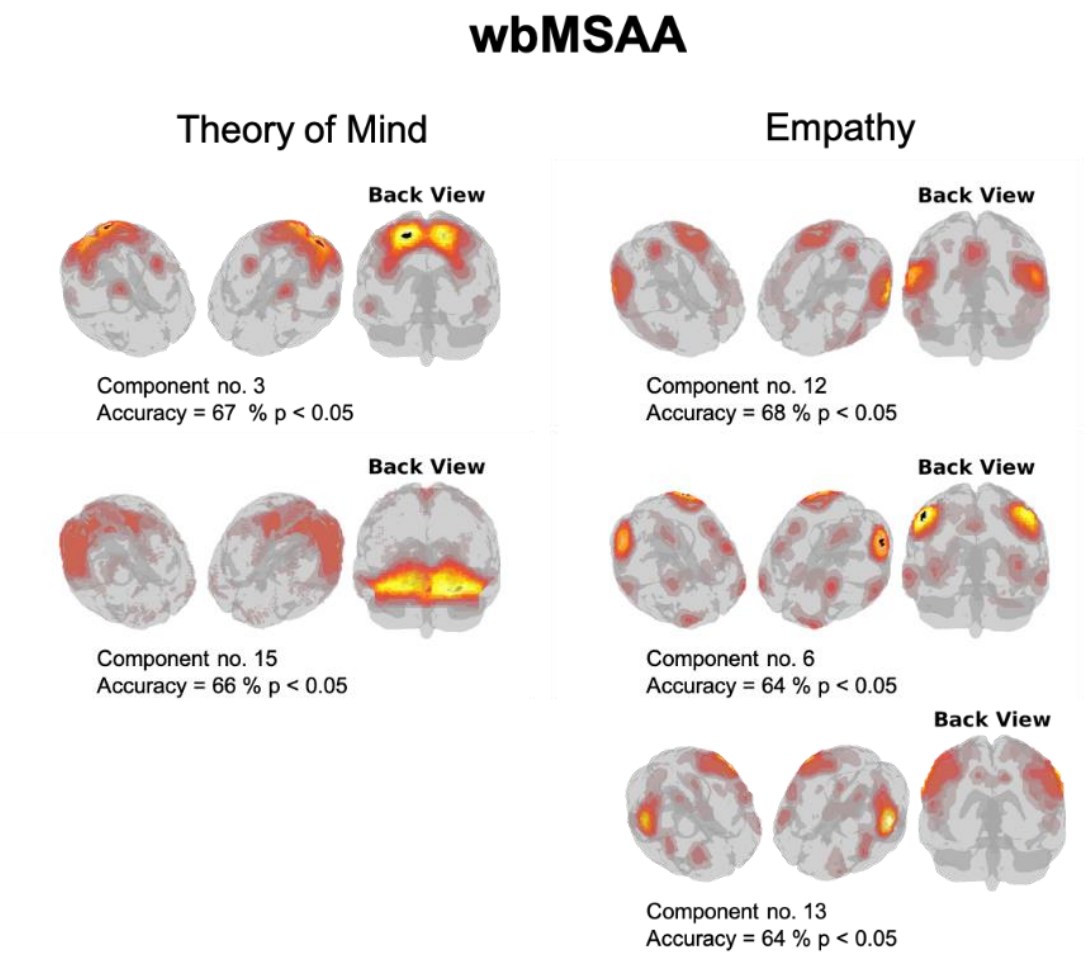
Supplementary figure 1: Significant task classification networks from SBA analysis. 3D visualization of all networks that obtained significant task classification for either the theory of mind (left and middle column) and empathy (right column) classification using the 25 networks coming from seed based analysis (SBA). The network number (no.) corresponds to the seed region number listed in supplementary table 1.



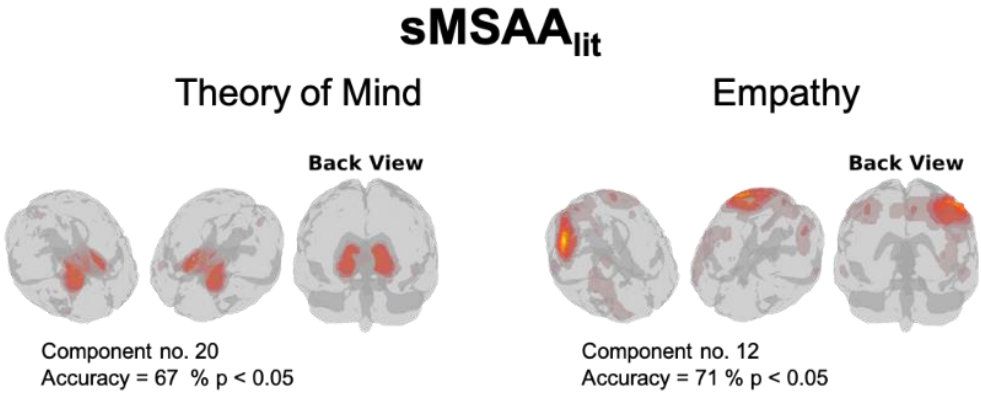
Supplementary figure 2: Significant task classification networks from ICA. 3D visualization of all networks that obtained significant task classification for either the theory of mind (left column) and empathy (right column) classification using the times series (TS) from the 25 components coming from the independent component analysis (ICA). The component number (no.) corresponds to the order of the networks when returned from the decomposition method, and corresponds to the order of the .nii files available at http://www.brain-fmri.com/MSAA/supplement/.



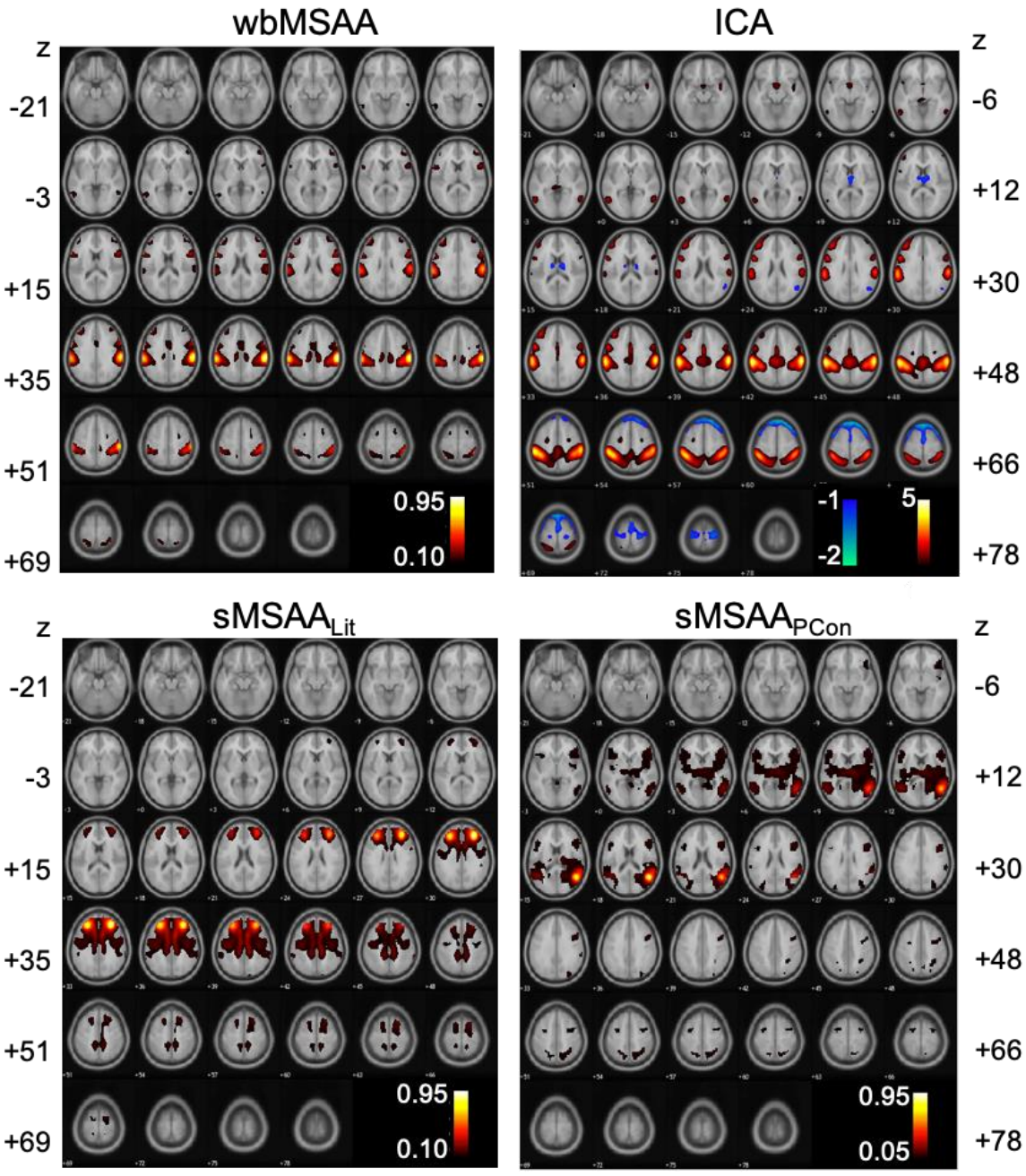
Supplementary figure 3: Significant task classification networks from wbMSAA. 3D visualization of all networks that obtained significant task classification for either the theory of mind (left column) and empathy (right column) classification using the times series (TS) from the 25 components coming whole brain multi subject archetypal analysis (wbMSAA). The component number (no.) corresponds to the order of the networks when returned from the decomposition method, and corresponds to the order of the .nii files available at http://www.brain-fmri.com/MSAA/supplement/.

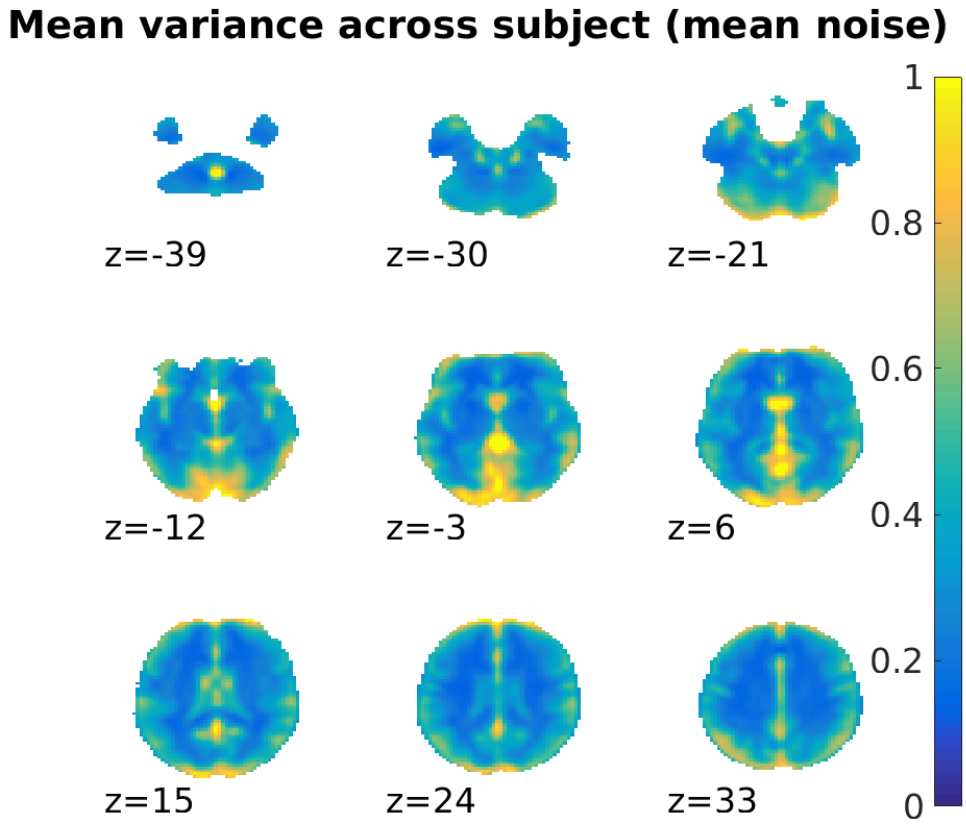


Supplementary figure 4: Significant task classification networks from $sMSAA_{Lit}$. 3D visualization of all networks that obtained significant task classification for either the theory of mind (left column) and empathy (right column) classification using the times series (TS) from the 25 components coming spotlight multi subject archetypal analysis ($sMSAA_{Lit}$). The component number (no.) corresponds to the order of the networks when returned from the decomposition method, and corresponds to the order of the .nii files available at http://www.brain-fmri.com/MSAA/supplement/.



Supplementary figure 5: Axial slices of the best HSA classifying networks determined by the decomposition methods; whole brain MSAA (top left), and spotlight MSAA (bottom) with center coordinates from the literature ($sMSAA_{lit}$) (left), and pooled condition analysis ($sMSAA_{Pcon}$) (right). Finally, axial slices from ICA (top right). For MSAA the visualization threshold was 10% fractional contribution for wbMSAA and $sMSAA_{Lit}$ and 5% for $sMSAA_{PCon}$. For visualization, the ICA map was thresholded at a z-score of 1.



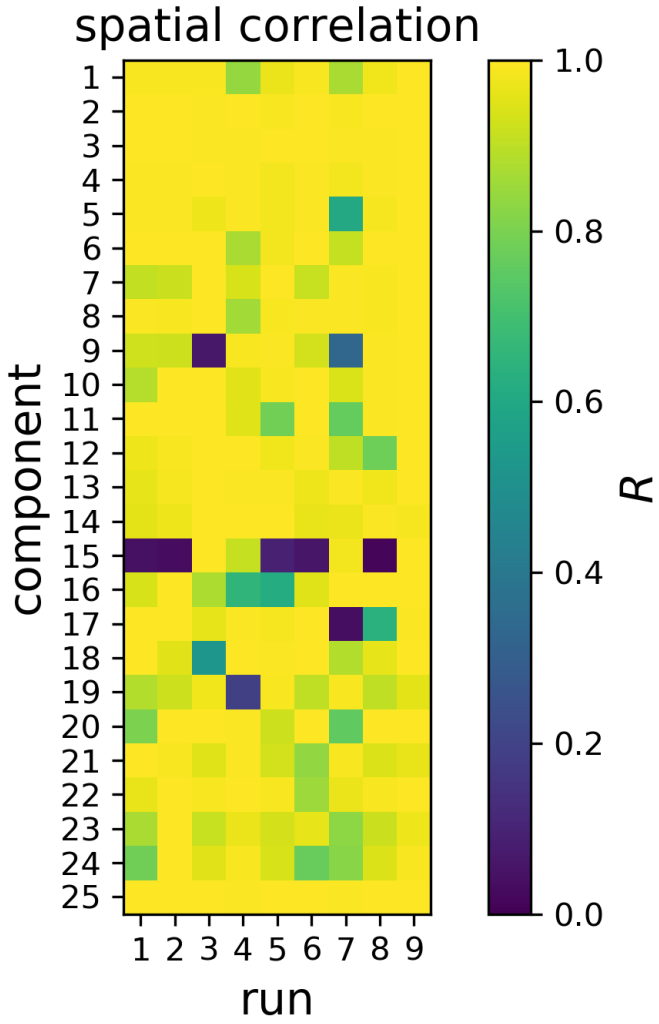


Supplementary figure 6: Mean variance (estimate of noise) between subjects using the wbMSAA algorithm. It is clearly seen that the algorithm determined most noise at the edges and close to big blood vessels, which likely reflect residual movement artifacts and noise due to blood pulsation respectively. Please note that the color scale is arbitrary scaled to 1.

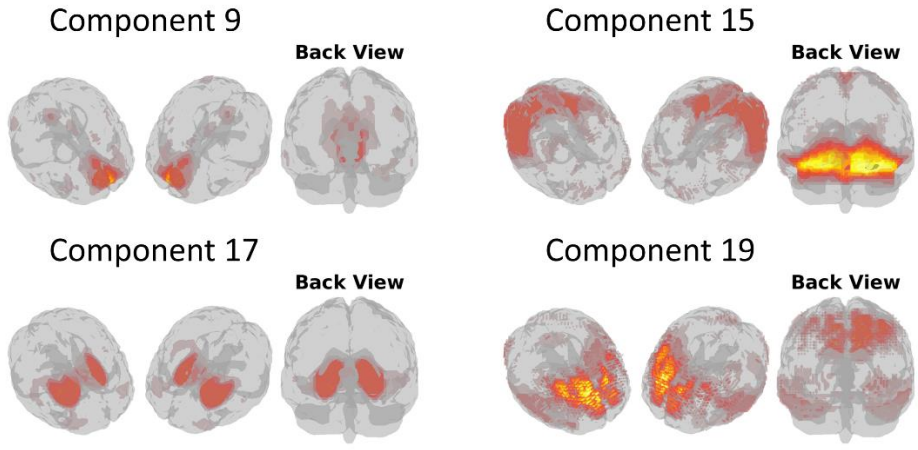
Stability of whole-brain multi-subject archetypical analysis

The cost function in multi-subject archetypical analysis is non-convex just like independent component analysis, hence it is not guaranteed that the same components will be identified across multiple runs with different random initializations. To alleviate this issue, the run which obtained the lowest value of the cost function out of 10 optimizations each with random initializations were chosen each time we used the algorithm. To further investigate the stability, we here rerun this procedure 10 times investigating the similarity of the spatial maps obtained. As there is a trivial ordering/permutation ambiguity of the components across components, they were matched across runs by successively pairing the components that were most correlated (based on correlation of the spatial maps). In this procedure we used the run that was used in the main article as reference, hence the ordering of the components is the same as in the main article.

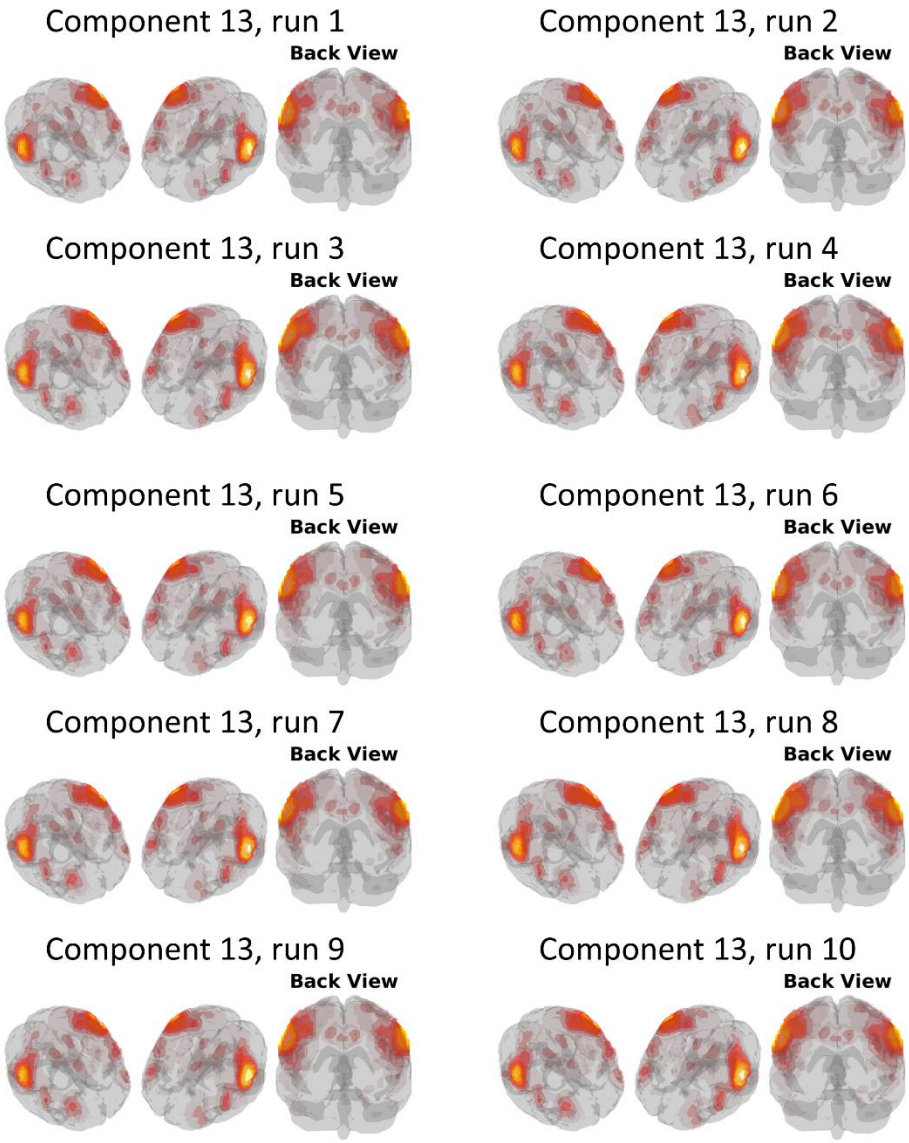
Supplementary figure 7 shows the spatial correlation of each of the components across the 10 runs. The reference run (run number 9 in supplementary figures 9 and 10) is used as reference and is therefore not shown. For most of the components the components are quite consistent across runs generally obtaining spatial correlations above 0.9, however some components (in particular component 15) are poorly matched across runs. While this indicates some instability, it is not too unexpected in case of model mismatch as some components (in particular unstable nuisance components) may not be identified in all runs. For the components obtaining significant classification rates we generally observe very high correlation across runs indicating high stability, in supplementary figure 8 component number 13 (which obtained significant classification of HSA) is displayed across the 10 runs as an example. Similarly, the least stable component number 15 is displayed in supplementary figure 9. Note, that due to the high dimensionality of the spatial components the correlation value can be low even if the components are visually quite similar.



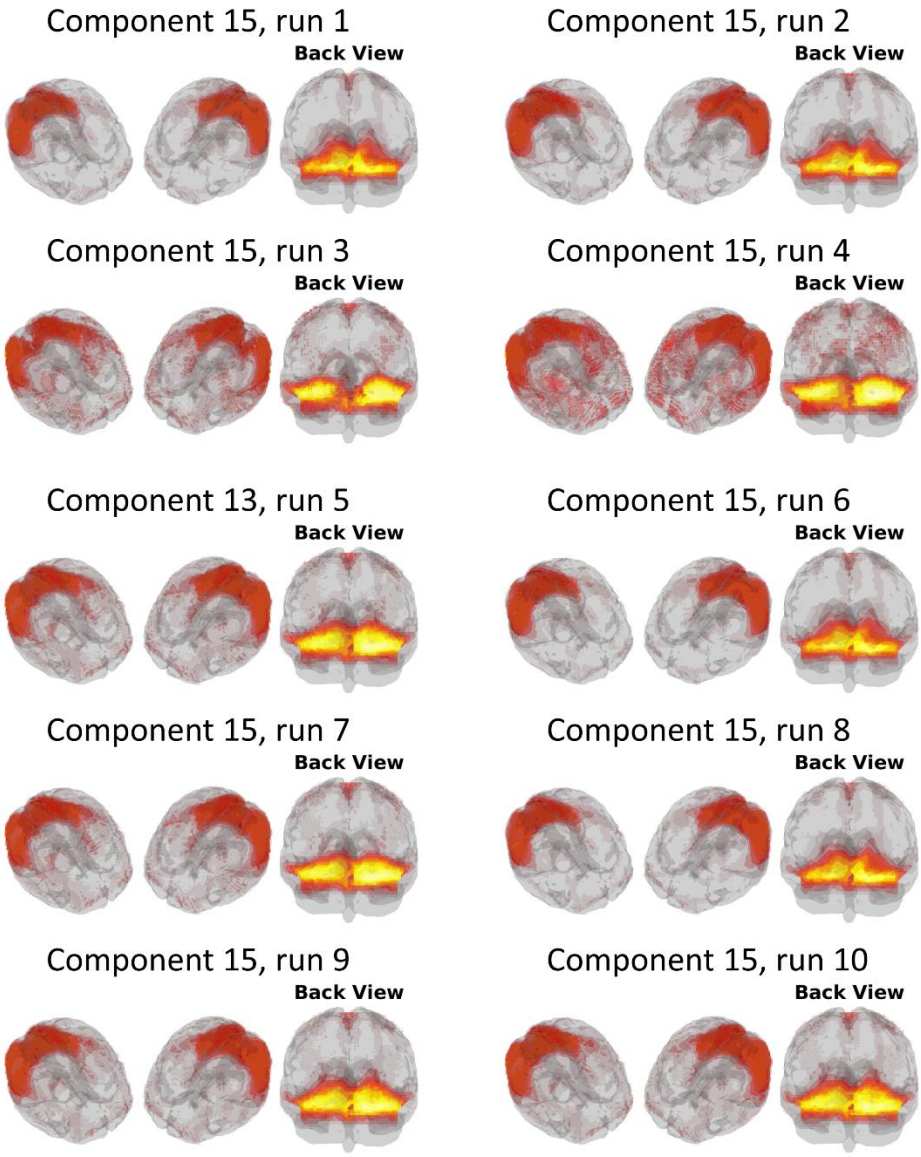
Supplementary figure 7: Correlation of spatial maps across runs and components.



Supplementary figure 8: Spatial maps of the least stable components. The figure shows the spatial maps of the components that were least stable across runs, for brevity only the refence run is shown.



Supplementary figure 9: Spatial maps of component 13. This component obtained significant HSA classification is displayed across the 10 runs. The components are extremely similar across runs, and repeating the HSA classification for each run also showed significant classification in all 10 runs.



Supplementary figure 10: Spatial maps of component, 15 which was the least stable across runs. Note that despite the low correlation value the components are visually similar across runs.

Interpretation of covariance features for HSA classification

To investigate which of the regions of interest (ROI) and functional connectivity between them were responsible for the significant classification of HSA using covariance features were significant we performed an analysis aiming at reveling the significant features. For this analysis the number of features was the diagonal the upper triangular part of the 25 by 25 ROI covariance matrix resulting in a total of 325 features.

As direct interpretation of weight maps in support vector classification is known to be ambiguous, we used the procedure suggested by Haufe et al. (Haufe et al., 2014) to invert the decoding model to identify an activation map. As statistical inference on activation maps is not immediately possible, we investigated the stability of these maps using repeated cross validation. To this end we used the split-half resampling approach suggested in (Strother et al., 2002) to identify reproducible Z-scored activation maps.

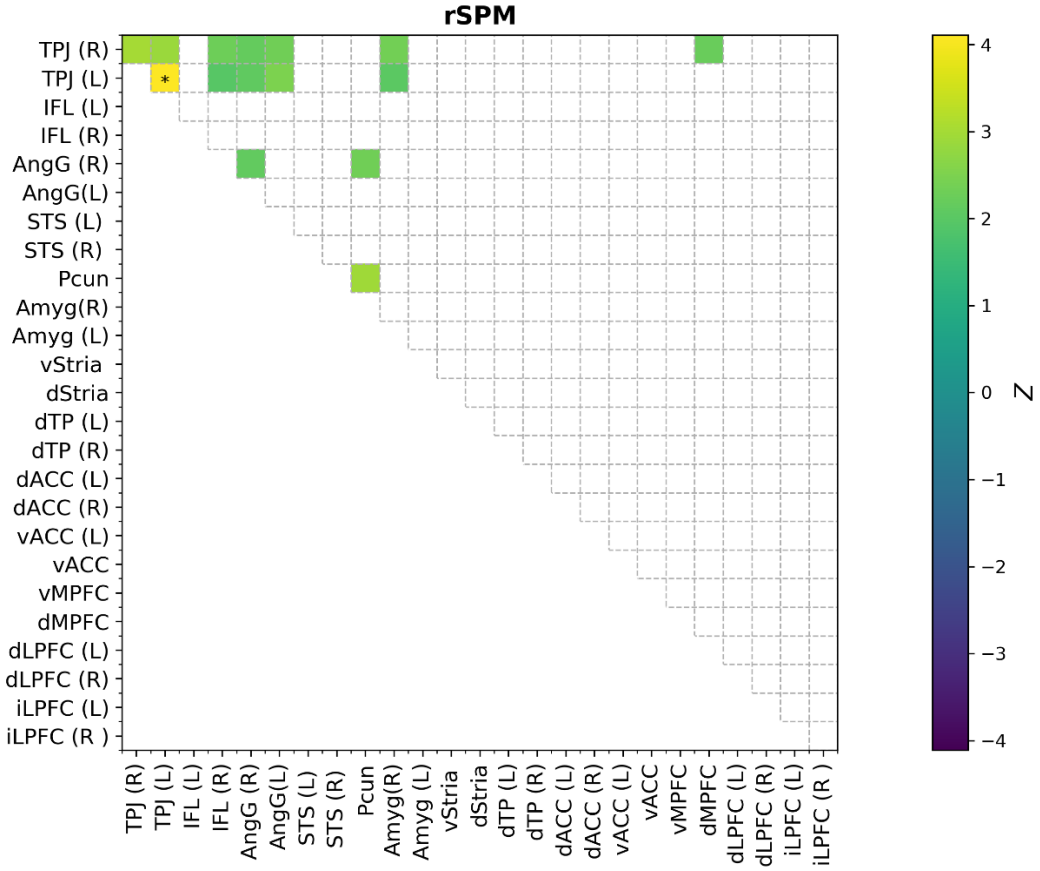
The procedure involved randomly splitting the data into two equally sized proportions (while keeping the proportions of high and low HSA approximately equal in the two proportions). Then the support vector classification where fitted on each of the splits thereby identifying two feature weight vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ in this case with dimensionality 325, these were then converted into activation maps $\boldsymbol{a}_{i,1}$ and $\boldsymbol{a}_{i,2}$ following the procedure suggested in (Haufe et al., 2014)

$$\boldsymbol{a}_i = \boldsymbol{C_i w_i}.$$

where $\boldsymbol{C}_i$ is the estimated 325 by 325 data covariance for split i. Note, that scaling of $\boldsymbol{a}_i$ is arbitrary. Then a reproducible statistical map was constructed as

$$\boldsymbol{Z} = \frac{\boldsymbol{a_1} + \boldsymbol{a_2}}{\sigma(\boldsymbol{a_1} - \boldsymbol{a_2})},$$

Where σ is the standard deviation operator (here subtracting the mean of $\boldsymbol{a}_1 - \boldsymbol{a}_2$ is allowed as for each split as an equivalent opposite split exists). The main insight behind this equation is that $\boldsymbol{a}_1 + \boldsymbol{a}_2$ is an estimate of the activation map while $\boldsymbol{a}_1 - \boldsymbol{a}_2$ is an unbiased estimate of variability of the activation map (due to the independent splits) see (Strother et al., 2002) and (Rasmussen, Hansen, Madsen, Churchill, & Strother, 2012) for further details. In this case we are assuming equal variance across features by using a scalar covariance estimate. The result is an approximately z-scored (under an assumption of normality) activation map. The estimate can be improved by averaging across repeated splits, in this study we averaged across 100 splits of the data. To do inference on the significant features we compared the averaged z-scored map to a cumulative Normal distribution and Bonferroni corrected across the 325 multiple comparisons to control two-sided family-wise type I error at the 5% level leading to a $\boldsymbol{Z}$ threshold of $\Phi^{-1}\left(\frac{0.05/2}{325}\right) \approx 3.78$, where Φ^{-1} denotes the inverse normal cumulative distribution function. Supplementary figure 11 shows the significant features. Note that only feature surviving the family-wise error correction is the variance within the left temporal parietal junction.



Supplementary figure 11: Reproducible statistical map of covariance features for classification of HSA, z-scores significant at the two-sided 5% significance level is shown and significance at 5% Bonferroni corrected for multiple comparisions is indicated by a *. Only the covariance of the left temporal parietal junction, TPJ (L) reach significance.

Supplementary table 3: Results from pooled condition (PCon) mass univariate analysis. Center coordinates from this analysis were used for the PCon spotlight mask used for MSAA. Table shows peak Z score, cluster size, and MNI coordinates for all significant clusters (Significance level $\alpha_{RFT} \leq 0.05$, where random field theory was used to correct for multiple comparison correction).

	Left hemisphere					**Right hemisphere**				
	Peak Z	**Cluster size**	**MNI coordinate** x	y	z	**Peak Z**	**Cluster size**	**MNI coordinates** x	y	z
Pooled condition analysis (PCon)										
Anterior middle temporal gyrus	6.15	109	-54	-9	-21	7.28	235	54	0	-21
Temporoparietal junction	7.30	287	-51	-69	24	7.12	332	51	-63	24
Lateral Superior temporal sulcus	4.75	6	-42	18	-33					
Medal superior temporal sulcus	4.62	1	-36	21	-33					
Cuneus	7.56	135	-12	-108	9	7.23	93	15	-105	12
Ventral medial prefrontal cortex						5	19	3	51	-15
Dorsal medial frontal gyrus	5.63	46	-12	51	48	4.7	3	12	57	42
Anterior parahippocampal gyrus	5.38	30	-24	15	-21					
Posterior parahippocampal gyrus	4.71	8	-21	-33	-18					
Fusiform gyrus	5.05	7	-36	-48	-21					
Middle occipital gyrus	4.81	7	-45	-84	0					
Precuneus	7.37	603	0	-57	39					

References

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. https://doi.org/10.1016/j.neuroimage.2013.10.067

Rasmussen, P. M., Hansen, L. K., Madsen, K. H., Churchill, N. W., & Strother, S. C. (2012). Model sparsity and brain pattern interpretation of classification models in neuroimaging. *Pattern Recognition*, *45*(6), 2085–2100. https://doi.org/10.1016/j.patcog.2011.09.011

Strother, S. C., Anderson, J., Hansen, L. K., Kjems, U., Kustra, R., Sidtis, J., … Rottenberg, D. (2002). The quantitative evaluation of functional neuroimaging experiments: the NPAIRS data analysis framework. *NeuroImage*, *15*(4), 747–771. https://doi.org/10.1006/nimg.2001.1034