Kim Dong-Youl (Orcid ID: 0000-0003-4844-0729)

Lee Jong-Hwan (Orcid ID: 0000-0002-8902-6009)

fMRI multivoxel pattern analysis reveals neuronal substrates for collaboration and competition with myopic and predictive strategic reasoning

Dong-Youl **Kim**[1], Eun Kyung **Jung**[1], Jun **Zhang**[2], Soo-Young **Lee**[3], Jong-Hwan **Lee**[1]

[1]Department of Brain and Cognitive Engineering, Korea University
[2]Department of Psychology, University of Michigan
[3]Department of Electrical Engineering and Dept. of Bio & Brain Engineering, KAIST

Word count: 6,899 (excluding a title page, abstract, references, one table, and eleven figures)

Running title: MVPA of fMRI data from collaboration and competition with strategic reasoning

All Correspondence to:
Jong-Hwan Lee, Ph.D.
Department of Brain and Cognitive Engineering
Korea University
Anam-ro 145
Seongbuk-gu, Seoul 02841
Republic of Korea
E-mail: jonghwan_lee@korea.ac.kr
Tel: +82-2-3290-5922
Fax: +82-2-3290-3667

Abstract

Competition and collaboration are strategies that can be used to optimize the outcomes of social interactions. Research into the neuronal substrates underlying these aspects of social behavior has been limited due to the difficulty in distinguishing complex activation via univariate analysis. Therefore, we employed multivoxel pattern analysis of functional magnetic resonance imaging to reveal the neuronal activations underlying competitive and collaborative processes when the collaborator/opponent used myopic/predictive reasoning. Twenty-four healthy subjects participated in 2×2 matrix-based sequential-move games. Searchlight-based multivoxel patterns were used as input for a support vector machine using nested cross-validation to distinguish game conditions, and identified voxels were validated via the regression of the behavioral data with bootstrapping. The left anterior insula (accuracy $= 78.5\%$) was associated with competition, and middle frontal gyrus (75.1%) was associated with predictive reasoning. The inferior/superior parietal lobules (84.8%) and middle frontal gyrus (84.7%) were associated with competition, particularly in trials with a predictive opponent. The visual/motor areas were related to response time as a proxy for visual attention and task difficulty. Our results suggest that multivoxel patterns better represent the neuronal substrates underlying the social cognition of collaboration and competition inter-mixed with myopic and predictive reasoning than do univariate features.

Keywords: Collaboration; competition; fMRI; multivoxel pattern analysis; strategic reasoning; visual

2

attention

1. Introduction

Social interaction influences both the self and others (Rilling et al., 2004) and can shape an individual's behavior in distinct ways depending on environmental demands, such as whether competition or collaboration with others is required to achieve the best outcomes. Previous neuroimaging studies attempting to identify the neuronal substrates underlying social interactions have found increased activation within the medial prefrontal cortex, posterior superior temporal sulcus, and temporoparietal junction (Coricelli and Nagel, 2009; Fareri and Delgado, 2014; McCleery et al., 2011; Rilling et al., 2004; Saxe and Kanwisher, 2003; van der Meer et al., 2011).

Despite these previous studies, there have been few systematic comparisons of the brain regions associated with competitive and collaborative decision-making (Becchio et al., 2012; Christopoulos and King-Casas, 2015; Decety et al., 2004; Le Bouc and Pessiglione, 2013; Ramsøy et al., 2015). Using a pattern completion game, Decety and colleagues (2004) found that the competitive decision-making process was encoded in the dorsomedial prefrontal cortex, whereas the collaborative decision-making process was encoded in the medial orbitofrontal cortex and limbic area (Decety et al., 2004). More recently, Le Bouc and Pessiglione (2013) employed a handgrip task to manipulate the physical effort between participants and a computerized agent in a competitive or collaborative context and found activation in the medial prefrontal cortex during competition and in the temporoparietal junction during collaboration (Le Bouc and Pessiglione, 2013).

Social interactions may also require higher-level reasoning for optimal outcomes, such as predicting the responses of others in specific situations (Coricelli and Nagel, 2009; Krill and Platek, 2012; Xiang et al., 2012; Yoshida et al., 2010). Coricelli and Nagel (2009) adopted the "beauty contest" game to investigate how a player's own strategies account for the strategic reasoning of other players

4

(Costa-Gomes and Crawford, 2006; Nagel, 1995) and found that high-level reasoning was associated with greater activation in the medial prefrontal cortex, as evidenced by the blood-oxygenation level-dependent (BOLD) response from functional magnetic resonance imaging (fMRI). Yoshida and colleagues (2010) found increased dorsolateral prefrontal activation under higher-level reasoning compared with lower-level reasoning during a two-player stag hunt game in which the participant and computerized agent had to choose whether to collaborate in hunting a highly valued stag or defer to hunting less valuable rabbits separately (Battalio et al., 2001; Skyrms, 2004). Reasoning level refers to the level of thinking that is employed in social interactions in which the actions of others affect the individual's own outcomes in order to predict these actions (Goodie et al., 2012; Hedden and Zhang, 2002). Therefore, we were interested in investigating both the effects of different reasoning levels on the decision-making process and the difference in neuronal activation patterns for each reasoning level in competitive and collaborative scenarios.

A univariate approach in which each voxel is treated as an independent unit (Becchio et al., 2012; Christopoulos and King-Casas, 2015; Decety et al., 2004; Le Bouc and Pessiglione, 2013; Ramsøy et al., 2015) has been widely adopted to estimate voxel-wise neuronal activation in imaging studies of social behavior. However, this univariate approach may be insufficient when attempting to identify the neuronal underpinnings of experimental variables and/or conditions if the spatial pattern of activation across neighboring voxels is also informative in explaining other variables and/or conditions. As an alternative, a multivoxel or multivariate pattern analysis (MVPA) strategy that employs neuronal activation patterns is widely considered to be a promising analytical tool for the investigation of representative patterns in two or more experimental conditions (Allefeld and Haynes, 2014; Haxby, 2012; Kay et al., 2008; Kriegeskorte et al., 2006; Mur et al., 2009; Wang et al., 2013), such as competition/collaboration and lower-/higher-level reasoning.

5

Therefore, in the present study, we examined the neuronal substrates of low- and high-level strategic decision-making in contexts requiring competition or collaboration using MVPA to distinguish activation patterns acquired during a modified matrix-based sequential-move game originally designed by Hedden and Zhang (Hedden and Zhang, 2002; Zhang et al., 2012). In contrast to other paradigms, such as the beauty contest game (Coricelli and Nagel, 2009; Costa-Gomes and Crawford, 2006; Nagel, 1995) and the stag hunt game (Battalio et al., 2001; Skyrms, 2004; Yoshida et al., 2010), the brain regions contributing to two distinct components of social interaction – (a) competition/collaboration and (b) depth of strategic reasoning – could be systematically investigated using this chosen paradigm. We hypothesized that distinct patterns of neuronal activation depending on the requirement for competition or collaboration and whether the social counterpart used low-level (myopic) or high-level (predictive) reasoning can be revealed using our experimental paradigm by applying MVPA analysis to acquired fMRI data. We also predicted that the prefrontal cortex, parietal lobe, and other brain regions would be uniquely altered under different social interaction when systematically evaluated using behavioral measurements.

2. Materials and Methods

2.1. Participants

The Institutional Review Board (IRB) at Korea University approved the overall study protocol. All participants provided written informed consent and were compensated based on IRB documents. Healthy volunteers with no neuropsychiatric or neurological disorders and no contraindications for magnetic resonance imaging (MRI) were recruited. A total of 24 right-handed volunteers participated (12 females, Edinburg's handedness score = 89.3 ± 5.8; age = 23.9 ± 3.3 years; see Table S1 for additional details on the participants).

6

2.2. Sequential-move games involving competition and collaboration with two levels of reasoning

Figure 1 shows the experimental paradigm for the 2×2 matrix-based sequential-move game originally proposed by Hedden and Zhang (2002) and modified to incorporate a competitive (CP) or collaborative (CB) mode of social interaction with a staff member (social counterpart; one of our lab members) employing two levels of reasoning. The matrix contains four cells (A, B, C, and D) with two rewards in each cell (the first reward for the participant and the second for the staff member). The game starts at cell A, and the participant needs to decide whether to move to cell B for the best final outcome depending on the task conditions. If the participant decides to move to cell B, the staff member then needs to decide whether to move to cell C. If the staff member decides to move to cell C, the participant can then decide to move to cell D. These decisions are based on rewards. If either the participant or staff member decides not to move to the next cell, the game ends and the participant and staff member receive the dollar amount written in that final (stop) cell. In collaboration mode, the participant was instructed to maximize both the sum of their rewards and that of the staff member, whereas in competition mode, the participant was instructed to maximize their reward while minimizing the staff member's reward. In addition, the participants were told that the staff member would use either low-level reasoning or high-level reasoning (referred to as *myopic* and *predictive*, respectively).

For example, in the collaboration mode game shown in Figure 1, the participant would know that if they move from A to B, (a) a myopic staff member will compare the sum of rewards at B and C ($4 vs. $3) and decide to stay at B to maximize the total reward, and (b) a predictive staff member will also compare the sum of the rewards at D versus B and C ($8 vs. $4 and $3) and thus will move to C (despite it having less money) because the staff member knows that the participant will then move to D to obtain the largest reward. Thus, the participant will stay at A for a myopic staff member (so that

7

the game stops at A and they get $5) or move to B for a predictive staff member (so that the game stops at cell D and they get $8).

In competition mode, the participant would know that (a) a myopic staff member will compare the difference of rewards at B and C ($4 − $3 = $1 vs. $3 − $1 = $2) and decide to move to C to obtain a greater reward than the participant, and (b) a predictive staff member will also compare the difference of the reward at D (i.e., $2 − $4 = −$2) and thus will not move from B to C despite the greater reward at C vs. B ($2 vs. $1), because the staff member knows that the participant will move from C to D to obtain a greater reward than the staff member (4$ − $2 = $2). Thus, the participants will move to B against a myopic staff member (so that the game proceeds to C and then D) or stay at A against a predictive staff member (so that the game stops at A and the participant gets $1 rather than stopping at B where the staff member obtains the larger reward).

In matched-control mode using the same set of games for the competition or collaboration condition, the participants were asked to maximize their own reward while ignoring the staff member's reward. In control mode using the same game for collaboration in Figure 1, the participant will not move to B when facing a myopic staff member (because the staff member will stay at B for $3 rather than move to C for $1) or will move to B when facing a predictive staff member (because the staff member will move to C knowing that the participant will then move to D for the higher reward, which is $4 for both).

In each game trial, two questions were also presented: (Q1) "Will OO (the staff member's initials) move to C?" and (Q2) "Will you move to B?" The next game trial started when the participants

8

answered "Yes" or "No" to these two questions by pressing a button on a fiber-optic response pad (Current Design, PA, USA; www.curdes.com).

2.3. Experimental setup

There were four game scenarios across the competition and collaboration game modes with either a myopic or predictive staff member (i.e., a 2×2 factorial design), and each scenario was performed in a separate fMRI run (i.e., four runs total). Each run consisted of 10 game trials in either the competitive or collaborative mode along with the 10 matched control trials with either a myopic or predictive staff member (i.e., 80 trials total; 20 trials per run). The game mode and the staff member's reasoning level were fixed in each scenario. Participants were told the game mode (competitive or collaborative) before each scenario began; however, they were not informed of the staff member's reasoning level in each scenario and thus had to infer it from the staff member's behavior as the trials proceeded in order to obtain the maximum reward. The participants were told that they would be compensated proportionally to the rewards they obtained from all of the game trials. Once the participant was placed in the MRI scanner, a computerized agent, instead of the staff member, assigned programmed strategies according to the game mode and reasoning level used. The participants performed the four scenarios in a pseudo-randomized order (i.e., inter-mixed order across four scenarios for each of the subjects; "randperm.m" function with a randomized seed in MATLAB [version 2016a] was used for the randomization). Prior to the experiment, subjects practiced the task during an interview day.

During the debriefing, the participants provided subjective scores on a 5-point Likert scale for (1) "how much do you agree that you were competing against the staff?" (or "how much do you agree that you were collaborating with the staff?") (22 of 24 participants) and (2) "how difficult were the competition games?" (or "how difficult were the collaboration games?") (all 24 participants). The

9

competition and collaboration scores were defined from the first question and the difficulty scores for each of the competition and collaboration game modes were defined from the second question. Then, the participants were informed that a computerized agent was employed on behalf of our staff member to ensure the consistency of the experiment.

2.4. fMRI data acquisition and preprocessing

BOLD fMRI data were acquired using a standard gradient-echo echo-planar imaging (EPI) pulse sequence on a 3-Tesla MRI scanner (Tim Trio; Siemens, Erlangen, Germany; TR/TE = 2000/30 ms; 35 interleaved axial slices parallel to the anterior and posterior commissural plane without a gap; voxel size = $3 \times 3 \times 4$ mm^3, matrix size = 80×80 voxels; flip angle = 90°; field-of-view = 240×240 mm^2).

The first five EPI volumes (10 s) of each run were excluded to allow the T_1 effect to equilibrate. The remaining EPI volumes were preprocessed using the SPM8 toolbox (www.fil.ion.ucl.ac.uk/spm) with standard options, including slice timing correction, realignment, and spatial normalization to the Montreal Neurological Institute template with 3-mm isotropic voxel size. Spatial smoothing was not performed to fully preserve task condition-dependent differences in the multivoxel patterns from the BOLD signals (Kriegeskorte et al., 2008). All participants ($n = 24$) exhibited head motion of less than 1 mm during realignment. The mean (± standard deviation) of translational and rotational head motions across all the subjects were 0.63 mm (± 0.32 mm) and 1.14 ° (± 0.41 °), respectively.

2.5. Estimation of neuronal activation

To create the features for MVPA, a general linear model (GLM) was applied to the preprocessed BOLD

10

fMRI data (Kamitani and Sawahata, 2010; Mahmoudi et al., 2012; Misaki et al., 2010; Swisher et al., 2010). The onset timing of each competition, collaboration, and control game trial and offset timing for the participants to make a decision regarding the second question (Q2) were used to model the reference hemodynamic response function (HRF) for each game trial using SPM8. Then, the trial-wise regressors across all the conditions (i.e., competition, collaboration, and control) were defined as the results of the convolved signals between the HRF and trial-wise information for onset and offset of the trial. In this study, the control trials were not used for further analysis and the response time was not included as a parametric modulation factor in the GLM. For each participant, neuronal activation patterns were estimated from the beta values of the GLM across voxels within the whole brain (i.e., 10 beta maps for each competition or collaboration game mode with the myopic or predictive staff member). If the response time of any trial was slower than 1.5 times the interquartile range of all response times for that game mode, the corresponding trial was treated as invalid and excluded from analysis (2 trials for two subjects and 1 trial for three subjects).

2.6. Classification of competition versus collaboration and myopic versus predictive reasoning

Figure 2(a) presents a flow diagram of the searchlight MVPA-based classification approach adopted for each subject. There were a total of six scenarios for binary classification: (i) competition vs. collaboration mode across all myopic and predictive staff member trials, (ii) competition vs. collaboration mode considering only myopic staff member trials, (iii) competition vs. collaboration mode considering only predictive staff member trials, (iv) myopic staff vs. predictive staff member trials across both game modes, (v) myopic staff vs. predictive staff for the competition mode trials, and (vi) myopic staff vs. predictive staff for the collaboration mode trials. Four-class classification across all four conditions (vii) was also conducted. The available trials for each of these seven classification scenarios were divided into five folds: three folds for training a support vector machine (SVM)

11

classifier, one fold for validation (to optimize the SVM parameter and searchlight size among 1-, 2-, and 3-voxel sizes), and one remaining fold for testing. Therefore, each fold included 8 trials for each of the classification scenarios from (i), (iv), and (vii) and 4 trials for the classification scenarios of (ii), (iii), (v), and (vi).

The beta values for the voxels within the searchlight area were normalized to fall between 0 and 1 across the trials in the training data (i.e., normalization with the range from 0 to 1 across the number of voxels × the number of trials used), and scaling factors for this normalization were applied to the validation and test data. A *v*-SVM classifier with a linear kernel was used as implemented in the LIBSVM software toolbox (www.csie.ntu.edu.tw/~cjlin/libsvm). The latent parameter *v* of the *v*-SVM was optimized via a grid search using uniformly distributed candidate values (from 0.1 to 0.8, with an interval of 0.1) (Chen et al., 2005). The nested cross-validation classification test (Allefeld and Haynes, 2014; Filzmoser et al., 2009; Hebart et al., 2015; Lindquist et al., 2017; Varoquaux et al., 2017) was repeated 10 times using 10 randomly shuffled training, validation, and test sets, and average classification accuracies are reported. We used custom-made MVPA code implemented in MATLAB, and the results were verified by comparing them with results from the decoding toolbox (TDT) (Hebart et al., 2015). To validate the results between custom-made MVPA code and the TDT toolbox, we randomly selected three subjects and employed the same procedure, such as nested cross-validation with parameter optimization. The difference in individual classification accuracy between the custom-made MVPA code and the TDT toolbox ranged from 0% to 0.21% for all voxels in the whole brain.

The classification results from each of the 24 subjects were the clusters of brain regions with a minimum of 20 voxels, accuracy ≥ 70% for binary classification, and accuracy ≥ 50% for four-class

12

classification. These results were subjected to one-sample *t*-tests for group inference (Eger et al., 2009). The statistical significance of group inference was estimated from 10,000 classification tests performed using randomly permuted training, validation, and test sets, with randomly assigned class labels for the test sets (Cauchoix et al., 2014; Liang et al., 2013; Stelzer et al., 2013).

2.7. Comparison of identified brain regions using the univariate approach and MVPA

The neuronal activations arising from confounding factors related to visual attention level may potentially be estimated when using a univariate approach because BOLD signals are known to be very sensitive to attention level (Bartels et al., 2008; Li et al., 2008; Logothetis, 2003; Ungerleider and G, 2000). We thus conducted univariate estimation and compared the results to those from MVPA classification. For voxel-wise univariate analysis, a linear mixed-effect test (Barr et al., 2013) was adopted for each of the six contrasts: (1) competition vs. collaboration across all myopic and predictive staff member trials, (2) competition vs. collaboration considering only myopic staff member trials, (3) competition vs. collaboration considering only predictive staff member trials, (4) myopic staff vs. predictive staff member trials across both game modes, (5) myopic staff vs. predictive staff for competition mode, and (6) myopic staff vs. predictive staff for collaboration mode. The statistical significance ($p < 0.05$) was estimated from a null distribution that was obtained from 10,000 tests performed using randomly permuted sets with randomly assigned contrast labels.

2.8. Evaluation of identified brain regions using voxel-wise multiple regression

The brain regions identified using the univariate approach and MVPA were further evaluated with voxel-wise multiple regression and permutation testing. For the univariate approach, the beta values were predicted using the competitive and collaborative scores, with response time, age, and sex

13

included as confounding factors. In particular, response time was included to remove possible visual attention components affecting our experimental conditions (i.e., CP vs. CB and predictive vs. myopic). In addition, reasoning time was defined as the period from the onset time of the second question to the participants making a decision. Because the subjects were assumed to have finished mathematical calculations prior to answering the first question, we expected the subjects to process only the opponent's strategy (i.e. myopic/predictive) when answering the second question. This reasoning time was used as the interest-of-regressor to predict myopic and predictive reasoning order by regressing out response time, age, and sex. Classification accuracy was used as the target variable for MVPA. The statistical significance of the prediction based on the regression analysis was evaluated using nonparametric significance testing. To this end, the regression analysis was conducted using subjective scores or reasoning time in a randomly shuffled order. This process was repeated 10,000 times to generate a null distribution with one-tailed significance testing using a Bonferroni-corrected p-value (Kim et al., 2015; Maysov and Kipyatkov, 2011; Mundfrom et al., 2006).

2.9. Cross-set classification to validate MVPA

The identified brain regions and corresponding classification accuracy may be affected by potential problems of overfitting arising from training on one set of trials and testing on an independent set of trials. Thus, the cross-set classification approach (Kaplan et al., 2015; Wisniewski, 2018; Wisniewski et al., 2016) was adopted by training the classifier to distinguish between collaboration and competition using data from trials with a myopic player. The trained classifier was then applied to the test set on data from trials where the opponent was a predictive player. In addition, the classifier was trained using data from trials with a predictive player and tested using data from trials with a myopic player. Similarly, to distinguish trials involving myopic and predictive staff members, the classifier was trained using data from collaboration trials and tested using data from competition trials. Additionally, the

14

classifier was trained using data from competition trials and tested using data from collaboration trials.

3. Results

3.1. Behavioral data

The proportion of participants that made the correct decision according to the context of competition/collaboration with the myopic or predictive staff member (stay at cell A or move to cell B) was 95.1% ± 3.1% (mean ± standard deviation), indicating that participants accurately inferred the staff member's reasoning level. The average rating for task difficulty was significantly higher during the competition trials than the collaboration and control trials (Table 1). Response time was significantly higher in competition mode (9.31 ± 1.79 s) than in collaboration mode (6.73 ± 1.31 s; $t(23) = 9.63$, Bonferroni-corrected $p = 5.03 \times 10^{-5}$, effect size $d = 1.64$) and with the predictive staff member (9.48 ± 2.31 s) than with the myopic staff member (7.28 ± 1.51 s; $t(23) = 8.92$, Bonferroni-corrected $p = 8.98 \times 10^{-5}$, effect size $d = 1.13$; Figure 3).

3.2. Brain regions identified from competition versus collaboration classification

Figure 4 shows the results from searchlight-based multivoxel pattern classification (Fig. 4a) and the univariate approach (Fig. 4c) to distinguish between the competition and collaboration modes. The subsequent voxel-wise regression analysis for accuracy and (i) subjective scores (Fig. 6a) or (ii) response time (Fig. 6b) is also presented. The visual, parietal, and frontal areas fulfilled the criteria for classification, but the left anterior insula and bilateral parietal activation dominantly predicted the subjective scores for competition without the confounding influence of visual attention components. These brain regions also fulfilled the criteria for distinguishing competition from collaboration when facing the predictive staff member. In this case, classification results from the supplementary motor area (SMA) and the left anterior/right posterior insula were able to predict the subjective scores for competition when the visual attention components were regressed out. In contrast, only small regions

15

of the cuneus, cerebellum, posterior cingulate cortex, and precuneus fulfilled the criteria for the classification of competition versus collaboration with the myopic staff members. There were few voxels with features that were able to predict subjective scores in this classification scenario. Response time, a proxy for visual attention, was predominantly identified across the visual and parietal areas.

3.3. Brain regions identified from myopic versus predictive reasoning order classification

Figure 5 presents the results of the classification of game trials involving myopic and predictive reasoning order, and the subsequent regression analysis is shown in Figure 7. The visual area, superior and inferior parietal lobules, middle frontal gyrus, and medial superior frontal gyrus exhibited a classification accuracy that was well above the level of chance. Of these brain regions, the brain features in the middle frontal gyrus and bilateral parietal areas predicted reasoning time in the game trials with a predictive staff member after controlling for visual attention component. Parts of the visual areas were associated with the visual attention components of the game trials with a myopic staff member. Brain regions distinguishing the two reasoning levels in competition mode were found. After regressing out visual attention component, classification features in the superior parietal lobule, inferior/middle frontal gyrus, and inferior/superior parietal lobule predicted the reasoning time for games with a predictive staff member. The middle occipital gyrus and supplementary motor area also showed an association with the reasoning time for games with a myopic staff member. However, binary classification of the reasoning level in collaboration mode found that only the superior parietal area had features that predicted the reasoning time for games with a predictive staff member. In this classification analysis, visual attention components were mainly found in the visual and motor areas and parts of frontal and parietal areas.

3.4. Comparison between the univariate approach and MVPA

16

Figure 4 displays the spatial patterns for competition and collaboration from (a) MVPA, (b) overlapping patterns from MVPA and the univariate approach, and (c) the univariate approach. Figures 6 and 8 present the subsequent regression analysis for MVPA and the univariate approach, respectively. For myopic versus predictive reasoning order, Figure 5 represents the spatial patterns from (a) MVPA, (b) overlapping patterns from MVPA and the univariate approach, and (c) the univariate approach. Figures 7 and 9 present the results from a subsequent regression analysis for MVPA and for the univariate approach, respectively. The univariate approach mostly covers sensory areas such as the visual and motor areas and the corresponding patterns mostly overlapped with the patterns from MVPA. The brain areas identified using the univariate approach covered parts of the frontal, motor, and subcortical areas. Of these brain areas, the beta values of a few scattered voxels in the visual areas predicted behavioral information for trials involving a myopic staff member or collaboration. From the subcortical, frontal, and parietal areas, the beta values predicted cognitive load in trials involving a predictive staff member or competition. Both MVPA and the univariate approach found that the visual and motor areas were most strongly associated with the visual attention components.

3.5. Classification across all four conditions

Figure 10 shows the brain regions identified using four-class classification. The visual, parietal, and frontal areas showed significantly greater classification accuracies than chance (Fig. 10a; Table S11). The classification accuracy for competition against a predictive staff member (i.e., CP1) was higher than the classification accuracy for collaboration with either a myopic or predictive staff member (i.e., CB0 or CB1; Fig. 10b), despite the fact that the beta values in these brain regions did not differ across game modes and reasoning levels (Fig. 10c). The accuracy of the left inferior and right superior parietal lobules and left middle frontal gyrus for a predictive staff member (CP1) predicted the competition scores (Fig. 10d) by excluding the visual attention component. The visual areas and supplementary

17

motor areas were able to predict collaboration scores for both myopic and predictive staff members.

3.6. Cross-set classification

Figure 11 presents the brain regions identified by cross-set classification, including the (a) results from adopting nested cross-validation mixing conditions across the training and test sets, (b) results from cross-set classification by training the classifier to distinguish between collaboration and competition using data from trials with a myopic player and testing the classifier on data from trials where the opponent was a predictive player, or (c) vice versa. Classification was otherwise conducted by training the classifier to distinguish between myopic and predictive order in collaboration trials and testing the classifier using data from trials involving competition (e) and vice versa (f). The results demonstrated a significant overlap between nested cross-validation schema and cross-set classification. Of the brain regions identified using nested cross-validation, the superior frontal gyrus, insula, and parts of the parietal areas were not identified using cross-set classification for the contrast between competition and collaboration. For the contrast between myopic and predictive order, the middle/superior frontal gyrus, precentral gyrus, and supplementary motor areas were not identified.

4. Discussion

4.1. Summary of the study

Our modified matrix-based, sequential-move game with a systematically controlled opponent/collaborator assignment and reasoning level combined with searchlight-based multivoxel pattern classification identified multiple brain regions in the visual, parietal, and frontal areas (including the insula) associated with social cognition. To the best of our knowledge, this is the first study to apply these techniques to a nested cross-validation framework, which increases the possibility of generalization to independent subjects and reduces the potential for the under-estimation of neuronal

18

substrates compared to univariate analysis.

Based on subsequent voxel-wise multiple regression, left anterior insula activation appears to be strongly involved in thought processes related to competition but not collaboration. In addition, the middle frontal gyrus, inferior and superior parietal lobules, middle occipital gyrus, and cuneus appear more strongly associated with predictive reasoning than myopic reasoning. Finally, the inferior and superior parietal lobules and the middle frontal gyrus exhibited a specific association with competition only against an opponent using higher-level reasoning.

4.2. Brain regions identified from the classification of competition versus collaboration

The brain regions identified from the classification of competition versus collaboration were mainly found in the visual, parietal, and frontal areas, which is in accordance with previous studies (Bechara et al., 2000; Decety et al., 2004; Fareri and Delgado, 2014; Jakab et al., 2012; Jones et al., 2011; Koban and Pourtois, 2014; Le Bouc and Pessiglione, 2013). For example, several studies have reported greater neuronal activation in the middle and superior frontal areas during competitive tasks than during collaborative tasks (Decety et al., 2004; Fareri and Delgado, 2014; Le Bouc and Pessiglione, 2013). When visual stimuli were used to control the cognitive processes related to competition and collaboration among participants as well as the response type of a computerized agent (Le Bouc and Pessiglione, 2013), greater activation in the visual areas was reported for competition than for collaboration, possibly because of the increased attentional focus during competition. The fact that the classification accuracy of the left anterior insula was greater than the level of chance and that the corresponding beta values predicted competition scores suggests that insular activation is related to the level of motivation, which may be greater under competitive than collaborative conditions (Jakab et al., 2012; Koban and Pourtois, 2014). The observed activity in the SMA may be related to the mental

19

simulation of the sequential moves and/or to the keypad responses to Q1 and Q2 (Nachev et al., 2007). These brain regions were also identified by pattern classification of competition versus collaboration when considering only trials with a predictive staff member (higher-level reasoning) but were weaker when considering only trials with a myopic staff member (lower-level reasoning). This suggests the engagement of more complex neural processing pathways in the former condition, which is consistent with (a) the longer response times in competitive trials versus collaborative trials with a predictive staff member and (b) the disappearance of this difference between competition and collaboration in trials with a myopic staff member. Furthermore, regression analysis with response time as a covariate revealed that sensory areas including the visual and motor areas had a weak association with cognitive load. Our results also indicated that visual attention components were highly associated with response time. These results are in line with previous studies that have reported the prediction of visual awareness by removing visual attention components (Webb et al., 2016). He and colleagues also reported the separation of valid and invalid trials due to neglect from stroke patients (He et al., 2007), in which the functional processing of the brain might be better investigated by removing the visual attention components.

4.3. Brain regions identified using the classification of reasoning level

The middle frontal area and inferior and superior parietal lobules were identified from the classification of game trials with myopic or predictive staff members across the two game modes. This extensive cortical activation pattern is consistent with the longer response times when playing against a predictive staff member compared with a myopic staff member, particularly in competition mode. Similarly, in previous studies, higher-level reasoning by an opponent increased the cognitive effort required by participants to predict their opponent's strategy (Allred et al., 2016; Brañas-Garza and Smith, 2016; Duffy and Smith, 2014; Hedden and Zhang, 2002). Furthermore, neuronal activation in the dorsolateral

20

prefrontal cortex, a region in close proximity to the middle frontal area identified in the present study, was greater in a high-level than a low-level reasoning condition (Coricelli and Nagel, 2009). Stone and colleagues reported that autistic patients with damage to the middle frontal gyrus exhibited lower accuracy in a false belief task (Stone et al., 1998). This suggests that the middle frontal gyrus may be important for understanding others' thoughts, particularly higher levels of reasoning, a faculty impaired in cases of autism.

The brain regions across the visual and motor areas identified from MVPA classification disappeared when the confounding effect of response time was removed from the brain features. However, the middle frontal areas and bilateral parietal areas remained (Figs. 6 and 7). This regression analysis thus reflects task-related cognitive processes while ignoring visual attention components (He et al., 2007; Webb et al., 2016) and task difficulty (Demb et al., 1995; Gilbert et al., 2012). For example, Gilbert et al. (2012) have suggested that task difficulty can be measured by response time and does not provide an adequate account of task-induced changes in signals from regions-of-interest. Further research is thus warranted to identify an approach that can accurately define task difficulty and visual attention components in cognitive tasks.

4.4. Comparison between the univariate approach and MVPA

Our study compared the spatial patterns identified using the univariate approach and MVPA. Many MVPA studies have reported that classical univariate analysis is less capable of determining whether an identified brain region is related to specific cognitive process or highly affected by visual attention components (Davis et al., 2014; Gilron et al., 2017; Mahmoudi et al., 2012; Todd et al., 2013; Wisniewski et al., 2016). MVPA also faces potential issues such as feature selection, the complexity of

21

dimensionality, and validation by selecting optimal parameters (Mahmoudi et al., 2012). Nevertheless, in line with other MVPA-based research, our study revealed that MVPA outperformed the univariate approach, revealing that the insula, frontal and parietal regions were associated with the specific cognitive process load for one of the forms of social cognition (i.e., competition under predictive reasoning). On the other hand, the univariate approach mainly identified the visual and motor areas, including other regions including the insula, subcortical areas, and parietal areas. However, few voxels in the visual and motor areas survived after regressing out response time. Consequently, the MVPA approach allows brain regions to be investigated in terms of their relationship with cognitive processing regardless of the presence of visual attention components. In addition, whole-brain MVPA using deep neural networks would be a potentially viable approach to further enhance the prediction performance (Jang et al., 2016; Kim et al., 2019a; Kim et al., 2016).

4.5. Interaction between competition/collaboration and strategic reasoning level

From our four-class classification results, the beta values of the bilateral parietal lobule and middle frontal gyrus predicted competition scores. Notably, the bilateral parietal lobule was also identified in the classification of myopic versus predictive staff members, and the beta values of the corresponding areas were significantly associated with response time in the predictive condition. The temporoparietal junction has been reported as a core brain region involved in the "Theory-of-Mind" (Decety et al., 2004; Gallagher et al., 2000; Marjoram et al., 2006; Saxe, 2006) and is particularly associated with the mental state of belief (Saxe, 2006). The parietal lobule regions identified in our study (x, y, z = −48, −47, 46 mm and 44, −48, 43 mm for the left and right hemispheres, respectively) are in proximity to the temporoparietal junction. The minimum Euclidean distance between the focus of our identified right superior parietal lobule and the focus of the temporoparietal junction (as provided by Mars and colleagues; x, y, z = 49, −46, 46 for the right hemisphere) (Mars et al., 2012) is approximately 2 voxels

22

(6 mm). The bilateral parietal lobules have also been reported as the neuronal substrate for competition in previous studies (Decety et al., 2004; Fareri and Delgado, 2014). In contrast, activation of the parietal areas may also be related to attentional and executive switching between competition and collaboration modes (Decety et al., 2004).

4.6. Evaluation of MVPA via cross-set classification

We conducted MVPA for each experimental factor first using all data and then separately for each of the two conditions for the other factor. For example, the classification of collaboration and competition was run using only data from trials involving interaction with a myopic staff member or those with a predictive staff member. Kaplan et al. (2015) suggested cross-set classification as a way to characterize abstraction in neural representations and to avoid circular analysis (Kaplan et al., 2015; Wisniewski, 2018; Wisniewski et al., 2016). Similarly, the comparison between the performance of nested cross-validation and cross-set classification in the present study indicated the possibility of overfitting or relatively low accuracy when using cross-set classification.

4.7. Potential weaknesses and future work

The number of participants in our study was relatively low. Thus, in order to reduce the false positive error rate and increase the generalizability of our results for independent subjects, we employed a *k*-fold cross-validation framework by dividing the data from the participants into training and test sets for both multivoxel pattern classification and random permutation testing for logistic regression.

In principle, it is also possible that these differences in activation reflect the cognitive load for task performance (addition vs. subtraction) rather than the task conditions. However, the response times

23

did not differ between collaboration with and competition against the myopic staff member, which require different arithmetic tasks (Fig. 3). Moreover, previous fMRI (Kawashima et al., 2004) and EEG (Zhou et al., 2006) studies have reported no significant difference in the activation patterns between addition and subtraction. Previous studies that have employed MVPA to examine addition vs. subtraction have reported high accuracy in the absence of activation differences (Haynes et al., 2007). Interestingly, Haynes and colleagues (2007) reported similar middle frontal gyrus and insula areas to those identified in the present study to distinguish competitive and collaborative decision-making processes. They used response time as a regressor to estimate activation levels during addition and subtraction trials. However, in our study, potential confounding factors arising from response time-related patterns were addressed by conducting an additional regression analysis to remove the visual attention components. Thus, the interpretations in this study may have removed the effect of these mathematical calculations; future research can confirm whether the mathematical processes used need to be considered in designing an experiment.

Previous research that has compared event-related and block-based designs (Chee et al., 2003; Mechelli et al., 2003; Petersen and Dubis, 2012; Tie et al., 2009) has demonstrated that event-related designs may capture hemodynamic response functions better than block-based designs, thus increasing the sensitivity to the condition-of-interest. In the present study, we excluded a few trials in which the response time was slower than 1.5 times the interquartile range of all response times for that game mode. This created an unbalanced number of trials, which might have led to classification results that were biased toward the class with the higher number of trials. In addition, the eye-tracking during the experiment would be helpful in ascertaining the participants' understanding of the purpose of the experimental design by monitoring their choice process (Polonio et al., 2015). For example, we expect participants to gaze at and focus more on cell D than on other cells when following predictive reasoning

24

in our experiment.

As an extension of our study, interesting future work could include (1) identification of the neuronal substrates for social interactions using real-time fMRI (Emmert et al., 2016; Kim et al., 2015; Kim et al., 2019b; Lee et al., 2012) and (2) examination of the activation changes when social interaction is modulated toward cooperative behavior, such as in a hyper-scanning paradigm (Montague et al., 2002). Further, social interaction with real-time neuroimaging may be useful in preclinical settings to facilitate competitive and/or collaborative decision-making in patients with autistic spectrum disorders and/or antisocial personality disorders (Andari et al., 2010; Bühler et al., 2011; Mier et al., 2013). Another potentially interesting future study is to investigate the neuronal substrates for collaboration and cooperation in the context of reinforcement learning because the participants in our study received a reward after each trial. Previous studies have reported common neuronal substrates for social interaction and monetary reward learning, such as the ventral medial prefrontal cortex (Hare et al., 2010; Joiner et al., 2017; Lin et al., 2011). It would be of interest to determine whether a monetary reward enhances the capability for social interaction through a reinforcement learning framework by tracking value and state parameters based on in-game rewards (Lin et al., 2011; Zaki et al., 2016).

25

5. Conclusions

We identified some of the neuronal circuitry underlying competitive and collaborative thought processes in healthy participants during interactions with competitors/collaborators using low-level (myopic) or high-level (predictive) reasoning. For this task, we adopted searchlight-based multivoxel pattern classification and subsequent voxel-wise multiple regression with behavioral scores for validation in a nested cross-validation framework by separating the training and test subjects to prevent "double-dipping" or circular analysis. The anterior insula and SMAs were associated with competition but not collaboration. The middle frontal gyrus, inferior and superior parietal lobules, middle occipital gyrus, and the cuneus were also associated with competition, especially against those using high-level reasoning. We believe that the presented method may prove useful for the investigation of abnormal neuronal circuitry in patients with various cognitive and anti-social disorders.

References

Allefeld, C., & Haynes, J.-D. (2014). Searchlight-based multi-voxel pattern analysis of fMRI by cross-validated MANOVA. *Neuroimage*, *89*, 345-357. https://doi.org/10.1016/j.neuroimage.2013.11.043

Allred, S., Duffy, S., & Smith, J. (2016). Cognitive load and strategic sophistication. *Journal of Economic Behavior & Organization*, *125*, 162-178. https://doi.org/10.1016/j.jebo.2016.02.006

Andari, E., Duhamel, J.-R., Zalla, T., Herbrecht, E., Leboyer, M., & Sirigu, A. (2010). Promoting social behavior with oxytocin in high-functioning autism spectrum disorders. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(9), 4389-4394. https://doi.org/10.1073/pnas.0910249107

Bühler, E., Bachmann, C., Goyert, H., Heinzel-Gutenbrunner, M., & Kamp-Becker, I. (2011). Differential diagnosis of autism spectrum disorder and attention deficit hyperactivity disorder by means of inhibitory control and 'theory of mind'. *Journal of Autism and Developmental Disorders*, *41*(12), 1718-1726. https://doi.org/10.1007/s10803-011-1205-1

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bartels, A., Logothetis, N.K., & Moutoussis, K. (2008). fMRI and its interpretations: an illustration on directional selectivity in area V5/MT. *Trends in Neurosciences*, *31*(9), 444-453. https://doi.org/10.1016/j.tins.2008.06.004

Battalio, R., Samuelson, L., & Van Huyck, J. (2001). Optimization incentives and coordination failure in laboratory stag hunt games. *Econometrica*, *69*(3), 749-764. https://doi.org/10.1111/1468-0262.00212

Becchio, C., Cavallo, A., Begliomini, C., Sartori, L., Feltrin, G., & Castiello, U. (2012). Social grasping: From mirroring to mentalizing. *Neuroimage*, *61*(1), 240-248. https://doi.org/10.1016/j.neuroimage.2012.03.013

Bechara, A., Damasio, H., & Damasio, A.R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex*, *10*(3), 295-307. https://doi.org/10.1093/cercor/10.3.295

Brañas-Garza, P., & Smith, J. (2016). Cognitive abilities and economic behavior. *Journal of Behavioral and Experimental Economics*, *64*, 1-4. http://dx.doi.org/10.1016/j.socec.2016.06.005

Cauchoix, M., Barragan-Jason, G., Serre, T., & Barbeau, E.J. (2014). The neural dynamics of face detection in the wild revealed by MVPA. *Journal of Neuroscience*, *34*(3), 846-854. https://doi.org/10.1523/JNEUROSCI.3030-13.2014

Chee, M.W., Venkatraman, V., Westphal, C., & Siong, S.C. (2003). Comparison of block and event-related fMRI designs in evaluating the word-frequency effect. *Human Brain Mapping*, *18*(3), 186-193. https://doi.org/10.1002/hbm.10092

Chen, P.H., Lin, C.J., & Schölkopf, B. (2005). A tutorial on ν-support vector machines. *Applied Stochastic Models in Business and Industry*, *21*(2), 111-136. https://doi.org/10.1002/asmb.537

Christopoulos, G.I., & King-Casas, B. (2015). With you or against you: Social orientation dependent learning signals guide actions made for others. *Neuroimage*, *104*, 326-335. https://doi.org/10.1016/j.neuroimage.2014.09.011

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(23), 9163-9168. https://doi.org/10.1073/pnas.0807721106

Costa-Gomes, M.A., & Crawford, V.P. (2006). Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, *96*(5), 1737-1768. https://doi.org/10.1257/aer.96.5.1737

Davis, T., LaRocque, K.F., Mumford, J.A., Norman, K.A., Wagner, A.D., & Poldrack, R.A. (2014). What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *Neuroimage*, *97*, 271-283. https://doi.org/10.1016/j.neuroimage.2014.04.037

Decety, J., Jackson, P.L., Sommerville, J.A., Chaminade, T., & Meltzoff, A.N. (2004). The neural bases of cooperation and competition: an fMRI investigation. *Neuroimage*, *23*(2), 744-751. https://doi.org/10.1016/j.neuroimage.2004.05.025

Demb, J.B., Desmond, J.E., Wagner, A.D., Vaidya, C.J., Glover, G.H., & Gabrieli, J. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: a functional MRI study of task difficulty and process specificity. *Journal of Neuroscience*, *15*(9), 5870-5878. https://doi.org/10.1523/JNEUROSCI.15-09-05870.1995

Duffy, S., & Smith, J. (2014). Cognitive load in the multi-player prisoner's dilemma game: Are there brains in games? *Journal of Behavioral and Experimental Economics*, *51*, 47-56. https://doi.org/10.1016/j.socec.2014.01.006

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering cortical number coding from human brain activity patterns. *Current Biology*, *19*(19), 1608-1615. https://doi.org/10.1016/j.cub.2009.08.047

Emmert, K., Kopel, R., Sulzer, J., Brühl, A.B., Berman, B.D., Linden, D.E., Horovitz, S.G., Breimhorst, M., Caria, A., & Frank, S. (2016). Meta-analysis of real-time fMRI neurofeedback studies using individual participant data: How is brain regulation mediated? *Neuroimage*, *124*, 806-812. https://doi.org/10.1016/j.neuroimage.2015.09.042

Fareri, D.S., & Delgado, M.R. (2014). Differential reward responses during competition against in-and out-of-network others. *Social Cognitive and Affective Neuroscience*, *9*(4), 412-420. https://doi.org/10.1093/scan/nst006

Filzmoser, P., Liebmann, B., & Varmuza, K. (2009). Repeated double cross validation. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *23*(4), 160-171. https://doi.org/10.1002/cem.1225

Gallagher, H.L., Happé, F., Brunswick, N., Fletcher, P.C., Frith, U., & Frith, C.D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, *38*(1), 11-21. https://doi.org/10.1016/S0028-3932(99)00053-6

Gilbert, S., Bird, G., Frith, C., & Burgess, P. (2012). Does "task difficulty" explain "task-induced deactivation?". *Frontiers in Psychology*, *3*, 125. https://doi.org/10.3389/fpsyg.2012.00125

Gilron, R., Rosenblatt, J., Koyejo, O., Poldrack, R.A., & Mukamel, R. (2017). What's in a pattern? Examining the type of signal multivariate analysis uncovers at the group level. *Neuroimage*, *146*, 113-120. https://doi.org/10.1016/j.neuroimage.2016.11.019

Goodie, A.S., Doshi, P., & Young, D.L. (2012). Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making*, *25*(1), 95-108. https://doi.org/10.1002/bdm.717

Hare, T.A., Camerer, C.F., Knoepfle, D.T., O'Doherty, J.P., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience*, *30*(2), 583-590. https://doi.org/10.1523/JNEUROSCI.4089-09.2010

Haxby, J.V. (2012). Multivariate pattern analysis of fMRI: the early beginnings. *Neuroimage*, *62*(2), 852-5. https://doi.org/10.1016/j.neuroimage.2012.03.016

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R.E. (2007). Reading hidden intentions in the human brain. *Current Biology*, *17*(4), 323-328. https://doi.org/10.1016/j.cub.2006.11.072

He, B.J., Snyder, A.Z., Vincent, J.L., Epstein, A., Shulman, G.L., & Corbetta, M. (2007). Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. *Neuron*, *53*(6), 905-918. https://doi.org/10.1016/j.neuron.2007.02.013

Hebart, M., Görgen, K., & Haynes, J. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, *8*, 88. https://doi.org/10.3389/fninf.2014.00088

Hedden, T., & Zhang, J. (2002). What do you think I think you think?: Strategic reasoning in matrix games. *Cognition*, *85*(1), 1-36. https://doi.org/10.1016/S0010-0277(02)00054-9

Jakab, A., Molnár, P.P., Bogner, P., Béres, M., & Berényi, E.L. (2012). Connectivity-based parcellation reveals interhemispheric differences in the insula. *Brain Topography*, *25*(3), 264-271. https://doi.org/10.1007/s10548-011-0205-y

Jang, H., Plis, S.M., Calhoun, V.D., & Lee, J.-H. (2016). Task-specific feature extraction and classification of fMRI volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks. *Neuroimage*. https://doi.org/10.1016/j.neuroimage.2016.04.003

Joiner, J., Piva, M., Turrin, C., & Chang, S.W. (2017). Social learning through prediction error in the brain. *npj Science of Learning*, *2*(1), 8. https://doi.org/10.1038/s41539-017-0009-2

Jones, C.L., Minati, L., Harrison, N.A., Ward, J., & Critchley, H.D. (2011). Under pressure: response urgency modulates striatal and insula activity during decision-making under risk. *PloS One*, *6*(6), e20942. https://doi.org/10.1371/journal.pone.0020942

Kamitani, Y., & Sawahata, Y. (2010). Spatial smoothing hurts localization but not information: pitfalls for brain mappers. *Neuroimage*, *49*(3), 1949-52. https://doi.org/10.1016/j.neuroimage.2009.06.040

Kaplan, J.T., Man, K., & Greening, S.G. (2015). Multivariate cross-classification: applying machine learning techniques to characterize abstraction in neural representations. *Frontiers in Human Neuroscience*, *9*, 151. https://doi.org/10.3389/fnhum.2015.00151

Kawashima, R., Taira, M., Okita, K., Inoue, K., Tajima, N., Yoshida, H., Sasaki, T., Sugiura, M., Watanabe, J., & Fukuda, H. (2004). A functional MRI study of simple arithmetic—a comparison between children and adults. *Cognitive Brain Research*, *18*(3), 227-233. https://doi.org/10.1016/j.cogbrainres.2003.10.009

Kay, K.N., Naselaris, T., Prenger, R.J., & Gallant, J.L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352-355. https://doi.org/10.1038/nature06713

Kim, D.-Y., Yoo, S.-S., Tegethoff, M., Meinlschmidt, G., & Lee, J.-H. (2015). The inclusion of functional connectivity information into fMRI-based neurofeedback improves its efficacy in the reduction of cigarette cravings. *Journal of Cognitive Neuroscience*, *27*(8), 1552-1572. https://doi.org/10.1162/jocn_a_00802

Kim, H.-C., Bandettini, P.A., & Lee, J.-H. (2019a). Deep neural network predicts emotional responses of the human brain from functional magnetic resonance imaging. *Neuroimage*, *186*, 607-627. https://doi.org/10.1016/j.neuroimage.2018.10.054

Kim, H.-C., Tegethoff, M., Meinlschmidt, G., Stalujanis, E., Belardi, A., Jo, S., Lee, J., Kim, D.-Y., Yoo, S.-S., & Lee, J.-H. (2019b). Mediation analysis of triple networks revealed functional feature of mindfulness from real-time fMRI neurofeedback. *Neuroimage*, *195*, 409-432. https://doi.org/10.1016/j.neuroimage.2019.03.066

Kim, J., Calhoun, V.D., Shim, E., & Lee, J.-H. (2016). Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance:

29

Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *Neuroimage*, *124*(Pt A), 127-146. https://doi.org/10.1016/j.neuroimage.2015.05.018

Koban, L., & Pourtois, G. (2014). Brain systems underlying the affective and social monitoring of actions: an integrative review. *Neuroscience and Biobehavioral Reviews*, *46*, 71-84. https://doi.org/10.1016/j.neubiorev.2014.02.014

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863-8. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis–connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*. https://doi.org/10.3389/neuro.06.004.2008

Krill, A.L., & Platek, S.M. (2012). Working together may be better: activation of reward centers during a cooperative maze task. *PloS One*, *7*(2), e30613. https://doi.org/10.1371/journal.pone.0030613

Le Bouc, R., & Pessiglione, M. (2013). Imaging social motivation: Distinct brain mechanisms drive effort production during collaboration versus competition. *The Journal of Neuroscience*, *33*(40), 15894-15902. https://doi.org/10.1523/JNEUROSCI.0143-13.2013

Lee, J.H., Kim, J., & Yoo, S.S. (2012). Real-time fMRI-based neurofeedback reinforces causality of attention networks. *Neuroscience Research*, *72*(4), 347-354. https://doi.org/10.1016/j.neures.2012.01.002

Li, X., Lu, Z.-L., Tjan, B.S., Dosher, B.A., & Chu, W. (2008). Blood oxygenation level-dependent contrast response functions identify mechanisms of covert attention in early visual areas. *Proceedings of the national Academy of Sciences*, *105*(16), 6202-6207. https://doi.org/10.1073/pnas.0801390105

Liang, M., Mouraux, A., Hu, L., & Iannetti, G. (2013). Primary sensory cortices contain distinguishable spatial patterns of activity for each sense. *Nature communications*, *4*, 1979. https://doi.org/10.1038/ncomms2979

Lin, A., Adolphs, R., & Rangel, A. (2011). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience*, *7*(3), 274-281. https://doi.org/10.1093/scan/nsr006

Lindquist, M.A., Krishnan, A., López-Solà, M., Jepma, M., Woo, C.-W., Koban, L., Roy, M., Atlas, L.Y., Schmidt, L., & Chang, L.J. (2017). Group-regularized individual prediction: theory and application to pain. *Neuroimage*, *145*, 274-287. https://doi.org/10.1016/j.neuroimage.2015.10.074

Logothetis, N.K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *Journal of Neuroscience*, *23*(10), 3963-3971. https://doi.org/10.1523/JNEUROSCI.23-10-03963.2003

Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., & Brovelli, A. (2012). Multivoxel pattern analysis for FMRI data: a review. *Computational and Mathematical Methods in Medicine*, *2012*. https://doi.org/10.1155/2012/961257

Marjoram, D., Job, D.E., Whalley, H.C., Gountouna, V.-E., McIntosh, A.M., Simonotto, E., Cunningham-Owens, D., Johnstone, E.C., & Lawrie, S. (2006). A visual joke fMRI investigation into Theory of Mind and enhanced risk of schizophrenia. *Neuroimage*, *31*(4), 1850-1858. https://doi.org/10.1016/j.neuroimage.2006.02.011

Mars, R.B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M.F. (2012). Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks. *Cerebral Cortex*, *22*(8), 1894-1903. https://doi.org/10.1093/cercor/bhr268

Maysov, A., & Kipyatkov, V.E. (2011). Contrasting dynamics of cold resistance traits in field-fresh Myrmica ants during the active season. *Journal of Thermal Biology*, *36*(1), 64-70. https://doi.org/10.1016/j.jtherbio.2010.11.004

McCleery, J.P., Surtees, A.D., Graham, K.A., Richards, J.E., & Apperly, I.A. (2011). The neural and cognitive time course of theory of mind. *The Journal of Neuroscience*, *31*(36), 12849-54. https://doi.org/10.1523/jneurosci.1392-11.2011

Mechelli, A., Henson, R.N., Price, C.J., & Friston, K.J. (2003). Comparing event-related and epoch analysis in blocked design fMRI. *Neuroimage*, *18*(3), 806-810. https://doi.org/10.1016/S1053-8119(02)00027-7

Mier, D., Lis, S., Esslinger, C., Sauer, C., Hagenhoff, M., Ulferts, J., Gallhofer, B., & Kirsch, P. (2013). Neuronal correlates of social cognition in borderline personality disorder. *Social Cognitive and Affective Neuroscience*, *8*(5), 531-537. https://doi.org/10.1093/scan/nss028

Misaki, M., Kim, Y., Bandettini, P.A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage*, *53*(1), 103-118. https://doi.org/10.1016/j.neuroimage.2010.05.051

Montague, P.R., Berns, G.S., Cohen, J.D., McClure, S.M., Pagnoni, G., Dhamala, M., Wiest, M.C., Karpov, I., King, R.D., & Apple, N. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *Neuroimage*, *16*(4), 1159-1164. https://doi.org/10.1006/nimg.2002.1150

Mundfrom, D.J., Perrett, J.J., Schaffer, J., Piccone, A., & Roozeboom, M. (2006). Bonferroni adjustments in tests for regression coefficients. *Multiple Linear Regression Viewpoints*, *32*, 1-6.

Mur, M., Bandettini, P.A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101-109. https://doi.org/10.1093/scan/nsn044

Nachev, P., Wydell, H., O'Neill, K., Husain, M., & Kennard, C. (2007). The role of the pre-supplementary motor area in the control of action. *Neuroimage*, *36*, T155-T163. https://doi.org/10.1016/j.neuroimage.2007.03.034

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, *85*(5), 1313-1326.

Petersen, S.E., & Dubis, J.W. (2012). The mixed block/event-related design. *Neuroimage*, *62*(2), 1177-1184. https://doi.org/10.1016/j.neuroimage.2011.09.084

Polonio, L., Di Guida, S., & Coricelli, G. (2015). Strategic sophistication and attention in games: An eye-tracking study. *Games and Economic Behavior*, *94*, 80-96. https://doi.org/10.1016/j.geb.2015.09.003

Ramsøy, T.Z., Skov, M., Macoveanu, J., Siebner, H.R., & Fosgaard, T.R. (2015). Empathy as a neuropsychological heuristic in social decision-making. *Social Neuroscience*, *10*(2), 179-191. https://doi.org/10.1080/17470919.2014.965341

Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., & Cohen, J.D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage*, *22*(4), 1694-703. https://doi.org/10.1016/j.neuroimage.2004.04.015

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235-239. https://doi.org/10.1016/j.conb.2006.03.001

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage*, *19*(4), 1835-42. https://doi.org/10.1016/S1053-8119(03)00230-1

Skyrms, B. (2004) The stag hunt and the evolution of social structure. Cambridge University Press.

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage*, *65*, 69-82. https://doi.org/10.1016/j.neuroimage.2012.09.063

Stone, V.E., Baron-Cohen, S., & Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*(5), 640-656. https://doi.org/10.1162/089892998562942

Swisher, J.D., Gatenby, J.C., Gore, J.C., Wolfe, B.A., Moon, C.-H., Kim, S.-G., & Tong, F. (2010). Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *Journal of Neuroscience*, *30*(1), 325-330. https://doi.org/10.1523/JNEUROSCI.4811-09.2010

Tie, Y., Suarez, R.O., Whalen, S., Radmanesh, A., Norton, I.H., & Golby, A.J. (2009). Comparison of blocked and event-related fMRI designs for pre-surgical language mapping. *Neuroimage*, *47*, T107-T115. https://doi.org/10.1016/j.neuroimage.2008.11.020

Todd, M.T., Nystrom, L.E., & Cohen, J.D. (2013). Confounds in multivariate pattern analysis: theory and rule representation case study. *Neuroimage*, *77*, 157-165. https://doi.org/10.1016/j.neuroimage.2013.03.039

Ungerleider, S.K., & G, L. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*(1), 315-341. https://doi.org/10.1146/annurev.neuro.23.1.315

van der Meer, L., Groenewold, N.A., Nolen, W.A., Pijnenborg, M., & Aleman, A. (2011). Inhibit yourself and understand the other: neural basis of distinct processes underlying Theory of Mind. *Neuroimage*, *56*(4), 2364-74. https://doi.org/10.1016/j.neuroimage.2011.03.053

Varoquaux, G., Raamana, P.R., Engemann, D.A., Hoyos-Idrobo, A., Schwartz, Y., & Thirion, B. (2017). Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *Neuroimage*, *145*, 166-179. https://doi.org/10.1016/j.neuroimage.2016.10.038

Wang, J., Baucom, L.B., & Shinkareva, S.V. (2013). Decoding abstract and concrete concept representations based on single-trial fMRI data. *Human Brain Mapping*, *34*(5), 1133-1147. https://doi.org/10.1002/hbm.21498

Webb, T.W., Igelström, K.M., Schurger, A., & Graziano, M.S. (2016). Cortical networks involved in visual awareness independent of visual attention. *Proceedings of the National Academy of Sciences*, *113*(48), 13923-13928. https://doi.org/10.1073/pnas.1611505113

Wisniewski, D. (2018). Context-dependence and context-invariance in the neural coding of intentional action. *Frontiers in Psychology*, *9*, 2310. https://doi.org/10.3389/fpsyg.2018.02310

Wisniewski, D., Goschke, T., & Haynes, J.-D. (2016). Similar coding of freely chosen and externally cued intentions in a fronto-parietal network. *Neuroimage*, *134*, 450-458. https://doi.org/10.1016/j.neuroimage.2016.04.044

Xiang, T., Ray, D., Lohrenz, T., Dayan, P., & Montague, P.R. (2012). Computational phenotyping of two-person interactions reveals differential neural response to depth-of-thought. *PLoS Computational Biology*, *8*(12), e1002841. https://doi.org/10.1371/journal.pcbi.1002841

Yoshida, W., Seymour, B., Friston, K.J., & Dolan, R.J. (2010). Neural Mechanisms of Belief Inference during Cooperative Games. *The Journal of Neuroscience*, *30*(32), 10744-10751. https://doi.org/10.1523/jneurosci.5895-09.2010

Zaki, J., Kallman, S., Wimmer, G.E., Ochsner, K., & Shohamy, D. (2016). Social cognition as reinforcement learning: feedback modulates emotion inference. *Journal of Cognitive Neuroscience*, *28*(9), 1270-1282. https://doi.org/10.1162/jocn_a_00978

Zhang, J., Hedden, T., & Chia, A. (2012). Perspective-Taking and Depth of Theory-of-Mind Reasoning in Sequential-Move Games. *Cognitive Science*, *36*(3), 560-573. https://doi.org/10.1111/j.1551-6709.2012.01238.x

Zhou, X., Chen, C., Dong, Q., Zhang, H., Zhou, R., Zhao, H., Chen, C., Qiao, S., Jiang, T., & Guo, Y. (2006). Event-related potentials of single-digit addition, subtraction, and multiplication.

32

Tables

Table 1. Subjective scores from participants.

Subject number		#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	Mean± SD	Statistic
Subjective scores	**CP**	n/a	n/a	4	4	4	5	5	3	3	2	5	3	4	2	3	4	3	5	5	4	5	3	5	4	3.86 ± 0.99	Kendall's $\tau = 0.47$ ($p = 0.006$)
	CB	n/a	n/a	4	5	3	4	5	2	3	4	4	2	5	2	3	3	3	5	5	5	5	5	5	4	3.91 ± 1.11	
Task difficulty	**CP**	4	4	4	4	4	3	3	4	4	3	4	2	4	1	3	3	5	3	5	5	2	3	4	4	3.33 ± 0.98	$F(2,71) = 26.83$ (Bonferroni corrected $^{*}p=1.65\times10^{-7}$)
	CB	2	2	1	1	2	2	2	1	2	2	2	2	2	1	3	1	2	4	2	1	2	1	2	1	1.67 ± 0.49	
	CTR	2	3	2	3	3	2	2	2	2	2	3	2	3	2	4	3	3	2	5	3	2	3	3	3	2.33 ± 0.49	

SD, standard deviation; n/a, not available; CP, competition; CB, collaboration; CTR, control; [*], one-way repeated-measures

ANOVA test.

Figure Legends

Figure 1. The experimental paradigm used for the competition and collaboration trials with two reasoning levels. The move from cell A to B was determined by the participant, the move from cell B to C by a staff member (initials OO), and the move from cell C to D was the participant's decision if the staff member decided to move from cell B to C. The first and second rewards in each set of parentheses are the payoffs for the participant and staff member, respectively. Participants answered two questions on (a) the move of the staff member (Q1) and (b) their own move (Q2) with either "Yes" or "No." The participant then observed how the game proceeded based on his/her decision and the staff member's decision. The two players received their payoffs according to the cell where the game ended. See the "**2.2. Sequential-move games involving competition and collaboration with two levels of reasoning**" subsection of the Methods for details.

Figure 2. Procedure for the within-subject searchlight-based multivoxel pattern analysis (MVPA) classification. Once the functional magnetic resonance imaging (fMRI) data were preprocessed, the neural activation of the voxels was estimated from the general linear model across the whole brain. The multivoxel activation patterns were prepared for each trial in competition (CP) and collaboration (CB) modes and the lower and higher reasoning levels of myopic (MS) and predictive (PS) staff members, respectively, to conduct each binary classification and the four-class classification. Using a five-fold nested cross-validation (CV) framework, the optimal parameter for the support vector machine (SVM) was chosen based on the maximum validation accuracy, and the trained SVM was employed to determine test accuracy using the one remaining fold. See the "**2.6. Classification of competition versus collaboration and myopic versus predictive reasoning**" subsection of the Methods for details.

Figure 3. Average (bar) and standard deviation (whisker) of the response time across all subjects (*n* =

35

24) in the game trials with collaboration (CB) or competition (CP) and a myopic staff member using lower-level reasoning or a predictive staff member using higher-level reasoning. In our study, a computerized agent was employed as the staff member. The *p*-values from a paired *t*-test were Bonferroni-corrected by dividing it by the total number of subjects.

Figure 4. Brain regions identified in the classification of competition (CP) and collaboration (CB). (a) Brain maps representing the spatial patterns from group inference (Bonferroni-corrected *p*-values < 10^{-3} from 10,000 random permutations) obtained by applying one-sample *t*-tests to the accuracy maps from individual classification tests (accuracy > 70% with 20 contiguous voxels) across all 24 subjects. (First row) Classification of CP vs. CB across the two reasoning levels. (Second row) Classification of CP vs. CB from the game trials with MS. (Third row) Classification of CP vs. CB from the game trials with PS. (See Table S3 for details). (b) Overlapping patterns for the univariate approach and MVPA. (c) Brain maps representing the spatial patterns from group inference (Bonferroni-corrected *p*-values < 0.05 from 10,000 random permutations) obtained by applying one-sample *t*-tests to the beta maps from individual analysis across all 24 subjects. (First row) Contrast of CP vs. CB across the two reasoning levels. (Second row) Contrast of CP vs. CB from the game trials with MS. (Third row) Contrast of CP vs. CB from the game trials with PS. (See Table S5 for details). See the "**3.2. Brain regions identified from competition versus collaboration classification**" and "**3.4. Comparison between the univariate approach and MVPA**" subsections of the Results for details. MS, myopic staff; PS, predictive staff; MOG, middle occipital gyrus; IOG, inferior occipital gyrus; IPL, inferior parietal lobule; SPL, superior parietal lobule; aINS, anterior insula; SFG, superior frontal gyrus; PrCG, precentral gyrus; PCC, posterior cingulate cortex; SOG, superior occipital gyrus; pINS, posterior insula; SMA, supplementary motor area; IFG, inferior frontal gyrus.

36

Figure 5. Brain regions identified by a myopic (MS) versus a predictive (PS) staff member. (a) Brain maps representing the spatial patterns from group inference (Bonferroni-corrected p-values $< 10^{-3}$ from 10,000 random permutations) obtained by applying one-sample t-tests to the accuracy maps from individual classification tests (accuracy $> 70\%$ with 20 contiguous voxels) across all 24 subjects. (First row) Classification of MS vs. PS across the two decision-making processes. (Second row) Classification of MS vs. PS from CP. (Third row) Classification of MS vs. PS from CB. (See Table S4 for details) (b) Overlapping patterns for the univariate approach and MVPA. (c) Brain maps representing the spatial patterns from group inference (Bonferroni-corrected p-values < 0.05 from 10,000 random permutations) obtained by applying one-sample t-tests to the beta maps from individual analysis across all 24 subjects. (First row) Contrast of MS vs. PS across the two decision-making processes. (Second row) Contrast of MS vs. PS from CP. (Third row) Contrast of MS vs. PS from CB. (See Table S6 for details). See the "**3.3. Brain regions identified from myopic versus predictive reasoning order classification**" and "**3.4. Comparison between the univariate approach and MVPA**" subsections of the Results for details. CP, competition; CB, collaboration; MOG, middle occipital gyrus; SPL, superior parietal lobule; IPL, inferior parietal lobule; PrCG, precentral gyrus; MFG, middle frontal gyrus; mSFG, medial-superior frontal gyrus; SMA, supplementary motor area; IFG, inferior frontal gyrus; PoCG, postcentral gyrus.

Figure 6. Brain regions identified by regression analysis for the competition (CP) versus collaboration (CB) from MVPA. (a) Brain maps representing the spatial patterns (Bonferroni-corrected p-values $<$ 0.05 from 10,000 random permutations) obtained by regression analysis of subjective scores. (First row) CP vs. CB across the two reasoning orders. (Second row) CP vs. CB from the game trials with MS. (Third row) CP vs. CB from the game trials with PS. (See Table S7 for details). (b) Spatial patterns related to visual attention components during the regression analysis of response time. See the "**3.2.**

37

Brain regions identified from competition versus collaboration classification" subsection of the Results for details. MS, myopic staff; PS, predictive staff; MOG, middle occipital gyrus; aINS, anterior insula; IOG, inferior occipital gyrus; PrCG, precentral gyrus; SPL, superior parietal lobule; SFG, superior frontal gyrus; IPL, inferior parietal lobule; PCC, posterior cingulate cortex; pINS, posterior insula; SMA, supplementary motor area.

Figure 7. Brain regions identified by regression analysis for a myopic (MS) versus a predictive (PS) staff member from MVPA. (a) Brain maps representing the spatial patterns (Bonferroni-corrected *p*-values < 0.05 from 10,000 random permutations) obtained by regression analysis of reasoning time. (First row) MS vs. PS across the two decision-making processes. (Second row) MS vs. PS from CP. (Third row) MS vs. PS from CB. (See Table S8 for more details). (b) Spatial patterns related to visual attention components during the regression analysis of response time. See the "**3.3. Brain regions identified from myopic versus predictive reasoning order classification**" subsection of the Results for details. CP, competition; CB, collaboration; MOG, middle occipital gyrus; IPL, inferior parietal lobule; mSFG, medial-superior frontal gyrus; SPL, superior parietal lobule; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; PoCG, postcentral gyrus; SMA, supplementary motor area.

Figure 8. Brain regions identified by regression analysis for competition (CP) versus collaboration (CB) from the univariate approach. (a) Brain maps representing the spatial patterns (Bonferroni-corrected *p*-values < 0.05 from 10,000 random permutations) obtained by regression analysis of subjective scores. (First row) CP vs. CB across the two reasoning orders. (Second row) CP vs. CB from the game trials with MS. (Third row) CP vs. CB from the game trials with PS. (See Table S9 for more details) (b) Results related to visual attention during the regression analysis of response time. See the "**3.4. Comparison between the univariate approach and MVPA**" subsection of the Results for

38

details. MS, myopic staff; PS, predictive staff; mOFC, medial orbitofrontal cortex; IFG, inferior frontal gyrus; SFG, superior frontal gyrus; SPL, superior parietal lobule; SMA, supplementary motor area; PrCN, precuneus; PoCG, postcentral gyrus; PrCG, precentral gyrus; IPL, inferior parietal lobule; aINS, anterior insula; THL, thalamus.
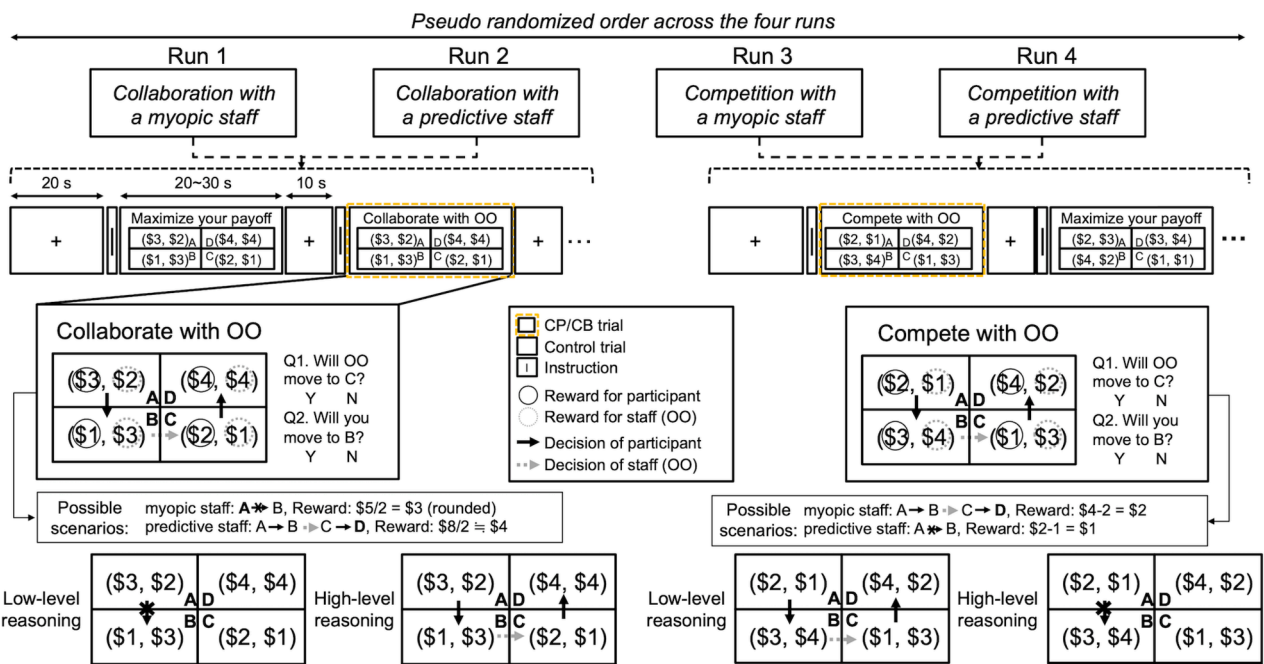
Figure 9. Brain regions identified by regression analysis for a myopic (MS) versus a predictive (PS) staff member from the univariate approach. (a) Brain maps representing the spatial patterns (Bonferroni-corrected p-values < 0.05 from 10,000 random permutations) obtained by regression analysis of reasoning time. (First row) MS vs. PS across the two decision-making processes. (Second row) MS vs. PS from CP. (Third row) MS vs. PS from CB. (See Table S10 for more details) (b) Results related to visual attention during the regression analysis of response time. See the "**3.4. Comparison between the univariate approach and MVPA**" subsection of the Results for details. CP, competition; CB, collaboration; MOG, middle occipital gyrus; PrCG, precentral gyrus; MCC, middle cingulate cortex; SPL, superior parietal lobule; IPL, inferior parietal lobule; THL, thalamus; PoCG, postcentral gyrus; SOG, superior occipital gyrus.

Figure 10. Brain regions identified by four-class classification. (a) Brain maps representing the spatial patterns from group-inference (Bonferroni-corrected p-values < 0.05 from 10,000 random permutations) obtained by applying one-sample t-tests to the accuracy maps from individual classification tests for four-class classification (accuracy > 50% with 20 contiguous voxels) across all 24 subjects. (See Table S11 for details). (b) Histogram of voxel-wise average classification accuracy within a cluster across subjects for each of the four classes. (c) Histogram of voxel-wise beta values within a cluster across subjects for each of the four classes. (d) The results from the regression using voxel-wise individual classification accuracy and behavioral information. See the "**3.5. Classification**
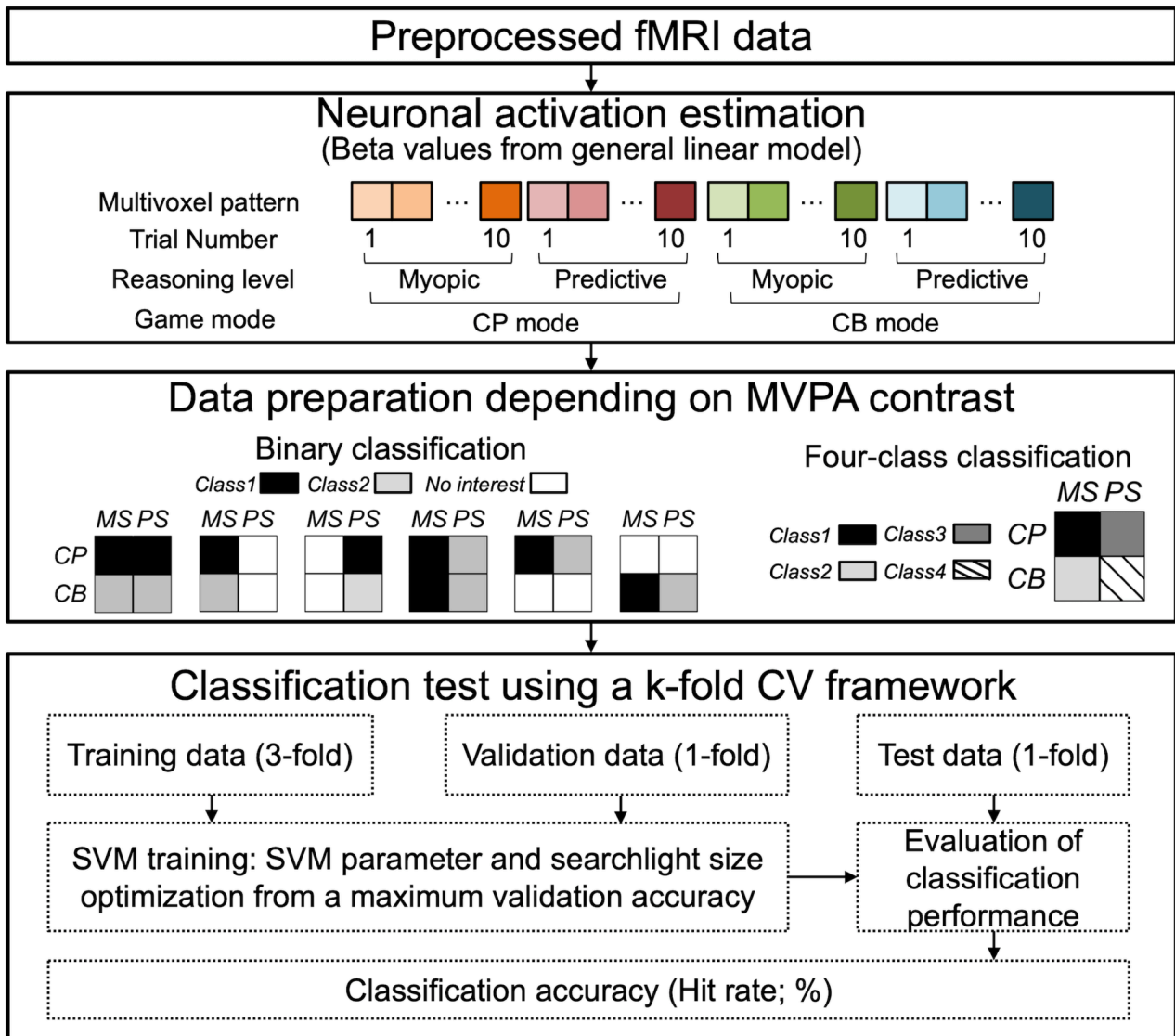
39

across all four conditions" subsection of the Results for details. CP, competition; CB, collaboration; MS, myopic staff; PS, predictive staff; IPL, inferior parietal lobule; SPL, superior parietal lobule; SMA, supplementary motor area; MFG, middle frontal gyrus; IFG, inferior frontal gyrus.
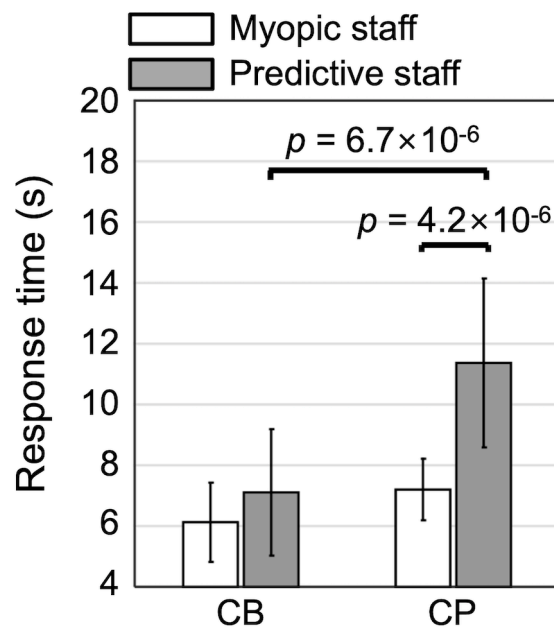
Figure 11. Brain regions identified from cross-set classification. (a) Classification of CP vs. CB with mixed training/test sets. (b) Classification of CP vs. CB with a training set from the game trials with MS and a test set from the game trials with PS. (c) Classification of CP vs. CB with a training set from the game trials with PS and a test set from the game trials with MS. (d) Classification of MS vs. PS with mixed training/test sets. (e) Classification of MS vs. PS with a training set from the game trials with CP and a test set from the game trials with CB. (f) Classification of MS vs. PS with a training set from the game trials with CB and a test set from the game trials with CP. See the "**3.7. Cross-set classification**" subsection of the Results for details. CP, competition; CB, collaboration; MS, myopic staff; PS, predictive staff; MOG, middle occipital gyrus; IOG, inferior occipital gyrus; aINS, anterior insula; PrCG, precentral gyrus; IPL, inferior parietal lobule; SPL, superior parietal lobule; SMA, supplementary motor area; MFG, middle frontal gyrus; SFG, superior frontal gyrus; mSFG, medial-superior frontal gyrus.
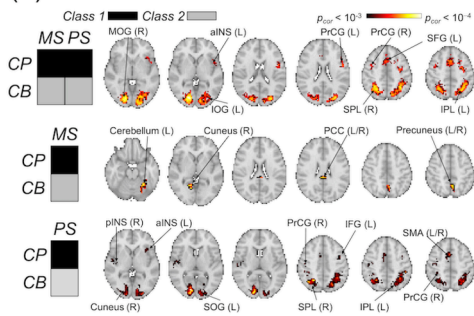
40

HBM_25127_fig1_overall_exp.tiff

Preprocessed fMRI data

Neuronal activation estimation
(Beta values from general linear model)

Multivoxel pattern
Trial Number
Reasoning level
Game mode

Data preparation depending on MVPA contrast

Binary classification

Class1 ■ Class2 ▨ No interest □

Four-class classification

Class1 ■ Class3 ▨ CP
Class2 □ Class4 ▨ CB

Classification test using a k-fold CV framework

Training data (3-fold) Validation data (1-fold) Test data (1-fold)

SVM training: SVM parameter and searchlight size optimization from a maximum validation accuracy

Evaluation of classification performance

Classification accuracy (Hit rate; %)

HBM_25127_fig2_overall_anal.tiff

Myopic staff
Predictive staff

$p = 6.7 \times 10^{-6}$

$p = 4.2 \times 10^{-6}$

Response time (s)

CB   CP

HBM_25127_fig3_rt.tiff

(a) MVPA of CP vs. CB (b) Overlap (c) Univariate approach
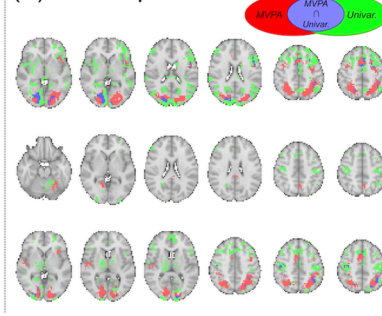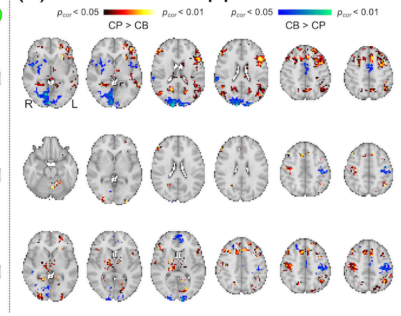
HBM_25127_fig4_cpcb_mvpa_unitv.tiff

(a) MVPA of MS vs. PS   (b) Overlap   (c) Univariate approach
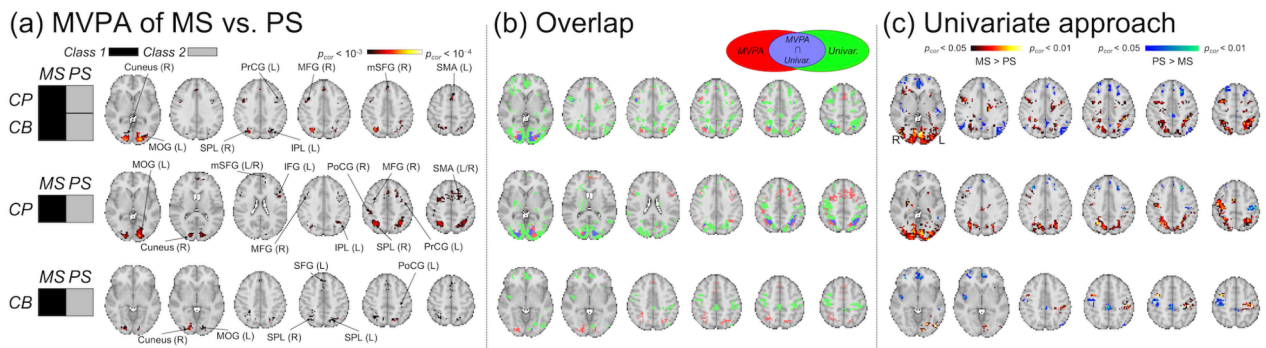
HBM_25127_fig5_msps_mvpa_unitv.tiff

MVPA of CP vs. CB

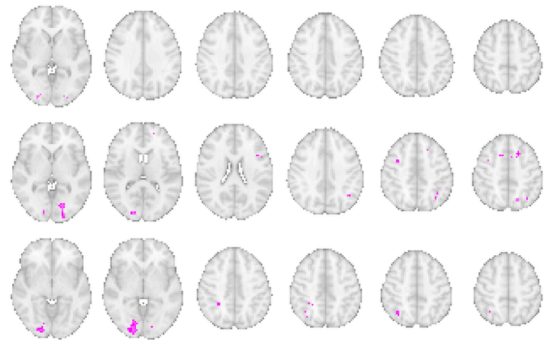### (a) Decision-making process component
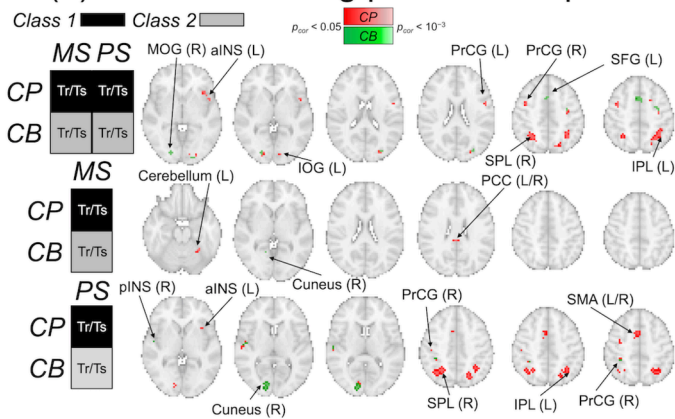
(b) Visual attention component
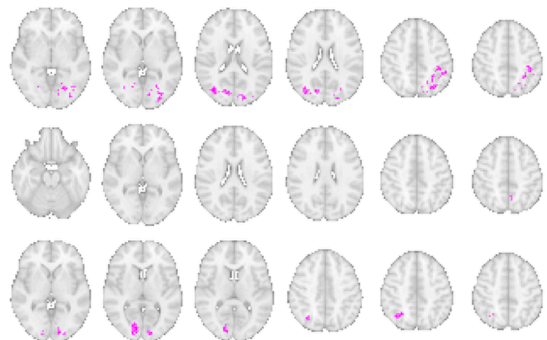
HBM_25127_fig6_cpcb_mvpa_visuatt.tiff

MVPA of MS vs. PS

(a) Reasoning process component
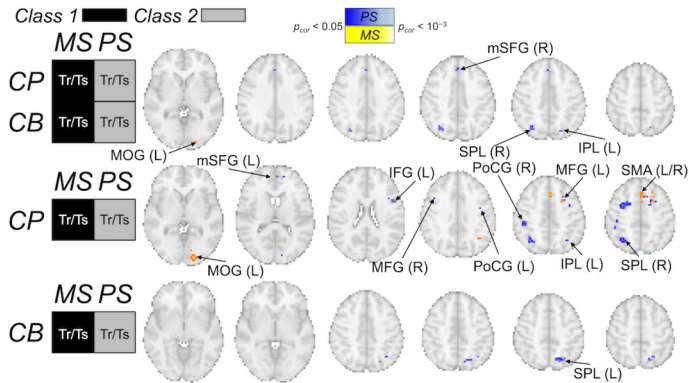
(b) Visual attention component

HBM_25127_fig7_msps_mvpa_visuatt.tiff

Univariate approach of CP vs. CB

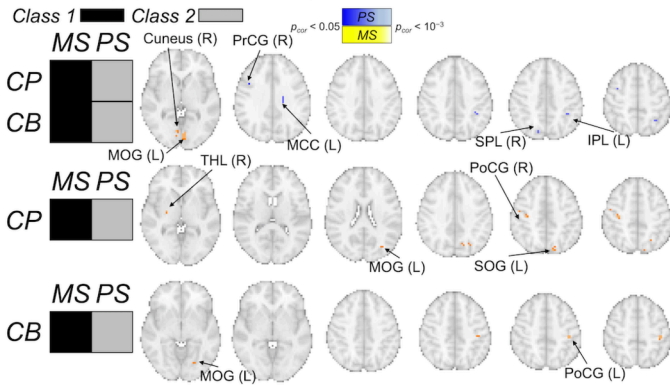(a) Decision-making process component
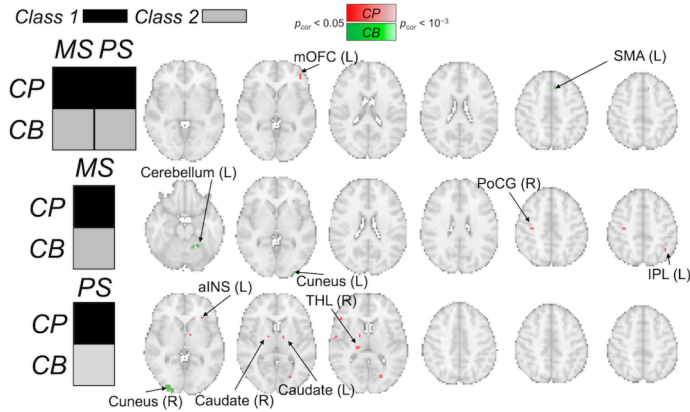
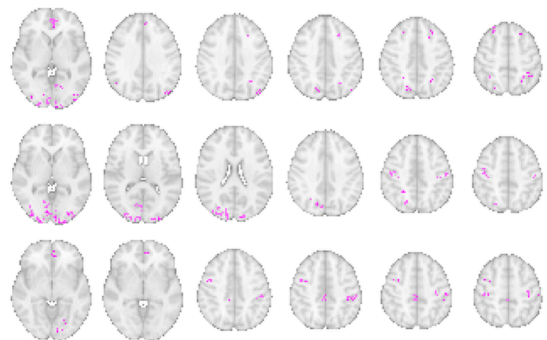(b) Visual attention component
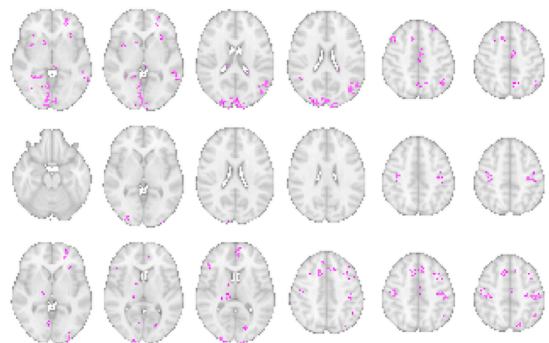
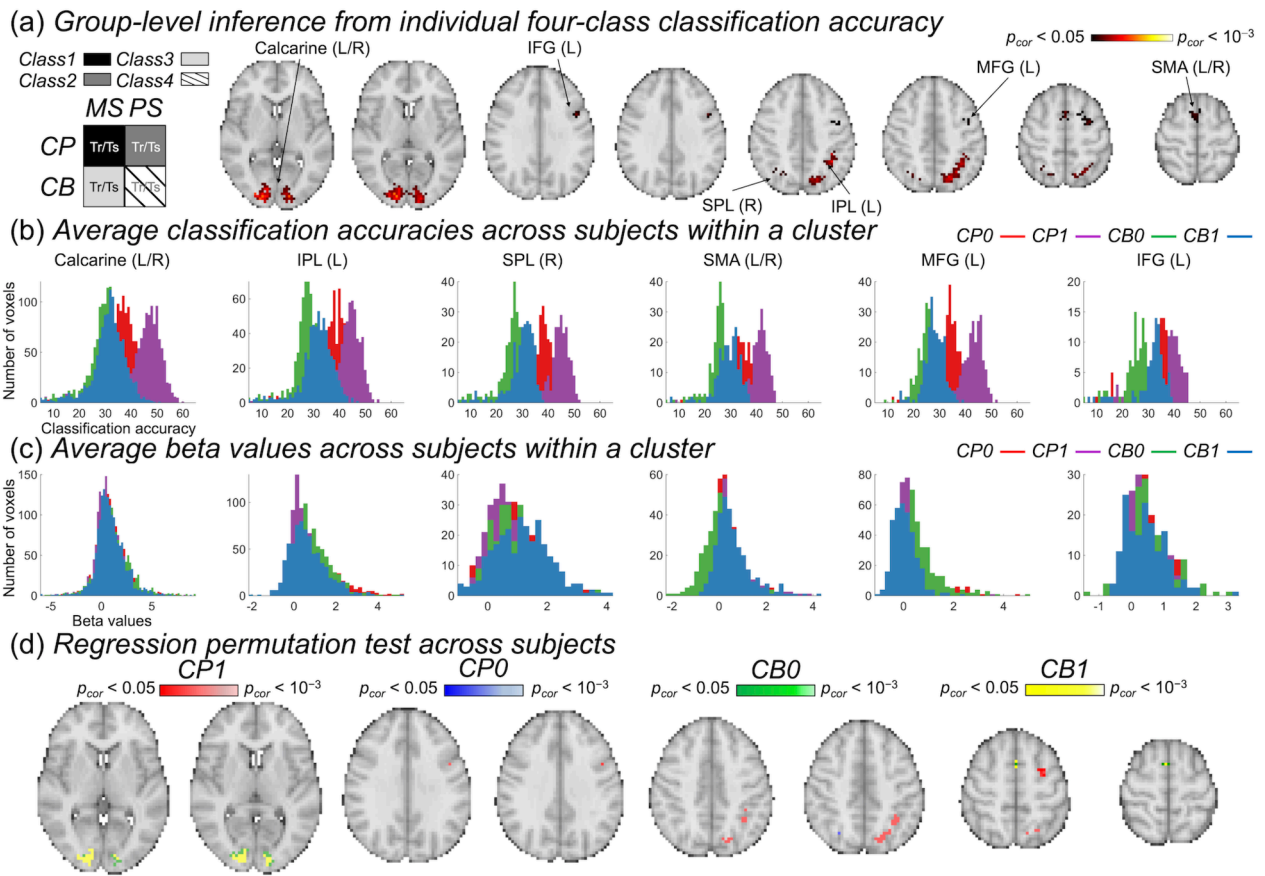HBM_25127_fig8_cpcb_unitv_visuatt.tiff

Univariate approach of MS vs. PS

(a) Reasoning process component | (b) Visual attention component

HBM_25127_fig9_msps_unitv_visuatt.tiff

(a) *Group-level inference from individual four-class classification accuracy*

(b) *Average classification accuracies across subjects within a cluster*

(c) *Average beta values across subjects within a cluster*

(d) *Regression permutation test across subjects*

HBM_25127_fig10_cpcb_msps.tiff
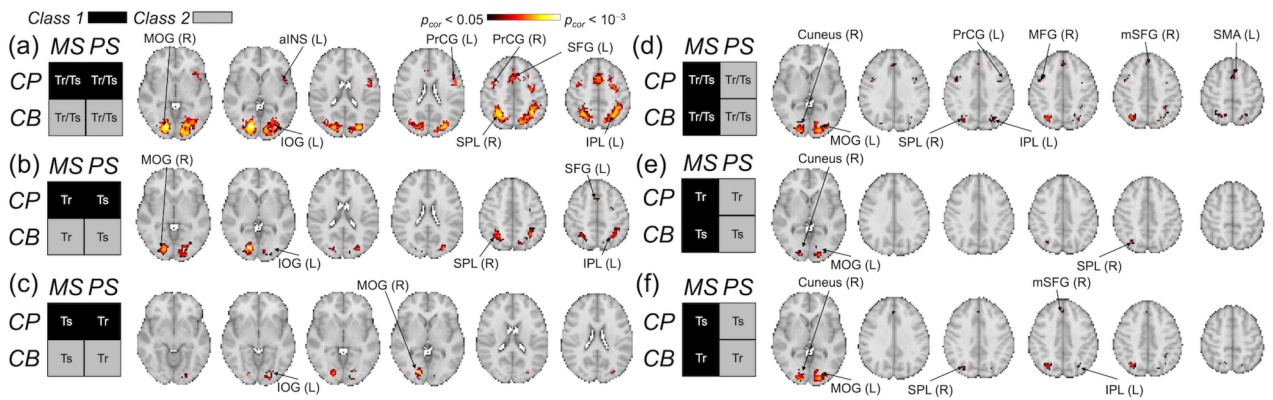
HBM_25127_fig11_cross_set_classification.tiff