doi:10.1093/scan/nsv006 SCAN (2015) 10, 1210–1218

Decoding covert motivations of free riding and cooperation from multi-feature pattern analysis of EEG signals

Dongil Chung,[1,2] Kyongsik Yun,[1] and Jaeseung Jeong[1]

[1]Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, South Korea and [2]Virginia Tech Carilion Research Institute, Roanoke, VA 24016, USA

Cooperation and free riding are among the most frequently observed behaviors in human social decision-making. In social interactions, the effects of strategic decision processes have been consistently reported in iterative cooperation decisions. However, the neural activity immediately after new information is presented, the time at which strategy learning potentially starts has not yet been investigated with high temporal resolution. Here, we implemented an iterative, binary public goods game that simulates cooperation/free riding behavior. We applied the multi-feature pattern analysis method by using a support vector machine and the unique combinatorial performance measure, and identified neural features from the single-trial, event-related spectral perturbation at the result-presentation of the current round that predict participants' decisions to cooperate or free ride in the subsequent round. We found that neural oscillations in centroparietal and temporal regions showed the highest predictive power through 10-fold cross-validation; these predicted the participants' next decisions, which were independent of the neural responses during their own preceding choices. We suggest that the spatial distribution and time–frequency information of the selected features represent covert motivations to free ride or cooperate in the next round and are separately processed in parallel with information regarding the preceding results.

INTRODUCTION

Cooperation and free riding are among the most frequently observed behaviors in human social decision-making (Isaac *et al.*, 1984; Andreoni, 1988; Ledyard, 1995; Fehr and Gachter, 2000; Camerer, 2003; Fehr and Fischbacher, 2003; Gächter *et al.*, 2010). Economic and psychological studies span various aspects of cooperation, including the mechanisms of evolution/cascades of cooperation (Nowak, 2006; Santos *et al.*, 2008; Fowler and Christakis, 2010; Perc and Szolnoki, 2010) and the effects of voluntary participation (Hauert *et al.*, 2002, 2007), costly punishments/rewards (O'Gorman *et al.*, 2009; Ule *et al.*, 2009; Boyd *et al.*, 2010; Janssen *et al.*, 2010; Sasaki *et al.*, 2012), and institutional designs (Palfrey and Rosenthal, 1991; Krajbich *et al.*, 2009). Despite a large body of evidence for cross-societal differences (Wu *et al.*, 2009; Gächter *et al.*, 2010), it has been consistently observed that strategic decision processes play a key role in repetitive decisions of cooperation or free riding (Andreoni, 1988, 1995; Camerer, 2003; Chung *et al.*, 2011a,b; Suzuki *et al.*, 2011). In other words, the outcomes of current decisions influence subsequent decisions via a decision mechanism instantiating a multi-round strategic algorithm. Recent neuroimaging studies have provided a great deal of information on the brain regions related to cooperation and defection (Rilling *et al.*, 2002; Frith and Singer, 2008; Rilling *et al.*, 2008; Baumgartner *et al.*, 2011). These regions are mainly responsible for social learning processes (King-Casas *et al.*, 2005; Behrens *et al.*, 2008; Hampton *et al.*, 2008; Ho *et al.*, 2008; Zhu *et al.*, 2012). However, the neural activity immediately after new information is presented, the time at which updating (or strategy learning) presumably starts has not yet been investigated with high temporal resolution.

Electroencephalograms (EEGs) are one of the most commonly used non-invasive neurophysiological methods in decision-making studies (Camerer, 2007; Mulert *et al.*, 2008; De Vico Fallani *et al.*, 2010; Polezzi *et al.*, 2010). Compared with functional magnetic resonance imaging (fMRI), EEG recordings can capture rich temporal dynamics during cognitive processes with high temporal resolution. Here, we utilized the EEG to investigate covert motivations underlying free riding and cooperation. In this study, we hypothesized that result presentation in an iterative decision sequence would not only induce cognitive or affective responses but also initiate strategy updates. The multivariate pattern analysis (MVPA) method with a support vector machine (SVM, a linear classifier) was used to identify neurophysiological markers of free riding and cooperation that reflected strategic decisions from the result presentation of the preceding decision round. There are several benefits of adopting these often-used methods in EEG-based brain–computer interface (BCI) studies (Wang *et al.*, 2004; Norman *et al.*, 2006; Lotte *et al.*, 2007; Schulz *et al.*, 2012). First, MVPA methods enhance the sensitivity to a particular mental state by using a pattern classification approach to multi-dimensional data (Norman *et al.*, 2006). This method can assist interpretation of time–frequency spectral signals during complex decision-making as cooccurring patterns. Taking advantage of this, we sought neural features from single-trial event-related spectral perturbation (ERSP) patterns (see Makeig *et al.*, 2004 for review), which reflect both rich temporal information and neural synchronization/desynchronization during decisions. Through the feature selection method, we were able to extract the sets of features (multi-feature combinations) that represented the neural signals most relevant to the decision to free ride or cooperate in the next round. Second, SVM determines which dimension of information vector is more valuable. Thus, projecting high-dimensional data to binary behavioral decision requires fewer a priori assumptions. We used the simplest linear kernel SVM to preserve the original relationships between variables (Schulz *et al.*, 2012), particularly when we combine two features for prediction [see unique combinatory performance (UCP) described in Materials and Methods section]. Third, we were able to perform *post hoc* confirmation on the neural data

Received 17 September 2013; Revised 11 August 2014; Accepted 9 February 2015
Advance Access publication 16 February 2015

The authors thank James Fowler (UCSD), Pearl Chiu (VTCRI) and Jacob Lee (VTCRI) for their valuable comments on our manuscript. The funders had no role in study design, data collection and analysis, the decision to publish, or the preparation of the manuscript. This work was supported by the CHUNG Moon Soul Research Center for Bio Information and Bio Electronics (CMSC) in KAIST and a Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MOST) (No. R01-2007-000-21094-0 and No. M10644000028-06N4400-02810; No. 20090093897 and No. 20090083561).

Correspondence should be addressed to Jaeseung Jeong, Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), 335 Gwahangno, Yuseong-gu, Daejeon 305-701, South Korea. E-mail: jsjeong@kaist.ac.kr.

through conventional statistics. The prediction accuracies obtained from the selected set of features validate the explanatory power of the correlated neural patterns on behavioral decisions.

This study searched for neural predictors of free riding and cooperation using EEG recordings. To test our hypothesis on strategy updates at the result presentation period, we used an iterative, binary, public goods game (PGG) that simulates simple social interactions. Three different conditions (one standard condition and two additional conditions with modulated incentives) were used to induce two main motivations (fear and greed) to free ride and independently examine their effects on cooperative behaviors (Dawes *et al.*, 1986; Chung *et al.*, 2011a). We analyzed EEGs that were time-locked to the result presentation of each round to search for neural predictors of subsequent free riding or cooperation decisions.

MATERIALS AND METHODS

Ethics statement

The experimental protocol and consent forms were reviewed and approved by the local institutional review board (IRB) at KAIST (KH2008-01).

Subjects

Web advertisements were used to recruit 65 healthy male subjects (age: 19–27; mean = 22.40 ± 1.97) from a local university (Choongnam National University, Daejeon, South Korea). All participants were right-handed and had no history of neurological or psychiatric disorders. Written informed consent was obtained from all participants after describing the experimental procedure. Participants were recruited in groups of five, and two participants from each group were randomly selected for EEG recordings during the task ($N = 26$, mean age = 22.27± 1.78). The other participants were prepared with the same procedure, but their EEGs were not recorded (sham recording). The participants were not informed whether they were in a real or sham EEG recording. All participants assigned to a group were seated face-to-face to enable group decision-making (with social interaction). However, no conversation between the participants was allowed during the game.

Experimental procedures: PGG

We used the binary PGG reported by Chung *et al.* (2011a,b) (for further information, see Materials and Methods therein). The participants were allocated a sum of money and required to choose whether to cooperate (give all money to the public good) or free ride (keep all money as their private good) on each round. Five participants were assigned to each group. The groups were given cards marked with either '5000' or '0'. The participants received 10% of the card's value in Korean currency ($0.50 or 500 Korean won per 5000-card) as a reward, providing a real financial incentive to the game. Each game consisted of 10 rounds, and before each round, each participant was given $5 worth of cards: one 5000-card and one 0-card. Each player had to choose whether to cooperate or free ride; i.e. they had to choose to submit a 5000-card or 0-card to the experimenter. Based on group members' cooperation, a success or failure result was decided for the group; success occurred if three or more of the five players cooperated, otherwise failure occurred. Previous studies showed that the same free-riding choices could occur from different underlying cognitive motivations based on the payoff structure (Dawes *et al.*, 1986; Chung *et al.*, 2011a). To identify neural predictors under these varying motivations, we implemented three conditions that have different rules of distributing the bonus for successful trials. Specific amounts of money were distributed among the players based on a predefined payoff matrix with three conditions: (i) a bonus was equally distributed regardless of each individual's decision if the group succeeded [condition-standard (CondS); there was no payback if the group failed], (ii) a bonus was equally distributed regardless of each individual's decision and paid back if the group failed [condition-no-fear (CondNF); the participants were assured that they would not lose their money in this condition], and (iii) a bonus was distributed to match each individual's decision (higher amounts went to cooperators) but not paid back if the group failed [condition-no-greed (CondNG); all group members were provided with a fair share of the money in this condition (Chung *et al.*, 2011a,b)]. Figure 1a depicts the payoff matrices for each condition based on the rules described above. The participants simultaneously and anonymously turned in their cards after a countdown from 5 to 0 was displayed on a monitor (5 s). After each round, the monitor displayed a fixation screen, a result whether the group had received a bonus, another fixation screen, and then the number of cooperators (each screen lasted for 5 s). To simulate a realistic environment, the bonus was distributed each round in the form of cards if the group succeeded in earning the bonus.

The order of the three conditions was counterbalanced, and instructions for each condition were provided just before the condition started. After the instructions, all players were provided with a four-question questionnaire specific for each condition to verify their understanding of the rules of the game. All decisions made by the players (including sham-recorded participants) and the group results for each round were recorded for further analysis.

EEG data acquisition and analysis

EEGs were recorded using two Neuroscan EEG-recording systems (Compumedics Neuroscan, USA) with 64-channel Quick-caps (Ag/AgCl Quick-cap, Compumedics Neuroscan, USA). Standard electrode sites based on the international 10–20 system were used. We used an electrode on the vertex of the head as a reference, and an extra electrode between Fz, FPz, F1 and F2 was used as a ground. The impedances of all electrodes were lower than 5 k Ω. EEGs were digitized at a sampling frequency of 1000 Hz and amplified with a 64-channel SynAmps^2 amplifier (Compumedics Neuroscan, USA). We recorded EEGs from two individual participants at the same time using two separate amplifiers and computers. These two recording systems were synchronized through a customized button box. Simultaneous recording of the EEG during the game, which is also known as EEG hyperscanning, was done for possible inter-brain synchronization analyses in future (Chung *et al.*, 2008; Yun *et al.*, 2008, 2012). However, we did not describe this in detail here, because the EEG hyperscanning analyses are out of the scope of this study.

EEG data analysis was conducted using EEGLAB software (http://sccn.ucsd.edu/eeglab) (Delorme and Makeig, 2004). In the preprocessing of the EEG, we applied a 0.1–100 Hz band-pass filter, a 55–65 Hz notch filter to remove 60 Hz AC noise, and baseline removal (correction) to remove gradual drift (i.e. to detrend the data). We used independent component analysis to detect and remove eye movement artifacts. Time–frequency, feature extraction, prediction, and validation analyses were conducted consecutively. EEGs were recorded during the entire PGG task, but we defined a time-of-interest (epoch) in this study from 200 ms before the result presentation to 1 s after the onset of each round of the PGG to extract neural predictors (of free riding and cooperating). For the time–frequency analysis, the ERSP was calculated for the defined time-of-interest, based on wavelet transformation [default wavelet cycle setting of [3, 0.5] was used (Delorme and Makeig, 2004)]. Spectral activity before the onset (200 to 0 ms) was used as baseline activity. The pad ratio was set to 16, and the alpha level was set to 0.001 based on the bootstrap method

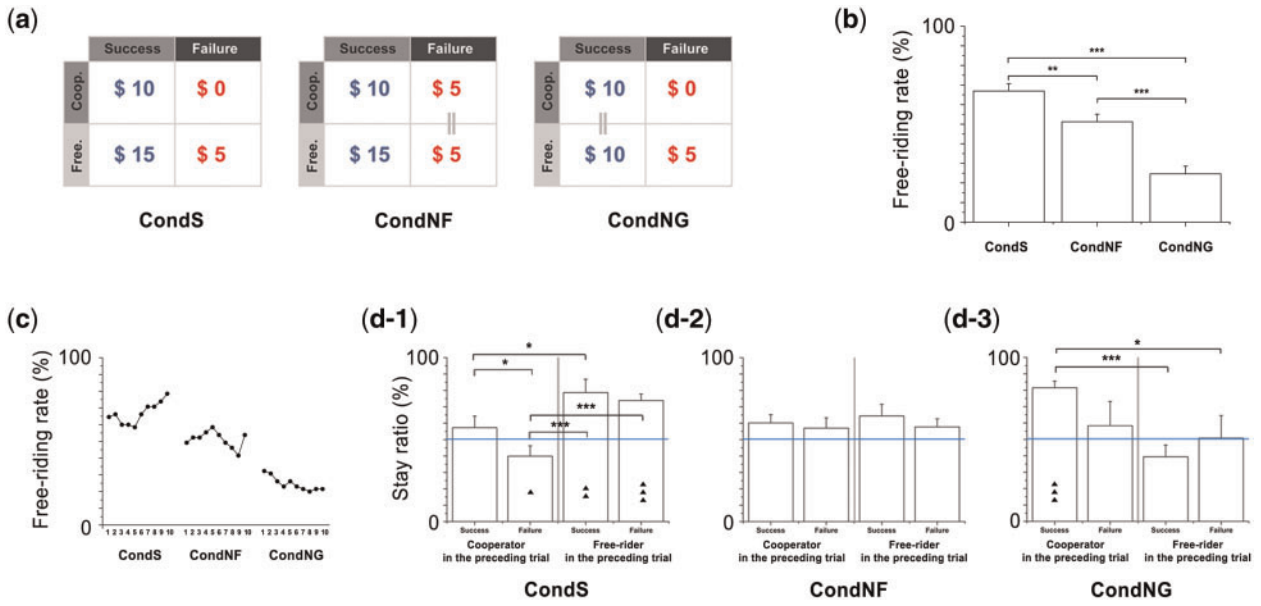

Fig. 1 Payoff matrices and behavioral decision performances in each condition. **(a)** Payoffs for free riders and cooperators are equal for failed cases of the CondNF, showing that possible loss is controlled. Payoffs are identical in successful cases, regardless of cooperation in CondNG, showing that greedy motivation is controlled. **(b)** Mean free riding rate differs between conditions. The participants exhibited significantly higher free riding rates in CondS than CondNF and CondNG. They showed the lowest free riding rate in CondNG. **(c)** In particular, free riding rate increased gradually over repeated trials in CondS, and decreased in CondNG. In CondNF, the participants showed free riding behavior that oscillated around the average. **(d)** Behavioral stay ratio following the success/failure result was also calculated to investigate participants' strategy. In each condition, all four cases (cooperator—success, cooperator—failure, free-rider—success, and free-rider—failure) were individually tested to determine whether the responses were biased away from a 50% chance of changing strategy (horizontal blue line). (d-1) In CondS, cooperators in the preceding round significantly shifted following failure, whereas cooperators shifted their choices randomly after success. In contrast, free riders always stayed regardless of the success/failure result. (d-2) In CondNF, both cooperators and free riders in the preceding round showed random shifts, regardless of the result. (d-3) In CondNG, cooperators in the preceding round significantly stayed only if the result was success. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{▲}P < 0.05$; $^{▲▲}P < 0.01$; $^{▲▲▲}P < 0.001$.

(2000 resamplings). From the wavelet cycle settings and the sampling rate, the lowest frequency limit was determined as 11.72 Hz. ERSPs between 12 and 50 Hz (including beta and gamma frequency ranges) were used for further analyses (see Supplementary text for event-related potential analysis).

The ERSPs during the presentation of the results were analyzed to investigate whether they reflected the presented results, and/or encoded each player's intention to free ride or cooperate in the subsequent round (i.e. whether they were neural predictors). We used sets of two consecutive rounds to examine the most immediate neural response to future action relationship (see Decision independency between trials section in Supplementary text). Thus, each participant contributed nine behavioral events to the neural predictor analysis (26 participants × 9 rounds = 234 rounds). The decision subsequent to the 10th round (the last round) could not be predicted and was excluded from the neural predictor analysis. Significant ERSPs for subsequent cooperation and free riding were calculated from pooled corresponding rounds. This step assisted us to investigate neural features that did not only predict within subject's decisions but also predict others' (prediction algorithms are described in the following paragraphs).

Differences in ERSPs were calculated between all instances of subsequent free riding and cooperation and used as a feature pool. We first measured ERSPs that significantly corresponded with subsequent free riding and cooperation ($P < 0.001$ as described above) and subtracted ERSPs for cooperation from free riding; thus, a positive ERSP indicated a significant activation associated with free riding, and a negative ERSP indicated a significant activation associated with cooperation. By using the feature pool, we were able to control out possible artifacts (e.g. individual eye movements or muscle activities) and reduce the dimensions of the data for the prediction step (Norman *et al.*, 2006; Lotte *et al.*, 2007). For feature extraction, we visually inspected the signals and excluded the electrodes that showed extreme spectral powers that spanned the entire frequency range (12–50 Hz), as those signals might result from electric noise or electromyogram (Onton and Makeig, 2009). The time (*X*-axis; $time_{max}$) and frequency (*Y*-axis; $frequency_{max}$) of the maximum absolute ERSP value within every discrete, non-zero ERSP cluster (Figure 2a) were measured, and the cluster size along the *X*- and *Y*-axes was calculated. Each cluster size was defined as the area of a rectangle with a width and height equal to twice the distance between $time_{max}$ (or $frequency_{max}$) and $time_{border}$ (or $frequency_{border}$), the edge of the neighboring non-zero ERSP point along the *X*-axis (or *Y*-axis). We selected the edge point (either $time_{max} > time_{border}$ or $time_{max} < time_{border}$) that was closer to $time_{max}$ (the same rule was used in the frequency dimension). We defined the clusters of non-zero ERSPs wider than 10 points in either the time or frequency dimension as features. The average ERSP value within the rectangular cluster was also measured and reflected the average response smoothed in the time and frequency dimensions. Each feature was defined in a five-dimensional space that included time, frequency, and spatial information; i.e. the maximum and average ERSP values for each fixed time, frequency, and electrode identity (location) were defined (Figure 2b). In conventional studies employing MVPA on fMRI data, the blood oxygenation level dependent signal change of a voxel is assumed to represent the value of a feature (LaConte *et al.*, 2005; Norman *et al.*, 2006). We defined each feature with five components of information as described above to best preserve time–frequency and spatial information. When we tested the prediction performance with a combination of two features ($feature^1$ at $electrode^1$, $time_{max}{}^1$ and $frequency_{max}{}^1$: [$ERSP_{max}{}^1$, $ERSP_{avg}{}^1$] and $feature^2$ at $electrode^2$, $time_{max}{}^2$ and $frequency_{max}{}^2$: [$ERSP_{max}{}^2$, $ERSP_{avg}{}^2$]), the information of the two features was concatenated (predictor: [$ERSP_{max}{}^1$, $ERSP_{avg}{}^1$, $ERSP_{max}{}^2$, $ERSP_{avg}{}^2$]; electrode, $time_{max}$ and $frequency_{max}$ are fixed information, which does not need to be included) and used as a predictor of the subsequent behavior (Figure 2b).

SVM, a supervised learning method for classification, was used to determine whether the defined features were suitable for classifying free riding and cooperation. We used the linear support vector

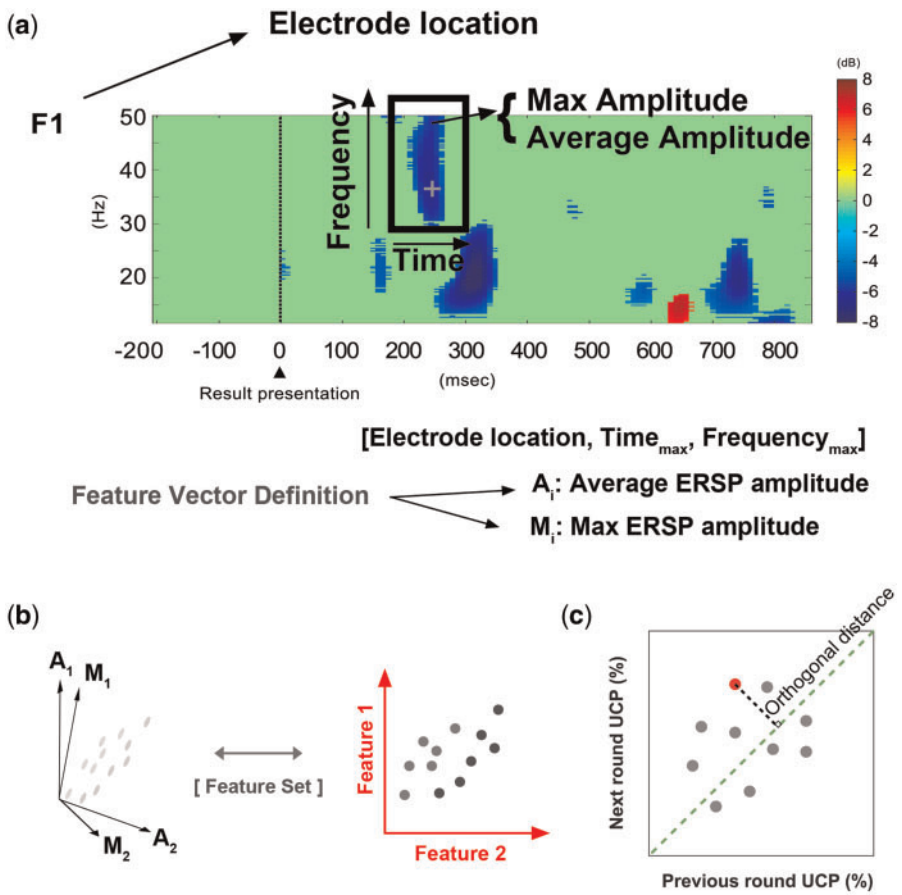

Fig. 2 An example of an EEG feature used to predict the subsequent free ride or cooperation decision. **(a)** The cross hair within the feature indicates the time and frequency of the maximum ERSP. **(b)** As depicted in the lower part of the figure, each feature vector was defined according to five types of information: electrode location, time$_{max}$, frequency$_{max}$, average ERSP amplitude, and max ERSP amplitude. When we tested prediction performance with a combination of two features (feature^1 at electrode^1, time$_{max}{}^1$, and frequency$_{max}{}^1$: [ERSP$_{max}{}^1$, ERSP$_{avg}{}^1$] and feature^2 at electrode^2, time$_{max}{}^2$, and frequency$_{max}{}^2$: [ERSP$_{max}{}^2$, ERSP$_{avg}{}^2$]), the information from the two features was concatenated in parallel (predictor: [ERSP$_{max}{}^1$, ERSP$_{avg}{}^1$, ERSP$_{max}{}^2$, ERSP$_{avg}{}^2$]).

classification from MATLAB R2010b (MathWorks, Natick, MA) to calculate prediction accuracy. Prediction performances were measured for all single features (Accuracy$_i$; $i = 1, 2, \ldots, n$ where n is the total number of features) and all sets of two separate features (Accuracy$_{ij}$; $i = 1, 2, \ldots, n$, and $j = 1, 2, \ldots, n$; when $i = j$, redundant information from the same feature was used). The free ride or cooperative decision was predicted using the EEGs within 1 s following the presentation of the results of the previous round. Rounds in which the group succeeded (success group) and rounds in which the group failed (failure group) were categorized into separate groups to test the predictive accuracy regardless of the preceding result. We used a 10-fold cross-validation analysis to verify the predictive accuracy obtained from each group of features. In other words, the group data were divided into 10 components, and one of the components was excluded from the training set. The excluded portion of the dataset was used as a testing set that enabled us to test the robustness of the prediction model and avoid over-fitting. This validation test was repeated 10 times (10-fold). To preserve equal ratio of cooperation and free riding decisions between training and test set, we randomly assigned subgroup number from 1 to 10, separately to the cooperation and free riding trials, then merged for further analyses. Because we investigated neurobehavioral relationship between two consecutive rounds (run base data set), the validation step confirmed both inter- and intra-individual decision predictions.

For each feature, we measured the UCP, which characterizes the average additional prediction accuracy when the feature i was combined with the other feature j in a prediction model (Hampton and O'Doherty, 2007; Clithero *et al.*, 2009; Carter *et al.*, 2012). The UCP of a feature i was defined as follows:

$$\mathrm{UCP}_i = \sum_{j=1}^{n} (\mathrm{Accuracy}_{ij} - \mathrm{Accuracy}_j)/n \quad (1)$$

where n is the number of features in the corresponding case (the success/failure group in each condition). We should note that the EEG features were extracted from the result presentation phase of the previous round. Thus, the features not only predicted cooperation in the next round but also reflected the participants' own decision in the previous round. To disentangle the features' predictive powers for the next round, we compared the next round UCP (i.e. the UCP in predicting the next round) with the previous round UCP (i.e. the UCP in predicting the previous round) (Figure 2c). The orthogonal distance of each feature on a two-dimensional representation of the two UCP measures was calculated to quantify each feature's predictive accuracy orthogonal to the recollection signal for previous decisions.

To report the neural features that have significant prediction power on subsequent decision, rather than reflecting the brain response on the previous round's result, we used two criteria: (i) features should have the next round UCP > the previous round UCP and (ii) the next round UCP > 0.

RESULTS

We tested whether the participants' free riding rates were significantly affected by conditions and/or interaction from the preceding round,

which reflect motivations and strategies under free riding. As we observed from previous studies (Chung *et al.*, 2011a), participants showed significant effect of condition and loss sensitivity (Figure 1b–d). On average, the standard condition (CondS) recruited the highest free riding, the condition without fear (CondNF) showed significantly less, and the condition without greed (CondNG) showed the least free riding rate (Figure 1b). In particular, free riding rate in CondS increased and it decreased in CondNG over repeated trials, whereas no statistical change was found in CondNF (Figure 1c). Statistically significant stay rate differences (whether a participant makes the same choice in the consecutive trial with the latest choice) based on the group success results in CondS [$\chi^2(3) = 24.0$, $P < 0.001$] and CondNG [$\chi^2(3) = 23.1$, $P < 0.001$] may account for the average changes of participants' decision over time (Figure 1d). These behavioral differences between conditions show that the manipulated payoff structure successfully induced differential cognitive motivations from the participants (see Supplementary text, Figures S1, and S2 for details).

Neural predictors of subsequent free riding and cooperation

By contrasting ERSP patterns in successive free riding from cooperation, we identified 302–578 features (Figure 2a; Figures S3–S5) that are specifically correlated with future decision in the corresponding condition and success/failure result of the previous round. Because free riding trials were contrasted with cooperating trials, in these features, positive activation represented a signal indicating free riding, and negative activation represented a signal indicating cooperation. All selected features were between 12 and 50 Hz in the frequency domain and 0–860 ms in the time domain (Figure S6).

Using the prediction and validation analysis (10-fold; see Materials and Methods section for details) based on the features from whole brain analysis associated with the success/failure result presentation, we measured the prediction accuracy of the features of the subsequent round. The prediction accuracies were compared with a baseline prediction rate that was defined based on participants' behavioral decisions following success or failure in the previous round (Table 1). In CondS, a feature set from C2 and FC3 showed a maximum prediction accuracy of 88.8% for the success group and FPz and PO4 showed 84.4% maximum prediction for the failure group. Compared with the defined baseline prediction (68.3%), any feature combination including C2, the selected feature, showed significantly higher prediction accuracy for the success group [$t(577) = 37.1$, $P = 6.3e\text{-}155$]. For the failure group, features including FPz showed significantly greater predictions than the baseline [78.2%; $t(301) = 7.2$, $P = 4.9e\text{-}12$]. In CondNF, the CP4 and PO3 set showed 78.9% maximum prediction accuracy for the success group, and CP3 and C3 showed 77.5% maximum prediction accuracy for the failure group. Feature combinations including CP3 showed significantly higher mean prediction accuracies than the baseline (58.8%) for the success group [$t(409) = 35.6$, $P = 2.0e\text{-}127$]. For the failure group, sets including CP3 predicted the next decision significantly higher than the baseline [62.5%; $t(338) = 18.7$, $P = 3.9e\text{-}54$]. In CondNG, the feature set of M2 and FC3 showed 88.5% maximum prediction for the success group. Compared with behavioral baseline (85%), any feature sets combined with M2 were significantly better predictors for the success group [$t(378) = 9.4$, $P = 6.6e\text{-}19$]. For the failure group, features from many electrodes combined with Cz (Table 1) showed 100% prediction accuracies (mean of all feature combination = 93.9%). However, we have to note that CondNG only had 14 failed trials (followed by 4 cooperation and 10 free riding), which was insufficient number of samples for cross validation. Thus, further results on the CondNG-failure should be interpreted carefully.

Measured prediction accuracies were converted to an UCP that summarizes each feature's average predictions. To distinguish the features that predicted the next round's free riding behavior orthogonally to the previous round, two different UCPs, the next round UCP and previous round UCP, were plotted in two-dimensional space for each condition and result (Figure 3). Both UCPs were measured using ERSPs at the result presentation phase (result of nth round). Individual data points located on the left-upper side of the diagonal lines (the green dashed lines in Figure 3) depict the features that predict upcoming decisions ($n+1$th round decision) better than the decisions of the previous round (nth round decision; CondS-success: 364 of 578 features, -failure: 25 of 302; CondNF-success: 166 of 401, -failure: 66 of 339; CondNG-success: 150 of 379, -failure: 164 of 494). UCPs from both CondS and CondNF, but not CondNG, showed significant positive correlation showing that features highly sensitive to the previous round result also predicts next round decision better (CondS-success: $r = 0.3$, $P = 6.8e\text{-}14$, -failure: $r = 0.2$, $P = 8.3e\text{-}04$; CondNF-success: $r = 0.3$, $P = 2.6e\text{-}09$, -failure: $r = 0.1$, $P = 0.04$; CondNG-success: $r = 0.03$, $P = 0.6$, -failure: $r = 0.07$, $P = 0.1$). These two-dimensional feature distributions show the dissociable prediction power of the ERSPs' time–frequency information for future cooperative behavior in parallel with the reflection of one's own decision in the previous round (the previous round UCP), such that the features with the longest orthogonal distance are the optimal neural predictors of free riding. Interestingly, UCPs following success is spread wide along x-axis, whereas UCPs following failure is spread wider along y-axis. Based on the definition of UCP [equation (1)], narrow spread features show comparable predictive power (or neural responses) between the features on the corresponding axis. Thus, this indicates that the extracted features following failed trials have comparable neural responses about the result, whereas features following succeed trials have comparable neural predictive power on the next trial.

Among the data selected above, only the features that had UCPs greater than 0% for the next round had significant prediction accuracy over all brain activity (average prediction accuracy with single feature). The time–frequency components of the features (next round UCP>previous round UCP and next round UCP > 0) were distributed as shown in Figure 4a (CondS-success: 42/578 features, -failure: 9/302; CondNF-success: 36/410, -failure: 29/339; CondNG-success: 61/379, -failure: 84/494). Interestingly, this pattern shows that participants' next decision following failure occurs slightly earlier (dark blue block, 30–35 Hz, 0–100 ms) than that following succeed (red tone blocks, 45–50, 35–40 Hz, 100–300 ms) (see Figure S7 for the pattern of predictors common to all conditions). Figure 4b depicts the orthogonal distance (Figure 2c) between each feature and the diagonal line (the green dashed line in Figure 3). Spatial patterns of neural predictors varied depending on the corresponding condition and the preceding results (Figure S7). Interestingly, two common patterns were found regardless of condition or results; centroparietal and frontotemporal regions showed long orthogonal distances on average, which indicate their high effective power on the prediction of next decision (see Figure S7 for the average pattern). In particular, we found that CondS and CondNF shared features in centroparietal region that predicting the next decision, whereas CondS and CondNG shared features from temporal region (Figure 4b). This inter-condition common feature patterns were only found when groups succeed in the preceding trial, but not following their failure. Together with figure 3, the missing joint pattern between conditions following failure may indicate that participants have more than one common strategy, in contrast to the trials following success that recruit common factors (i.e. processes) across individuals. These results show maps of the brain regions along to their time–frequency activities that are related to high

Table 1 Selected features at success/failure presentation with prediction accuracies greater than baseline. We measured prediction accuracies for all paired feature combinations. The features that showed higher average prediction, which is averaged through all feature sets including each corresponding feature, than the baseline prediction accuracy (chance level) were listed

Condition	Result	Number of cooperators/free riders	Max. prediction accuracy (%)	Electrodes combination with maximum prediction accuracy
CondS	Success	19/41 (68.3%[a])	88.8	C2 + FC3
	Failure	38/136 (78.2%)	84.4	FPz + PO4
CondNF	Success	47/67 (58.8%)	78.9	CP4 + PO3
	Failure	45/75 (62.5%)	77.5	CP3 + C3
CondNG	Success	187/33 (85%)	88.5	M2 + FC3
	Failure	4/10 (71.4%)	100.0	Cz + [AF4, C1, C2, C3, C4, C5, CP2, CP3, CP4, CP5, CPz, Cz, F1, F2, F3, F4, F5, F6, F7, F8, FC1, FC2, FC3, FC4, FC5, FC6, FCz, FP2, FT8, Fz, M1, O1, P2, P3, P4, P5, P6, P7, P8, PO3, PO5, PO6, PO7, PO8, POz, Pz, T7, T8, TP8][b]

[a]The baseline prediction rate as calculated based on the number of cooperation/free riding behavioral decisions in each condition and result. The more frequent strategy is presented as a proportion in the parentheses. [b]Any combination of the features from Cz and one of the other features reported within the bracket has the same prediction performance.

prediction accuracy for subsequent cooperation and/or free riding, independent of the previous decision.

DISCUSSION

This study investigates whether covert intentions of free riding and cooperation could be decoded from a short, single-trial EEG signal. We applied multivariate analysis to extract prominent features (i.e. neural markers) and used SVM to predict subsequent cooperation or free riding decisions at the inter- and intra-individual levels. Because the repeated binary PGG we implemented has a payoff matrix similar to that of the Stag Hunt game (a coordination game [Camerer, 2003]), free riding and cooperation decisions rely on adaptive learning, regarding the cooperativeness of the group members. Thus, we hypothesized that strategic decisions for each round are updated based on the result of the preceding round. In this study, we found that the multi-feature patterns from the ERSP signals measured within 1 s following the result presentation predicted the participants' subsequent free riding and cooperative behaviors at greater than chance levels.

From when and where the neural predictors occurred, we can draw three major interpretations on the neural processes during the decision-making. First, the finding that these neural predictors are observed within 1 s of result presentation suggests that covert decision-making processes begin immediately after receiving information. Previous studies showed not only monetary reward learning but also social interaction can be structured within learning paradigm (King-Casas *et al.*, 2005; Behrens *et al.*, 2008; Zhu *et al.*, 2012). In other words, one should recalibrate the expectation on others in the repeated social learning. Because this study used one of the repeated social interaction games, it was expected to observe information update regarding the group cooperation at every round. Interestingly, the subsequent free riding or cooperative decision was predicted not only from previous history of decisions but also from initial 1 s of neural activities. This rapid (or automatic) covert decision might have occurred due to simplicity of the implemented decision strategies, such as win-stay lose-shift (Nowak and Sigmund, 1993). However, based on the behavioral stay ratio, the participants' decision patterns were more sophisticated. Alternative explanation is that participants might have weighted previous history (or prediction error) as important component compared with other internal motivations. Because the most influential information is collected, future choices could be predicted with high accuracy. In this study, we observed that subsequent cooperative decision starts immediately after result presentation of the previous round. Decision model that embeds evidence accumulator, decision module that makes choice when enough evidence is accumulated (as in perceptual decision-making; Ploran *et al.*, 2007; Heekeren *et al.*, 2008), and/or weighted learning algorithm (e.g. reinforcement learning; Philiastides *et al.*, 2010; Fischer and Ullsperger, 2013) would assist better understanding on complex human decision process with finer temporal resolution.

Second, spatial distribution of the features that met the selection criteria gives us some insights on participants' decision processes. The CondS and CondNG shared features from bilateral temporal region that are highly predicting the next free riding decision, whereas CondS and CondNF shared neural predictors from centroparietal region. Interestingly, temporoparietal junction has been shown in numerous neuroimaging studies to play an important role in reading the minds of others (TOM) (Saxe and Kanwisher, 2003; Apperly *et al.*, 2004; Samson *et al.*, 2004; Carter *et al.*, 2012). That the features around this region had the longest orthogonal distances may support participants' usage of TOM, which is consistent with one of our previous studies on cognitive motivations in PGG (Chung *et al.*, 2011a). Based on the payoff structure (Figure 1a), cognitive motivation to free ride, shared between CondS and CondNG can be defined as 'fear of losing money'. Whether a participant loses his/her money after cooperation highly depends on the number of cooperators within the group. In contrast, CondNF guarantees a minimal reward regardless of participants' action, which may rather link participants' primary process on their own decision. We suggest that early features that predict participants' next decision support common and differential cognitive motivations between conditions, including the use of TOM.

Finally, time–frequency characteristics of the selected features relative to the unselected features show dynamic thought processes. The selected predictors of cooperation were not only distinguishable in the spatial domain, but they were also discrete in the temporal and spectral dimensions. According to the methods used for measuring prediction performance, only some portion of the predictors met the criteria, even though the other features were only separated by a few hundred milliseconds and/or were within the comparable frequency range. Furthermore, selected neural predictors' time–frequency pattern showed distinctive differences depends on the group result that may reflect differential functional processes (e.g. response speed, cognitive complexity, strategy; Figure 4; Figure S7). EEG microstates at the subsecond level have been broadly investigated by analyzing mental states during meditation or sleep (Cantero *et al.*, 2002; Lehmann *et al.*, 2006). In addition, several studies using various cognitive tasks have shown

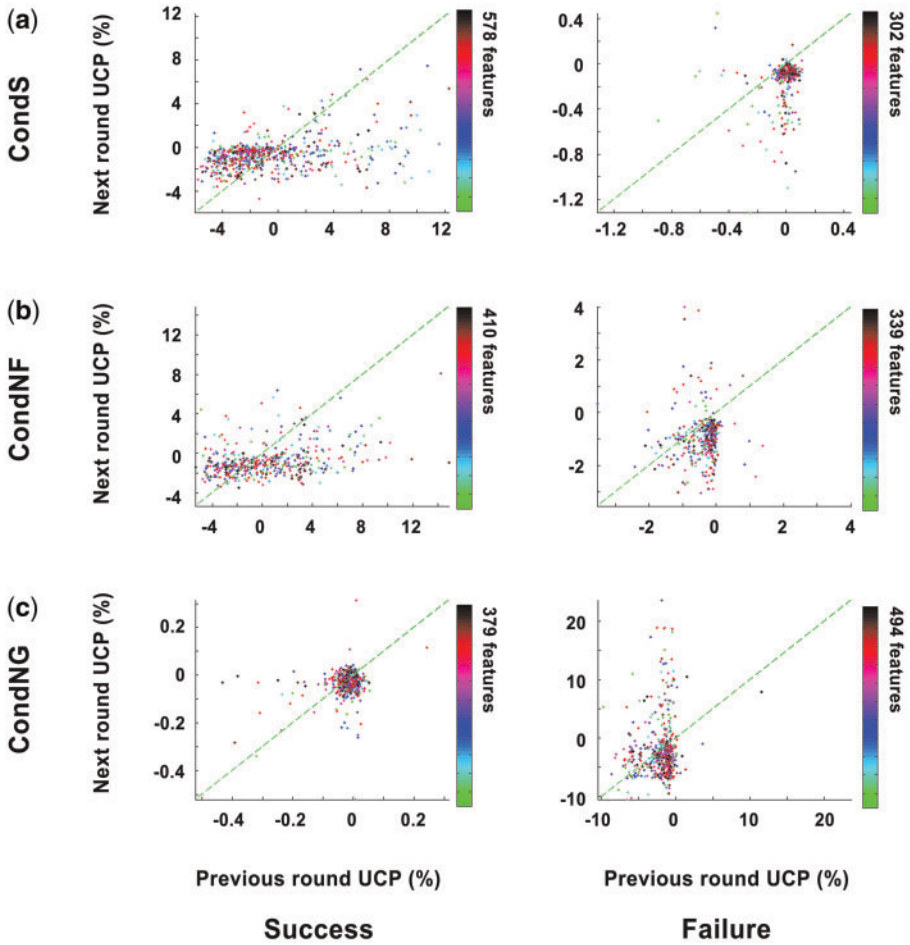

Fig. 3 Two-dimensional distribution of feature performances. To dissociate the features that only predicted decisions in the next round, the UCPs of the next round ($n+1$th round decision) were compared with the UCPs of the previous round (inverse prediction; nth round decision). UCPs were measured using ERSPs at the result presentation phase (result of nth round). Features located to upper-left side of the 45° lines (green dashed line) showed better performance in predicting the next decision than in reflecting participants' neural responses to their own decision from the previous round.

that rapid EEG and/or event-related potentials can distinguish different chunks of processes (Khateb *et al.*, 2000; Schnider *et al.*, 2002). This study showed that the prominent features that predict the next social decision were separable at scales of 100 ms and 5 Hz in the temporal and spectral dimensions, respectively. We suggest that the time–frequency characteristics in this study reflect dynamically changing mental states during complex social cognition.

This study has some limitations, and thus, the results should be interpreted with caution. First, free riding is not a dominant incentive in binary PGGs that have a threshold for a success or failure result (mixed strategy). The repeated free riding and/or cooperation choices in this study should depend more on coordinating with group members (i.e. adaptive learning) compared with those of a linear PGG (Houser and Kurzban, 2002; Bayer *et al.*, 2010). Although the game design falls within broad descriptions of PGG (Camerer, 2003), we should take this limitation into account for generalized interpretation of cooperation and free riding. Second, because the participants made decisions for 10 repeated rounds in each condition, we had a limited number of cases (samples) in certain conditions (e.g. in CondNG, success was the dominant result; there were 220 and 14 cases of success and failure, respectively). Thus, we should take this limitation into account when interpreting the predictability of the feature combinations. Third, reported neural predictors were restricted within 12–50 Hz frequency range that was determined by time–frequency analysis settings (wavelet). We have to note that the current results do not rule out the existence of relevant low frequency EEG activations on cooperation and free riding decisions (e.g. Babiloni *et al.*, 2007; Cohen *et al.*, 2007). Finally, the spatial information of selected neural features had low resolution due to the characteristics of the EEG method. We used 64-channel EEG caps for the experiment and selected the neural features from the electrode set. This study focused on the time–frequency information that represented the motivations to cooperate or free ride, but further source localization analysis might provide additional spatial dissociation between cognitive and affective motives.

To the best of our knowledge, this is the first EEG study to investigate neural predictors of free riding and/or cooperation behavior using a PGG. We observed evidence of rapid initiation of decision-making processes and monitored distinct neural features using SVMs and feature selection criteria (UCP). The results suggest a method to capture covert motivation in social decision-making within a group. Furthermore, we would like to emphasize that further analyses on selected feature patterns and their temporal dynamics would shed light on the microstates of the rapid, complex decision-making processes of both healthy populations and patients with neuropsychological impairments (Latchoumane *et al.*, 2007, 2009, 2012). Regarding practical uses, the short latency (less than 1 s) decoding of underlying intentions with single-trial ERSPs could potentially be applied to BCIs that assist paralyzed patients in expressing complicated mental states (Birbaumer *et al.*, 2008).

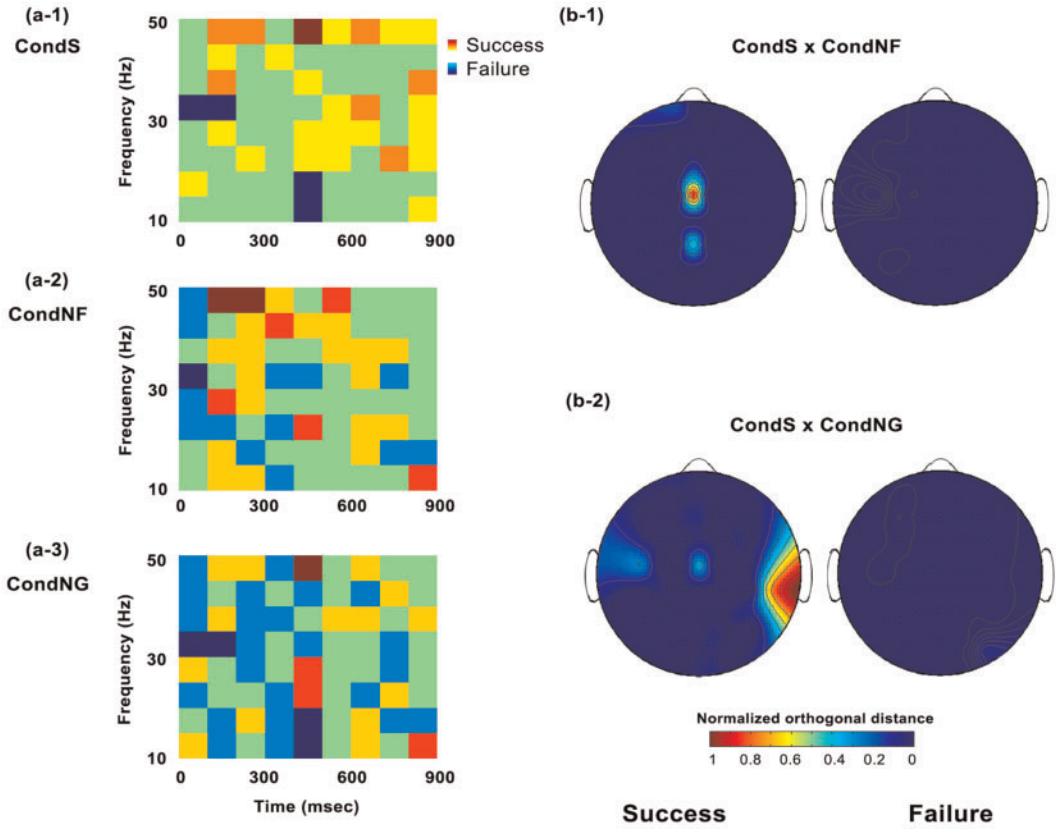

Fig. 4 Time–frequency characteristics and normalized orthogonal distances of the selected features. **(a, b)** Features above the predefined selection criteria (next round UCP > previous round UCP and next round UCP > 0) were plotted. **(a)** Time and frequency information of the selected features were depicted separately for succeeded (red scale) and failed (blue scale) trials, presenting frequency of feature occurrences color-coded. Each bin size was set 5 Hz and 100 ms. **(b)** Orthogonal distances were normalized within each subcondition (success and failure groups separately). To show spatial characteristics of common features between conditions, normalized orthogonal distances from each condition were multiplied and depicted as joint distribution. (b-1) CondS and CondNF share centroparietal feature, whereas (b-2) CondS and CondNG share features from temporal region.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

Conflict of Interest

None declared.

REFERENCES

Andreoni, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics*, *37*(3), 291–304.

Andreoni, J. (1995). Cooperation in public-goods experiments: kindness or confusion? *The American Economic Review*, *85*(4), 891–904.

Apperly, I.A., Samson, D., Chiavarino, C., Humphreys, G.W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience*, *16*(10), 1773–84.

Babiloni, F., Astolfi, L., Cincotti, F., et al. (2007). Cortical activity and connectivity of human brain during the prisoner's dilemma: an EEG hyperscanning study. In: Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, August 2007. IEEE, 4953–6.

Baumgartner, T., Götte, L., Gügler, R., Fehr, E. (2011). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping*, *33*(6), 1452–69.

Bayer, R.C., Renner, E., Sausgruber, R. (2010). Confusion and learning in the public goods game. School of Economics Working Papers, Research paper No.: 2010-24. October 2010. School of Econonics, The University of Adelaide.

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., Rushworth, M.F.S. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–9.

Birbaumer, N., Murguialday, A.R., Cohen, L. (2008). Brain-computer interface in paralysis. *Current Opinion in Neurology*, *21*(6), 634.

Boyd, R., Gintis, H., Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, *328*(5978), 617–20.

Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.

Camerer, C.F. (2007). Neuroeconomics: using neuroscience to make economic predictions. *The Economic Journal*, *117*(519), C26–42.

Cantero, J.L., Atienza, M., Salas, R.M. (2002). Human alpha oscillations in wakefulness, drowsiness period, and REM sleep: different electroencephalographic phenomena within the alpha band. *Neurophysiologie Clinique/Clinical Neurophysiology*, *32*(1), 54–71.

Carter, R.M.K., Bowling, D.L., Reeck, C., Huettel, S.A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science*, *337*(6090), 109–11.

Chung, D., Kim, Y.T., Jeong, J. (2011a). Cognitive motivations of free riding and cooperation and impaired strategic decision making in schizophrenia during a public goods game. *Schizophrenia Bulletin*, *39*(1), 112–9.

Chung, D., Yun, K., Jeong, J. (2008). Neural mechanisms of free-riding and cooperation in a public goods game: an EEG hyperscanning study. In: International Conference of Cognitive Science, The Korean Society for Cognitive Science, Seoul, South Korea, July 2008.

Chung, D., Yun, K., Kim, J.H., Jang, B., Jeong, J. (2011b). Different gain/loss sensitivity and social adaptation ability in gifted adolescents during a public goods game. *PLoS One*, *6*(2), e17044.

Clithero, J.A., Carter, R.M., Huettel, S.A. (2009). Local pattern classification differentiates processes of economic valuation. *Neuroimage*, *45*(4), 1329–38.

Cohen, M.X., Elger, C.E., Ranganath, C. (2007). Reward expectation modulates feedback-related negativity and EEG spectra. *Neuroimage*, *35*(2), 968.

Dawes, R.M., Orbell, J.M., Simmons, R.T., Kragt, A. (1986). Organizing groups for collective action. *The American Political Science Review*, *80*(4), 1171–85.

De Vico Fallani, F., Nicosia, V., Sinatra, R., et al. (2010). Defecting or not defecting: how to "read" human behavior during cooperative games by EEG measurements. *PLoS One*, *5*(12), e14187.

Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Fehr, E., Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785–91.

Fehr, E., Gachter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–94.

Fischer, A.G., Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron*, *79*(6), 1243–55.

Fowler, J.H., Christakis, N.A. (2010). Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(12), 5334.

Frith, C.D., Singer, T. (2008). The role of social cognition in decision making. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *363*(1511), 3875–86.

Gächter, S., Herrmann, B., Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *365*(1553), 2651–61.

Hampton, A.N., Bossaerts, P., O'Doherty, J.P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(18), 6741.

Hampton, A.N., O'Doherty, J.P. (2007). Decoding the neural substrates of reward-related decision making with functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(4), 1377–82.

Hauert, C., De Monte, S., Hofbauer, J., Sigmund, K. (2002). Volunteering as red queen mechanism for cooperation in public goods games. *Science*, *296*(5570), 1129–32.

Hauert, C., Traulsen, A., Brandt, H., Nowak, M.A., Sigmund, K. (2007). Via freedom to coercion: the emergence of costly punishment. *Science*, *316*(5833), 1905–7.

Heekeren, H.R., Marrett, S., Ungerleider, L.G. (2008). The neural systems that mediate human perceptual decision making. *Nature Reviews Neuroscience*, *9*(6), 467–79.

Ho, T.H., Wang, X., Camerer, C.F. (2008). Individual differences in EWA learning with partial payoff information. *The Economic Journal*, *118*(525), 37–59.

Houser, D., Kurzban, R. (2002). Revisiting kindness and confusion in public goods experiments. *The American Economic Review*, *92*(4), 1062–9.

Isaac, R.M., Walker, J.M., Thomas, S.H. (1984). Divergent evidence on free riding: an experimental examination of possible explanations. *Public Choice*, *43*(2), 113–49.

Janssen, M.A., Holahan, R., Lee, A., Ostrom, E. (2010). Lab experiments for the study of social-ecological systems. *Science*, *328*(5978), 613–7.

Khateb, A., Michel, C.M., Pegna, A.J., Landis, T., Annoni, J.M. (2000). New insights into the Stroop effect: a spatiotemporal analysis of electric brain activity. *Neuroreport*, *11*(9), 1849.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C.F., Quartz, S.R., Montague, P.R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science*, *308*(5718), 78–83.

Krajbich, I., Camerer, C., Ledyard, J., Rangel, A. (2009). Using neural measures of economic value to solve the public goods free-rider problem. *Science*, *326*(5952), 596–9.

LaConte, S., Strother, S., Cherkassky, V., Anderson, J., Hu, X. (2005). Support vector machines for temporal classification of block design fMRI data. *Neuroimage*, *26*(2), 317–29.

Latchoumane, C., Kim, I., Sohn, H., Jeong, J. (2012). Dynamical nonstationarity of resting EEGs in patients with attention-deficit/hyperactivity disorder (AD/HD). *IEEE Transactions on Bio-Medical Engineering*, *60*(1), 159–163.

Latchoumane, C.F.V., Chung, D., Kim, S., Jeong, J. (2007). Segmentation and characterization of EEG during mental tasks using dynamical nonstationarity. In: Proceedings of the Computational Intelligence in Medical and Healthcare (CIMED 2007), University of Plymouth, Plymouth, Devon, United Kingdom, July 2007.

Latchoumane, C.F.V., Vialatte, F.B., Jeong, J., Cichocki, A. (2009). EEG classification of mild and severe Alzheimer's disease using parallel factor analysis method. *Advances in Electrical Engineering and Computational Science*, *39*, 705–15.

Kagel, J.H., Roth, A.E. (1995). *The handbook of experimental economics*, Princeton, NJ: Princeton University Press.

Lehmann, D., Faber, P.L., Gianotti, L.R.R., Kochi, K., Pascual-Marqui, R.D. (2006). Coherence and phase locking in the scalp EEG and between LORETA model sources, and microstates as putative mechanisms of brain temporo-spatial functional organization. *Journal of Physiology-Paris*, *99*(1), 29–36.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, *4*, R1.

Makeig, S., Debener, S., Onton, J., Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, *8*(5), 204–10.

Mulert, C., Seifert, C., Leicht, G., et al. (2008). Single-trial coupling of EEG and fMRI reveals the involvement of early anterior cingulate cortex activation in effortful decision making. *Neuroimage*, *42*(1), 158–68.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–30.

Nowak, M., Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, *364*(6432), 56–8.

Nowak, M.A. (2006). Five rules for the evolution of cooperation. *Science*, *314*(5805), 1560–3.

O'Gorman, R., Henrich, J., Van Vugt, M. (2009). Constraining free riding in public goods games: designated solitary punishers can sustain human cooperation. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *276*(1655), 323.

Onton, J., Makeig, S. (2009). High-frequency broadband modulations of electroencephalographic spectra. *Frontiers in Human Neuroscience*, *3*(61), 1–18.

Palfrey, T.R., Rosenthal, H. (1991). Testing for effects of cheap talk in a public goods game with private information. *Games and Economic Behavior*, *3*(2), 183–220.

Perc, M., Szolnoki, A. (2010). Coevolutionary games—a mini review. *BioSystems*, *99*(2), 109–25.

Philiastides, M.G., Biele, G., Vavatzanidis, N., Kazzer, P., Heekeren, H.R. (2010). Temporal dynamics of prediction error processing during reward-based decision making. *Neuroimage*, *53*(1), 221–32.

Ploran, E.J., Nelson, S.M., Velanova, K., Donaldson, D.I., Petersen, S.E., Wheeler, M.E. (2007). Evidence accumulation and the moment of recognition: dissociating perceptual recognition processes using fMRI. *The Journal of Neuroscience*, *27*(44), 11912–24.

Polezzi, D., Sartori, G., Rumiati, R., Vidotto, G., Daum, I. (2010). Brain correlates of risky decision-making. *Neuroimage*, *49*(2), 1886–94.

Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., Kilts, C.D. (2002). A neural basis for social cooperation. *Neuron*, *35*(2), 395–405.

Rilling, J.K., King-Casas, B., Sanfey, A.G. (2008). The neurobiology of social decision-making. *Current Opinion in Neurobiology*, *18*(2), 159–65.

Samson, D., Apperly, I.A., Chiavarino, C., Humphreys, G.W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, *7*(5), 499–500.

Santos, F.C., Santos, M.D., Pacheco, J.M. (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature*, *454*(7201), 213–6.

Sasaki, T., Brännström, Å., Dieckmann, U., Sigmund, K. (2012). The take-it-or-leave-it option allows small penalties to overcome social dilemmas. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(4), 1165–9.

Saxe, R., Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in. *Neuroimage*, *19*(4), 1835–42.

Schnider, A., Valenza, N., Morand, S., Michel, C.M. (2002). Early cortical distinction between memories that pertain to ongoing reality and memories that don't. *Cerebral Cortex*, *12*(1), 54–61.

Schulz, E., Zherdin, A., Tiemann, L., Plant, C., Ploner, M. (2012). Decoding an individual's sensitivity to pain from the multivariate analysis of EEG data. *Cerebral Cortex*, *22*(5), 1118–1123.

Suzuki, S., Niki, K., Fujisaki, S., Akiyama, E. (2011). Neural basis of conditional cooperation. *Social Cognitive and Affective Neuroscience*, *6*(3), 338–47.

Ule, A., Schram, A., Riedl, A., Cason, T.N. (2009). Indirect punishment and generosity toward strangers. *Science*, *326*(5960), 1701.

Wang, T., Deng, J., He, B. (2004). Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns. *Clinical Neurophysiology*, *115*(12), 2744–53.

Wu, J.J., Zhang, B.Y., Zhou, Z.X., et al. (2009). Costly punishment does not always increase cooperation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(41), 17448.

Yun, K., Chung, D., Jeong, J. (2008). Emotional interactions in human decision making using EEG hyperscanning. International Conference of Cognitive Science, Seoul, Korea.

Yun, K., Watanabe, K., Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific Reports*, *2*(959), 1–8.

Zhu, L., Mathewson, K.E., Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1419–24.