Inter-brain synchronization is weakened by the introduction of external punishment

Jianbiao Li,[1,2,3] Jingjing Pan,[2] Chengkang Zhu, [1] and Yiwen Wang [4]

[1]School of Economics, Institute for Study of Brain-Like Economics, Shandong University, Jinan 251000, China
[2]China Academy of Corporate Governance, Reinhard Selten Laboratory, Nankai University, Tianjin 300071, China
[3]Nankai University Binhai College, Tianjin 300270, China
[4]School of Economics and Management, Fuzhou University, Fuzhou 350108, China
Correspondence should be addressed to Chengkang Zhu, School of Economics, Institute for Study of Brain-Like Economics, Shandong University, Jinan 250100, China. E-mail: zhuchengkang1989@126.com.
Jianbiao Li and Jingjing Pan contributed equally to this study.

Abstract

Punishment is a popular institution to enforce social norms in human society. However, how the punishment institution impacts the inter-brain neural signatures of two-person social interactions is still an open question. By performing electroencephalography recording of brain activity in two interacting parties as they simultaneously played both the revised repeated ultimatum game (rrUG) and the revised repeated dictator game (rrDG), this study focused on exploring how the introduction of external punishment influences inter-brain synchronization between the two parties. The data showed a significant negative effect of external punishment on inter-brain synchronization, with greater inter-brain synchronization observed in the rrDG than in the rrUG. We proposed a possible mechanism underlying this result. In the rrDG, the similar moral motivation of both proposers and responders results in inter-brain synchronization between them. However, in the rrUG, the introduction of external punishment crowds out the intrinsic moral motivation of the proposers, thereby undermining the inter-brain synchronization. Moreover, we found a significant positive correlation between the rejection rate from responders for disadvantageous inequal offer and inter-brain synchronization in the rrDG. These findings contribute to understanding the negative effect of punishment institution and shed light on the inter-brain mechanism underlying social interaction.

Key words: moral motivation; introduction of external punishment; inter-brain synchronization; EEG hyperscanning

Introduction

Punishment is a popular institution to enforce social norms in human society. Abundant behavioral and neuroimage literature has shed light on the great effect of punishment institution on the behaviors (Camerer and Thaler, 1995; Fehr and Gächter, 2000; Fehr *et al.*, 2002; Camerer, 2003; Fehr and Fischbacher, 2003) as well as on the underlying cognitive and neural procedure (Spitzer *et al.*, 2007; Weiland *et al.*, 2012; Chen *et al.*, 2017). However, how punishment institution impacts the inter-brain neural signatures of two-person social interactions is still an open question.

Recently, hyperscanning has been used to explore inter-brain synchronization during interactive decision-making in order to shed light on neuronal correlations between interacting dyads (Astolfi *et al.*, 2010, 2011, 2015; Tang *et al.*, 2016; Jahng *et al.*, 2017; Hu *et al.*, 2017; Zhang *et al.*, 2019). Hyperscanning has been applied to several neuroimaging techniques, including electroencephalography (EEG), functional near-infrared spectroscopy and functional magnetic resonance imaging (fMRI), in order to

record brain activity in two or more individuals simultaneously (Montague *et al.*, 2002; Koike *et al.*, 2015). Using these techniques, numerous studies have explored the neural mechanisms underlying two-person interactions (see Liu *et al.*, 2018 for a review). These studies indirectly reveal that inter-brain synchronization derives from the similar cognitive and emotional processes (i.e. imitation, empathy, mentalization, mutual cooperation) in both parties.

Based on previous hyperscanning studies, by recording EEG from two parties simultaneously playing both the revised repeated ultimatum game (rrUG) and the revised repeated dictator game (rrDG), this study focused on exploring how the introduction of external punishment influences inter-brain synchronization between the two parties. In the rrDG, proposers offer a division of money. When they provide unfair/fair offers, they feel moral disgust/satisfaction (Elster, 1999; Gintis, 2000). Responders decide whether to accept or reject the offer. Although the responder's choice has no monetary effect on both parties, they acquire moral satisfaction by expressing feelings of

anger/disgust to unfair offers (Yamagishi *et al.*, 2009; Krupka and Weber, 2013). As a result, the common factor that affects the decision-makings of proposers and responders is moral motivation that individuals avoid/pursue moral disgust/satisfaction. Moreover, both proposers and responders experience moral disgust/satisfaction when the distribution is inequitable/equitable. These similar cognitive and emotional processes will result in significant inter-brain synchronization between the two parties.

In the rrUG, the rejection of responders leads to nothing earned by both players. Allowing the rejection of responders to be monetary effective in the rrUG should be viewed as introducing an external punishment institution (Güth *et al.*, 1982; Camerer and Thaler, 1995). When punishment institution is introduced, moral motivation of the prosper will be crowded out. This phenomenon is called motivation crowding-out effect (Deci *et al.*, 1999; Gneezy and Rustichini, 2000a,b; Frey and Jegen, 2001; Lin and Yang, 2006; Holmås *et al.*, 2010). The proposers no longer consider whether they will experience moral disgust/satisfaction but instead think about whether the responders impose external punishment. A series of studies using fMRI have demonstrated that external punishment evokes prosocial behaviors by inducing cognitive activity associated with thinking about the punishing behaviors of others, as well as emotional activity associated with the fear of monetary loss (Spitzer *et al.*, 2007; Weiland *et al.*, 2012; Chen *et al.*, 2017). For responders, moral motivation is still the main factor affecting their decision-making behaviors (Yamagishi *et al.*, 2009). Thus, the introduction of punishment might reduce the inter-brain synchronization between the interacting parties. Moreover, in the rrUG, proposers and responders make decisions relying on their beliefs (Weiland *et al.*, 2012; Chen *et al.*, 2017). Therefore, there is more mentalizing process in the rrUG than in the rrDG, which might increase inter-brain synchronization between the two parties. As a result, the overall effect of external punishment on inter-brain synchronization is not clear. By comparing the difference in inter-brain synchronization between the rrUG and rrDG, this study attempts to reveal the effect of punishment institution on inter-brain synchronization.

Materials and methods
Participants
Forty-four healthy male participants (aged 19.52 ± 1.89 years, mean $\pm$ s.d.) were recruited from Fuzhou University. Four subjects were excluded from further analysis because of inappropriate behaviors during the task (i.e. simply pressing the same key or dozing off due to sleep deprivation in the previous night). The sample size was preliminarily determined according to previous successful EEG hyperscanning studies (Jahng *et al.*, 2017; Hu *et al.*, 2017; Zhang *et al.*, 2019). Using G*Power software, we calculated the power of our inter-brain synchronization data and found that the calculated power values of our main results were all >0.8, suggesting that our sample size was sufficient. All participants were right-handed and had no history of neurological or psychiatric disorders. Written informed consent was obtained from each participant in accordance with the Declaration of Helsinki. The experimental procedure was approved by the University Committee on Human Research Protection, Fuzhou University. Participants unacquainted in advance were randomly paired with one another and prepared for the experiment in the preparation room. The role of each participant was assigned by lottery and was fixed throughout the whole experiment. Participant pools were restricted to men, because sex differences have been reported for the UG and the DG (Chew *et al.*, 2013; Eckel and Grossman, 1996). After the experiments, participants were paid for their participation: each participant received a base payment of 30 Chinese yuan (CNY, roughly equal to $4.50), plus a bonus of 20–30 CNY based on the decisions they had made during the experiment.

Task
During the tasks, participants in each pair were seated comfortably in separate experimental rooms. Each pair played two games: the rrUG and the rrDG. Each game was played 120 times. Participants did not play another game until they had finished the first one. The sequence of the two games was counterbalanced. In each trial, the proposers and responders faced a distributive offer of either 2:8, 5:5 or 8:2, chosen at random by a computer (40 trials per offer). The first number of the offer indicates the amount earned by the proposer, while the second number is that earned by the responder (an offer of 2:8, for example, means the proposer earns 2 CNY and the responder earns 8 CNY). Then, the proposer was asked to confirm that they wanted to make such an offer at the same time as the responder was asked if they would like to accept such an offer. The choice of the proposer was always executed, while that of the responder was only executed in the rrUG. If the proposer chose to make the offer, it meant that they indeed proposed such an offer to the responder in this trial. In this case, the acceptance of the responder led to the offer being executed and to both players earning money from this offer. In contrast, the choice made by the responder to reject had a different effect in each of the two games. In the rrUG, the choice to reject led to the deal being broken; hence, both players would earn nothing in this trial. In the rrDG, however, the choice to reject had no effect, with the offer executed regardless of the disgust the responder expressed at this offer. Another possible situation involved the proposer choosing to give up the offer, meaning that they would like to propose another offer. In this situation, whatever the choice made by the responder in this trial, the trial was excluded when their earnings were calculated. However, this trial was not excluded from the data analysis. It was emphasized to the participants that they would receive twice the mean amount of what they earned over the entire game as a reward.

Procedure
After the EEG electrodes had been attached, the participants were seated in a comfortable chair ~100 cm in front of a 23-inch computer monitor. Before each game began, all the participants read the instructions carefully and were asked to complete six practice trials (Figure 1 shows the timeline of a single trial). As illustrated in Figure 1, when the game began, a white fixation cross appeared in the center of a black screen for 800 ms, followed by a black screen for 500–700 ms. Afterwards, a divided color pie representing 10 CNY was presented for 2000 ms to indicate the offer (with the red and green parts, respectively representing the amounts earned by the proposer and the responder). The length of the response period was not fixed, but was <2000 ms (if at least one subject didn't response during 2000 ms, they would perform the same trial again). After both players had pressed a key, a black screen presented for 500–700 ms. Finally, the outcome was presented on the screen for 2000 ms. Participants could therefore see how much money each player had earned if the proposers accepted the offer. Otherwise, they saw two hash symbols with a colon between them ('#: #'). Then the fixation cross appeared again to begin the next trial, and trials continued until the game ended.
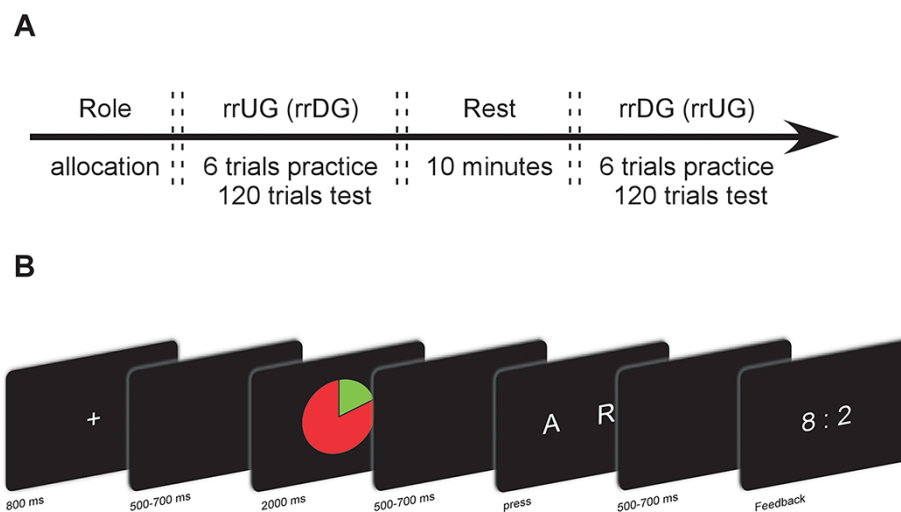
Fig. 1. (A) Procedure of the experiment. (B) Overview of the task. A white cross appeared in the center of the screen for 800 ms. Afterwards, a divided color pie was presented for 2000 ms to indicate the offer (with the red and green parts, respectively representing the amounts earned by the proposer and the responder), followed by a black screen for 500–700 ms. Then the proposer and the responder responded to the offer simultaneously. The length of the response period was not fixed but was <2000 ms. After both of the players had pressed a key, the outcome was presented on the screen for 2000 ms, followed by a black screen for 500–700 ms.

EEG data acquisition and pre-processing

The EEG signals and behavioral responses of each dyad were recorded continuously and simultaneously using two 32-channel Neuroscan portable EEG systems (Compumedics Neuroscan, Victoria, Australia). Pre-processing of the EEG data was conducted using EEGLAB (Delorme and Makeig, 2004) and custom MATLAB (MathWorks, Natick, MA, USA) scripts. Offline EEG time series were band-pass filtered from 0.1–50 Hz with slopes of 24 dB. EEG data were reset to the average of the left and right mastoids. EEG epochs were extracted from –1000 to +2000 ms relative to the timings of the offer and outcome presentations, respectively. A manual artifact correction procedure was applied to eliminate trials with artifacts based on visual inspection. An independent component analysis (ICA) was run to remove eye movement, with the ICA components related to eye movement being manually selected (Sejnowski, 1996). Signals containing EEG amplitudes greater than ±150 μV were excluded. Only those epochs without artifacts in either participant were considered for further analysis.

Brain synchronization analysis

Time–frequency analysis was performed using the built-in ft_freqanalysis function in the Fieldtrip toolbox, based on complex Morlet wavelet convolution (Oostenveld *et al.*, 2011; 5 cycles, 4–50 Hz, 47 spaced frequencies). The 2000-ms epochs were extracted at the onsets of the presentations of the offer and outcome, respectively (2000 time points per epoch). Because of the poor spatial resolution of EEG analyses, we clustered the scalp electrodes according to six corresponding brain regions (Jahng *et al.*, 2017): (i) frontal (FP1, FP2, F7, F3, FZ, F4 and F8), hereafter referred to as F; (ii) frontocentral (FC3, FCZ, FC4, C3, CZ and C4), hereafter referred to as FC; (iii) parietal (CP3, CPZ, CP4, P3, PZ and P4), hereafter referred to as P; (iv) left temporoparietal (FT7, T3, TP7 and T5), hereafter referred to as LTP; (v) right temporoparietal (FT8, T4, TP8 and T6), hereafter referred to as RTP; and (vi) occipital (O1, OZ and O2), hereafter referred to as O. The EEG data for each brain region was calculated by averaging data from the corresponding electrodes.

Inter-brain synchronization was estimated using the phase locking value (PLV; Lachaux *et al.*, 1999) for all 36 pairs of brain regions (6 × 6) between each proposer and responder, both in the rrUG and in the rrDG, based on specific time periods and frequency bands of interest. Four frequency bands were considered: theta (4–7 Hz), alpha (8–12 Hz), beta (13–30 Hz) and gamma (31–50 Hz). These four frequency bands have been identified as typical frequency ranges in previous EEG hyperscanning studies (Astolfi *et al.*, 2011; Hu *et al.*, 2017). Two time periods were extracted: 0–2000 ms after the offer presentation and 0–2000 ms after the outcome presentation. The PLV is a measure of the consistency of the phase difference and is associated with the inter-trial variance of the phase difference. It is defined as:

$$PLV_{i,j} = T^{-1}N^{-1} \left| \sum_{t=1}^{T} \sum_{n=1}^{N} \exp^{\varphi_i(t,n) - \varphi_j(t,n)} \right|$$

where N represents the number of trials, T represents the number of time points, φ is the phase, | | represents the complex modulus and i and j indicate the brain regions of the proposer and the responder in a dyad, respectively. Phases were extracted from signals using the Morlet wavelet transform (Delorme and Makeig, 2004; Mu *et al.*, 2016).

Before examining brain synchronization for the rrUG *vs* the rrDG, we conducted a statistical test to differentiate significant $PLV_{i,j}$ values against background fluctuations (Lachaux *et al.*, 1999). We generated a series of 200 new $PLV_{i, shuffle(j)}$ values by shuffling the region j variable for trials 200 times and computing the corresponding PLV each time. We then averaged the $PLV_{i,j}$ series at the subject level to obtain an averaged $PLV_{i,j}$, and averaged each series of $PLV_{i, shuffle(j)}$ at the subject level to obtain 200 averaged $PLV_{i, shuffle(j)}$s. We defined the phase-locking statistic (PLS) as the proportion of surrogate averaged $PLV_{i, shuffle(j)}$s exceeding the original average $PLV_{i,j}$. Significant synchrony existed between the pairs of regions i and j only if the $PLS_{i,j}$ was <0.05 after false discovery rate (FDR) correction. Our regions of interest (ROIs) at specific time periods and frequency bands were
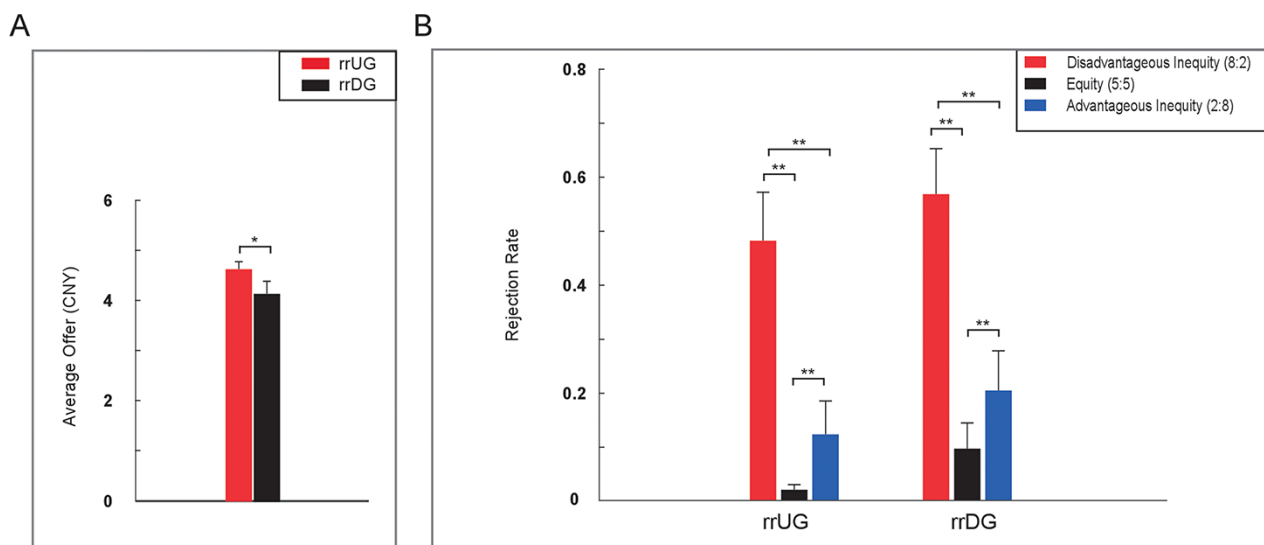
Fig. 2. Behavioral results. (A) Average offers proposed in the rrUG and in the rrDG. (B) The rejection rate of proposers was analyzed using a two-way rmANOVA with the introduction of punishment (rrUG *vs* rrDG) and condition (disadvantageous inequity, equity and advantageous inequity) as the within-subject factor. $*P < 0.05$, $**P < 0.01$.
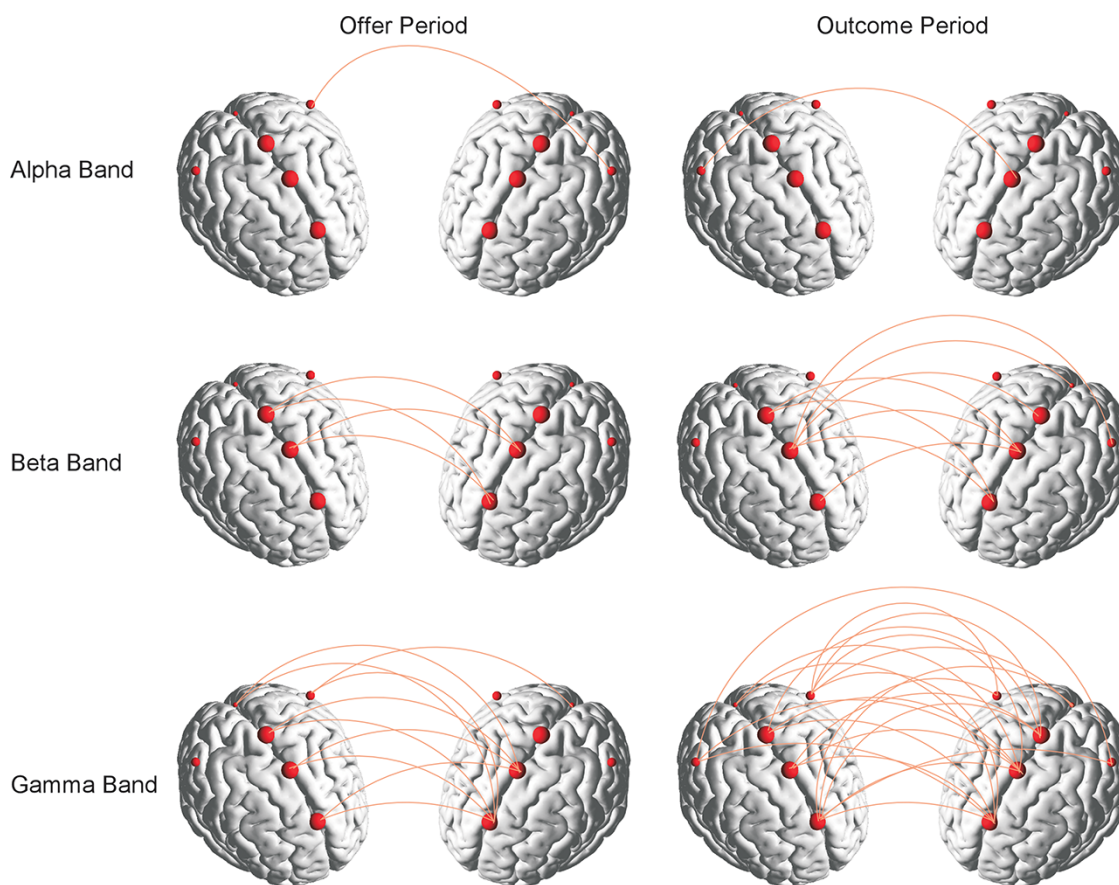

Fig. 3. ROIs of inter-brain synchronization. 576 PLSs (FDR-corrected) were examined for all 36 pairs of regions at two time periods and in four frequency bands for both of the games. In total, 47 ROIs with significant inter-brain synchronization were found in the alpha, beta and gamma bands in the rrUG or in the rrDG (all Ps < 0.05). For each pair of brains shown in the figure, the left one represents that of the proposer and the right one represents that of the responder.

therefore set to those pairs of regions with significant synchrony in either the rrUG or the rrDG. Further comparisons of inter-brain synchronization were then performed solely based on synchronization within the ROIs.

Results
Behavioral results

For each game, the mean offer was defined as the average of the offers provided by the proposer to the responder. A paired *t*-test

was conducted on the mean offers of the rrUG and rrDG. As illustrated in Figure 2, a significantly higher mean offer was made in the rrUG (mean ± SE, 4.64 ± 0.15 CNY) than that in the rrDG (mean ± SE, 4.16 ± 0.25 CNY; $P = 0.031$; $N = 20$).

The rejection rate of responders was defined as the proportion of the offers to be rejected by responders. We conducted a two-way repeated measures analysis of variance (rmANOVA) with the introduction of punishment (rrUG *vs* rrDG) and condition (disadvantageous inequity, equity and advantageous inequity) as within-subject factors and found a significant main effect of condition [$F (2, 38) = 33.93$, $P < 0.001$; partial $\eta^2 = 0.641$; $N = 20$]. However, we found no significant main effect of the introduction of punishment ($P > 0.1$), nor any significant interaction effect ($P > 0.1$).

Inter-brain synchronization results

PLVs were used to measure potential inter-brain synchronization between the EEG signals of the two interacting parties in the rrUG and the rrDG. The PLVs ranged from 0 to 1, with 1 indicating perfect phase synchrony. We examined all 576 PLSs (FDR-corrected). Based on the PLSs, 47 ROIs where significant inter-brain synchronization was found in either game were selected for further analyses. As illustrated in Figure 3, the ROIs selected in our study were as follows: (i) alpha band, offer period: LTP–LTP (with the former indicating the brain area of the proposer and the latter that of the responder); (ii) alpha band, outcome period: RTP–FC; (iii) beta band, offer period: FC–F, FC–FC, P–F and P–FC; (iv) beta band, outcome period: F–FC, FC–F, FC–FC, FC–P, FC–LTP, FC–O, P–F and P–FC; (v) gamma band, offer period: F–F, F–FC, FC–F, FC–FC, P–F, P–FC, LTP–F, LTP–O, RTP–F, O–F and O–FC; (vi) gamma band, outcome period: F–F, F–FC, F–P, F–LTP, F–RTP, F–O, FC–F, FC–FC, FC–P, P–F, P–FC, P–P, LTP–F, LTP–FC, LTP–P, LTP–RTP, LTP–O, RTP–F, RTP–FC, RTP–RTP, O–F and O–FC.

Using these ROIs, we first investigated whether the introduction of punishment impacted the inter-brain synchronization between the interacting parties. A two-way rmANOVA was conducted with the introduction of punishment and the ROI as the within-subject factors. We found a significant main effect of the introduction of punishment [$F(1, 19) = 15.899$, $P = 0.001$; partial $\eta^2 = 0.456$; power = 0.966], with a higher PLV for the rrDG (mean ± SE, 0.157 ± 0.007) than for the rrUG (mean ± SE, 0.126 ± 0.003; $P = 0.001$). In addition, there was a significant main effect of ROI [$F(46, 874) = 48.974$, $P < 0.001$; partial $\eta^2 = 0.720$; power = 1.000] and a significant interaction effect [$F(46\,874) = 3.919$, $P = 0.035$; partial $\eta^2 = 0.171$; power = 1.000].

We then performed paired *t*-tests for every ROI in the rrUG and the rrDG. Because the EEG data were recorded in six regions, FDR correction for multiple comparisons was applied. As illustrated in Figure 4, for 43 ROIs higher PLVs were found for the rrDG than for the rrUG (all Ps < 0.05). For the remaining four ROIs, the PLVs in the rrDG were similar to those in the rrUG.

The correlation between IBS and behavioral results

We correlated the rejection rate for disadvantageous inequal offer from responders with inter-brain synchronization in the rrUG and in the rrDG, respectively. As illustrated in Figure 5, we found a significant positive correlation between the rejection rate and inter-brain synchronization in the rrDG (Spearman $r = 0.449$, $P = 0.047$). However, we found no significant correlation between the rejection rate and inter-brain synchronization in the rrUG (Spearman correlation, $P > 0.1$). In addition, we correlated the rejection rate for advantageous inequal offers and equal offers

from responders with inter-brain synchronization in the rrUG and in the rrDG, respectively. However, we did not find any significant correlation (Spearman correlation, $P > 0.1$).

Discussion

Social norms are an essential part of our lives, because almost every unique moral behavior in human society is motivated by following social norms. This universal intrinsic motivation was termed 'moral motivation' by early economists (Bowles and Hwang, 2008). A large number of social psychological and public economic studies have demonstrated incentive policies such as external punishment crowd out moral motivation (Deci *et al.*, 1999; Gneezy and Rustichini, 2000a,b; Frey and Jegen, 2001; Lin and Yang, 2006; Holmås *et al.*, 2010). The inter-brain synchronization derives from the similar cognitive and emotional processes, such as moral motivation or mentalization. Therefore, that the introduction of external punishment reduced the moral motivation of proposers lowers inter-brain synchronization in the rrUG. Moreover, there is more mentalizing process in the rrUG than in the rrDG (Weiland *et al.*, 2012; Chen *et al.*, 2017), which increases the inter-brain synchronization between the two parties. As a result, the overall effect of external punishment on inter-brain synchronization is not clear. In the current study, we employed an EEG-based hyperscanning technique during the rrUG and the rrDG to investigate how the introduction of external punishment influences inter-brain synchronization between interacting dyads.

In line with a large number of studies in experimental economics, our behavioral results showed that the introduction of external punishment increased the monetary amounts offered by the proposers (see Camerer and Thaler, 1995; Camerer, 2003 for reviews). However, we found no significant effect of the introduction of punishment on the rejection rate of responders. This finding was different from Yamagishi *et al.* (2009), in which the monetary effect of rejection increased the rejection rate of responders for unequal offers (i.e. with the monetary effect of rejection, 48.1% of the participants rejected disadvantageous unfair offers in our study and 50–70% of the participants rejected disadvantageous unfair offers in their study; without the monetary effect of rejection, 56.8% of the participants rejected disadvantageous unfair offers in our study and 30–40% of the participants rejected disadvantageous unfair offers in their study). Such discrepancy might come from the fact that non-monetary effective rejection was without cost in our rrDG; however, it was costly in their private impunity game. Since responders rejected unequal offers without any cost in the rrDG, they increased their rejection rate, offsetting the negative effect from the non-monetary effect of rejection.

Importantly, significant inter-brain synchronization was observed in 47 ROIs in the alpha, beta and gamma bands. In ∼91.5% of the ROIs, inter-brain synchronization was significantly greater in the rrDG than in the rrUG, while in the remaining 8.5% of ROIs, inter-brain synchronization was comparable in the rrDG and the rrUG. These neural findings revealed a robust effect of the introduction of external punishment on inter-brain synchronization between the two interacting parties and provided insight into the potential mechanism underlying the generation and change of inter-brain synchronization during interactive decision-making. In the rrDG, the similar moral motivation of proposers and responders produces inter-brain synchronization between them. However, in the rrUG, the introduction of external punishment crowds out the moral motivation of proposers.
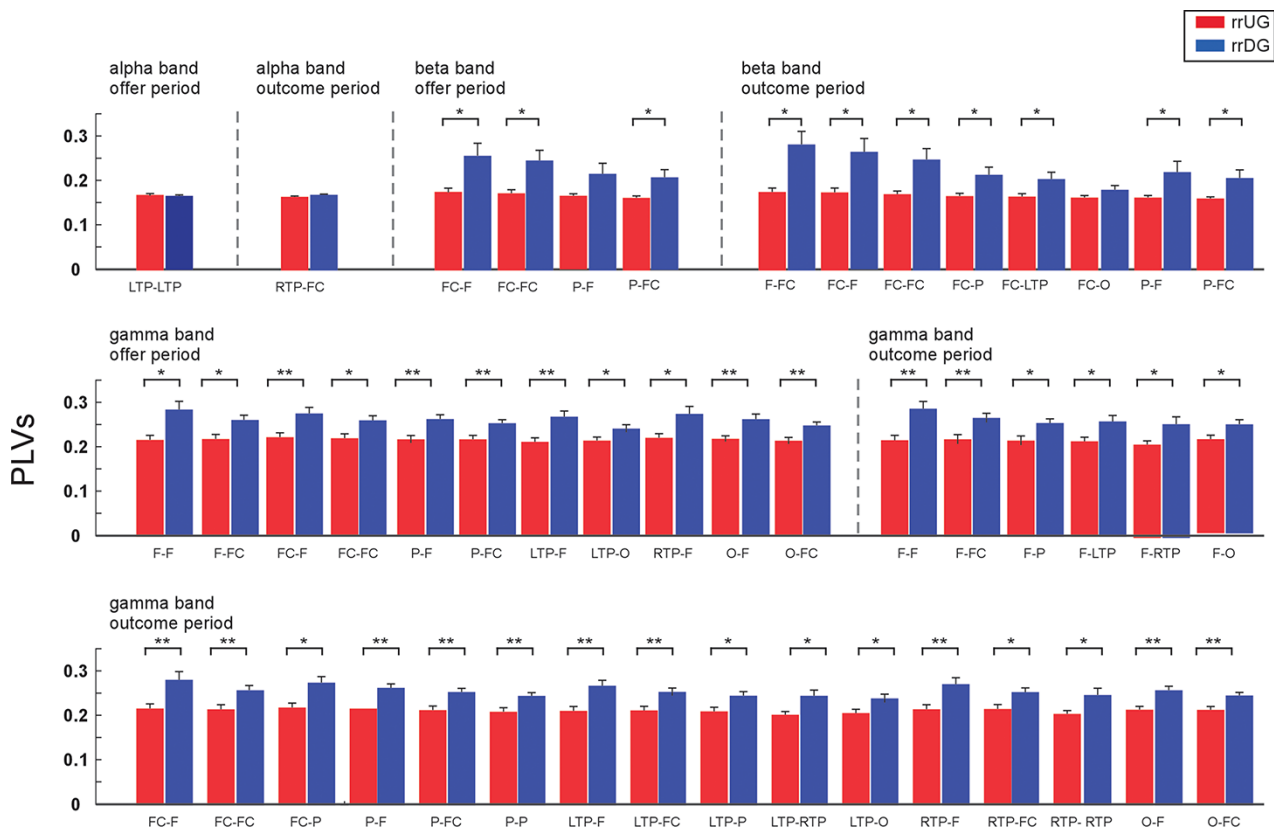
Fig. 4. Inter-brain synchronization of 47 ROIs in the rrUG and in the rrDG. For 43 of the ROIs (beta band, offer period: FC–F, FC–FC and P–FC; beta band, outcome period: F–FC, FC–F, FC–FC, FC–P, FC–LTP, P–F and P–FC; gamma band, offer period: F–F, F–FC, FC–F, FC–FC, P–F, P–FC, LTP–F, LTP–O, RTP–F, O–F and O–FC; gamma band, outcome period: F–F, F–FC, F–P, F–LTP, F–RTP, F–O, FC–F, FC–FC, FC–P, P–F, P–FC, P–P, LTP–F, LTP–FC, LTP–P, LTP–RTP, LTP–O, RTP–F, RTP–FC, RTP–RTP, O–F and O–FC), higher PLVs were found for the rrDG than for the rrUG. All Ps were FDR corrected. *$P < 0.05$, **$P < 0.01$.
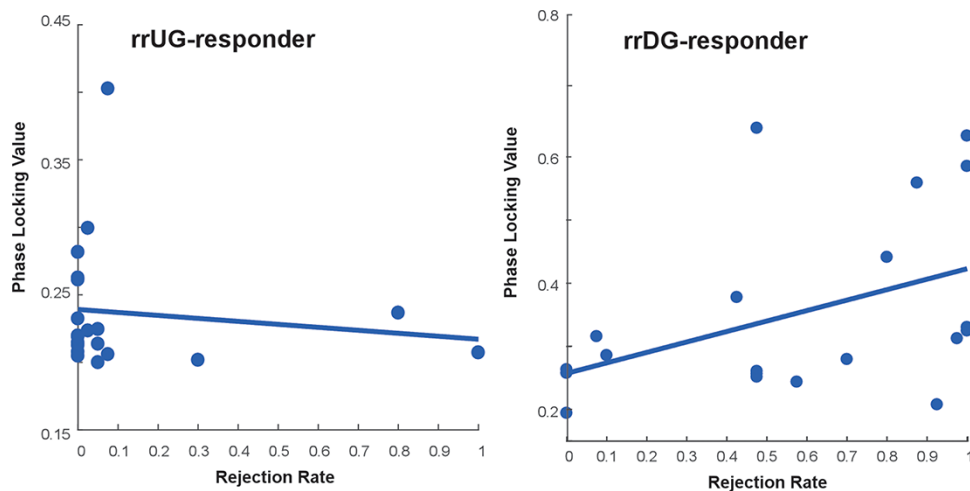

Fig. 5. The correlation between the rejection rate from responders for disadvantageous inequal offer and inter-brain synchronization in the rrDG and in the rrUG.

Previous imaging experiments have provided neural evidence that external punishment changes the cognitive and emotional processes of proposers in the UG (see Lee, 2008 for a review). A series of studies demonstrated that external punishment leads to proposers making decisions motivated by a fear of monetary loss, instead of a concern for complying with internalized social norms (Spitzer *et al.*, 2007; Weiland *et al.*, 2012; Chen *et al.*,

2017). As this fear is not the motivation with which the responder makes decisions, the introduction of external punishment in the rrUG undermined inter-brain synchronization. These results also suggested that the overall effect of punishment institution on inter-brain synchronization is negative even through there is more mentalizing process in the rrUG. In other words, the negative effect of moral motivation crowded out on inter-brain

synchronization exceeds the positive effect of mentalization on inter-brain synchronization.

We found a significant positive correlation between the rejection rate from responders for disadvantageous inequal offers and inter-brain synchronization in the rrDG. In the rrDG, the only reason for responders to reject the disadvantageous inequal offer is the moral motivation. The higher inter-brain synchronization therefore represents the stronger moral motivation, providing an evidence to support the idea that the moral motivation of proposers and responders leads to inter-brain synchronization between them. This mechanism was also supported by our event-related potential analysis (see Supplementary Figure S2). However, we did not find any significant correlation between rejection rate for disadvantageous inequal offers from responders and inter-brain synchronization in the rrUG. In the rrUG, responders were not only motivated by moral disgust/satisfaction but also motivated by self-interest. As the rejection rate is impacted by the heterogeneous preference for money, we hardly found a clear correlation between the rejection rate and inter-brain synchronization.

To our knowledge, three other hypotheses have been proposed to explain inter-brain synchronization during interactive decision-making: the mutual phase resetting hypothesis, the cooperative interaction hypothesis and the similar task hypothesis. The mutual phase resetting hypothesis proposes that salient social signals produced by each partner act as synchronization triggers to reset the phase of ongoing oscillations in the other partner and increase interpersonal neural synchronization within dyads (Leong *et al.*, 2017). If the mutual phase resetting hypothesis were able to explain our results, we would have found a significantly higher inter-brain synchronization during the outcome period than during the offer period, because the outcome is a more salient social signal than the offer provided by the computer. The cooperative interaction hypothesis suggests that neural activity is more synchronized when the partners participate in cooperative interactions (Balconi and Vanutelli, 2016; Mu *et al.*, 2016, 2017; Jahng *et al.*, 2017; Hu *et al.*, 2017). Given that there is no cooperative interaction between the proposers and responders in the rrUG and the rrDG, the cooperative interaction hypothesis cannot be used to explain our results. Moreover, the similar task hypothesis suggests that inter-brain synchronization can be induced by performing the same task, such as listening to the same music (Abrams *et al.*, 2013) or watching the same movie (Nummenmaa *et al.*, 2012). However, the decision tasks performed by proposers and responders are quite different both in the rrUG and in the rrDG. As a result, the similar task hypothesis is also unable to explain our results. Taken together, none of these hypotheses can explain the higher inter-brain synchronization in the rrDG than in the rrUG.

The mechanism proposed to underlie inter-brain synchronization in this study can also explain the findings of other interactive decision-making hyperscanning studies. Using cooperative tasks, a number of studies have found that the inter-brain synchronization of dyadic partners adopting cooperative strategies is greater than that of dyadic partners employing defection-based strategies (Jahng *et al.*, 2017; Hu *et al.*, 2017; Zhang *et al.*, 2019). In cooperative tasks, conditional cooperation is regarded as a social norm that drives people to cooperate with each other (see Fehr and Schurtenberger, 2018 for a review). When partners cooperate in the Prisoner's Dilemma Game (PDG) or in other tasks, they are driven by the moral motivation to follow conditional cooperation norms; when they defect, however, they are driven by self-interest. Based on their finding, it should be reasoned that

the enhanced moral motivation increased the inter-brain synchronization of two players. Jahng *et al.* (2017) investigated the effect of face-to-face contact on inter-brain synchronization in the PDG. In the face-to-face condition, the wallboard was removed to allow participants to face each other, while in the face-blocked condition, a wallboard remained in place. They found that the cooperation level and inter-brain synchronization of dyadic partners were significantly higher in the face-to-face condition than in the face-blocked condition. A previous experimental study demonstrated that face-to-face contact strengthens the moral motivations of the interacting parties (Hoffman *et al.*, 1994). These findings suggest that the enhanced moral motivation increased inter-brain synchronization in the two parties. Hu *et al.* (2017) manipulated the payoff contest in terms of the Cooperation index (CI) and found a higher inter-brain synchronization in the high CI condition than in the low CI condition. Compared with the high CI condition, participants earned more by defecting in the low CI condition. Given that the monetary incentive to defect crowds out the moral motivation, participants were more motivated by cooperative norms in the high CI condition than in the low CI condition. As a consequence, they acted more cooperatively and their neural activities were more synchronized in the high CI condition. A number of other studies have focused on the effect of sociality on inter-brain synchronization (Astolfi *et al.*, 2015; Mu *et al.*, 2016, 2017; Hu *et al.*, 2017). Astolfi *et al.* (2015) investigated a third-party punishment paradigm involving three subjects: the dictator, the receiver and the observer. The dictator and the receiver played a DG, in which the decision of the dictator was shown to both the receiver and the observer, and then the observer was able to exact a costly punishment on the dictator. The role of the dictator was played by a computer (PC condition) for half of the trials and by an actor (agent condition) for the other half of the trials. Participants were informed whether the dictator role was being played by the computer. Mu *et al.* (2016, 2017) used a coordination game in which two players either played with each other (coordination task) or played separately with a computer (control task). Hu *et al.* (2017) set up the interaction to involve either another human partner (H–H condition) or a machine (H–M condition), although the actions of the 'computer' were actually still carried out by their partner in the H–M condition. In all of these studies, a higher inter-brain synchronization was found when participants interacted with each other (i.e. in the agent condition, the coordination task and the H–H condition). Social norms make sense when people interact with others in a group, as opposed to interacting with a computer. As a result, participants were motivated to simultaneously comply with social norms only when they were interacting with each other or when they were observing a social interaction between two partners, and in such situations, their neural activities were more synchronized. Given the general ability to explain inter-brain synchronization of our mechanism in the field of social interaction, it should be widely used in future studies.

Despite the use of a number of controls in the design of the present study, there were several limitations. First, only male participants were recruited for our study. Therefore, the interpretation of our results is restricted to male participants and caution should be taken in generalizing the results to female or mixed-sex participants. Future research is warranted to test the effect of external punishment on both female and mixed-sex participants. Second, since EEG hyperscanning provides a limited spatial resolution, future neuroimaging studies could help to reveal more precise information regarding the neural locations of the inter-brain synchronization. Third, the mechanism we used to

explain the effect of external punishment on inter-brain synchronization is only one possible mechanism. Future research should aim to find behavioral indexes of emotional or cognitive states to test this mechanism directly.

In summary, our study provided evidence that the introduction of external punishment could reduce inter-brain synchronization during social interaction. Moreover, we found a significant positive correlation between the moral motivation of responders and inter-brain synchronization. We propose a possible mechanism underlying these results. In the rrDG, the similar moral motivation of both proposers and responders results in inter-brain synchronization. However, in the rrUG, the reduction of inter-brain synchronization was driven by the crowd-out effect of punishment institution on the moral motivation of proposers. These findings contribute to understanding the negative effect of punishment institution and shed light on the inter-brain mechanism underlying social interaction.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary data

Supplementary data is available at *SCAN* online.

Data availability

The datasets generated during the current study are available in Mendeley data.

Author Contributions

C.Z., J.P., Y.W. and J.L. designed the experiment. C.Z., J.L. and J.P. carried out the experiment, analyzed the data and wrote the paper. C.Z., J.P., J.L. and Y.W. revised the paper. J.P. and J.L. contributed equally to this work.

References

Abrams, D.A., *et al.* (2013). Inter-subject synchronization of brain responses during natural music listening. *European Journal of Neuroscience*, **37**, 1458–69.

Astolfi, L., *et al.* (2010). Simultaneous estimation of cortical activity during social interactions by using EEG hyperscannings. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Buenos Aires, Argentina, IEEE, 2814–7.

Astolfi, L., *et al.* (2011). Imaging the social brain by simultaneous hyperscanning during subject interaction. *IEEE Intelligent Systems*, **26**, 38–45.

Astolfi, L., *et al.* (2015). Investigating the neural basis of empathy by EEG hyperscanning during a third-party punishment. In: 2015 Annual International Conference of the Engineering in Medicine and Biology Society. Milano, Italy, IEEE, 5384–7.

Balconi, M., Vanutelli, M.E. (2016). Competition in the brain. The contribution of EEG and fNIRS modulation and personality effects in social ranking. *Frontiers in Psychology*, **7**, 1587.

Bowles, S., Hwang, S.H. (2008) Social preferences and public economics: mechanism design when social preferences depend on incentives. *Journal of Public Economics*, **92**(8–9), 1811–20.

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

Camerer, C.F., Thaler, R.H. (1995). Anomalies: ultimatums, dictators and manners. *Journal of Economic Perspectives*, **9**, 209–19.

Chen, Y.H., Chen, Y.C., Kuo, W.J., Kan, K., Yang, C.C., and Yen, N.S. (2017). Strategic motives drive propsers to offer fairly in Ultimatum games: an fMRI study. *Scientific reports*, **7**(1), 1–11.

Chew, S.H., *et al.* (2013). Sex-hormone genes and gender difference in ultimatum game: experimental evidence from China and Israel. *Journal of Economic Behavior and Organization*, **90**, 28–42.

Deci, E.L., *et al.* (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, **125**(6), 627.

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**, 9–21.

Dumas, G., *et al.* (2010). Inter-brain synchronization during social interaction. *PLoS One*, **5**, e12166.

Eckel, C.C., Grossman, P.J. (1996) Altruisminanonymous dictator games. *Games and Pconomic behavior*, **16**(2), 181–91.

Elster, J. (1999). *Alchemies of the Mind. Rationality and the Emotions*. Cambridge: Cambridge University Press.

Fehr, E., Fischbacher, U., Gächter, S. (2002) Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, **13**(1), 1–25.

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

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

Fehr, E., Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, **2**, 458–68.

Frey, B.S., Jegen, R. (2001) Motivation crowding theory. *Journal of Economic Surveys*, **15**(5), 589–611.

Gintis, H. (2000) Strong reciprocityand human sociality. *Journal of theoretical biology*, **206**(2), 69–179.

Gneezy, U., Rustichini, A. (2000a) A fine is a price. *The Journal of Legal Studies*, **29**(1), 1–17.

Gneezy, U., Rustichini, A. (2000b) Pay enough or don't pay at all. *Quarterly Journal of Economics*, **115**(3), 791–810.

Güth, W., *et al.* (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, **3**, 367–88.

Hoffman, E., *et al.* (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, **7**(3), 346–80.

Holmås, T.H., *et al.* (2010). Does monetary punishment crowd out pro-social motivation? A natural experiment on hospital length of stay. *Journal of Economic Behavior and Organization*, **75**(2), 261–7.

Hu, Y., *et al.* (2017). Inter-brain synchrony and cooperation context in interactive decision making. *Biological Psychology*, **133**, 54–62.

Jahng, J., *et al.* (2017). Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the prisoner's dilemma game. *NeuroImage*, **157**, 263–74.

Koike, T., *et al.* (2015). Hyperscanning neuroimaging technique to reveal the 'two-in-one' system in social interactions. *Neuroscience Research*, **90**, 25–32.

Krupka, E.L., Weber, R.A. (2013) Identifying social norms using coordination games: why does dictator game sharing vary? *Journal of the European Economic Association*, **11**(3), 495–524.

Lachaux, J.P., *et al.* (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, **8**, 194–208.

Lee, D. (2008). Game theory and neural basis of social decision making. *Nature Neuroscience*, **11**, 404–9.

Leong, V., *et al.* (2017). Speaker gaze increases information coupling between infant and adult brains. *Proceedings of the National Academy of Sciences*, **114**, 13290–5.

Lin, C.C., Yang, C.C. (2006) Fine enough or don't fine at all. *Journal of Economic Behavior and Organization*, **59**(2), 195–213.

Liu, D., *et al.* (2018). Interactive brain activity: review and progress on EEG-based hyperscanning in social interactions. *Frontiers in Psychology*, **9**, 1862.

Montague, P.R., *et al.* (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *NeuroImage*, **16**, 1159–64.

Mu, Y., *et al.* (2016). Oxytocin enhances inter-brain synchrony during social coordination in male adults. *Social Cognitive and Affective Neuroscience*, **11**, 1882.

Mu, Y., *et al.* (2017). The role of gamma interbrain synchrony in social coordination when humans face territorial threats. *Social Cognitive and Affective Neuroscience*, **12**, 1614–23.

Nummenmaa, L., *et al.* (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proceedings of the National Academy of Sciences*, **109**, 9599–604.

Oostenveld, R., *et al.* (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, **2011**, 1.

Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In: Advances in Neural Information-Processing Systems 8: Proceedings of the 1995 Conference, Vol. 8. MIT press, 145.

Spitzer, M., *et al.* (2007). The neural signature of social norm compliance. *Neuron*, **56**(1), 185–96.

Tang, H., *et al.* (2016). Interpersonal brain synchronization in the right temporo-parietal junction during face-to-face economic exchange. *Social Cognitive and Affective Neuroscience*, **11**, 23–32.

Weiland, S., *et al.* (2012). Neural correlates of fair behavior in interpersonal bargaining. *Social Neuroscience*, **7**, 537–51.

Yamagishi, T., *et al.* (2009). The private rejection of unfair offers and emotional commitment. *Proceedings of the National Academy of Sciences*, **106**(28), 11520–3.

Zhang, D., *et al.* (2019). The dynamics of belief updating in human cooperation: findings from inter-brain ERP hyperscanning. *NeuroImage*, **198**, 1–12.