Social Cognitive and Affective Neuroscience, 2020, 1315–1324

doi:10.1093/scan/nsaa154
Advance Access Publication Date: 13 November 2020
Original Manuscript

Interpersonal brain synchronization under bluffing in strategic games

Zhihao Wang,[1] Yiwen Wang,[2,3,*] Xiaolin Zhou,[4,5,6,7,8,*] and Rongjun Yu[9,10,11,*]

[1]School of Psychology, Center for Studies of Psychological Application and Key Laboratory of Mental Health and Cognitive Science of Guangdong Province, South China Normal University, Guangzhou, 510631, China, [2]School of Economics and Management, Fuzhou University, Fuzhou, 350108, China, [3]Institute of Psychological and Cognitive Sciences, Fuzhou University, Fuzhou, 350108, China, [4]School of Psychology and Cognitive Science, East China Normal University, Shanghai, 200062, China, [5]School of Psychological and Cognitive Sciences, Peking University, Beijing, 100871, China, [6]Beijing Laboratory of Behaviour and Mental Health, Peking University, Beijing, 100871, China, [7]Institute of Psychological and Brain Sciences, Zhejiang Normal University, Zhejiang, 321004, China, [8]PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China, [9]Department of Management, Hong Kong Baptist University, Hong Kong, 1123, China, [10]Department of Sport, Physical Education and Health, Hong Kong Baptist University, Hong Kong, 1123, China, and [11]Department of Physics, Hong Kong Baptist University, Hong Kong, 1123, China

Correspondence should be addressed to Rongjun Yu, Department of Management, Hong Kong Baptist University, Hong Kong, 1123, China. E-mail: rongjunyu@hkbu.edu.hk; Yiwen Wang, School of Economics and Management, Fuzhou University, Fuzhou 350108, China. E-mail: wangeven@126.com; Xiaolin Zhou, School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China. E-mail: xz104@pku.edu.cn

Abstract

People commonly use bluffing as a strategy to manipulate other people's beliefs about them for gain. Although bluffing is an important part of successful strategic thinking, the inter-brain mechanisms underlying bluffing remain unclear. Here, we employed a functional near-infrared spectroscopy hyperscanning technique to simultaneously record the brain activity in the right temporal-parietal junction in 32 pairs of participants when they played a bluffing game against each other or with computer opponents separately. We also manipulated the penalty for bluffing (high vs low). Under the condition of high relative to low penalty, results showed a higher bluffing rate and a higher calling rate in human-to-human as compared to human-to-computer pairing. At the neural level, high relative to low penalty condition increased the interpersonal brain synchronization (IBS) in the right angular gyrus (rAG) during human-to-human as compared to human-to-computer interaction. Importantly, bluffing relative to non-bluffing, under the high penalty and human-to-human condition, resulted in an increase in response time and enhanced IBS in the rAG. Participants who bluffed more frequently also elicited stronger IBS. Our findings support the view that regions associated with mentalizing become synchronized during bluffing games, especially under the high penalty and human-to-human condition.

Key words: bluffing; fNIRS hyperscanning; rAG; mentalizing

Received: 22 July 2019; **Revised:** 2 September 2020; **Accepted:** 12 November 2020

Introduction

Bluffing is a commonly used strategy to manipulate other people's beliefs about ourselves, usually for personal gains (Bhatt *et al.*, 2010). Bluffing takes place often in negotiation and in the context of a game, such as poker, where deception is consented to in advance by the players (Carson, 1993). Specifically, deceiving other players into thinking one has a higher or lower ranking hand than the actual one is often socially acceptable and is commonly expected as a tactic in social negotiations (Carson, 1993). For instance, one can tell a prospective house-buyer that the minimum acceptable price for the house is $90 000, although in reality it is as less as $80 000. Although it is ubiquitous in social life, the neural mechanisms of bluffing remain to be explored.

Few studies have examined the neural activations of bluffing in bargaining tasks (Bhatt *et al.*, 2010; Carter *et al.*, 2012; Piva *et al.*, 2017). In one study, participants were divided into three groups based on strategic depth: incrementalists (honest group), conservative (neutral group) and strategists (bluffing group). Results showed that the bluffing group had stronger activation than other groups in the right temporal-parietal junction (rTPJ) when they bluffed, suggesting a critical role of the rTPJ in bluffing in bargaining games (Bhatt *et al.*, 2010). Although single-brain neuroimaging studies can reveal regional activity in an individual brain, investigation of dual-brain networks can further reveal the neural coupling between two brains and capture the dynamic features that are unique to social interactions (Hari *et al.*, 2015). We thus studied bluffing in interactive situations by examining interpersonal brain synchronization (IBS).

The rTPJ is a candidate brain area to generate IBS, given that rTPJ has been demonstrated to track the signal of deception (Baumgartner *et al.*, 2009; Jenkins *et al.*, 2016; Zhang *et al.*, 2017). In deception, the rTPJ has been implicated in theory of mind (ToM) processes, such as representing others' beliefs and future actions (Jenkins *et al.*, 2016). It is also true that the rTPJ plays a crucial role in bluffing due to its unique link with ToM reasoning (Saxe and Wexler, 2005; Bhatt *et al.*, 2010). The rTPJ encompasses a large area and can be roughly divided into four subregions: the angular gyrus (AG), supramarginal gyrus (SMG), superior temporal gyrus (STG) and inferior parietal lobe (IPL; Carrington and Bailey, 2009; Carter and Huettel, 2013; Donaldson *et al.*, 2015; Piva *et al.*, 2017). However, the specific functions of each subarea underlying bluffing are still not well known.

The power of hyperscanning in understanding brain synchronizations has been evidenced in social neuroscience research (Babiloni and Astolfi, 2014; Hari *et al.*, 2015). In the past two decades, many researchers have sought to identify the neural mechanisms in dual-brain or multi-brain networks (Duan *et al.*, 2015). They used a technique developed in 2002, termed 'hyperscanning', to simultaneously record two or more brain activities (Montague *et al.*, 2002). Furthermore, functional near-infrared spectroscopy (fNIRS) hyperscanning is not only more convenient and cost-effective relative to functional magnetic resonance imaging (fMRI) hyperscanning but also has better spatial resolution than electroencephalogram (EEG) hyperscanning (Cui *et al.*, 2012). Importantly, this technique has better ecological validity and can be applied to measure personal interactions in a natural context (Liu *et al.*, 2017). Many fNIRS hyperscanning studies used IBS as an important index to assess mentalizing behavior (Jiang *et al.*, 2012; Wang *et al.*, 2018). For example, deception induced stronger IBS than honesty in real face-to-face situations, suggesting that more mentalizing processes were involved during deception (Zhang *et al.*, 2017). Similarly, stronger IBS may be involved in bluffing relative to non-bluffing.

In addition to the investigation of inter-brain mechanism under bluffing, some factors may affect bluffing behaviors. Given that the outcome for bluffing is self-concerned and the goal of bluffing is personal gains (Bhatt *et al.*, 2010), it is of great importance to examine the impact of penalty on bluffing. Penalty, a punishment imposed for deception being caught, has shown great effects on deception (Gneezy, 2005; Tomash and Reed, 2015). As a monetary motivation to avoid penalty, people show greater skin conductance responses in face of high penalty, reflecting greater motivation to avoid penalty in deception (Tomash and Reed, 2015). On the other hand, manipulation of others' beliefs is the key component of bluffing (Bhatt *et al.*, 2010). Rapid identification of the interacting agent (herein, competing with human or computer) in the social context has long been considered crucial for social cognition (Carter *et al.*, 2012; Hu *et al.*, 2018; Hackel *et al.*, 2020). It has been demonstrated that learning from humans relative to non-humans (slot machines) involves more abstract trait inferences during reinforcement learning (Hackel *et al.*, 2020). It is especially the case when bluffing is involved, revealing different bluffing patterns between human-to-human pairing and human-to-computer pairing (Carter *et al.*, 2012). Particularly, the rTPJ tracks the decisions of social agents (Carter *et al.*, 2012) and is involved in IBS in social interactions (Piva *et al.*, 2017). To advance our understanding of bluffing, we investigate how penalty and agents modulate bluffing behavior and the associated inter-brain mechanisms.

In the current study, we simultaneously recorded brain activity in the rTPJ of each pair playing a poker game (Carter *et al.*, 2012). We aim to examine how penalty and agents modulate bluffing behavior and its inter-brain mechanisms by manipulating the penalty for bluffing (either high or low) and the type of opponent (human or computer). We hypothesized that human pairs, under penalty, would synchronize best than in other contexts. We expected that bluffing would elicit increased reaction time (RT) and enhanced IBS in the rTPJ.

Methods

Participants

Sixty-four healthy adults (age: mean ± s.d. = 21.53 ± 1.95 years) participated in this study. Participants were randomly assigned to 32 same-gender pairs (16 female pairs), and the members of each pair were strangers. The sample size was determined based on the medium effect size by G*power v3.1 (Faul *et al.*, 2009). Thirty pairs were needed to detect a reliable effect [Cohen $f = 0.25$, $\alpha = 0.05$, $1 - \beta = 0.9$, analysis of variance (ANOVA): repeated measures, within factors; Faul *et al.*, 2009]. All participants were right-handed (according to the Edinburgh Handedness Inventory; Oldfield, 1971), had normal or corrected-to-normal vision, and reported no history of neurological or psychiatric conditions. All participants provided written informed consent according to protocols approved by the institutional review board of the South China Normal University.

Tasks and procedures

We adopted a modified single-card poker game (Carter *et al.*, 2012). A pair of participants sat opposite to each other at either side of the table, each facing a computer display.

The two computer screens were either separated by a board (face-blocked) or not (face-to-face; between-subject variable; Donaldson et al., 2015). In the face-to-face situation, participants may not necessarily look at one another the whole time throughout the experiment (Donaldson et al., 2015). Consistent with previous hyperscanning studies using two-person games (Donaldson et al., 2015), participants were randomly assigned to one of two different fixed roles. The 'banker' who had a card (high or low) had to decide whether to bet or fold, whereas the 'follower' had to decide whether to call or fold after knowing the banker's decision. The definitions of poker terms are as follows. Bet is an action of bankers. Once a bet has been made, bankers have a chance to win. Fold is an action of both bankers and followers. Players who fold discard their cards and cannot win. Call is an action of followers. Followers can call if bankers have bet. In this game, each participant was instructed to win as many tokens as possible. As an incentive, the participant who won more tokens received 40 RMB (∼$6), whereas the loser received 20 RMB (∼$3). Game rules are shown as follows (Figure 1B). The banker has two options (bet or fold), while the follower has two options (call or fold). When the banker received a high card, if the banker bets, they remain in the game and they can win regardless of the follower folding or calling. If the banker folds, they lose and the follower wins without action. When the banker received a low card, if the banker bets, they remain in the game and they can win if the follower folds; if the follower calls, the banker will lose. If the banker folds, they cannot win, but they also do not risk losing.

In each trial (Figure 1A), a central fixation was first shown on the screen for 800−1200 ms for the banker and follower. Subsequently, a card, either high or low, was randomly presented to the banker for 4 s while the follower waited (a 'Wait' word was shown in the center of the screen). Then, the banker had to decide to bet or fold (a 'Bet?' word was shown in the center of the screen for both the banker and follower). Then the banker's choice (bet or fold) was shown to the banker while a 'Call?' word was shown to the follower if the banker bets. Otherwise, a 'Bet?' word was shown again for the follower. Specifically, if the banker placed a bet, the follower had to decide to call or fold. If the banker folded, the follower would see the feedback after 1 s. The display of the feedback (i.e., 'lose 18' for the banker and 'win 18' for the follower) would last for 2 s and be followed by a blank screen for another 2 s for both the banker and follower. Please note that there was no time limitation in both the betting and the calling period. We used high/low card rather than specific numbers in cards to control for potential confound of different subjective values of probability (Rangel et al., 2008). To be consistent with the previous study (Carter et al., 2012), we defined 'bluffing' as placing a bet on receiving a low card, and 'not bluffing' as folding on receiving a low card. All experimental procedures were presented using E-prime 2.0.

We designed a 2 (bluffing penalty: high vs low) × 2 (opponent: human-to-human vs human-to-computer) within-subjects experiment to explore the moderating context for bluffing. We operated the bluffing penalty by controlling the monetary ratio between disclosed and successful bluffing. Specifically, we defined b/a value of 9.0 as the high bluffing penalty and b/a value of 1.5 as the low bluffing penalty (for details, see *Supplementary material*).

In the human-to-human condition, both the banker and the follower were human participants playing against each other in the same game. In the human-to-computer condition, each member of the pair played separately against a computer, with one member assigned to be a banker playing against one computer (follower) and the other member assigned to be a follower playing against another computer (banker). Please note that we defined a trial as a bluffing or non-bluffing trial according to the banker's choice. That is, when the banker bluffed, it is a bluffing trial, and when the banker did not bluff, it is a non-bluffing trial.

Participants were given instructions about the rules of the game. They completed several practice rounds before the formal experiment started. For each pair of participants, the baseline was a 3 min resting-state block, during which participants were required to relax their mind and keep still with their eyes closed. (Jiang et al., 2012 2015; Hu et al., 2017). Four blocks of trials followed: (i) high penalty and human-to-human, (ii) low penalty and human-to-human, (iii) high penalty and human-to-computer and (iv) low penalty and human-to-computer. Each block, lasting ∼6 min, included 30 trials (randomly generated with 15 high-card trials and 15 low-card trials). Note that each block commenced with an instruction cue about the type of opponent and the amplitude of penalty. The order of the four blocks was counterbalanced across each pair. Before and after each block, a blank screen was displayed for 30 s to keep participants in a steady state. No discussion was allowed during the whole experiment.

Subjective measurements

Although no discussion was allowed during the whole experiment, participants may use their first impressions to guide their choices. It has been documented that facial attractiveness influenced first impression and deception (Zebrowitz et al., 2018). To control for the potential confound of the first impression, upon arrival, we asked the participants to rate the relative appearance and competence of their opponent using a nine-point Likert scale ranging from '1' (not very much) to '9' (very much). For example, 'I think the opponent is handsome/beautiful'. They also completed the interpersonal reactivity index (IRI; Davis, 1983) scale and the risk-taking propensity scale (Wang et al., 2019). The IRI measures the individuals' ability to empathize with partners' perspectives and feelings. Of the four seven-item subscales of IRI, comprising perspective taking (PT), empathetic concern (EC), fantasy (F) and personal distress (PD), we focused mainly on PT and EC. The scores of each subscale were summed for each participant. After experiments, participants were debriefed about their beliefs about whether the opponent was a human or a computer in the specified condition.

fNIRS data acquisition

An ETG-4000 optical topography system (Hitachi Medical Corporation, Japan) was used to record oxyhemoglobin (HBO) and deoxyhemoglobin (HBR) concentrations for each participant. The absorption of near-infrared light (wavelengths: 695 and 830 nm) was measured at a sampling rate of 10 Hz. Sixteen optodes (eight emitters and eight detectors, 3 cm optode separation) were separately attached to the right hemisphere of each participant in a 4 × 4 patch to cover the rTPJ (Donaldson et al., 2015), forming 24 measurement channels (CHs) where P6 and CP6 were located at an emitter and a detector respectively (Figure 2) according to the 10–20 EEG system.

The anatomical positions of optodes in relation to standard head landmarks, including inion; nasion; top center, Cz; AL

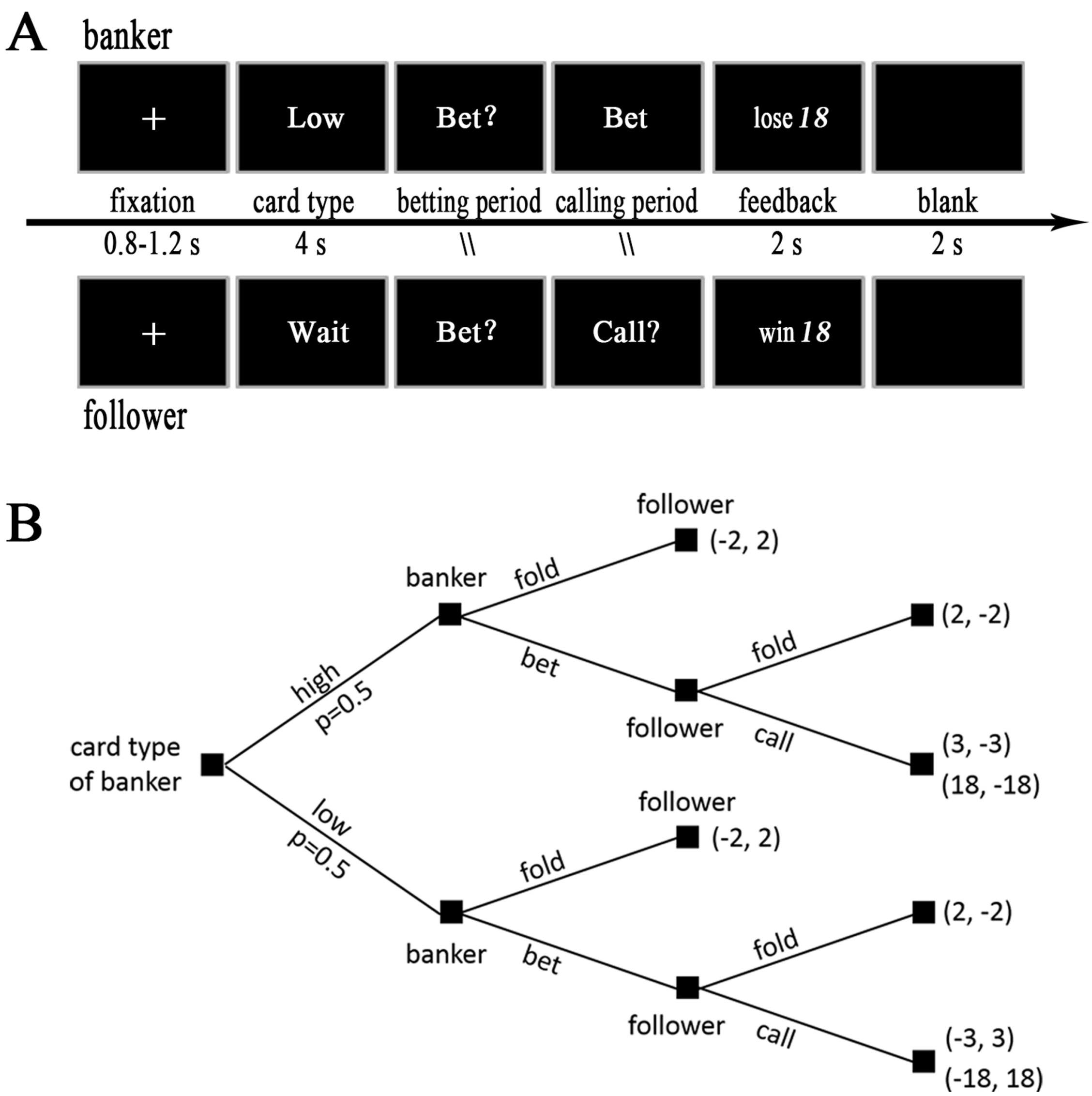

Fig. 1. (A) Experimental procedure. Events and time flow in a trial. (B) Game rules. The sequences of actions were represented in a game tree format. First, the banker had an equally likely chance to get either a high card or a low card. Then, the banker had to decide whether to bet or fold. Next, if the banker decided to bet, the follower had to decide to call or fold. Otherwise, no further decision was required. Finally, the outcomes were displayed. Each array represented the payoff for the banker (the former) and the follower (the latter).

(left pre-auricular point); and AR (right pre-auricular point), were measured by a 3D electromagnetic tracking device (FASTRAK; Polhemus, USA) after the experiment. We then used SPM_NIRS software (Singh *et al.*, 2005; YE *et al.*, 2009) with MATLAB to obtain the Montreal Neurological Institute coordinates. Dividing covered regions into several cerebral cortexes, we identified the regions of interest (ROIs: AG, SMG, IPL and STG), which are listed in Table 1. We focused on HBO analyses since HBO is more sensitive to changes in cerebral blood flow than HBR (Cui *et al.*, 2011).

Table 1. Regions of interest (ROIs), channel numbers, and mean Montreal Neurological Institute (MNI) coordinates of regions of interest

		Mean MNI coordinates		
ROIs	CHs	x	y	z
Superior temporal gyrus	8, 9, 12, 13	68	−43	22
Inferior parietal gyrus	19	59	−25	55
Supramarginal gyrus	15, 16	65	−30	44
Angular gyrus	20, 21	41	−69	53

Data analysis

Response time (RT), bluffing rate, calling rate and the questionnaires were collected. We used SPSS 23.0 to conduct statistical analyses, with alpha set to $P < 0.05$ (two-tailed). Partial eta squared ($\eta_p{}^2$) and Cohen *d* were used to assess effect size.

IBS extraction. Following previous studies on fNIRS hyperscanning (Cui *et al.*, 2012; Jiang *et al.*, 2015; Donaldson *et al.*, 2015; Liu *et al.*, 2016), the data were not pre-processed when computing wavelet transform coherence (WTC). WTC has been widely used in fNIRS hyperscanning studies (Cui *et al.*, 2012;

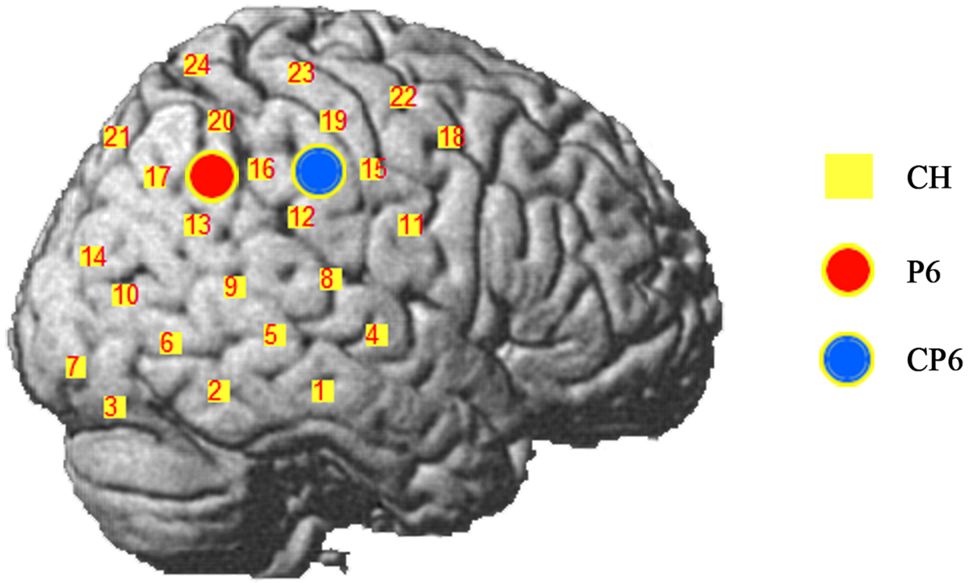

Fig. 2. fNIRS setup. A 4 × 4 patch placed on and cover the right temporal-parietal junction, forming 24 measurement channels (CHs) where P6 and CP6 were located at an emitter and a detector respectively according to the 10–20 EEG system.

Jiang *et al.*, 2015; Donaldson *et al.*, 2015; Liu *et al.*, 2016). Through time–frequency decomposition, WTC measures cross-correlation between two time series to capture synchronous activities, which might not be discoverable with traditional time series analysis [i.e., Fourier analysis; for details, see Grinsted *et al.*, (2004)]. Before data analysis, we checked data quality by visual inspection and did not remove any CHs or participant pairs. Participants were kept in a steady state during resting periods lasting 30 s before the start of each block. Signals during each of the resting periods were removed during pre-processing.

Considering the RT of bankers (averaged RT across four conditions: mean ± s.d. = 1129.8 ± 917.87 ms; for RT in each condition, see *Supplementary material*), the period of interest was identified as between 5 and 7 s (0.143 and 0.2 Hz, respectively) within the betting period. Specifically, we focused on the 5–7 s segments in the frequency domain and the whole trial period in the time domain. This is consistent with previous studies of fNIRS hyperscanning (Cui *et al.*, 2012; Donaldson *et al.*, 2015; Zhang *et al.*, 2017; Wang *et al.*, 2019). For example, one study focused on the period of 2–7 s within the betting period in the frequency domain and the whole trial in the time domain (Zhang *et al.*, 2017). Focusing on this frequency band also enabled the removal of high- and low-frequency artificial noise such as that related to cardiac pulsation (∼1 Hz). We then used WTC to calculate the fNIRS signal relationship between members of the same pair (separately for the resting-state condition, high penalty and human-to-human condition, low penalty and human-to-human condition, high penalty and human-to-computer condition, and low penalty and human-to-computer condition). We generated a 2D WTC map (Figure 3A). The average WTC values in this frequency band during the four blocks and the resting-state were calculated. Please note that we computed the IBS value for each trial and then they were averaged for each pair across trials.

Consistent with previous fNIRS hyperscanning studies (Sanfey *et al.*, 2003; Cui *et al.*, 2012), IBS values were Fisher-z-transformed after calculating differences between conditions and the resting state. For each condition, one-sample *t*-tests of IBS values among all pairs were performed for each ROI [Table 1; false (FDR) discovery rate corrected]. We generated four *t*-maps of IBS values (Figure 3B) using the spline method. Furthermore, in the condition with the highest IBS among these four conditions and specified ROIs, we compared IBS values between bluffing and not bluffing.

Moderating effect of psychological variables on bluffing. A 2 (bluffing penalty: high vs low) × 2 (opponent: human-to-human vs human-to-computer) repeated measured ANOVA on the bluffing rate of bankers, calling rate of followers and IBS for each condition was performed separately to explore the moderating effect of penalty and agent on bluffing.

Bluffing vs not bluffing. In light of the moderating effect of the motivation on the cognitive mechanisms of deception (Suchotzki *et al.*, 2017; Gerlach *et al.*, 2019) and IBS (Liu *et al.*, 2017), we thus conducted paired samples *t*-test between bluffing and not bluffing on RT and IBS in the specified ROIs separately within the highest IBS condition. Moreover, we divided pairs into two groups: pairs that tended to bluff more (HB; 11 pairs) and pairs that tended to bluff less (LB; 10 pairs) using the median of bluffing rate (46.7%). We used two samples *t*-test to compare two groups on IBS in low-card trials.

Results

All participants confirmed that they believed they were playing with the human/computer in the corresponding condition. There was no significant difference between bankers and followers in subjective ratings of participants' opponents on appearance and competence [appearance: $t_{(31)} = -0.615$, $d = 0.166$; competence: $t_{(31)} = -0.341$, $d = 0.081$]. There were also no significant differences between bankers and followers in risk taking propensity, PT and EC [risk taking: $t_{(31)} = 0.321$, $d = 0.087$; PT: $t_{(31)} = -0.104$, $d = 0.030$; EC: $t_{(31)} = 0.869$, $d = 0.170$]. These findings

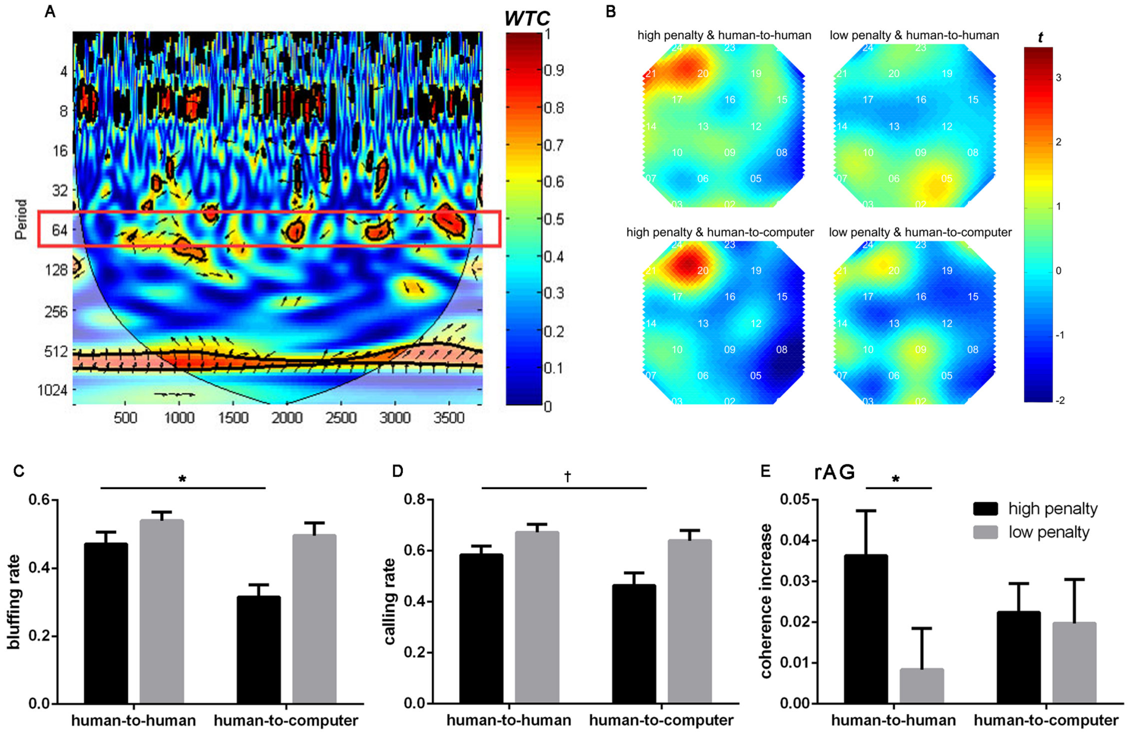

Fig. 3. (A) Frequency band of interest. IBS as measured by WTC that is based on raw HBO signal from channel 21 from banker and follower in the high penalty and human-to-human condition. The red border represents the frequency band of interest (5–7s). The value of WTC is denoted by the color bar. **(B)** One sample *t*-test maps of IBS for each condition. **(C–E)** The bluffing rate, calling rate and IBS in rAG. Error bars indicate standard errors. $^{\dagger}P<0.1$, $^{*}P<0.05$. IBS: interpersonal brain synchronization; WTC: wavelet transform coherence; rAG: right angular gyrus.

suggested that bankers and followers were matched in terms of attractiveness and individual traits.

We did not find any difference between the face-to-face and face-blocked conditions. We thus combined the data from those two groups when conducting our analyses. In each of the four conditions, compared with the resting-state session, significant coherences were identified at right angular gyrus (rAG), including CH20 and CH21 in the high penalty and human-to-human condition [$t_{(31)}=3.3515$, $P=0.002$, $d=0.592$; FDR corrected; Figure 3B]. No other significant effect in the ROI was found [STG (CH 8, CH9 CH12 and CH13), $t_{(31)}=0.3844$, $P=0.703$, $d=0.0679$; IPG (CH19), $t_{(31)}=0.3863$, $P=0.703$, $d=0.0683$; SMG (CH15 and CH16), $t_{(31)}=0.5143$, $P=0.6107$, $d=0.0909$].

For bluffing rate of bankers, the main effects of bluffing penalty and opponent were both significant [penalty: $F_{(1,31)}=19.815$, $P<0.001$, $\eta_p^2=0.390$, high < low; opponent: $F_{(1,31)}=11.904$, $P=0.002$, $\eta_p^2=0.277$, human-to-human > human-to-computer]. The interaction effect was also significant [$F_{(1,31)}=6.437$, $P=0.016$, $\eta_p^2=0.172$; Figure 3C]. After testing for simple effects, the bluffing rate in the human-to-human condition was significantly higher than that in the human-to-computer under high bluffing penalty [$F_{(1,31)}=17.070$, $P<0.001$, $\eta_p^2=0.355$], but there was no significant difference in the low bluffing penalty condition [$F_{(1,31)}=1.552$, $P=0.222$, $\eta_p^2=0.048$]. The same effects were found for calling rate of followers [penalty: $F_{(1,31)}=17.416$, $P<0.001$, $\eta_p^2=0.360$, high < low; opponent: $F_{(1,31)}=8.801$, $P=0.006$, $\eta_p^2=0.221$, human-to-human > human-to-computer; penalty × opponent (marginally significant): $F_{(1,31)}=3.211$, $P=0.083$, $\eta_p^2=0.094$; Figure 3D]. Simple effects tests revealed that the calling rate in the human-to-human condition was significantly higher than that in the human-to-computer under high bluffing penalty [$F_{(1,31)}=12.515$, $P=0.001$, $\eta_p^2=0.288$], but there was no significant difference in the low bluffing penalty condition [$F_{(1,31)}=0.799$, $P=0.378$, $\eta_p^2=0.025$].

With regard to the IBS, there was a marginally significant main effect of bluffing penalty [$F_{(1,31)}=4.108$, $P=0.052$, $\eta_p^2=0.128$, high > low]. There was also a significant interaction effect between bluffing penalty and opponent [$F_{(1,31)}=4.707$, $P=0.039$, $\eta_p^2=0.144$; Figure 3E]. Testing for simple effects revealed a higher IBS in the high bluffing penalty condition compared to that in the low bluffing penalty condition in the human-to-human condition [$F_{(1,31)}=8.254$, $P=0.008$, $\eta_p^2=0.228$], but not in the human-to-computer condition [$F_{(1,31)}=0.042$, $P=0.840$, $\eta_p^2=0.001$]. No significant main effect of opponent was found [$F_{(1,31)}=0.087$, $P=0.770$, $\eta_p^2=0.003$]. Please note that all 30 trials for each condition were included for statistical analyses. The highest IBS in rAG was found in the high penalty and human-to-human condition. We, therefore, focused on these conditions in the next analysis.

After paired-sample $t\times$tests were conducted, we observed that bluffing increased RT and IBS compared with not bluffing under the high penalty within human-to-human pairings [RT: $t_{(31)}=2.256$, $P=0.031$, $d=0.399$, Figure 4A; IBS: $t_{(31)}=2.125$, $P=0.042$, $d=0.376$, Figure 4B]. There were 14.125 ± 5.813 trials for bluffing (range: 2–14) and 15.875 ± 5.813 trials for not bluffing (range: 1–13) per participant (trial number: mean ± s.d.). In addition, pairs who tended to bluff more (HB; 11 pairs) produced a significantly stronger IBS in rAG than pairs who tended to bluff

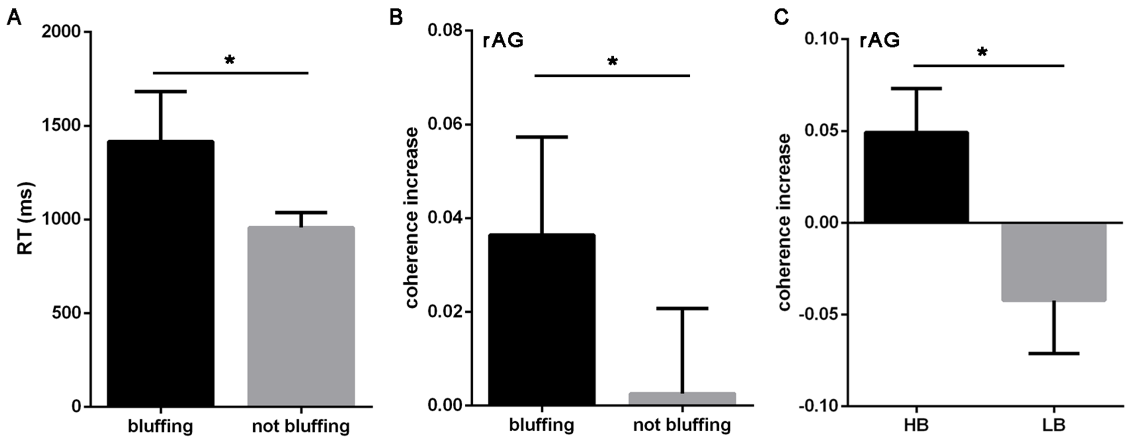

Fig. 4. Bluffing vs not bluffing under high pealty within human-to-human pairings. (A) Behavioral result. (B) IBS result. (C) Group difference result. $^{*}P < 0.05$. rAG: right angular gyrus. HB: pairs that tended to bluff more; LB: pairs that tended to bluff less. IBS: interpersonal brain synchronization.

less under the high penalty in human-to-human pairings [LB; 10 pairs; IBS: $t_{(19)} = 2.455$, $P = 0.024$, $d = 1.068$, Figure 4C].

Discussion

Our study investigated the inter-brain mechanism of bluffing in real social interaction by fNIRS hyperscanning. IBS in the rAG was highest for human partners when the penalty for bluffing is high, identifying the most sensitive context for bluffing. Importantly, we demonstrated that bluffing increased RT and IBS in the rAG, compared with not bluffing. Pairs that tended to bluff more (HB) showed increased IBS than pairs that tended to bluff less (LB). Taken together, our work demonstrates the key role of rAG in bluffing and supports the view that regions associated with mentalizing become synchronized during bluffing games, especially under the high penalty and human-to-human interaction.

We found that the rAG, instead of other subregions among rTPJ, was sensitive in the bluffing game. The rTPJ has long been considered to be crucial for attention, memory, language and social processing (Carter and Huettel, 2013; Donaldson et al., 2015; Schurz et al., 2017). Given that few studies have tested the functional distinction of its subregions, Carter and Huettel (2013) proposed a nexus model for TPJ to account for distinct functions within the TPJ. They dissociated psychological processes by brain regions using fMRI meta-analyses (including ~4000 studies). Specifically, activation peaks within the TPJ that were correlated with mentalizing fell largely in the AG (in the posterior TPJ). For instance, a meta-analysis study demonstrated that task requiring mentalizing, such as story comprehension, engages the AG (Mar, 2011). Therefore, the identified rAG in the current study may reflect that mentalization is engaged in the bluffing process.

We observed the increased RT in bluffing vs not bluffing. This was consistent with previous deception research (Suchotzki et al., 2017; Gerlach et al., 2019), suggesting that the measurement of RT can differentiate lying from the truth. The longer RT may suggest that deception is more cognitively demanding with greater engagement of TPJ (Rapoport et al., 1997; Carrington and Bailey, 2009; Carter et al., 2012; Carter and Huettel, 2013) than truth-telling (Suchotzki et al., 2017). We also found the increased IBS in rAG for bluffing (vs not bluffing), which is consistent with a previous study showing that deception elicited the stronger IBS in TPJ than honesty (Zhang et al., 2017).

Recently, many researchers have paid attention to IBS, which has been widely used in hyperscanning research (Babiloni and Astolfi, 2014; Wang et al., 2018). The functional significance of IBS remains to be understood. When recording the neural activities of each pair of bats (mammal with its high level of sociality) simultaneously, the synchronized brains covaried with the degree of social interactions, most prominent in the gamma-band (Zhang and Yartsev, 2019). The gamma-band has been implicated in social interaction and played important roles in information integration (Gerlach et al., 2019; Zhang and Yartsev, 2019). Therefore, the IBS may originate from the interpretation of others' intentions after integrating multisensory inputs (Dai et al., 2018; Zhang and Yartsev, 2019). The two brains may become correlated as they form a closed motor-sensory loop to guide future actions (Zhang and Yartsev, 2019). In addition to the mentalization (Mar, 2011), the rAG has been recruited as a cross-modal integrative hub (Sperduti et al., 2011; Seghier, 2013). Therefore, the increased IBS in the rAG may serve as an integrative hub of mentalizing others' intentions. These mutual thoughts about other's intentions make the two brains synchronized in the rAG.

Specific to our turn-based bluffing context, two stages of psychological processes may elicit the IBS. First, both parties perceived and integrated their common external stimuli (i.e., decision from the last trial). Second, mentalizing each other's mental state (for bankers: will the follower call; for followers: will the banker bluff me) made two brains form a closed loop, which elicited high IBS. The follower would have the same experience regardless of the banker's intention, since the follower did not know in advance whether the banker would bluff. The difference between bluffing and not bluffing was in the second stage. The synchrony we observed was driven possibly by changes in the banker's neural activity in relation to a relatively constant neural state in the follower. However, it is unknown that the observed synchrony derived from both high or both low mentalizing activity in the two players. Since the RT was increased in bluffing (Suchotzki et al., 2017; Gerlach et al., 2019), it is probably that both the banker and the follower may have high mentalizing activity in bluffing compared to not bluffing. In sum, consistent with the function of integration and mentalization in rAG (Seghier, 2013), bluffing in our study may elicit higher level of mentalization, and higher level of synchronization with the follower, than not bluffing.

Next, IBS was highest for human partners when the penalty for bluffing is high. In light of the moderating effect of the motivation on the cognitive mechanisms of deception (Suchotzki *et al.*, 2017; Gerlach *et al.*, 2019), this result suggests that human-to-human interaction in face of high penalty is the most motivating condition for bluffing. Please note that the mentioned motivation referred to the motivation to avoid penalty. In the field of deception, many studies have examined the consequence of deception from the perspective of reward maximization (Gneezy, 2005). It has been demonstrated that people are more prone to lie when there is a higher potential reward, and this is accompanied by stronger activation in the TPJ (Bhatt *et al.*, 2010). In our study, there was less bluffing under the high penalty condition in addition to the highest IBS, possibly indicating an inhibition effect of penalty. No significant IBS in low penalty during human-to-human pairings (compared with rest), to some extent, further supports the motivate explanation. These findings suggest that high penalty within human-to-human pairings is the motivating context for bluffing. There was stronger physiological arousal (i.e., skin conductor response) in deception as compared to honesty in face of high penalty (Tomash and Reed, 2015). Penalty, like reward, moderates people's tendency to lie (Gerlach *et al.*, 2019). Therefore, the observed stronger IBS in rAG may suggest that more resources are devoted to mentalization of others' intentions.

Comparing human-to-human with human-to-computer condition, we observed different patterns at both behavioral and neural levels. Given that the observed effect on IBS may be driven by extraneous features, we kept the timing of gambling and the visual stimuli consistent for each trial between human-to-human and human-to-computer conditions (Piva *et al.*, 2017). Here, we only focused on the betting period when the banker was deciding whether to bluff or not (Cui *et al.*, 2012; Zhang *et al.*, 2017), and the follower was waiting for the decision. In addition, we also replicated our results by using the decision period in the betting stage, i.e., 5–7 s after the onset of the betting, according to the average RT (*Supplementary material*). On the other hand, it is widely acknowledged that the TPJ is implicated in social interaction (Carter and Huettel, 2013). For example, the TPJ selectively tracks the upcoming decisions of the human opponent rather than the computer opponent (Carter *et al.*, 2012). Human-to-human interaction indeed shows a different behavioral (Sanfey *et al.*, 2003; Carter *et al.*, 2012) and IBS patterns from human-computer interaction (Hirsch *et al.*, 2017; Piva *et al.*, 2017). The increased IBS in rAG during the human-to-human condition in the current study may indicate the importance of AG during the social context.

We did not observe any difference in effect between face-to-face and face-blocked arrangements. Using hyperscanning techniques (Babiloni and Astolfi, 2014; Wang *et al.*, 2018), many studies have employed the face-to-face arrangement to increase ecological validity (Jiang *et al.*, 2012; Donaldson *et al.*, 2015; Jahng *et al.*, 2017; Zhang *et al.*, 2017). Importantly, face-to-face interaction can provide substantial non-verbal cues considered crucial in social interaction (i.e., facial expression and node) to decipher the opponent's intentions and predict the impending behavior (Jahng *et al.*, 2017), such as deception (Zhang *et al.*, 2017). We, therefore, provided face-to-face and face-blocked arrangements in the current study. However, no effect was found. Future studies could further explore the effect of face-to-face interaction on bluffing using more sensitive behavioral measurements (e.g., eye-tracking).

Several limitations are worth noting. First, the narrow sampling window of rTPJ limits the investigation of complex neural events. For example, we did not cover the medial pre-frontal cortex which has been implicated in mentalizing and social decision-making (Coricelli and Nagel, 2009). Although many studies found the importance of rTPJ in deception (i.e. meta-analysis; Lisofsky *et al.*, 2014), future studies should cover more brain areas to study the neural correlates of bluffing at the whole-brain level. Second, we did not observe any significant activation in the banker's brain (*Supplementary material*). It is worth mentioning that TPJ is in the default mode network. Thus, it is 'active' during rest, and so there might not be a significant difference in TPJ between rest and betting because of high baseline activity (Buckner and DiNicola, 2019). In addition, it is not surprising considering that the activation in the single brain did not necessarily coincide with inter-brain coherence (Cui *et al.*, 2012; Baker *et al.*, 2016; Zhang *et al.*, 2017). For example, in spite of having no significant change in signal amplitude during the cooperation task relative to the resting-state in either of the participants, there was a significant increase in the inter-brain coherence (Cui *et al.*, 2012). Measurement of inter-brain coherence is an assessment of dynamic interaction between brains (Babiloni and Astolfi, 2014). Another reason for the lack of activation in the single brain may be due to fNIRS' limited sensitivity to measure a deep-brain signal (>3 cm; Scholkmann *et al.*, 2013). Although previous studies using fMRI have demonstrated the importance of TPJ in bluffing (Carter *et al.*, 2012), no significant activation in the current study was detected. Future studies could examine the relationship between the activation in the single brain and inter-brain coherence from multi-modalities. Finally, given the analysis of pairs with high vs low rates of bluffing is underpowered (the small number of pairs in each group), further replication is required.

In conclusion, using the fNIRS hyperscanning technique combined with a novel bluffing paradigm, our findings support the view that regions associated with mentalizing become more synchronized during bluffing, especially in human-to-human interaction when the penalty is high. The current evidence further advances our understanding of bluffing and its inter-brain mechanisms Our work is a major step toward understanding the functional significance of rAG and the neural correlates of bluffing.

Funding

X.Z. was supported by a research grant from National Natural Science Foundation of China (71942001). This work was supported by the National Social Science Foundation of China (19ZDA361) and the National Natural Science Foundation of China (71942002, 31771238).

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at *SCAN* online.

Reference

Babiloni, F., Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: past, present and future. *Neuroscience and Biobehavioral Reviews*, *44*, 76–93.

Baker, J.M., Liu, N., Cui, X., et al. (2016). Sex differences in neural and behavioral signatures of cooperation revealed by fNIRS hyperscanning. *Scientific Reports*, *6*, 1–11.

Baumgartner, T., Fischbacher, U., Feierabend, A., et al. (2009). The neural circuitry of a broken promise. *Neuron*, *64*, 756–70.

Bhatt, M.A., Lohrenz, T., Camerer, C.F., et al. (2010). Neural signatures of strategic types in a two-person bargaining game. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 19720–25.

Buckner, R.L., DiNicola, L.M. (2019). The brain's default network: Updated anatomy, physiology and evolving insights. *Nature Reviews Neuroscience*, *20*, 593–608.

Carrington, S.J., Bailey, A.J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, *30*, 2313–35.

Carson, T. (1993). Second thoughts about bluffing. *Business Ethics Quarterly*, *3*, 317–42.

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

Carter, R.M.K., Huettel, S.A. (2013). A nexus model of the temporal–parietal junction. *Trends in Cognitive Sciences*, *17*, 328–36.

Coricelli, G., Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 9163–68.

Cui, X., Bray, S., Bryant, D.M., et al. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *NeuroImage*, *54*, 2808–21.

Cui, X., Bryant, D.M., Reiss, A.L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *NeuroImage*, *59*, 2430–37.

Dai, R., Liu, R., Liu, T., et al. (2018). Holistic cognitive and neural processes: a fNIRS-hyperscanning study on interpersonal sensorimotor synchronization. *Social Cognitive and Affective Neuroscience*, *13*, 1141–54.

Davis, M.H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113–26.

Donaldson, P.H., Rinehart, N.J., Enticott, P.G. (2015). Noninvasive stimulation of the temporoparietal junction: a systematic review. *Neuroscience and Biobehavioral Reviews*, *55*, 547–72.

Duan, L., Dai, R.-N., Xiao, X., et al. (2015). Cluster imaging of multi-brain networks (CIMBN): a general framework for hyperscanning and modeling a group of interacting brains. *Frontiers in Neuroscience*, *9*, 1–8.

Faul, F., Erdfelder, E., Buchner, A., et al. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–60.

Gerlach, P., Teodorescu, K., Hertwig, R. (2019). The truth about lies: a meta-analysis on dishonest behavior. *Psychological Bulletin*, *145*, 1–44.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95*, 384–94.

Grinsted, A., Moore, J.C., Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, *11*, 561–66.

Hackel, L.M., Mende-Siedlecki, P., Amodio, D.M. (2020). Reinforcement learning in social interaction: The distinguishing role of trait inference. *Journal of Experimental Social Psychology*, *88*, 103948.

Hari, R., Henriksson, L., Malinen, S., et al. (2015). Centrality of social interaction in human brain function. *Neuron*, *88*, 181–93.

Hirsch, J., Zhang, X., Noah, J.A., et al. (2017). Frontal temporal and parietal systems synchronize within and across brains during live eye-to-eye contact. *NeuroImage*, *157*, 314–30.

Hu, Y., Pan, Y., Shi, X., et al. (2018). Inter-brain synchrony and cooperation context in interactive decision making. *Biological Psychology*, *133*, 54–62.

Hu, Y., Hu, Y., Li, X., et al. (2017). Brain-to-brain synchronization across two persons predicts mutual prosociality. *Social Cognitive and Affective Neuroscience*, *12*, 1835–44.

Jahng, J., Kralik, J.D., Hwang, D.-U., 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.

Jenkins, A.C., Zhu, L., Hsu, M. (2016). Cognitive neuroscience of honesty and deception: A signaling framework. *Current Opinion in Behavioral Sciences*, *11*, 130–37.

Jiang, J., Chen, C., Dai, B., et al. (2015). Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 4274–79.

Jiang, J., Dai, B., Peng, D., et al. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience*, *32*, 16064–69.

Herrmann, C.S., Fründ, I., Lenz, D. (2010). Human gamma-band activity: a review on cognitive and behavioral correlates and network models. *Neuroscience and Biobehavioral Reviews*, *34*, 981–92.

Lisofsky, N., Kazzer, P., Heekeren, H.R., et al. (2014). Investigating socio-cognitive processes in deception: a quantitative meta-analysis of neuroimaging studies. *Neuropsychologia*, *61*, 113–22.

Liu, N., Mok, C., Witt, E.E., et al. (2016). NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative Jenga game with face-to-face communication. *Frontiers in Human Neuroscience*, *10*, 1–11.

Liu, T., Saito, G., Lin, C., et al. (2017). Inter-brain network underlying turn-based cooperation and competition: A hyperscanning study using near-infrared spectroscopy. *Scientific Reports*, *7*, 1–12.

Mar, R.A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, *62*, 103–34.

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

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.

Piva, M., Zhang, X., Noah, J.A., et al. (2017). Distributed neural activity patterns during human-to-human competition. *Frontiers in Human Neuroscience*, *11*, 1–14.

Rangel, A., Camerer, C., Montague, P.R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*, 545–56.

Rapoport, A., Erev, I., Abraham, E.V., et al. (1997). Randomization and adaptive learning in a simplified poker game. *Organizational Behavior and Human Decision Processes*, *69*, 31–49.

Rilling, J.K., Sanfey, A.G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, *62*, 23–48.

Sanfey, A.G., et al. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*, 215–22.

Saxe, R., Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, *43*, 1391–99.

Scholkmann, F., Holper, L., Wolf, U., et al. (2013). A new methodical approach in neuroscience: Assessing inter-personal brain coupling using functional near-infrared imaging (fNIRI) hyperscanning. *Frontiers in Human Neuroscience*, *7*, 1–6.

Schurz, M., Tholen, M.G., Perner, J., et al. (2017). Specifying the brain anatomy underlying temporo-parietal junction activations for theory of mind: a review using probabilistic atlases from different imaging modalities. *Human Brain Mapping*, *38*, 4788–805.

Seghier, M.L. (2013). The angular gyrus: Multiple functions and multiple subdivisions. *The Neuroscientist*, *19*, 43–61.

Singh, A.K., Okamoto, M., Dan, H., et al. (2005). Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. *NeuroImage*, *27*, 842–51.

Sperduti, M., Delaveau, P., Fossati, P., et al. (2011). Different brain structures related to self- and external-agency attribution: a brief review and meta-analysis. *Brain Structure and Function*, *216*, 151–57.

Suchotzki, K., Verschuere, B., Bockstaele, B.V., et al. (2017). Lying takes time: a meta-analysis on reaction time measures of deception. *Psychological Bulletin*, *143*, 428–53.

Tang, H., Mai, X., Wang, S., 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.

Tomash, J.J., Reed, P. (2015). Using conditioning to elicit skin conductance responses to deception. *Learning and Motivation*, *49*, 31–37.

Wallach, M.A., Kogan, N. (2007). Aspects of judgment and decision making: Interrelationships and changes with age. *Behavioral Science*, *6*, 23–36.

Wang, C., Zhang, T., Shan, Z., et al. (2019). Dynamic interpersonal neural synchronization underlying pain-induced cooperation in females. *Human Brain Mapping*, *40*, 3222–32.

Wang, M.-Y., Luan, P., Zhang, J., et al. (2018). Concurrent mapping of brain activation from multiple subjects during social interaction by hyperscanning: a mini-review. *Quantitative Imaging in Medicine and Surgery*, *8*, 819–37.

Ye, J., Tak, S., Jang, K., et al. (2009). NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. *NeuroImage*, *44*, 428–47.

Zebrowitz, L.A., Ward, N., Boshyan, J., et al. (2018). Older adults' neural activation in the reward circuit is sensitive to face trustworthiness. *Cognitive, Affective & Behavioral Neuroscience*, *18*, 21–34.

Zhang, M., Liu, T., Pelowski, M., et al. (2017). Gender difference in spontaneous deception: a hyperscanning study using functional near-infrared spectroscopy. *Scientific Reports*, *7*, 1–13.

Zhang, W., Yartsev, M.M. (2019). Correlated neural activity across the brains of socially interacting bats. *Cell*, *178*, 413–28.e22