Frontal EEG activities reflect social interactive behaviors during a deception game relying on facial expressions

Haoming Qi[1], Dongcheng He[2]

1. School of Information Network Security, People's Public Security University of China, Beijing, China
2. Institute of Computational Social Science, Institute of Dataspace, Hefei Comprehensive National Science Center, Hefei, Anhui, China

Corresponding Author: Dongcheng He

hedch1995@gmail.com

288 Chuangxin Blvd, High-Tech District, Hefei, Anhui, China, 230071

Abstract

The frontal cortex plays a critical role in the deception. While previous neuroimaging studies have examined the functions of various subregions within the frontal cortex, particularly in relation to truth inhibition and lie production, there is a limited understanding of frontal activities during deception in the context of social interaction. To bridge this gap, we investigated an EEG dataset collected during a deception game involving two participants (one player and one observer) who engaged in social interaction. In the game, the player was instructed or chose spontaneously whether to lie or tell the truth, while the observer decided whether to doubt or trust the player based on the player's facial expressions. Our analyses on the behavioral and EEG data revealed two major findings: (1) The observers' responses were directly impacted by the players' facial expressions. Without being informed about the trial conditions, the observer was more likely to doubt the player in instructed conditions compared to the spontaneous ones, a tendency that was evident in both the behavioral data and the observer's ventrolateral prefrontal cortex activities. (2) The observer's responses were found to depend on the player's neural activities. As evidence, both the observer's ventrolateral prefrontal cortex and the player's dorsomedial-ventrolateral prefrontal cortex activities reflected the observer's responses (trust versus doubt). According to these results and previous research, we conclude that the frontal EEG activities could encode socially interacted behaviors. We also propose a neural coupling model driven by facial expressions from the player's emotional regulation system to the observer's emotion perception system during the deception game.

1. Introduction

A large body of neuroimaging studies has emphasized the contributions of frontal cortex in deception (Spence et al., 2001; Spence et al., 2004; Abe et al., 2006; Kireev et al., 2006; Ito et al., 2012; Scheuble and Beauducel, 2020). These studies have focused on the process of generating false information while suppressing the truth, identifying distinct functional subregions within the frontal cortex. Specifically, the dorsolateral prefrontal cortex has been implicated in the production of lies (Langleben et al., 2005; Nunez et al., 2005; Priori et al., 2008; Ito et al., 2012). Within this area, Priori et al. (2008) found that activities were distinctive by different types of lies, such as denying a fact and making up an event that never happened. Additionally, Ito et al. (2012) found that the left dorsolateral prefrontal cortex was more activated during lie preparation than truth preparation, a pattern not observed on the right side, suggesting a bilateral distinction in the dorsolateral prefrontal cortex. Conversely, the inferior frontal gyrus has been suggested to act in the truth inhibition during deception (Abe, 2011). This region has been reported to activate when delivering false information (Kireev et al., 2017; Kireev et al., 2015; Verschuere et al., 2012; Vartanian et al., 2013) and in studies involving typical response inhibition tasks (Aron et al., 2003; Obeso et al., 2013; Sebastian et al., 2013).

Although the functional specializations of subregions in the deceivers' frontal cortex have been widely reported, little is known about the simultaneous cortical dynamics of both the deceiver and the deceived, as well as their social interactions during deception. Deception, inherently a type of social behavior, often involves emotions as a critical medium in these interactions. The role of frontal activities in emotion has been investigated for decades. Early studies on people with cortical lesions have suggested the asymmetrical involvement of prefrontal regions to positive and negative emotions (Goldstein, 1944), which has been confirmed by subsequent studies (Gainotti, 1972; Robinson and Price, 1982). In the context of social interaction like deception, the roles of frontal cortex on social modulation involve various of functions (Harmon-Jones et al., 2009). Particularly, previous studies have demonstrated the relation between the left frontal electrical activities and the emotion states of facial expression, with increased activation during joyful expressions (Ekman and Davidson, 1993; Graham and Cabeza; 2001). Similar asymmetrical frontal activities were also found to reflect the emotional motivation directions. For example, promotion was found to be related to higher asymmetrical bilateral frontal activations than prevention (Amodio et al., 2004). The dynamics of such emotional motivation may play an active role during deception. For example, guilt, an important emotional component in deception (Battigalli et al., 2013; Seiter and Bruschke, 2007), has been linked to both increased and reduced motivation varying over time (Amodio et al., 2007).

To sort out the frontal activities during deception at a social interaction aspect, this study analyzed an EEG dataset collected from a player and an observer simultaneously during a task designed to encourage deception. At each time, the player decided to relay either the true or a

false information to the observer, who watched the player's facial expression before choosing to trust or doubt the player. By examining the signals from the frontal regions and analyzing participants' behaviors, our primary aim is to determine if there are neural or behavioral linkages between the player and the observer based on the player's facial expressions. To answer this question, on both roles' frontal EEG activities, we examined the effects from either role's behaviors. We hypothesized that the frontal activities representing socially interacted behaviors in one role would be contingent on the behaviors of the other role.

2. Methods

2.1. Description of the Dataset

As illustrated by Chen, Fazil, and Wallraven (2024), the dataset we used in the present investigation was curated from a data description study published in *Scientific Data.* This dataset contains a complete set of raw behavioral data and EEG data, as well as a preprocessed EEG dataset that were collected in a two-player deception experiment. Twenty-four participants (12 females, age: 25±4.34 yrs) with normal or corrected-to-normal vision and no reported history of neurological disorder participated in this experiment for monetary rewards of around USD10/h. This experiment was approved by the Institutional Review Board with the number KUIRB-2019-0043-01. All participants were naïve to the experimental paradigm and gave written informed consent prior to the experiment.

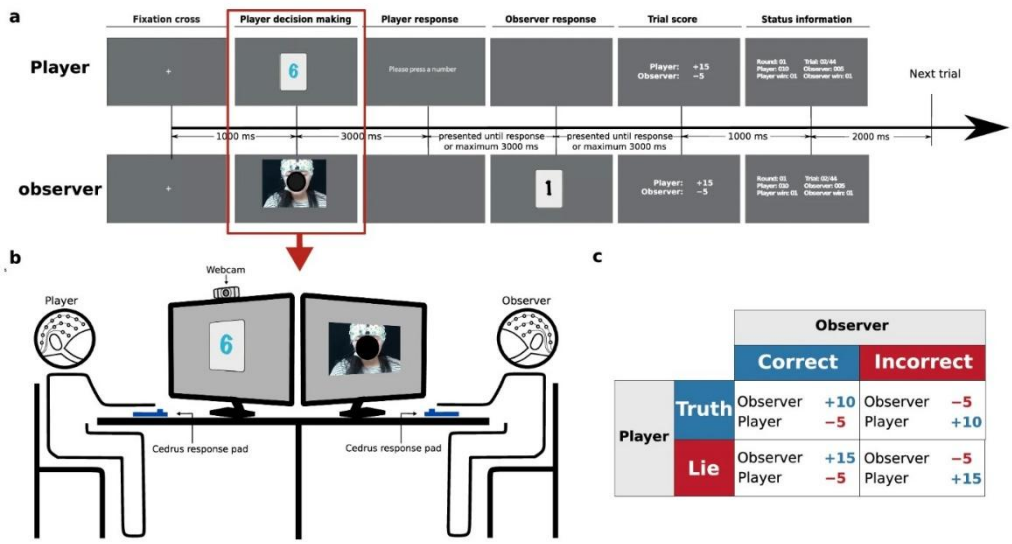

Figure 1. Illustration of the experiment. a) Schematic of a trial. b) Setup of the experiment. c) Scoring system used in the experiment. This figure was copied from Figure 1 in Chen, Fazil, and Wallraven (2024).

2.2. Experimental Procedures

Depicted in Figure 1, the task of this experiment involved two participants serving as the player

and the observer, respectively, at each time. The two participants faced each other and sat in front of two separated 24-inch monitors (resolution: 1920×1080 px^2, fresh rate: 60Hz; produced by LG, South Korea), with a HD pro C920 webcam (Logitech, Switzerland) installed on the top of the player's monitor (Figure 1 b). Schematic in Figure 1 a, a trial began with a fixation period for 1000ms, during which both participants were asked to fixate on a fixation cross appeared in the center of their respective monitors. Then, a card contained a colored digit (ranging from 1 to 6) presented on the monitor of the player's side, whereas on the observer's side, the monitor showed the live stream of the player's face. This stage lasted for 3000ms, and during this period, the player had to decide what number to be relayed to the observer depending on the color (black, purple, or blue) of the presented digit. The three colors cued three different strategic conditions: (1) the instructed truth condition that the player had to relay the same number presented in the card to the observer; (2) the instructed lie condition that the player had to relay a different number than the presented one; and (3) the spontaneous condition that the player chose any number within the range 1-6 to relay. Succeeding this player decision stage, the player was asked to relay the number by pressing a correspondent button on a RB-740 response pad (Cedrus Corporation, USA) within 3000ms, during which only the background was presented on the observer's side. Triggered by the player's response, a card showing the relayed number in black presented on the observer's monitor, and the observer was asked to decide whether the information was a lie (to doubt) or the truth (to trust) by pressing one of two buttons on the response pad within 3000ms. Upon the observer's response, feedback was given to both participants according to a score system as illustrated in Figure 1 c. The score system, explained before the experiment, was designed to encourage active participation of both participants by successful deceiving and detecting in exchange for higher scores. At the end of each trial, for the monitors of both sides, the trial scores that both participants acquired were presented for 1000ms and followed by a summary of status showing the accumulated scores, number of trials and rounds won by each side, as well as the progress of experiment for 3000ms.

The 24 participants were separated into 12 pairs, and in each pair, two participants took turns to serve as the player and the observer. This resulted to 24 paired cases, and 23 of them successfully finished the experiment. For each case, the entire experiment consisted of 11 sessions, in which each session contained 44 trials (spontaneous: 22, instructed lie: 11, instructed truth: 11) and these trials were presented in a randomly shuffled order. During all these experiments, EEG and responses of both participants were recorded. This experiment was programmed in Python using PsychoPy.

2.3. EEG Apparatus

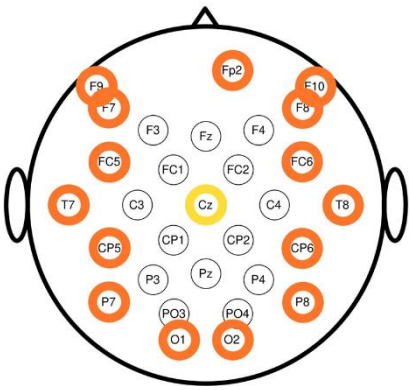

Figure 2. Electrode locations relative to the scalp. This figure was reproduced from Figure 3 in Chen, Fazil, and Wallraven (2024). Color rings around the labels in this figure have no specific meanings.

According to the data description, the EEG data was recorded with two BrainAmp amplifiers (Brain Products, Germany) from 30 EEG electrodes and an EOG electrode. Electrode locations on the caps are visualized in Figure 2. Due to connection issues, the Oz channel was removed from the dataset. The data was digitized at 500Hz, nose-referenced online, and forehead grounded to the electrode Fpz.

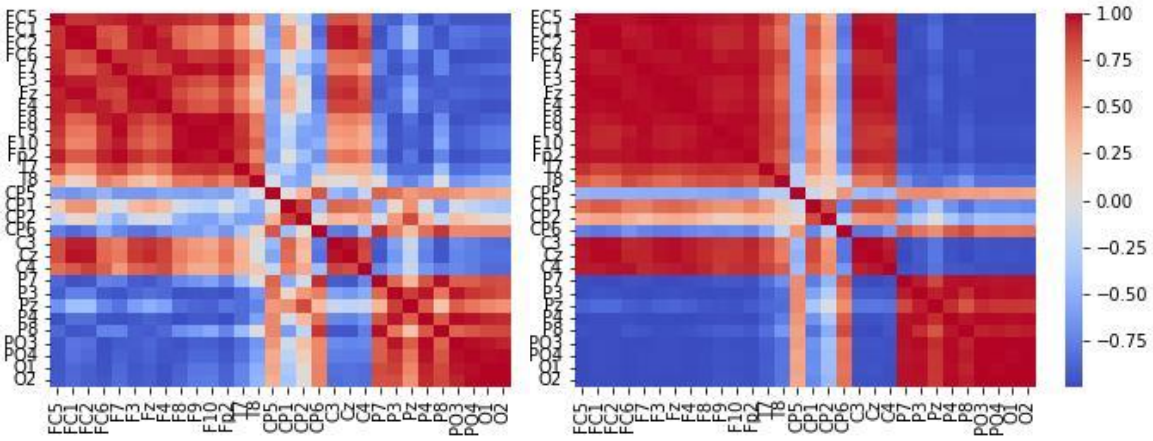

Figure 3. Pairwise Pearson correlation coefficients among the signals from all electrodes presented in heat maps, with the left and right maps representing the player and the observer, respectively. The electrodes in both maps are arranged in a grouping manner based on the cortical regions.

2.4. Data Analysis

2.4.1. Data Preprocessing and Regions of Interest (ROIs)

In the preprocessed dataset, the data was down-sampled to 100Hz and filtered at 1-49Hz. Artifacts were detected and rejected using EEGLAB functions and ICA-based algorithms with Matlab (The MathWork, USA). Data from each channel was then baseline corrected by 500ms

prior to the stimulus onset. Further details about the algorithms during preprocessing can be found in the data description (Chen, Fazil, and Wallraven, 2024). This study aimed to investigate the neural or behavioral linkages between the player and the observer based on the player's facial expressions. Therefore, we only used the data from the "player decision making" stage (as labeled in Figure 1 a), which lasted for 3500ms beginning from 500ms prior to the onset of the digit stimulus (for the player) and the facial expression display (for the observer), and ending 3000ms afterwards.

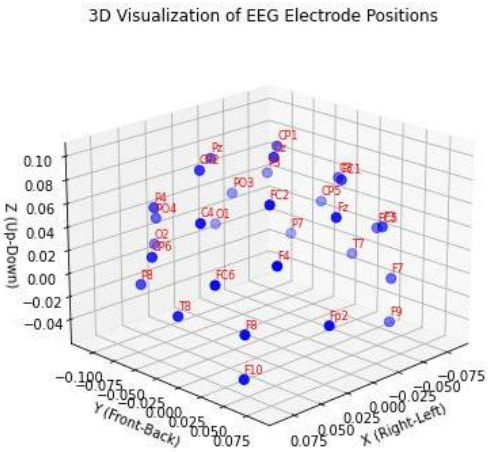

Figure 4. Template electrode positions on the scalp. The coordinates are in meters.

Figure 3 shows the pairwise Pearson correlation coefficients among the signals that were collected from all the electrodes for both the player and the observer during the "player decision making" stage. The data presented in this figure used averaged signals across all the trials and all the subjects. As validated by previous studies, such data-driven signal clustering based on inter-ROI dependencies could effectively segment the within- and cross-system signal sources (Blank et al., 2014). In Figure 3, we found system-level processing in mainly two cortex regions, the frontal region and the posterior region, as reflected by the overall high positive consistencies among the signal channels within these two regions and the negative correlations between them. As highlighted in the Introduction, this study targeted on the neural activities in the frontal cortex. Therefore, we selected the electrodes from the frontal region. Further, as detected in Figure 2 and 3, the electrodes placed in neighboring positions on the scalp could have collected signals with high correlations due to the inherent limited spatial resolution of EEG techniques. We thus grouped the electrodes in the frontal region into multiple ROIs to enhance the signal quality according to the following criteria: two electrodes were linked if their signals' Pearson correlation coefficient was above 0.9 and the distance between them was below 5cm. Since the original EEG cap information was not available, we instead used a template montage, "standard_1020", acquired from the MNE-Python package (Gramfort et al., 2013) to determine the electrode positions and visualized them in Figure 4. In the frontal region, using the averaged

signals combining the player and the observer, we found the following electrode pairs satisfy the criterion: F7-F9, F7-FC5, F3-FC5, F3-FC1, Fz-FC1, Fz-FC2, F4-FC6, F8-FC6, and F8-F10. These linked electrode pairs were then arranged into three ROIs: F7-F9-FC5, F8-F10-FC6, and Fz-FC1-FC2, with two exceptions: F3 and F4 who seemed to associate with these ROIs ambiguously. We then labeled these ROIs based on their anatomical correspondence to underlying physiological brain regions, such that F7-F9-FC5, F8-F10-FC6, and Fz-FC1-FC2 were labeled as left ventrolateral prefrontal cortex (left vlPFC), right ventrolateral prefrontal cortex (right vlPFC), and dorsomedial prefrontal cortex (dmPFC), respectively.

In the following analysis, we analyzed the EEG measurements from our selected channels according to the experimental conditions and participants' responses by four variables: (1) result: correct versus incorrect; (2) player's task: instructed versus spontaneous; (3) player's behavior: truth versus lie; and (4) observer's response: trust versus doubt.

2.4.2. ERP Analysis

In the ERP analysis, for each participant pair and each role during the experiment, EEG signals were averaged across trials but discriminated according to the specific variable to be examined. Then, we compared the ERP signals at each sampling time instant using repeated-measures ANOVAs (RM-ANOVAs) with the variable as the main factor. Multiple test corrections by a window of 50ms were performed on the p-values. Since both central-lateral and bilateral differences could be recognized by the subsequent correlation analysis, we didn't discern signals from the left and right lateral regions at this stage, and instead we averaged the signals from these two ROIs and examined the effects on the lateral regions at an overall level. Also, neither did we analyze the difference in magnitude between the central and lateral ROIs herein.

2.4.3. Inter-ROI Correlation Analysis

To investigate neural interactions across regions, we conducted an inter-ROI correlation analysis. Firstly, in a bilateral vlPFC correlation analysis, we calculated the moving-window Pearson correlation coefficients between the ERP signals from the left and right vlPFC with a fixed window length. Then, in a dmPFC-vlPFC correlation analysis, we calculated the moving-window Pearson correlation coefficients between the ERP signals from the vlPFC and the dmPFC with a same window length. The optimal window length was determined by a sensitivity analysis considering the information (positively correlated with the variance of the signal), the smoothness (negatively related with the variance of the first-order difference of the signal), and the loss in signal length due to the window size. An effectiveness index was calculated using Equation 1 and tested with six different window lengths ranging from 50ms (5 points) to 1600ms (160 points).

$$EI = [Var(series) - Var(\Delta series)] * (1 - \frac{Win}{Len(series)}) \quad (1)$$

Where EI stands for effectiveness index, Var indicates the variance of the input series, Δ indicates the first-order difference given a series, Win indicates the window size, and Len indicates the length of the input series. We used the moving window Pearson correlation between the signals from left and right vlPFC as the series input and acquired the following results, as illustrated in Figure 5.

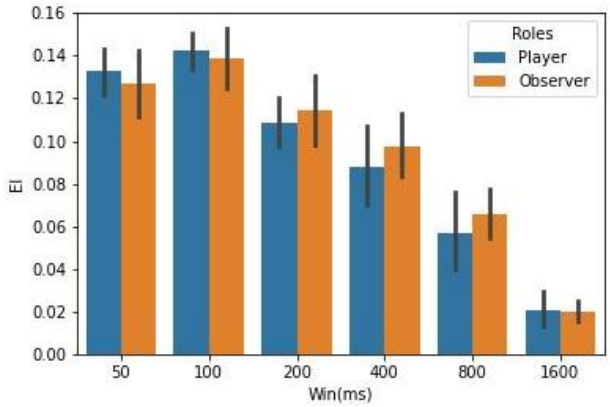

Figure 5. Effective index with respect to the window size grouped by subject roles. The bars indicate the mean across subjects and the error bars indicate 95% of the confidence interval across subjects.

A RM-ANOVA with the window size and subject's role as the main factors detected a significant effect from the window size on the effective index ($F(5,110)=195.77$, $p<0.0001$). Of all the window sizes tested, we found 100ms was the optimal one resulting the highest EI, and thus 100ms was used as the window length in the inter-ROI analysis.

Finally, we compared the correlation coefficients at each sampling time instant using RM-ANOVAs with each of the four variables (result, task, player's behavior, and observer's response) as the main factor, and we multiple-test corrected the resulted p-values by a window of 50ms.

2.4.4. Bilateral vlPFC Phase Analysis

The purpose of this analysis was to examine if the patterns we detected from the inter-ROI correlations was due to the phase differences in signals between the left and the right vlPFC. We hope this analysis could provide further explanations about the inter-ROI dependencies, as either amplitude variations or phase shifting can influence the Pearson correlations, but they undergo distinctive neural activities. Herein, subsequent to the observations from the inter-ROI correlation analysis, we derived the phase shifts in signals between the left and the right vlPFC through the Hilbert transformation, using Equation 2-3 (Kschischang, 2006).

$$\hat{g}(t) = F^{-1}(-j \cdot sgn(f) \cdot G(f)) \quad (2)$$

Where $\hat{g}(t)$ is the Hilbert transform of a given time-series $g(t)$, F^{-1} is the inverse Fourier transform of the input signal, $G(f)$ is the Fourier transform of $g(t)$. $\phi_g(t)$, the instantaneous phase of $g(t)$, can then be calculated using Equation 3.

$$\phi_g(t) = tan^{-1}(\frac{\hat{g}(t)}{g(t)}) \quad (3)$$

3. Results

3.1. Behavioral Results

We first examined if the observers' responses were balanced based on results (correctness) and responses, respectively. For the 484 trials each participant ran as an observer, the mean (std) number of correct and incorrect trials were 241.13 (13.56) and 242.74 (13.5), respectively. A RM-ANOVA with the result (correct versus incorrect) as the main factor on the count of trials found no significant effect ($F(1,22)=0.08$, $p=0.78$). The observers' responses by two choices were also not significantly uneven as indicated by a RM-ANOVA with the response (trust verses doubt) as the main factor on the count of trials ($F(1,22)=1.8$, $p=0.19$). Next, we examined if the observers' responses were unbalanced across different experimental conditions. We used the ratio between the counts of trials responded with "doubt" and "trust" to determine the degree of incredibility under each condition. The mean (std) ratios for the instructed lie, instructed truth, spontaneous lie, and spontaneous truth conditions were 1.14 (0.27), 1.31 (0.45), 0.86 (0.22), and 0.81 (0.24), respectively. A RM-ANOVA with the player's task (instructed versus spontaneous) and behavior (truth versus lie) as the main factors on the count of trials found a significant effect of the task ($F(1,22)=15.72$, $p<0.001$) and a significant interaction between the two main factors ($F(1,22)=5.89$, $p<0.05$), but didn't find any significant effect of the player's behavior ($F(1,22)=1.52$, $p=0.23$). These results indicate that, although the observers' responses were generally even by both correctness and judgements, they tended to be more skeptical when the players were instructed, regardless of what types of instructions were given. Note that the observers were not explicitly instructed about the experimental conditions during the entire experiment, therefore these results suggest that the observers could detect, at a certain degree, if the players were instructed or not. Moreover, the observer understood that the number of trials were even between the instructed lie and truth conditions. Therefore, the observers' biased responses indicate that they were directly influenced by the players' facial expressions.

3.2. ERP Results

Figure 6 shows the average ERP signals from both roles and the three ROIs. We observed that these ERPs were different in magnitude between the two roles and/or across the three ROIs at certain periods within the first 1000ms. We thus determined the duration from 0 to 1000ms as the critical period to be examined for the effects from those variables we illustrated previously.

In all the analyses, we only detected one major effect that was from the player's task (instructed

versus spontaneous) on their vlPFC regions lying mainly between 450ms and 540ms, shown in Figure 7 (left panel), as indicated by RM-ANOVAs on a temporal sample basis followed by multiple-test corrections. We collected all the samples during this period and ran another RM-ANOVA with the player's task as the main factor on the sample collections and it confirmed the significant effect ($F(1,22)=17.87$, $p<0.001$). Such major effect was controlled by other conditional settings, where no major effects were detected on ERPs of both roles and both ROIs, from results, players' behaviors, and observers' responses, as shown in Figure 4 (right panel) and Figure A1, A2 in the Appendix. These results suggest that, for the players, the execution of task instructions could be represented by the activities that recorded at the vlPFC.

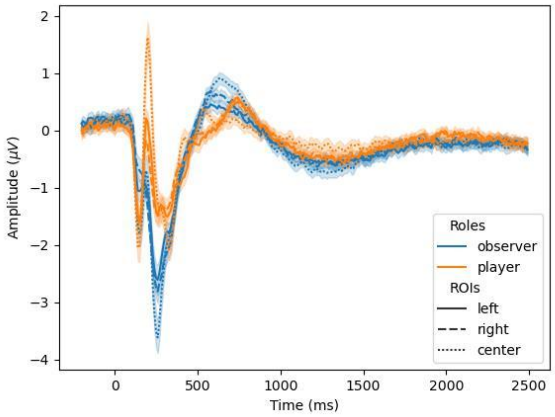

Figure 6. Averaged ERP signals for both roles and all three ROIs during the "player decision making" stage. Time 0 marks the stimulus onset. Shadings indicate the 95% confidence intervals across participants.

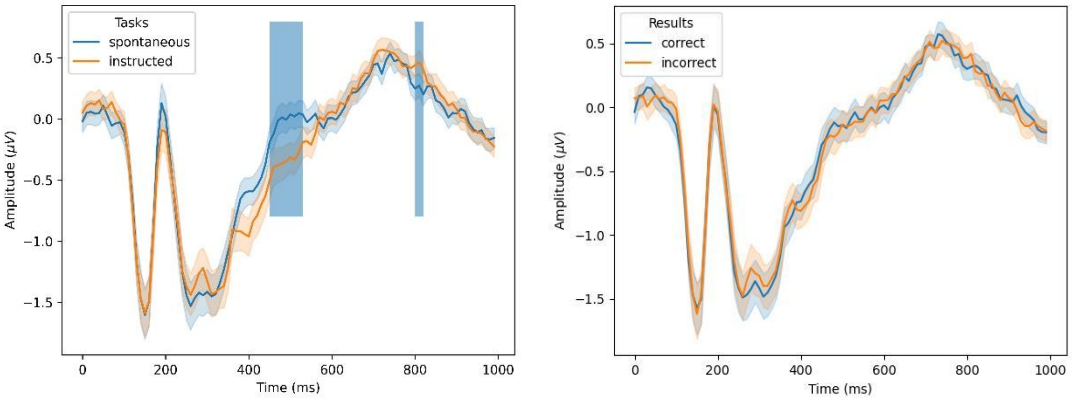

Figure 7. ERPs from vlPFC for the players conditioned by tasks (left panel) and results (right panel). Blue boxes highlighted the temporal duration where RM-ANOVAs detected significant effects ($p<0.05$) from the condition on the ERPs after multiple test corrections. Time 0 marks the stimulus onset. Shadings indicate the 95% confidence intervals across participants.

3.3. Bilateral vlPFC Correlation Analysis

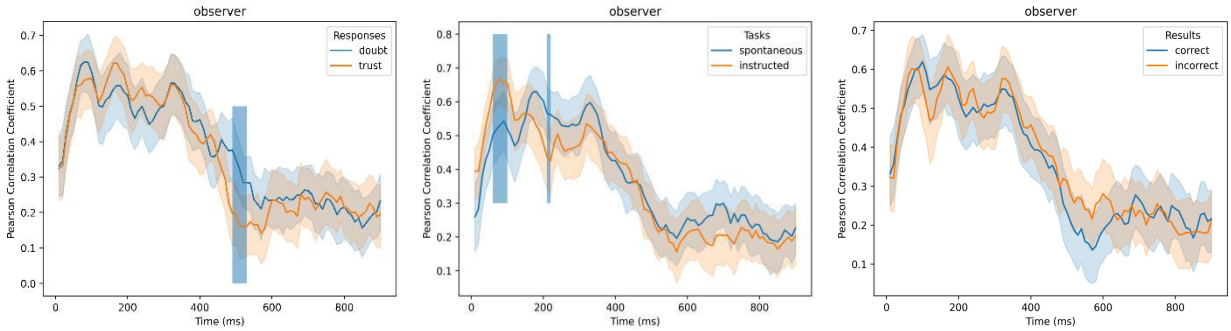

Figure 8. Pearson correlation coefficients between ERPs from left and right vlPFCs for the observers, conditioned by responses (left panel), tasks (middle panel), and results (right panel). Blue boxes highlighted the temporal duration where RM-ANOVAs detected significant effects ($p<0.05$) from the condition on the correlation coefficients after multiple test corrections. Time 0 marks the stimulus onset. Shadings indicate the 95% confidence intervals across participants.

During the critical period beginning from 0ms to 1000ms, we observed two major effects from those four variables on the correlation dynamics between the left and the right vlPFCs, both of which were from the observer's side. As shown in Figure 8, it was detected by the RM-ANOVAs on a temporal sample basis that the observer's response (doubt versus trust) had a significant effect on their inter-lateral dependencies between 480ms and 530ms, as confirmed by another RM-ANOVA with the observer's response as the main factor on the collection of correlation coefficients within this duration ($F(1,22)=11.24$, $p<0.01$). Moreover, the player's task also had a significant effect on the observers' inter-lateral dependencies mainly between 50ms and 100ms indicated by a same statistical testing manner ($F(1,22)=8.89$, $p<0.01$). These results suggest that the observers' responses, towards the players' facial expressions, could be encoded by the dynamics of their bilateral asymmetry in the signals recorded from the vlPFC. Moreover, the player's task was also reflected by the observer's vlPFC activities in as early as the first 100ms, which indicates an early awareness of the player's task from the observer's side through facial expression observations. These effects were controlled by other conditions as shown in Figure 5 (panel right) and Figure A3 in the Appendix.

3.4. dmPFC-vlPFC Correlation Analysis

In all the results, we observed one effect on dmPFC-vlPFC correlations, as indicated by the RM-ANOVA on a temporal sample basis. Shown in Figure 9, the observer's response had a significant effect on the player's central-lateral connections between 650ms and 790ms, as confirmed by another RM-ANOVA with the observer's response (doubt versus trust) as the main factor on the collection of correlation coefficients within this duration ($F(1,22)=21.07$, $p=0.0001$). This result provides direct evidence for the linkage from the player's neural activities in the

frontal cortex, specifically in the connections between signals recorded from the dmPFC and vlPFC, to the observer's responses. Such effect was controlled by all other conditional settings as shown in Figure 9 (panel middle and right) and Figure A4 in the Appendix.

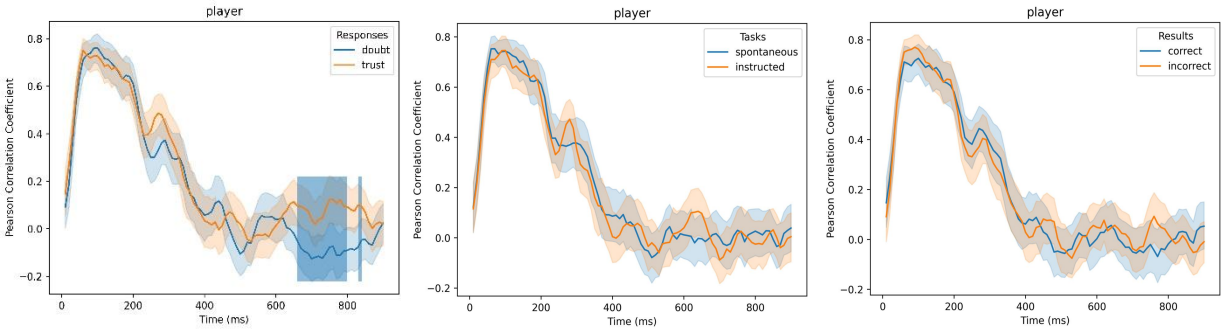

Figure 9. Pearson correlation coefficients between ERPs from dmPFC and vlPFC for the players, conditioned by responses (left panel), tasks (middle panel), and results (right panel). Blue boxes highlighted the temporal duration where RM-ANOVAs detected significant effects ($p<0.05$) from the condition on the correlation coefficients after multiple test corrections. Time 0 marks the stimulus onset. Shadings indicate the 95% confidence intervals across participants.

3.5. **Bilateral vlPFC Phase Analysis**

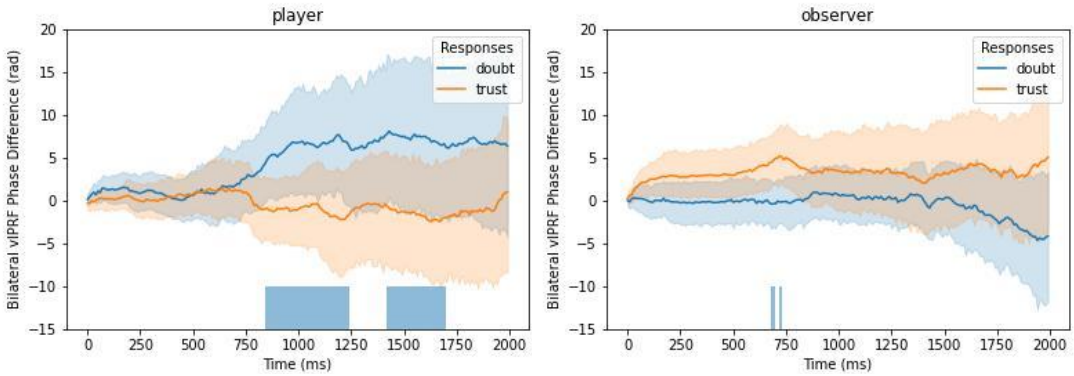

Figure 10. Phase of the signals in left vlPFC subtracted by that in the right, conditioned by responses. Blue boxes on the x-axis highlighted the temporal duration where RM-ANOVAs detected significant effects ($p<0.05$) from the condition on the phase difference. Time 0 marks the stimulus onset. Shadings indicate the 95% confidence intervals across participants.

In the inter-ROI correlation analysis, we found major effects from the observers' responses (doubt or trust) on both the player's and the observer's frontal EEG activities, and both effects occurred during a time period when the correlations reduced. As a linear approach, the Pearson correlation decreases not only when the input series exhibit non-collinear trends in amplitude changes but also when the series oscillate with phase shifts. To address the issue about if the detected effects were induced by the uneven bilateral phase difference conditioned on the

observers' responses, we subtracted the instantaneous phase of the signals in the left vlPFC with respect to time by that in the right vlPFC and show the phase differences in either response conditions in Figure 10. We found major effects detected by the RM-ANOVA from the observers' responses on the players' bilateral vlPFC phase difference. However, such effects were minor on the observers' neural activities. For the player, bilateral asymmetrical activities were observed and such asymmetry reflected different valences between the "trust" and "doubt" conditions. As shown in Figure 10, player's left vlPFC lead the right vlPFC temporally if the observer doubted the player, whereas the right lead the left if the observer trusted the player. These findings provide further evidence to support that the player's emotional modulation could impact the observer's behaviors through facial expressions. However, at this stage, we couldn't associate the inter-ROI correlations with such phase differences, as they were incurred in non-overlapped temporal duration.

4. Discussion

In the behavioral analysis, we found the numbers of correct and incorrect trials were relatively even across all participants. This raised the question about whether the observers made their decisions by random guesses or if they utilized the information conveyed through the players' facial expressions. Regarding this question, we had two major findings in this study. First, not only the players but also the observers could be consciously aware of the player's task of whether being instructed or spontaneous. In the EEG data, we found the related information was represented by the activities recorded from the vlPFC, by which both roles' behaviors were encoded and/or impacted. As the evidence, our analyses detected significant effects of the player's task on both the magnitude of ERPs from the players' lateral ROIs and the degree of observers' bilateral frontal asymmetries. It was also confirmed by the significantly higher doubt-to-trust ratio for the instructed conditions than spontaneous ones. This indicates the observers' responses were directly influenced by the player's facial expressions. Second, the observers' responses were directly impacted by the players' neural activities, as indicated by the significant effect of the observer's response (doubt versus trust) on the players' dmPFC-vlPFC correlations and bilateral vlPFC phase differences. A simple and straightforward postulate to this phenomenon is that the observer's decision to trust or doubt the player depended on their ability to detect or recognize specific facial behaviors of the player influenced by the dmPFC-vlPFC and bilateral vlPFC consistencies. Whereas from the observer's side, detection or recognition to these facial expressions might be encoded by the dynamics of their bilateral asymmetries in the vlPFC, as indicated by the significant effect of the observer's response (doubt versus trust) on their inter-lateral ROI correlations.

Our results indicate that the observers responded based on the players' facial expressions rather than guessing. We propose a neural coupling model, as illustrated in Figure 11, to explain how the facial expressions were cortically processed by both roles and how their behaviors were

linked to their neural activities. We discuss the neural correlates involved in this model in the next paragraph.

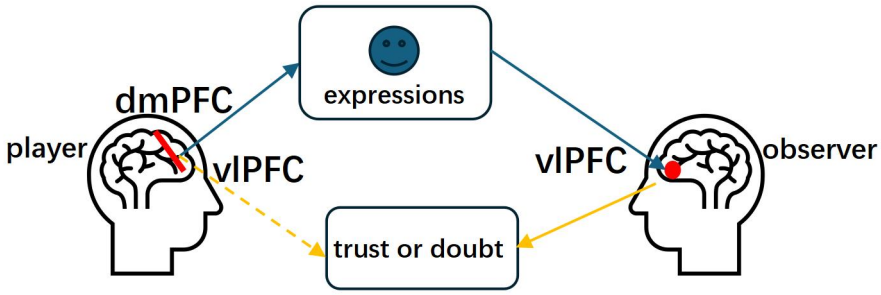

Figure 11. Illustration of a neural coupling model between the player and the observer. The blue arrows illustrate the information path, and the yellow arrows illustrate the effects observed in the data.

Our findings involved the EEG signals collected from our three ROIs that mainly covered the dmPFC and the vlPFC. These areas were documented to play critical roles during the emotion regulation and perception. Recent research has implicated that the dmPFC functioned in regulating emotions by influencing other prefrontal regions (Scharnowski et al., 2020; Baumgartner et al., 2011; Koush et al. 2019). The involvement of dmPFC in emotional regulation has been reported by a series of studies, which found the increased activation of dmPFC with respect to increased emotional levels in contrast to the control (Domes et al., 2010; Ochsner et al., 2004; Ochsner et al., 2009). In addition, the exerting role of dmPFC in emotional regulation has also been emphasized by uncovering the functional connections between the dmPFC and other regions including superior frontal gyrus (SFG), ventral medial prefrontal cortex (vmPFC), and amygdala (Koush et al., 2019; Braunstein et al., 2017; Berboth and Morawetz, 2021; Banks et al. 2007). Given the other functions that were recognized, including interacting emotions with memory (Kensinger et al., 2021), action controls (Domenech and Koechlin, 2015), and integrating external information during social cognition (Martin et al., 2017), dmPFC has been considered to serve as the key hub for emotional regulation. As for the vlPFC, neuroimaging studies have shown that this region activated in response to emotional facial expressions by associating with the amygdala (Meta-analysis: Fusar-Poli et al., 2009). For example, Liu et al. (2017) observed weaker modulations between vlPFC and amygdala with appeasing faces compared to the control in macaques. Also, evidenced in an TMS study, subjects' sensitivity to happy and angry faces were mediated after their vlPFC was interrupted (Chick et al., 2020). The vlPFC has thus been considered to be responsible for facial emotion perception.

Consistent to the previous findings, this study found that the EEG signals recorded from vlPFC were sensitive to two factors: the player's task (instructed versus spontaneous) and the observer's response (doubt versus trust), and all the information was conveyed by the players' facial expressions. Furthermore, the player's dynamics of consistency between dmPFC and vlPFC

were sensitive to the observer's responses. This indicates the linkage between the player's emotional regulations and the observer's behaviors at a social interaction aspect. Taken together, a neural coupling from the player's emotional regulation system to the observer's emotion perception system was likely to be built based on the player's facial expressions during the deception experiment.

5. Conclusions

In this study, we investigated an EEG dataset regarding the social interaction by facial expressions during the deception. Our analysis led to two major findings, suggesting the active role that the frontal cortex played in encoding social interactions: (1) Despite not being explicitly informed of the specific conditions, the observer was more likely to doubt the player in instructed conditions, with the effect reflected by both behavioral data and the observer's vlPFC activities. (2) Both the observer's (vlPFC) and the player's (dmPFC-vlPFC, bilateral vlPFC) neural activities reflected effects of the observers' responses (trust versus doubt). According to the previous research, dmPFC-vlPFC connectivity and vlPFC activities were related to emotional regulation and emotion perception, respectively. Consequently, our findings suggest the presence of a neural coupling process based on facial expressions from the player's emotional regulation system to the observer's emotion perception system.

Data Availability Statement

The data used in the paper is readily accessible on Figshare at https://doi.org/10.6084/m9.fgshare.24760827.v3 (Chen, Fazil, and Wallraven, 2024).

Acknowledgments

The author declared no funding source in support of this study and no conflict to interest.

References

Abe, N. (2011). How the brain shapes deception: An integrated review of the literature. *The Neuroscientist*, *17*(5), 560-574.

Abe, N., Suzuki, M., Tsukiura, T., Mori, E., Yamaguchi, K., Itoh, M., & Fujii, T. (2006). Dissociable roles of prefrontal and anterior cingulate cortices in deception. *Cerebral cortex*, *16*(2), 192-199.

Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological science*, *18*(6), 524-530.

Amodio, D. M., Shah, J. Y., Sigelman, J., Brazy, P. C., & Harmon-Jones, E. (2004). Implicit regulatory focus associated with asymmetrical frontal cortical activity. *Journal of Experimental Social Psychology*, *40*(2), 225-232.

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature neuroscience*, *6*(2), 115-116.

Banks, S. J., Eddy, K. T., Angstadt, M., Nathan, P. J., & Phan, K. L. (2007). Amygdala–frontal connectivity during emotion regulation. *Social cognitive and affective neuroscience*, *2*(4), 303-312.

Battigalli, P., Charness, G., & Dufwenberg, M. (2013). Deception: The role of guilt. *Journal of Economic Behavior & Organization*, *93*, 227-232.

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature neuroscience*, *14*(11), 1468-1474.

Berboth, S., & Morawetz, C. (2021). Amygdala-prefrontal connectivity during emotion regulation: A meta-analysis of psychophysiological interactions. *Neuropsychologia*, *153*, 107767.

Blank, I., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. Journal of neurophysiology, 112(5), 1105-1118.

Braunstein, L. M., Gross, J. J., & Ochsner, K. N. (2017). Explicit and implicit emotion regulation: a multi-level framework. *Social cognitive and affective neuroscience*, *12*(10), 1545-1557.

Chen, Y., Fazli, S., & Wallraven, C. (2024). An EEG Dataset of Neural Signatures in a Competitive Two-Player Game Encouraging Deceptive Behavior. *Scientific data*, *11*(1), 389.

Chick, C. F., Rolle, C., Trivedi, H. M., Monuszko, K., & Etkin, A. (2020). Transcranial magnetic stimulation demonstrates a role for the ventrolateral prefrontal cortex in emotion perception. *Psychiatry research*, *284*, 112515.

Domenech, P., & Koechlin, E. (2015). Executive control and decision-making in the prefrontal

cortex. *Current opinion in behavioral sciences*, *1*, 101-106.

Domes, G., Schulze, L., Böttger, M., Grossmann, A., Hauenstein, K., Wirtz, P. H., ... & Herpertz, S. C. (2010). The neural correlates of sex differences in emotional reactivity and emotion regulation. *Human brain mapping*, *31*(5), 758-769.

Ekman, P., & Davidson, R. J. (1994). Affective science: A research agenda. *The nature of emotion: Fundamental questions*, 411-430.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., ... & Politi, P. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of psychiatry and neuroscience*, *34*(6), 418-432.

Gainotti, G. (1972). Emotional behavior and hemispheric side of the lesion. *Cortex*, *8*(1), 41-55.

Goldstein, K. (1944). The mental changes due to frontal lobe damage. *The Journal of Psychology*, *17*(2), 187-208.

Graham, R., & Cabeza, R. (2001). Event-related potentials of recognizing happy and neutral faces. *Neuroreport*, *12*(2), 245-248.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. Frontiers in Neuroinformatics, 7, 267.

Harmon-Jones, E., Gable, P. A., & Peterson, C. K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological psychology*, *84*(3), 451-462.

Ito, A., Abe, N., Fujii, T., Hayashi, A., Ueno, A., Mugikura, S., ... & Mori, E. (2012). The contribution of the dorsolateral prefrontal cortex to the preparation for deception and truth-telling. *Brain research*, *1464*, 43-52.

Kschischang, F. R. (2006). The hilbert transform. University of Toronto, 83, 277.

Kensinger, E. A., & Ford, J. H. (2021). Guiding the emotion in emotional memories: The role of the dorsomedial prefrontal cortex. *Current Directions in Psychological Science*, *30*(2), 111-119.

Kireev, M., Korotkov, A., Medvedeva, N., Masharipov, R., & Medvedev, S. (2017). Deceptive but not honest manipulative actions are associated with increased interaction between middle and inferior frontal gyri. *Frontiers in neuroscience*, *11*, 482.

Kireev, M., Slioussar, N., Korotkov, A. D., Chernigovskaya, T. V., & Medvedev, S. V. (2015). Changes in functional connectivity within the fronto-temporal brain network induced by regular and irregular Russian verb production. *Frontiers in Human Neuroscience*, *9*, 36.

Koush, Y., Pichon, S., Eickhoff, S. B., Van De Ville, D., Vuilleumier, P., & Scharnowski, F. (2019). Brain networks for engaging oneself in positive-social emotion regulation. *Neuroimage*, *189*, 106-115.

Langleben, D. D., Loughead, J. W., Bilker, W. B., Ruparel, K., Childress, A. R., Busch, S. I., & Gur, R. C. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Human brain mapping*, *26*(4), 262-272.

Liu, N., Hadj-Bouziane, F., Moran, R., Ungerleider, L. G., & Ishai, A. (2017). Facial expressions evoke differential neural coupling in macaques. *Cerebral Cortex*, *27*(2), 1524-1531.

Martin, A. K., Dzafic, I., Ramdave, S., & Meinzer, M. (2017). Causal evidence for task-specific involvement of the dorsomedial prefrontal cortex in human social cognition. *Social Cognitive and Affective Neuroscience*, *12*(8), 1209-1218.

Nunez, J. M., Casey, B. J., Egner, T., Hare, T., & Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *Neuroimage*, *25*(1), 267-277.

Obeso, I., Robles, N., Marrón, E. M., & Redolar-Ripoll, D. (2013). Dissociating the role of the pre-SMA in response inhibition and switching: a combined online and offline TMS approach. *Frontiers in human neuroscience*, *7*, 150.

Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., & Gross, J. J. (2004). For better or for worse: neural systems supporting the cognitive down-and up-regulation of negative emotion. *Neuroimage*, *23*(2), 483-499.

Ochsner, K. N., Ray, R. R., Hughes, B., McRae, K., Cooper, J. C., Weber, J., ... & Gross, J. J. (2009). Bottom-up and top-down processes in emotion generation: common and distinct neural mechanisms. *Psychological science*, *20*(11), 1322-1331.

Priori, A., Mameli, F., Cogiamanian, F., Marceglia, S. A. R. A., Tiriticco, M., Mrakic-Sposta, S., ... & Sartori, G. (2008). Lie-specific involvement of dorsolateral prefrontal cortex in deception. *Cerebral cortex*, *18*(2), 451-455.

Robinson, R. G., & Price, T. R. (1982). Post-stroke depressive disorders: a follow-up study of 103 patients. *Stroke*, *13*(5), 635-641.

Scharnowski, F., Nicholson, A. A., Pichon, S., Rosa, M. J., Rey, G., Eickhoff, S. B., ... & Koush, Y. (2020). The role of the subgenual anterior cingulate cortex in dorsomedial prefrontal–amygdala neural circuitry during positive-social emotion regulation. *Human brain mapping*, *41*(11), 3100-3118.

Scheuble, V., & Beauducel, A. (2020). Individual differences in ERPs during deception: Observing vs. demonstrating behavior leading to a small social conflict. *Biological Psychology*, *150*, 107830.

Sebastian, A., Baldermann, C., Feige, B., Katzev, M., Scheller, E., Hellwig, B., ... & Klöppel, S. (2013). Differential effects of age on subcomponents of response inhibition. *Neurobiology of aging*, *34*(9), 2183-2193.

Seiter, J. S., & Bruschke, J. (2007). Deception and emotion: The effects of motivation, relationship type,

and sex on expected feelings of guilt and shame following acts of deception in United States and Chinese samples. *Communication Studies*, *58*(1), 1-16.

Spence, S. A., Farrow, T. F., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport*, *12*(13), 2849-2853.

Spence, S. A., Hunter, M. D., Farrow, T. F., Green, R. D., Leung, D. H., Hughes, C. J., & Ganesan, V. (2004). A cognitive neurobiological account of deception: evidence from functional neuroimaging. Philosophical Transactions of the Royal Society B: Biological Sciences, 359(1451), 1755.

Vartanian, O., Kwantes, P. J., Mandel, D. R., Bouak, F., Nakashima, A., Smith, I., & Lam, Q. (2013). Right inferior frontal gyrus activation as a neural marker of successful lying. *Frontiers in Human Neuroscience*, *7*, 616.

Verschuere, B., Schuhmann, T., & Sack, A. T. (2012). Does the inferior frontal sulcus play a functional role in deception? A neuronavigated theta-burst transcranial magnetic stimulation study. *Frontiers in Human Neuroscience*, *6*, 284.

Appendix

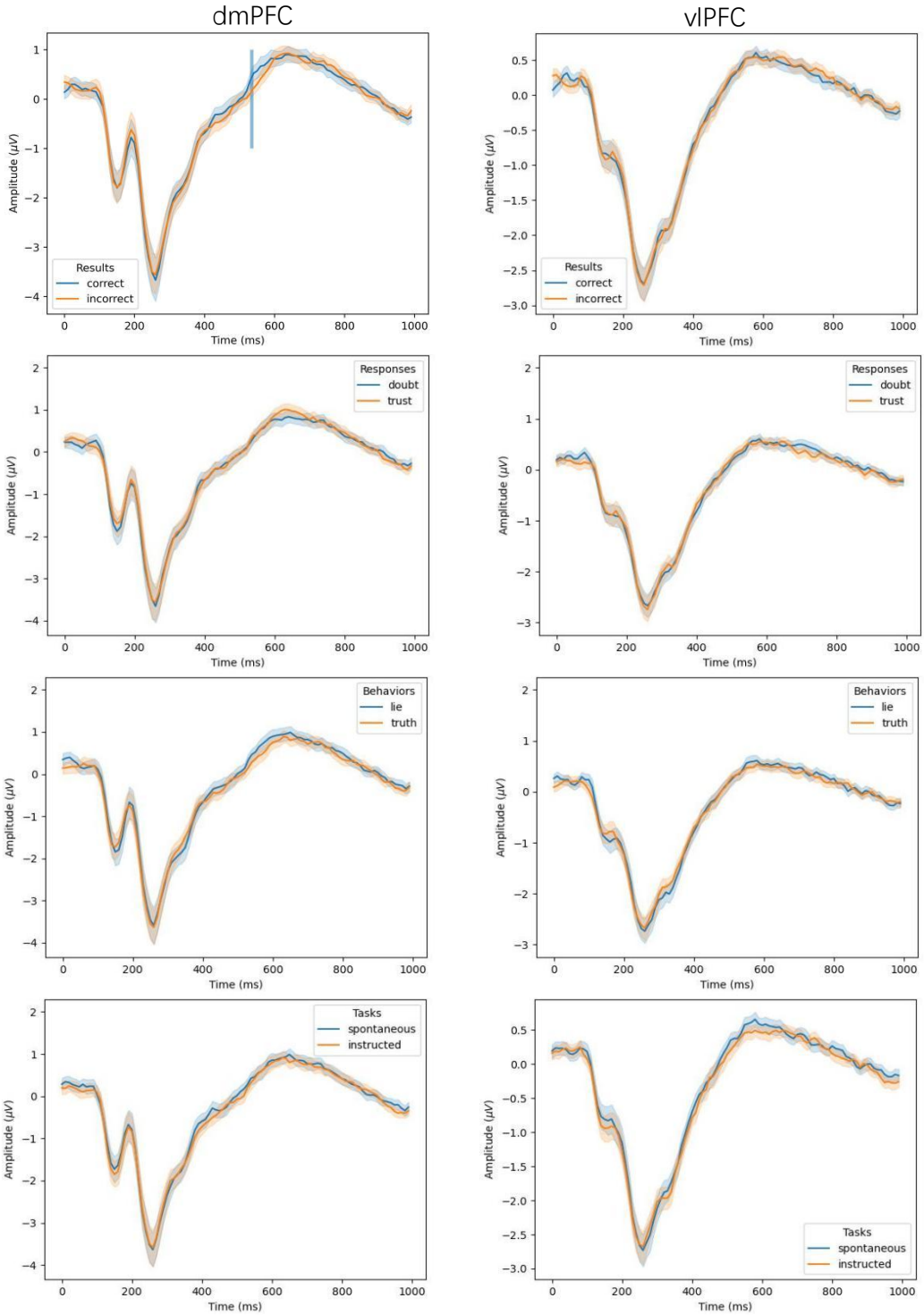

Figure A1. Center and lateral frontal ERPs for the observers conditioned by different variables (indicated by the legends). Blue boxes highlighted the temporal duration where RM-ANOVAs detected significant effects ($p<0.05$) from the condition on the ERPs after multiple test corrections. Time 0 marks the stimulus onset. Shadings indicate the 95% confidence intervals across participants.

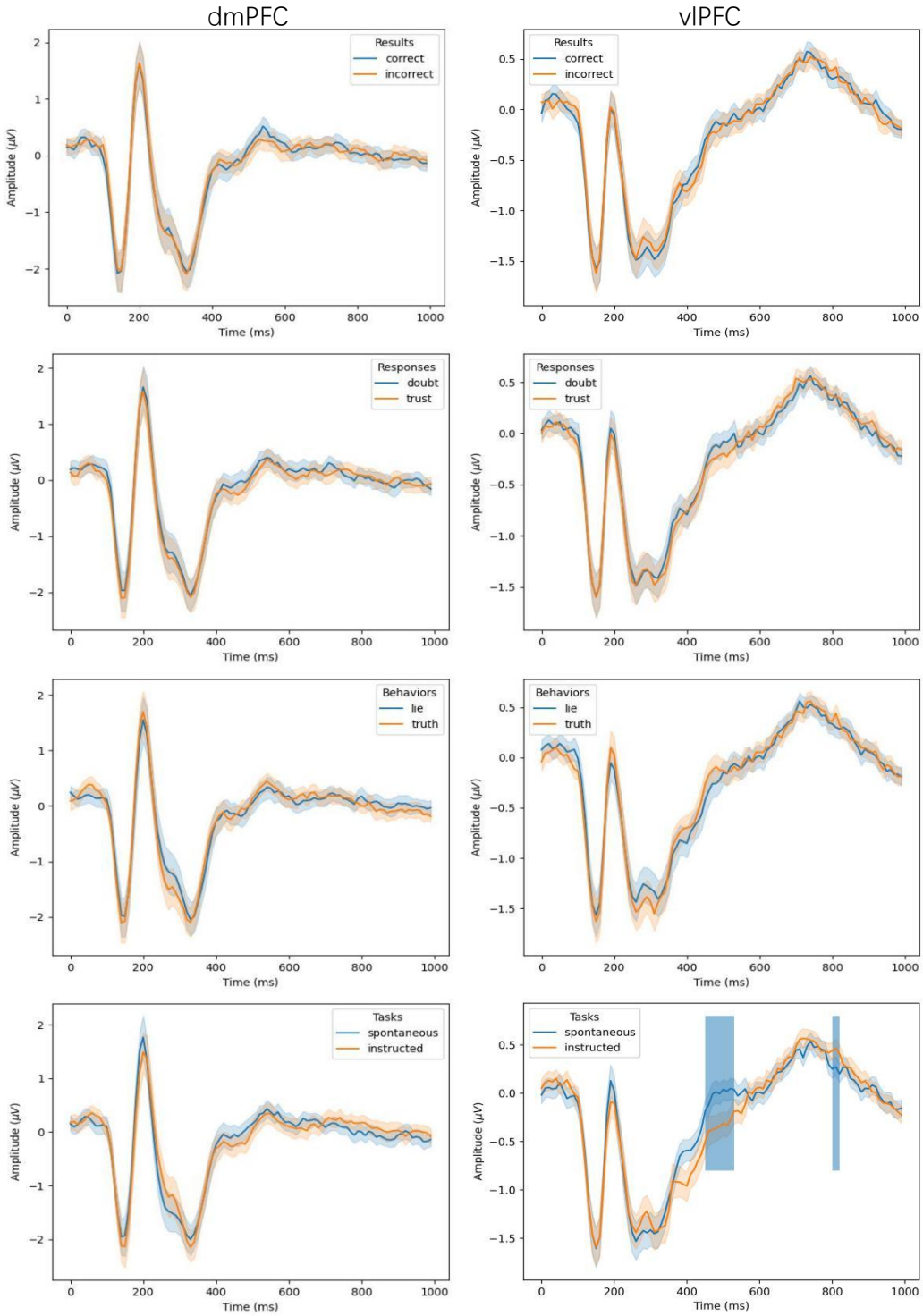

Figure A2. Center and lateral frontal ERPs for the players conditioned by different variables (indicated by the legends). Blue boxes highlighted the temporal duration where RM-ANOVAs detected significant effects ($p<0.05$) from the condition on the ERPs after multiple test corrections. Time 0 marks the stimulus onset. Shadings indicate the 95% confidence intervals across participants.

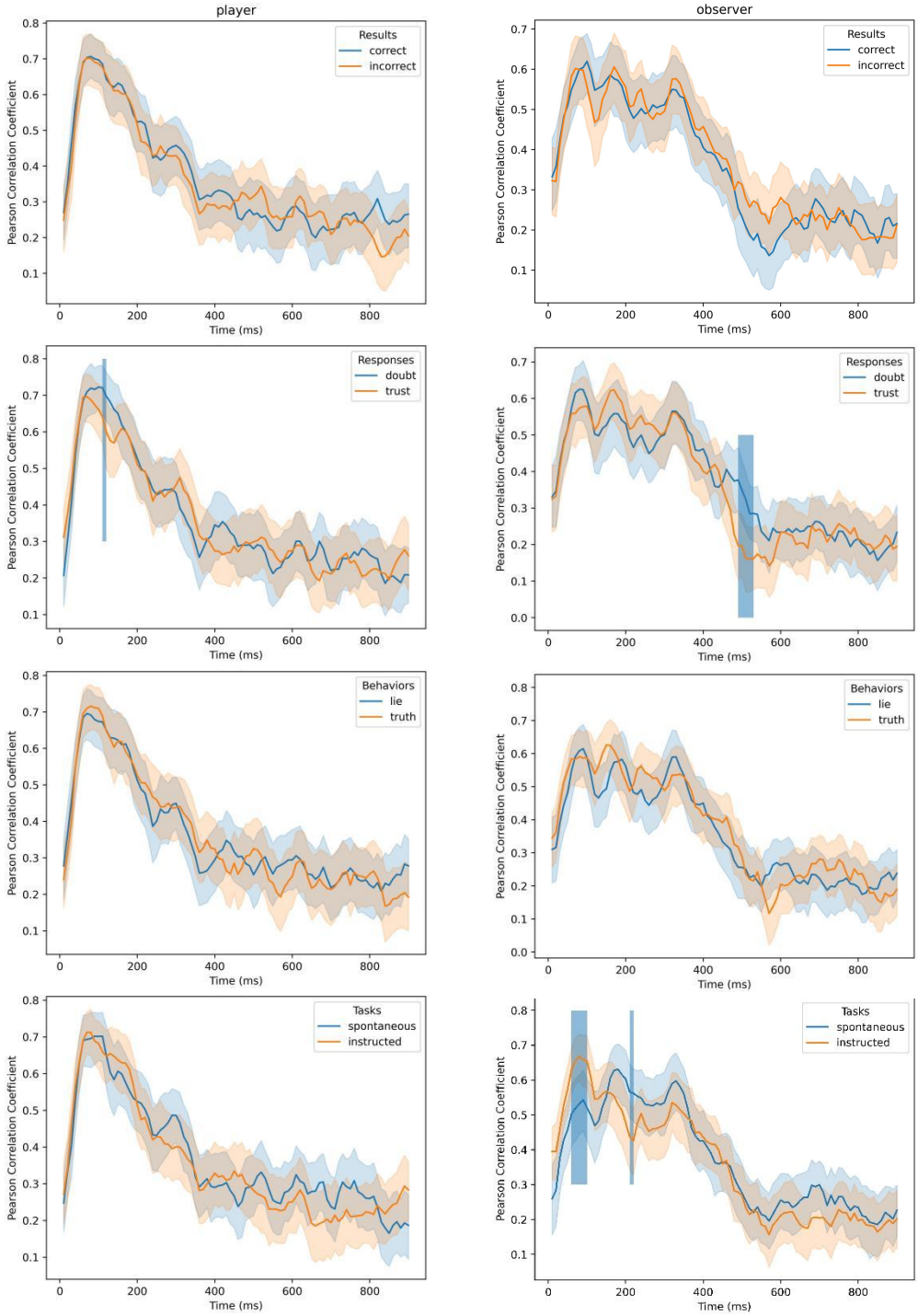

Figure A3. Pearson correlation coefficients between ERPs from left and right vlPFCs for the observers (right panel) and players (left panel) conditioned by different variables (indicated by the legends). Blue boxes highlighted the temporal duration where RM-ANOVAs detected significant effects ($p<0.05$) from the condition on the correlation coefficients after multiple test corrections. Time 0 marks the stimulus onset. Shadings indicate the 95% confidence intervals across participants.

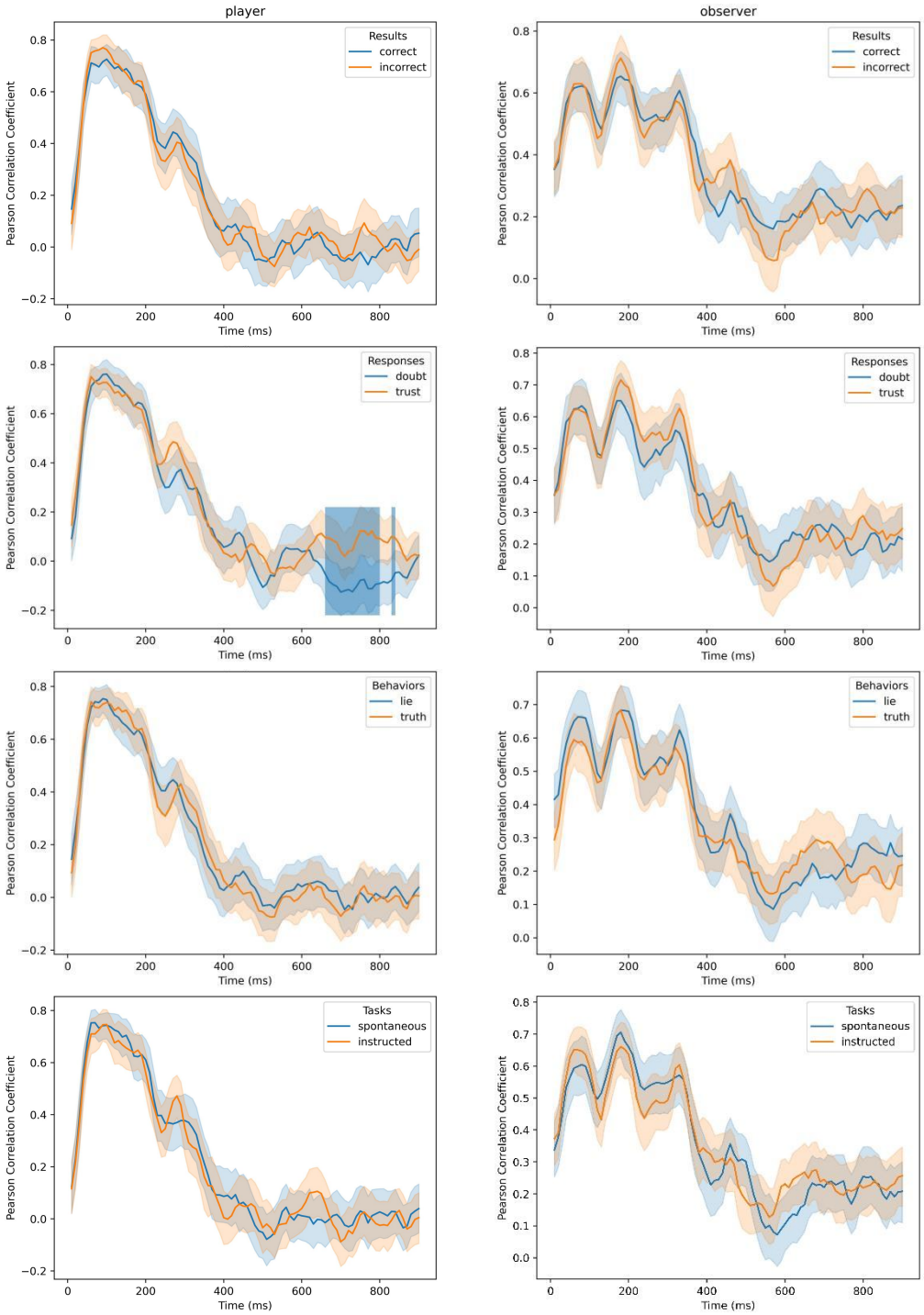

Figure A4. Pearson correlation coefficients between ERPs from dmPFC and vlPFC for the observers (right panel) and players (left panel) conditioned by different variables (indicated by the legends). Blue boxes highlighted the temporal duration where RM-ANOVAs detected significant effects ($p<0.05$) from the condition on the correlation coefficients after multiple test corrections. Time 0 marks the stimulus onset. Shadings indicate the 95% confidence intervals across participants.