The role of anterior prefrontal cortex (area 10) in face-to-face deception measured with fNIRS

P. Pinti[1,2], Andrea Devoto[2], Isobel Greenhalgh[2], I. Tachtsidis[1], P. Burgess[2], A. Hamilton[2]

[1]Department of Medical Physics and Biomedical Engineering, University College London, Gower Street, Malet Place Engineering Building, WC1E 6BT, London, United Kingdom
[2]Institute of Cognitive Neuroscience, University College London, Alexandra House, WC1N 3AZ, London, United Kingdom

Corresponding author: Antonia Hamilton, a.hamilton@ucl.ac.uk

Running Title: Anterior prefrontal cortex in deception.

Words count: 8174

Abstract

Anterior prefrontal cortex (PFC; Brodmann area 10) activations are often, but not always, found in neuroimaging studies investigating deception, and the precise role of this area remains unclear. To explore the role of PFC in face-to-face deception, we invited pairs of participants to play a card game involving lying and lie detection while we used functional near infrared spectroscopy (fNIRS) to record brain activity in PFC. Participants could win points for successfully lying about the value of their cards or for detecting lies. We contrasted patterns of brain activation when the participants either told the truth or lied, when they were either forced into this or did so voluntarily, and when they either succeeded or failed to detect a lie. Activation in anterior PFC was found in both lie production and detection, unrelated to reward. Analysis of cross-brain activation patterns between participants identified areas of PFC where the lead player's brain activity was synchronized their partner's later brain activity. These results suggest that during situations that involve close interpersonal interaction, anterior PFC supports processing widely involved in deception, possibly relating to the demands of monitoring one's own, and other people's behaviour.

1. Introduction

Deception is a common aspect of social interactions and is defined as "a deliberate attempt to mislead others" (DePaulo et al., 2003). There is a long history of research into the behaviours, cognitive processes and brain systems which might be linked to deception, but several important questions remain unanswered. In particular, many studies of deception have used simple, non-interactive tasks performed by a single participant, which may lack ecological validity. Here, we used a naturalistic interactive card game to enable participants to lie and detect lies in a face-to-face context, while brain activity in prefrontal cortex was monitored with functional Near Infrared Spectroscopy (or fNIRS). To explain the rationale behind our study, we first review previous

neuroimaging studies of lying, and then describe the specific task and cognitive manipulations which we use here.

Studies of the human ability to tell lies have come to a general consensus that lying is a more effortful task than telling the truth because it requires the engagement and integration of more cognitive processes, such as mentalising to monitor the other person's ongoing beliefs and inhibition to suppress the true response. It is believed that prefrontal cortex (PFC) plays a particular role in deception (Spence et al., 2001; Kozel et al., 2004a; Kozel et al., 2004b; Spence at al., 2004; Kozel et al., 2005; Langleben et al., 2005; Abe et al., 2006; Abe et al., 2007; Gamer et al., 2007; Spence et al., 2008; Bhatt et al., 2009), and different subregions of PFC may have different roles. In a recent meta-analysis of studies of lying, strong activation of dorsal anterior cingulate, right TPJ and temporal poles was found in "socially interactive" studies of deception, compared to non-interactive studies (Lisofsky et al., 2014). In addition, anterior PFC (i.e. the most anterior part of the brain, often also referred to as rostral prefrontal or the frontal pole, corresponding to BA10) activations are often reported in "socially interactive" deception paradigms involving face-to-face interaction or considerable personal contact (Abe et al., 2007; Ganis et al., 2009; Sip et al., 2010) in contrast to "non-interactive" protocols with no direct interpersonal exchange (Abe et al., 2014; Langleben et al., 2005). These findings motivate us to study the role of prefrontal cortex during lying and lie detection in an interactive context.

Studying the neural substrates of deception is challenging because the contexts where real world lying occurs cannot easily be reproduced in the lab. To obtain good ecological validity, studies need to be close in format, demands, and structure to situations that the participant may encounter in everyday life. Part of this naturalistic format is that lying commonly takes place within the context of an interaction between two people. This means that the liar has to produce a suitable deceptive message while controlling his/her behaviour avoiding any deception cues (e.g., changes in voice pitch, speech disturbances, etc.) and appearing natural (Zuckerman et al., 1981). Meanwhile, the person listening to a liar tries to detect the lie, by identifying relevant cues in both speech and

behaviour and interpreting such cues to judge whether the other person is lying or telling the truth. However, neuroimaging modalities like fMRI do not normally permit the kind of natural social behaviours (eye contact, smiling, head turns, etc.) that may be important in face-to-face deception. Thus, we use fNIRS as our neuroimaging modality of choice.

fNIRS is an optical-based neuroimaging method that shines near infrared light into the brain and measures the concentration changes of oxygenated (HbO_2) and deoxygenated (HbR) haemoglobin quantifying the changes in light attenuation. Like fMRI, fNIRS is based on neurovascular coupling and monitors brain activity detecting the changes in cerebral blood flow (i.e., the hemodynamic response) that follows neuronal activity (Scholkmann et al., 2014; Pinti et al., 2018a). It is well-suited for experiments that require natural movement, speaking or interaction with other people (Pinti et al., 2018b). However, fNIRS monitors only the superficial layers of the cortex and the head coverage is determined by the number of optodes (i.e., light sources and detectors) that the instrument is equipped with (Quaresima & Ferrari, 2019). Therefore, we focus on recording brain activity within anterior and lateral PFC bilaterally. fNIRS was previously used to explore the neural mechanisms underlying deception in naturalistic situations, but not extensively. In general and in agreement with previous fMRI studies, greater activations in PFC are also found with fNIRS when deceptive conditions are compared to control and truth-telling conditions (Bunce et al., 2005; Ding et al., 2013; Ding et al., 2014; Tian et al., 2009), also at the single-trial level (Bhutta et al., 2015). Our study builds on these results by examining the cognitive processes engaged in both lie production and detection in a novel interactive task.

In a face-to-face communication between two people, where one person tells a lie to another, there are distinct cognitive processes and behaviours which are engaged in both the person telling the lie and the person trying to detect it. Lie production is believed to involve inhibiting the tendency to tell the truth, theory of mind and reward processes (Spence et al., 2004; Sip et al., 2008). In terms of brain systems, the inhibitory processes which are required to suppress the tendency to tell the truth and to produce a lie instead, have been linked to lateral PFC (Abe et al., 2006). Mentalising is required

to consider what another person knows and how to effectively communicate false information to that person. This can include second order mentalising, where one tells the truth with the intent to deceive the other (Sip et al., 2010) and has been linked to medial PFC (Amodio & Frith, 2006; Sip et al., 2010; Yu et al., 2019). Finally, in a trial-by-trial experimental context, deception may involve reward processes (when the liar feels they have been successful) or punishment (if they are caught). Such processes are supported by the orbitofrontal cortex (Fellows, 2011; Wallis, 2012) and the basal ganglia (Hikosaka et al., 2006; Foti et al., 2011).

Thus, a situation involving face-to-face lying can engage several different cognitive processes linked to different brain systems. We aimed to design a task that would allow natural social interaction, and spontaneous production of lies or truthful statements while also distinguishing these different aspects of lying. We were particularly interested in the difference between telling the truth with an intent to deceive, and telling the truth without such an intent, because the former requires second order mentalising processes but neither involves the inhibition of a truthful response. In order to create such a task, we draw on perhaps the most common situation where two people (legally) deceive each other repeatedly and spontaneously while face to face with each other - playing card games. Dice games have previously been used to study spontaneous social lying (Sip et al., 2010), and here we follow a similar logic, creating a simple card game where participants can win points for a successful lie when in the role of Informer, or can win points for detecting a lie when in the role of Guesser (Figure 1).

This context also allows us to study the brain systems engaged in the Guesser as they try to detect their partner's lies. Whilst substantial research has been carried out on the neural correlates of lie production, we still know very little about the detection of lies in face-to-face interaction. Many studies have examined how we judge if another person is trustworthy in the context of economic games (King-Casas et al., 2005) and find evidence of involvement of anterior cingulate and caudate brain regions. For example, anterior cingulate is engaged when someone receives unfair behaviour (Fett et al., 2012). In addition, detecting someone else is lying may alter our impression of that person,

and medial PFC was found to be involved when we form a first impression or when we try to determine the psychological and mentality traits of another person (Mitchell et al., 2005). However, few previous studies have directly investigated how people detect if another person is lying.

Finally, the use of wearable fNIRS allowed us to collect data from pairs of participants at the same time in a hyperscanning configuration. Recently, there has been an explosion of interest in understanding fNIRS hyperscanning (Hirsch et al., 2018; Gvirts & Perlmutter, 2020). Several studies suggest that there is cross-brain coherence in prefrontal cortex when participants engage in competition or cooperative tasks (Cui et al., 2012; Fishburn et al., 2018; Liu et al., 2016; Piva et al., 2017) but the patterns of cross-brain coherence during competitive tasks like our card game is less clear. The precise localisation and timing of these coherence patterns is also rather variable, with many studies reporting coherence on long time scales that are hard to link to specific cognitive processes. Here, we analyze the corresponding patterns of brain activity both in the single brain using classical inference methods like the General Linear Model (GLM) (Friston et al., 1994) and we use an exploratory analysis to test if there is any evidence for cross-brain correlations during our competitive task, and if we can identify a particular time lag for the correlations. To do this, we apply a novel method proposed by Stephens et al. (2010) as a proof-of-principle to look at the brain-to-brain coupling between the two people (Stephens et al., 2010; Kingsbury et al., 2019). In this method, we fit the brain activity in each participant with a traditional design matrix approach and add an additional regressor to the design matrix which represents the brain activity at the same location in their partner. As our participants have clearly defined Informer / Guesser roles where one person makes a statement and the other responds, we can also test the direction of any cross-brain effects. That is, we can time-shift this additional regressor by different amounts and then test which time-lag gives the strongest relationship between prefrontal cortex of the Informer and Guesser. This is an exploratory method and our aim here is to test if it can be applied to fNIRS data as a proof-of-principle. If a relationship between the Informer brain and Guesser brain can be identified, further study will be needed to determine what this pattern means in terms of cognitive processes.

To summarise, this study aims to investigate lying and lie detection in an interactive, two-person context within the domain of 'second person neuroscience' (Schilbach et al., 2013). We use a simple card game to create trials where participants spontaneously lie or tell the truth, with or without the intent to deceive. This means that we can distinguish processes of reward, inhibition and theory of mind between different trials. We can also examine, for the first time, the brain mechanisms involved when a participant tries to detect that their partner is lying, and we can explore the time lags between patterns of brain activity in the two participants engaged in a communicative and competitive interaction.

Our hypotheses were: (i) the production of lies is associated with an increase in anterior PFC activity respect to truth-telling reflecting the recruitment of the executive system in lying; (ii) telling the truth with the intent to deceive will engage medial anterior PFC more than truth telling without an intent to deceive, because the former engages more mentalising processes. Further exploratory analyses will examine differences between instructed and spontaneous lies and neural mechanisms engaged when detecting lies in one's partner.
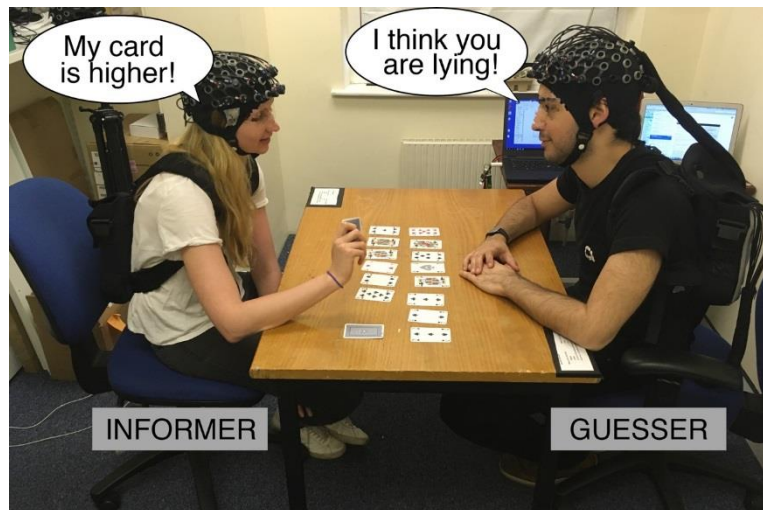
2. Materials and Methods

2.1 Participants

Seventy-eight healthy adults (26.4±10.1 years of age, 33 male/45 female) were recruited from the University College London's psychology subject pool and participated in a two-person interactive card game experiment. Participants were all strangers and were arranged in pairs for a total of 39 dyads. All participants had normal or corrected-to-normal vision and were required not to have a history of gambling addiction. In addition, only individuals with previous experience with card games were included in the study. All participants provided written informed consent in accordance with the guidelines approved by the University College London local research ethics committee.

One dyad was excluded because one of the participants could not play card games. Three dyads were further excluded due to technical issues (no video recordings for two dyads; no fNIRS

data for one dyad). Data from 35 pairs of participants were thus analysed. Among these, 12 dyads were monitored with one fNIRS device and 23 dyads with two fNIRS instruments due to equipment availability.

2.2. Experimental protocol

Participants were sitting at the opposite sides of a table facing each other (Figure 1). A game called 'Tell', inspired by the card game Cheat, was developed for the purpose of the experiment.

Figure 1. Experimental setup. Two rows of 8 cards each were placed in front of the participants. In this representative round, the participant on the left has the role of Informer and makes a truthful or untruthful statement on the value of her card respect to the other participant's corresponding card; the participant on the right is the Guesser and has to guess whether the Informer told the truth or lied. fNIRS (LIGHTNIRS, Shimadzu Corp., Kyoto, Japan) was used to monitor both participants' brain activity while playing cards.

The card came was comprised of 12 rounds for the dyads monitored with one fNIRS device, and 24 rounds for the dyads with two fNIRS systems. One dyad had 22 rounds and two dyads 23 due to delays in the testing session and time constraints for the next testing slot. Each round consisted of 8 trials or played cards. The experimenter distributed two rows of 8 cards each on the table between

the two participants at each round, with one row containing cards facing up and the other row cards facing down. The player closest to the facing-down cards was referred to as the Informer, while the other player was referred to as the Guesser. In each trial, the Informer picked up a faced-down card and 'informed' the Guesser whether his/her card was 'higher' or 'lower' than the Guesser's card. In particular, the 'ace' was considered the highest card and the 'two' the lowest. The Guesser had then to guess if the Informer was making a truthful claim or was bluffing. After the Guesser's guess, the Informer revealed his/her card and gave the Guesser feedback whether he/she was 'correct' or 'incorrect'. The Guesser would win the trial by guessing correctly whether the Informer was being truthful or not. On the other hand, the Informer would win by 'fooling' the Guesser. At the end of the round, the cards were removed from the table by the experimenter and the new set of cards was arranged on the table. Participants swapped roles after each round. Participants finishing the game with the most won trials got an extra monetary reward, in order to encourage the Informer to deceive the Guesser and to ensure a balanced occurrence of lies and truths.

Control trials were included to have situations where the Informer is forced and instructed by the experimenter to lie or tell the truth rather than do so voluntarily. In particular, there were two types of control trials: 1) Instructed lie: in case both the Informer and the Guesser's cards had the same value, any statement made by the Informer ('higher' or 'lower') counted as a lie; 2) Instructed truth: when the Guesser's card was either the highest (i.e., the ace) or the lowest (i.e., the two), the Informer was forced to tell the truth as it would have been impossible to deceive the Guesser successfully. The latter allowed us to have trials where truth is told with no intention to deceive.

As a result, there were 7 possible trial outcomes based on the Informer and Guesser statements. The possible scenarios and related outcomes are summarized in Table 1. It is worth stressing that in 6 types of trial (TT, TF, FT, FF, I-LT, I-LF) both true and false claims are made by the Informer with the intention to deceive the Guesser, while the I-T condition is the only case where true statements are made with no deceptive intention.

9

| | Scenario | | | Example hand | | | |
| | | | Cards | | Play | | |
Scenario	ID	Outcome	Informer	Guesser	Informer	Guesser
The Informer tells the truth; The Guesser guesses right.	TT	The Guesser wins.	6♣	K♥	'Lower'	'True'
The Informer tells the truth; The Guesser guesses wrong.	TF	The Informer wins.	K♥	8♦	'Higher'	'False'
The Informer lies; The Guesser guesses wrong.	FT	The Informer wins.	4♣	10♦	'Higher'	'True'
The Informer lies; The Guesser calls the bluff.	FF	The Guesser wins.	Q♣	J♠	'Lower'	'False'
Instructed Truth	I-T	Tie	9♣	A♥	'Lower'	'True'
Instructed Lie 1: The Informer lies; The Guesser guesses wrong.	I-LT	The Informer wins.	6♦	6♥	'Higher'	'True'

Instructed Lie 2: The Informer lies; The Guesser guesses right.	I-LF	The Guesser wins.	J♥	J♣	'Lower'	'False'

Table 1. Summary of the possible scenarios for a trial or played pair of cards related to their outcomes. The 'Example hand' column shows how a potential game trial looks like. The Cards column shows the cards owned by the participants and the Play column what the players would have to say to lead to the respective outcome.

A total of 96 card pairs were used for the game. The cards for each round were prepared in advance and set out on the table in a fixed order which appeared random to the participants. This meant that over the whole study, the card pairs were selected to have an equal number of trials for the following six different possible combinations: 1) Same cards pairs (instructed lie trials); 2) Higher-Lower, small difference pairs (a difference in value smaller or equal to 4); 3) Higher-Lower, large difference pairs (a difference in value larger than 4); 4) Lower-Higher, small-difference pairs (a difference in value smaller or equal 4); 5) Lower-Higher, large difference pairs (a difference in value larger than 4); 6. Extremity-card to the Guesser (either a 2 or an Ace; instructed truth trials). Because participants could freely choose to lie or tell the truth on some trials, and could guess correctly or incorrectly, we could not precisely control the trial numbers for each of the conditions, but this manipulation of card order balanced the likely trial types as closely as possible. The rounds were presented in the same order for all pairs.
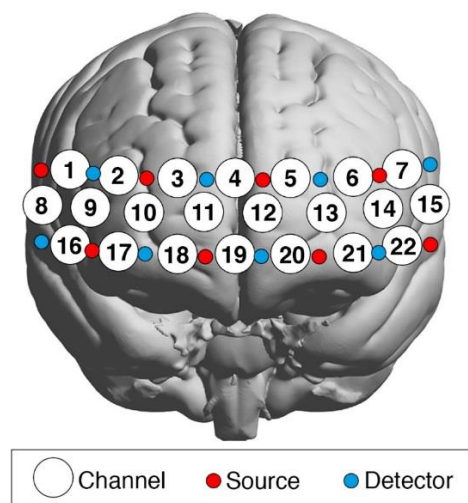
2.3. Equipment

A video camera was used to record the entire experimental session in order to recover the time of onset and the outcome of each trial in post-processing.

Brain oxygenation and hemodynamic signals were recorded using the wearable LIGHTNIRS system (Shimadzu Corp., Kyoto, Japan). For 23 dyads, two wearable LIGHTNIRS devices were used and both participants were monitored simultaneously. In 12 dyads, fNIRS data were acquired on one participant at a time for half of the rounds with one LIGHTNIRS instrument; then the device was placed on the other participant for the second half of the rounds. This was due to limited equipment availability.

The LIGHTNIRS system is equipped with 8 semiconductor lasers (light sources) and 8 avalanche photodiodes (detectors) arranged in an alternated configuration as shown in Figure 2, creating 22 measurement channels covering PFC. Each light source emitted light at three wavelengths (780, 805, and 830 nm) and raw intensity signals were recorded at a sampling frequency of 13.33 Hz.

In order to place the fNIRS cap in a reliable way across all the participants, we followed a similar procedure used in our previous study (Pinti et al., 2015). More precisely, channel 19 was placed in correspondence of the Fpz point of the 10-20 electrode placement system, which was previously marked on the participant's head as the 10% of the Nasion-Inion distance from the Nasion (Pinti et al., 2015).

Figure 2. Schematic representation of sources (red circles) and detectors (blue circles) placement. Sources and detectors are arranged in a 8 x 2 configuration creating 22 measurement channels (white circles).

For one participant, we recorded the coordinates of the fNIRS optodes and anatomical landmarks based on the 10-20 electrode placement system (Nasion, Inion, right and left pre-auricular points, Cz) using a 3D magnetic digitizer (Patriot, Polhemus, Vermont). The fNIRS channel locations were then co-registered onto a standard brain template, and the corresponding MNI coordinates and their anatomical locations were then estimated using the NIRS-SPM package (Ye at al., 2009). The resulting MNI coordinates, their anatomical locations, and the atlas-based probabilities are listed in Supplementary Table 1. Based on our previous studies, we are confident that the procedure described in Pinti et al. (2015) provides a consistent and similar placement of the fNIRS probe over the PFC across participants. Therefore, here we consider the one participant's 3D-digitized data as representative of the fNIRS probe placement in our experimental cohort. To make our discussions easier to follow, we describe our results in relation to the anatomical location of the activation rather than in terms of channel number. We have thus divided the probed brain area (i.e., PFC) into three sub-regions using the anatomical locations obtained through the co-registration procedure (Supplementary Table 1): (i) Right PFC: Channels 1, 2, 8, 9, 16, 17; (ii) Anterior PFC: Channels 3, 4, 5, 10, 11, 12, 13, 18, 19, 20; Left PFC: Channels 6, 7, 14, 15, 21, 22.

2.4. Data analysis

Video recordings of each experimental session were coded using the ELAN software (Wittenburg et al., 2006). In particular, we recovered the outcome of the trials (Table 1) and the start and end times of each trial, where each trial starts when the Informer picks up a card and ends when the Informer gives feedback to the Guesser (correct/incorrect). This was done to recover the onsets of each trial per condition or outcome (Table 1) to use for the analysis of fNIRS data. The number of TT, TF, FT, and FF trials were variable across participants as the Informer chose when to lie and when to tell the truth spontaneously, while the number of control trials (I-LT, I-LF, I-T) was fixed.

In Table 2, we present the average number of trials across participants occurring for each possible scenario and for the two roles, and the corresponding average trial duration.

Role	Scenario ID	Number of trials (mean ± std. dev.)	Duration of trials (s) (mean ± std. dev.)
Informer	TT	24.6 ± 8.3	4.5 ± 1.0
	TF	9.4 ± 5.5	5.4 ± 1.7
	FT	10.9 ± 6.0	4.8 ± 1.4
	FF	12.2 ± 7.0	5.3 ± 1.6
	I-T	10.9 ± 2.8	3.8 ± 1.0
	I-LT	6.2 ± 2.0	4.9 ± 1.4
	I-LF	5.7 ± 2.5	5.5 ± 1.7
Guesser	TT	23.8 ± 9.0	4.4 ± 1.0
	TF	9.8 ± 5.1	5.4 ± 1.6
	FT	11.9 ± 5.4	4.7 ± 1.3
	FF	12.1 ± 6.8	5.3 ± 1.6
	I-T	10.5 ± 3.3	3.8 ± 1.0
	I-LT	5.7 ± 2.3	5.0 ± 1.4
	I-LF	5.5 ± 2.7	5.3 ± 1.6

Table 2. Group average number of duration of trials for each experimental scenario and role. Mean and standard deviation across the 70 analysed participants are reported.

The pre-processing of fNIRS data was carried out using the Homer2 toolbox (Huppert et al., 2009) following similar pre-processing pipeline steps described in our recent publication on the standardization of fNIRS analysis procedures (Pinti et al., 2018c). Raw intensity data were first visually inspected to assess signals' quality. Channels showing detectors saturation or poor optical

coupling as marked by a lack of the heart beat frequency (~1 Hz) in the signal's power spectrum (Pinti et al., 2018c) were excluded from further analyses. Raw intensity signals were then converted into changes in optical density (function: *hmrInteensity2OD*) and corrected for motion artifacts using the wavelet-based method (function: *hmrInteensity2OD*; *iqr=1.5*; (Molavi & Dumont, 2012)). A band-pass filter (function: *hmrBandpassFilt*; order: 3rd; band-pass frequency range [0.01 0.4] Hz) was used to remove physiological noise such as heart rate, and low frequency noise and slow trends in the data. The modified Beer-Lambert law was then applied to compute the concentration changes of HbO_2 and HbR (function: *hmrOD2Conc*; assuming a fixed DPF of 6 (Yücel et al., 2016)). The pre-processed HbO_2 and HbR were combined into the 'activation signal' (Scholkmann et al., 2014) by means of the correlation based signal improvement (CBSI) method (Cui et al., 2010) in order to localize functional activation on the basis of one signal including the contribution of both HbO_2 and HHb at the same time. The inclusion of both fNIRS signals has the potential to reduce false positives at the statistical inference stage (Tachtsidis & Scholkmann, 2016).

2.4.1. Contrast effects analysis

A first-level (or single-subject) channel-wise GLM (Friston et al., 1994) analysis was applied to the 22 channels activation signals, down-sampled to 3 Hz, using the SPM for fNIRS toolbox (https://www.nitrc.org/projects/spm_fnirs/) to localize functional brain activity occurring in response to the possible scenarios of the game (Table 1).

For each participant, the design matrix modelled the 7 possible outcomes listed in Table 1 for both the rounds as Informer and for the rounds as Guesser, resulting in 14 regressors in total (7 for the Informer role and 7 for the Guesser role). Regressors were computed convolving the event epochs corresponding to each trial of the possible outcomes (i.e., starting when the Informer picked up a card and ending when the Informer provides feedback to the Guesser) with the canonical Hemodynamic

Response Function (HRF) and were used to fit the fNIRS activation signals. Single-subject beta-values were estimated for the following contrasts of interest:

Lying contrasts

Informer: All lying > all truth - [(*FT+FF+I-LF+I-LT*) > (*TT+TF+I-T*)]: to reveal which parts of PFC are more involved in lie production than telling the truth;

Informer: Only spontaneous deception - [(*FT+FF*) > (*TT+TF*)]: to reveal which parts of PFC are more involved in lie production than telling the truth when the Informer decides to do so according to his/her own will;

Informer: Only forced deception - [(*I-LT+I-LF*) > (*I-T*)]: to reveal which parts of PFC are more involved in lie production than telling the truth when the Informer is forced to do so;

Informer: Truth to deceive - [(*TT+TF*) > (*I-T*)]: to assess if there are any differences between telling the truth with the intention to deceive or not;

Effects of instructions

Informer: Chosen trials > Instructed trials - [(*TT+TF+FT+FF*) > (*I-LT+I-LF+I-T*)]: to investigate if activation patterns in the Informer differ in the voluntary trials respect to forced trials;

Informer: Chosen trials > Instructed trials (only lying) - [(*FT+FF*) > (*I-LF+ I-LT*)]: to test if there are any differences between lying voluntarily or when forced;

Overall task contrasts

Guesser: Reward - [(*TT+FF+I-LF*) > (*TF+FT+I-LT*)]: to assess the patterns of brain activity when the Guesser is successful at detecting to when he/she is deceived by the Informer;

Informer: Reward - [(*TF+FT+I-LT*) > (*TT+FF+I-LF*)]: to localize the brain activation patterns to the informer successfully deceiving the Guesser with respect to when he fails in doing so.

Effect of role - [(*TT+TF+FT+FF+ I-LT+I-LF+I-T*)INF > (*TT+TF+FT+FF+ I-LT+I-LF+I-T*)GUESS]: to assess which brain regions are involved more when the participant assumes the Informer role with respect to the Guesser role.

To localize brain activity at the group level, channel-wise one sample t-tests vs 0 were performed on the group beta-estimates for each contrast. To correct for multiple comparisons, we considered as significant the activation patterns where two or more adjacent channels show a significant effect for $p<0.05$ (Southgate et al., 2014).

2.4.2. Brain-to-brain analysis

In order to explore the brain-to-brain coupling between the Informer and the Guesser, we used the procedure introduced by Stephens et al. (2010) as a proof-of-principle. Other approaches have been used previously to analyse fNIRS hyperscanning data, with the wavelet-based cross-brain coherence approach being the most common (Cui et al., 2012; Hirsch et al., 2017; Liu et al., 2016). Here, our design involves short events with variable timing (Table 2) and an asymmetric relationship between participants. This makes the use of conventional coherence-based approaches unsuitable as these may require longer epochs to have enough time-frequency resolution. Therefore, we followed the procedure described in Stephens et al. (2010) to measure the direct coupling between the Informer and the Guesser. This method applies general linear models (Eq. 1) to the fNIRS signals to test if the Guesser's brain activity (Y_{GUESS}) is spatially and temporally coupled to the Informer's brain activity (X_{INF}):

$$Y_{GUESS} = X_{INF} \cdot \beta + \varepsilon \qquad \text{(Eq. 1)}$$

This analysis was carried out on the 23 dyads (i.e., 46 subjects) where participants were monitored using two fNIRS instruments simultaneously. More precisely, the model in Eq. 1 was applied at different time lags τ ($\tau = [-\tau_{max}, \tau_{max}]$) by temporally shifting the fNIRS time series of one channel (Ch) of one participant (i.e., the Informer) and using that to predict the time series of the corresponding channel in the other participant's brain (i.e., the Guesser). Signals are shifted both backward (from 0 to $-\tau_{max}$, Informer precedes) and forward (from 0 to τ_{max}, Guesser precedes) with respect to the start of each round ($\tau = 0$) (Stephens et al., 2010). This corresponds to:

$$Y_{GUESS}(t, Ch) = X_{INF}(t + \tau, Ch) \cdot \beta(Ch) + \varepsilon \qquad \text{(Eq. 2)}$$

where $\tau = [-\tau_{max}, \tau_{max}]$ with 1 second-steps and $Ch = 1, 2, ..., 22$. Here, we chose $\tau_{max} = 8$ seconds. β-values are then estimated minimizing the root mean squared error. Fitting the model at different time lags (Eq. 2) allows us to explore the temporal dynamics of the brain-to-brain coupling. In this way, we can identify at which time lag or delay the Informer's brain is maximally coupled to the Guesser's brain, i.e., when the Informer's brain hemodynamic activity best predicts the Guesser's brain hemodynamic activity.

Eq. 2 was applied to the fNIRS activation signals resulting from the concatenation of the activation signal portions of all the rounds as Informer (X_{INF}) and as Guesser (Y_{GUESS}) separately. For each of the 17 time shifts and each channel, we applied t-tests on the group β-values to quantify the significance of the relationship between the Informer and the Guesser's brain activities. We used the false discovery rate (FDR) at 0.05 level to account for multiple comparisons and identified the time lag at which X_{INF} best predicts Y_{GUESS}, i.e., the highest significant β-value. In addition, in order to investigate the strength of the brain-to-brain coupling without making any assumption on the distribution underlying the β-value sample, we also estimated the mean and the confidence interval using non-parametric permutations. These non-parametric models were produced by randomly Fourier-scrambling the original Y_{GUESS} and X_{INF} fNIRS signals, i.e. by Fourier-transforming the

original data, randomly shuffling its phase, and back reconstruct the time-series, and applying Eq. 2 at each scramble. This was repeated 1000 times for each τ and channel, and in order to create pseudo-data which lacks the temporal structure of the signals while preserving their frequency content. The time lag at which the β-value exceeded the 95% confidence interval (i.e., 2.5% and 97.5% percentiles) were considered significant.
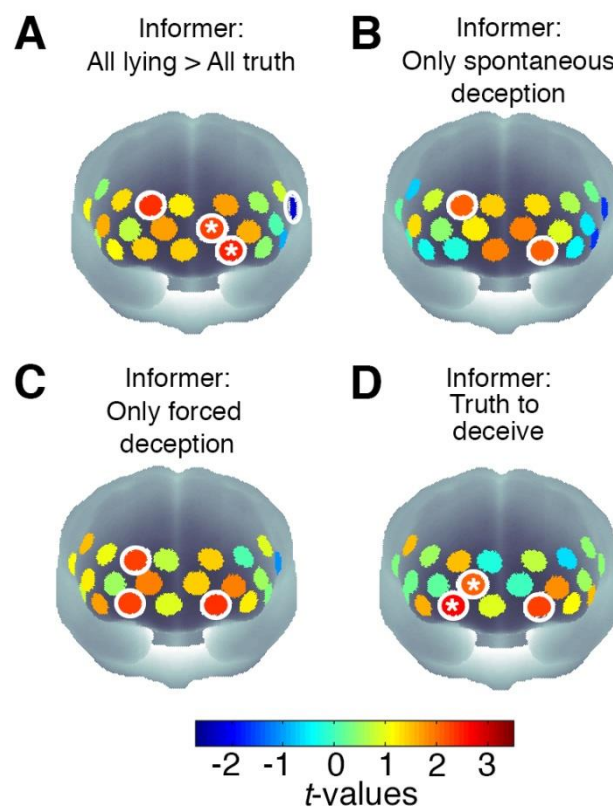
3. Results

3.1. Contrast effects results

In this section, we report the group-level results for the contrasts listed in Section 2.4.1. In particular, we report both the uncorrected and multiple comparison-corrected results according to method described in Southgate et al. (2014). In Figure 3 and 4 below, significant channels for p<0.05 at the uncorrected level are circled in white, while significant channels surviving the correction for multiple comparisons (i.e., two or more significant adjacent channels) are marked with asterisks. These results are obtained from the GLM analysis on the activation signal (i.e., the CBSI-combined HbO_2 and HbR signal). Corresponding results for HbO_2 and HbR are presented in Supplementary Figure 1 and 2.

<u>Lying contrasts</u>

Concerning the production of lies, in general we found that lying is associated with an increase in brain hemodynamic activity in anterior PFC (Channels 3, 12, 20) compared to being truthful (Figure 3 A). To further confirm this, we also observed larger changes in the activity of anterior PFC either if the act of lying is voluntary (Figure 3 B) or forced (Figure 3 C). However, in the comparisons between spontaneous deception and forced deception trials, the significant channels in anterior PFC (Channel 3 and 20 for spontaneous; Channel 3, 18, 20 for forced) did not survive multiple comparison correction.

Our experimental design allowed us also to assess whether the increased levels of activation in PFC that we observed in these contrasts are related to the intention to deceive another person or just to the production of false statements. Therefore, we have compared the conditions where the Informer made true claims with the intention to deceive the Guesser to the trials where he/she told the truth with an intention to deceive (Figure 3 D). Anterior PFC (Channel 11, 18) was significantly more engaged when the Informer told the truth in order to deceive the Guesser, suggesting that the cognitive processes supported by anterior PFC are recruited when we have the intention to deceive another person, face-to-face, even if we tell the truth.

Figure 3. *t*-values maps corresponding to the group level GLM results on the activation signal for the 4 lying contrast. Significant channels ($p < 0.05$) at the uncorrected level are circled in white, while significant channels surviving the correction for multiple comparisons (i.e., two or more adjacent channels) are marked with asterisks (Southgate et al., 2014).
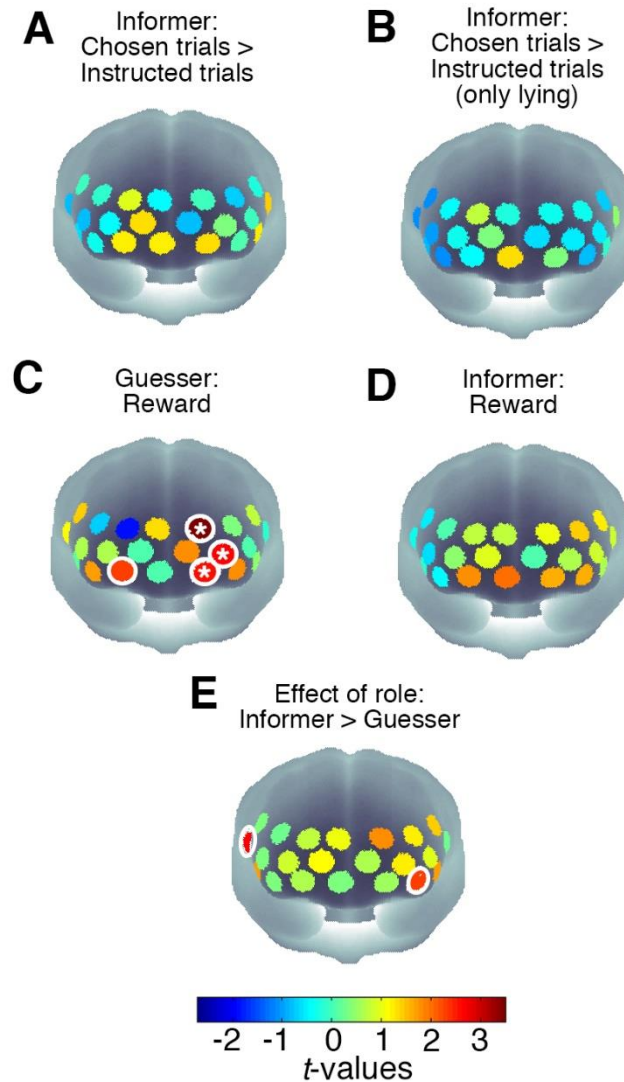
Effects of instruction

To further investigate if hemodynamic activity in PFC is modulated by the instructions (spontaneous vs forced deception), we compared conditions where the Informer produced deceptive statements spontaneously to conditions where the Informer was forced and instructed by the experimenter to either lie or tell the truth. We did not find significant differences in the activation patterns in PFC when the Informer either tells the truth and lies (Figure 4 A) for his/her choice and when he/she is forced to do so, and between spontaneous and forced lie production (Figure 4 B). This confirms the result in Figure 3 D, showing that the increase in anterior PFC (Channel 11, 18, 20) activity is mainly related to the intention to deceive another person, regardless of whether it is our choice or we are instructed to do so.

Overall task contrasts

In terms of the neural correlates of detection of deception, the group-level GLM analysis revealed that when the Guesser is successful at detecting lies (Figure 4 C), we observe an increase in anterior PFC activity in the Guesser's brain, including Channel 5, 13, 20 ($p<0.05$, corrected for multiple comparisons). This could be related to the fact that these trials are rewarded for the Guesser, but when the Informer received a reward for successful lying, we did not observe the same pattern. That is, when we compared the trials when the Informer wins the trial or loses against the Guesser (i.e., when he/she she successfully deceives or not the Guesser), we did not find a significant effect of reward (Figure 4 D). Finally, we tested whether there are any differences in the recruitment of sub-regions within PFC when the participants undertake the role of Informer or Guesser. We found that the Informer role activates the right and left lateral PFC (Channel 8, 21) more than when the
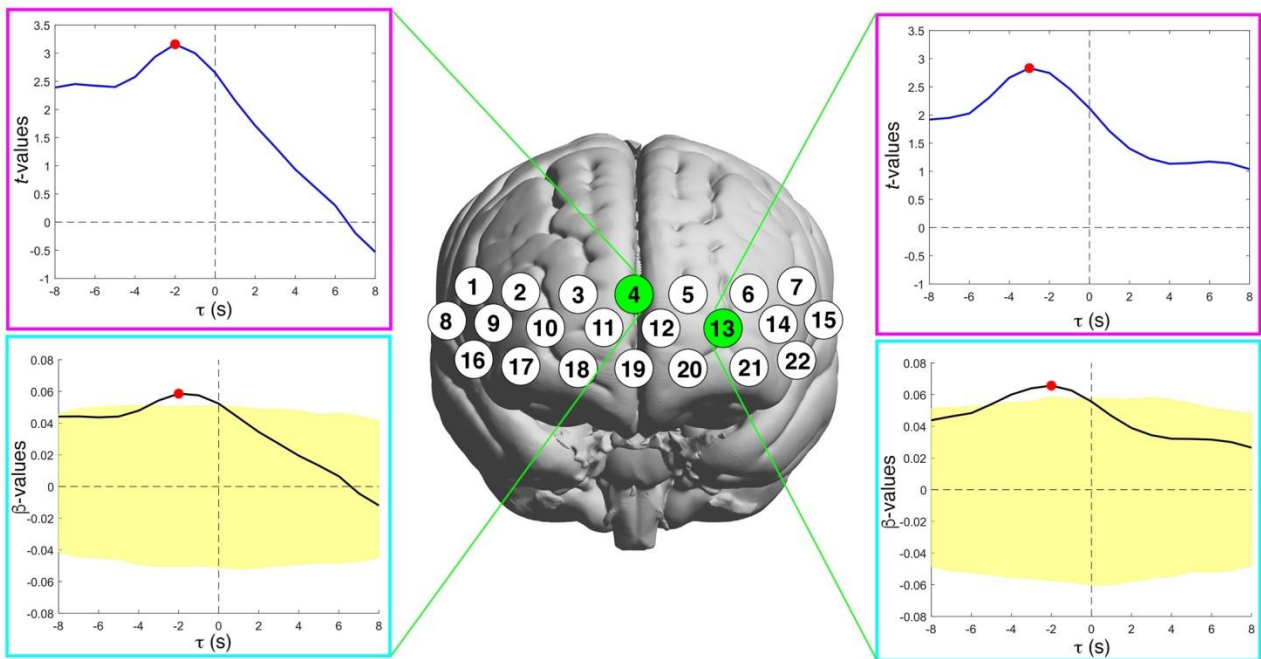
participant is playing the Guesser role (Figure 4 E). However, channels in right (Channel 8) and left

PFC (Channel 21) did not meet the multiple comparison correction criteria.

Figure 4. *t*-values maps corresponding to the group level GLM results on the activation signal for the effect of instructions (A-B) and the overall task contrasts (C-E). Significant channels (p<0.05) at the uncorrected level are circled in white, while significant channels surviving the correction for multiple comparisons (i.e., two or more adjacent channels) are marked with asterisks (Southgate et al., 2014).

3.2. Brain-to-brain analysis results

The brain-to-brain analysis described in Section 2.4.2 revealed a significant (p<0.05) brain-to-brain coupling between the Informer's brain activity and the Guesser's brain activity in anterior and left lateral rostral PFC (Channel 4 and 13; principally, BA 10), as shown in the parametric t-test results in Figure 5. In particular, this happened at $\tau = -2\ s$. However, these results did not survive the FDR correction.

Figure 5. Brain-to-brain analysis group results. The magenta boxes include the t-test results on the group β-values at different time lags (τ) that revealed a significant coupling between the Informer and the Guesser's brain in anterior and left lateral rostral PFC (Brodmann Area 10) (Channel 4 and 13; red dot; p<0.05). The blue line represents the t-values at different time lags (τ). The significant Informer-Guesser coupling was found at $\tau = -2\ s$, which however did not survive the FDR correction. The light blue boxes include the non-parametric permutation (N=1000) results. The black line represents the average β-values across the 1000 permutations and at different time lags (τ), while the yellow shaded area highlights the 95% confidence interval (2.5% and 97.5% percentiles). The significant Informer-Guesser coupling corresponds to the β-value exceeding the confidence interval.

This occurs for $\tau = -2\,s$ (red dot) in anterior and left lateral rostral PFC (Channel 4 and 13; principally, Brodmann Area 10).

Results were also confirmed by the non-parametric procedure (Figure 5), highlighting the significant coupling between the Informer and the Guesser's brains in both anterior and left lateral rostral PFC (Channel 4 and 13) for $\tau = -2\,s$, as shown by the corresponding β-values exceeding the 95% confidence interval (yellow shaded areas in Figure 5 B).

4. Discussion

The present study identified the neural circuitry associated with both the production of lies and the detection of deception. Using fNIRS to record from prefrontal cortex, we were able to study pairs of participants who took part in an interactive card game in a face-to-face setting. This has more ecological validity than some other settings, and our experimental design allowed the participants to decide when to lie or tell the truth spontaneously. Results show the engagement of anterior prefrontal cortex when lying but also when telling the truth with deceptive intent. We interpret these data in relation to previous studies of prefrontal function.

In agreement with previous studies (Kozel et al., 2005; Langleben et al., 2005; Abe et al., 2006; Abe et al., 2007; Gamer et al., 2007; Spence et al., 2008; Bhatt et al., 2009), the GLM analysis showed that activation in anterior PFC increases during the production of lies when compared to the production of true statements (Figure 3 A-C). A possible interpretation of this result is that lie production is a harder task than telling the truth because it recruits several processes supported by PFC such as inhibition, emotion processing, mentalising (Spence et al., 2004; Ganis et al., 2003; Abe et al., 2006; Abe et al., 2007; Abe et al., 2008; Christ et al., 2009; Bhatt et al., 2009; Karim et al., 2010; Sip et al., 2010; Kireev et al., 2013; Yin et al., 2016). However, interpreting the activity patterns of anterior regions of prefrontal cortex recorded with fNIRS is challenging because this brain area

24

has several possible functions, including executive and social cognition. For instance, the gateway hypothesis of rostral prefrontal cortex (area 10) proposed by Burgess and colleagues (2007) suggests that different subregions of the most rostral parts of anterior PFC support stimulus-oriented processing and stimulus-independent attending. In general, activation in the medial aspects of anterior PFC is associated with increases in stimulus-oriented attending, and lateral anterior PFC (BA 10) activation increases accompany increases in stimulus-independent thought (e.g., Henseler et al., 2011; for review see Burgess and Wu, 2013). Stimulus-oriented attending processes would be involved when attending very closely to another person's face, or looking for "tells" in their behaviour, whereas stimulus-independent processes would be used in problem-solving in open-ended situations, such as deciding for yourself whether now is the right time to lie, or how likely it is that another person might. These two anterior PFC (BA 10) systems are thought to work in concert with each other, with activation increases in the medial region often (but not always) accompanied by decreases in the other, at least for some forms of human cognition (Burgess et al., 2011). Caudal to, but anatomically adjacent to these regions is a medial subregion of anterior PFC that overlaps with the areas of medial PFC which have been linked to mentalising and social cognition (Gilbert et al., 2006). It is likely that the "attending" and "mentalising" subregions interact highly in face-to-face social situations (Gilbert et al., 2007; see also Hartwright, et al., 2014) because of the demand for both mentalising and direction of attention when watching other people and thinking about what they might do, and what one might do in response. So, both these sets of cognitive skills are highly relevant to lying. According to this framework, there are two possible interpretations of the medial anterior PFC activations in our data. One is that, when lying, the Informer might attend more towards the environment to attend to the social cues from the Guesser and the card-pair configuration, and the activity of anterior PFC reflects this stimulus-oriented problem solving. The other is that, when lying, the Informer might mentalise about the Guesser's perspective on the task and the activity of anterior PFC might reflect these social processes (Sip et al, 2008). Future studies using high-resolution fNIRS or with control conditions which factor-out stimulus-oriented attending might be useful to distinguish these

possibilities. For instance, one could parametrically vary the degree of utility of stimulus-oriented attending by restricting the degree to which each participant can see each other (e.g. by the use of screening), so that participant would not be able to observe the cues coming from the other person. At the same time one could distinguish between this activity related to stimulus-oriented attending, and those involved with mentalising and stimulus-independent attending, by temporally separating out periods when one might speculate about the other person's intentions, or make a decision about whether you were going to deceive the other person on a particular trial, and/or encouraging that particular form of cognition through instruction or question (e.g. "what do you think the other person is thinking?"). This kind of design (i.e. parametric varying of demands within an experiment to contrast regions of activation in contiguous anatomical regions involved in different forms of cognition) have already been successfully used to study the functions of anterior PFC (e.g., Benoit et al., 2012).

In the current study, our experimental design allowed us to test whether the increased activation we observed in anterior PFC is related to the intention to deceive the other person or is just due to the production of false statements. Specifically, we compared the trials where the Informer told the truth to deceive the Guesser with the trials where the Informer also told the truth but without the intention to deceive. The results show an increase in anterior PFC activity related to the deceptive intent, with higher signal when we intend to deceive the other person rather than when statements are made with no deceptive purposes (Figure 3 D). In contrast, there were no differences in the Informer's brain between conditions with a forced lie compared to a spontaneous lie. These results imply that the activation of anterior PFC seen here was linked to the intent to deceive rather than either the task constraints or the lie itself.

As our task gave participants a reward for good performance, it is also useful to consider if the patterns of brain activity could reflect processing of rewards. Specifically, Informers received a reward if they successfully deceived the Guesser (TF, FT and I-LT conditions), while the Guesser received a reward if they were not deceived (TT, FF and I-LF conditions). Several studies link

prefrontal cortex to value-based decision making and rewards (Gläscher et al., 2008), though these conditions typically engage more orbital areas of PFC (Fellows, 2011; Wallis, 2012) that are not accessible to fNIRS. In our data, there was no effect of reward in the Informer's brain. This implies that the results we report above in relation to the production of lies and of deceptive intent are not biased by the reward received on some trials.

However, we found robust activation patterns in the brain of the Guesser when s/he successfully detected a lie (rewarded trials). To our knowledge, this is the first neuroimaging study of lie detection. Across all three conditions where the Guesser gave an accurate answer, there was more activity in anterior PFC (Figure 4 C). In interpreting this data, we suggest this is not solely a reward-related activation because no comparable reward pattern was seen in Informers (Figure 4 D). However, as outlined, we cannot easily determine if this is related to executive functions such as attending toward external stimuli that might have aided the Guesser to detect lies, or to mentalising processes or social learning about the performance of the Informer. As this finding was one of the most robust, it would be interesting to follow up and define what factors in successful lie detection are most important. Finally, we observed when the participants have the role of the Informer, right and left lateral PFC are engaged more than when the participant is playing the Guesser (Figure 4 E), possibly related to the involvement of the "monitoring within working memory" processes (Barbey et al., 2013) as the Informer considers whether to produce a deceptive statement.

In addition to the GLM analysis of one brain at a time, we carried out a brain-to-brain analysis to test if it is possible to measure brain-to-brain synchrony during naturalistic social communications, as previously shown by Stephens et al. (2010) for fMRI. Our proof of principle was successful, and we found that the anterior PFC channels of the Informer are temporally coupled to the Guesser's brain activity with a delay of 2 seconds (Figure 5). Again, the significant Informer-Guesser coupling occurs in anterior PFC, which is in line with both previous neuroimaging studies but also with our GLM results showing the key involvement of anterior PFC in interpersonal deception. The coupling pattern means that, if a particular trial engages anterior PFC in the Informer, then a similar activation pattern

is found in the Guesser 2 seconds later. This brain-to-brain coupling might arise from the use of prefrontal neural circuits to encode one's own behaviour and that of the other person and to coordinate with each other within a shared social situation (Kingsbury et al., 2019). Therefore, the temporal synchrony we found might reflect the social experience that the Informer and the Guesser were sharing and the ongoing interactive behaviours between the two while playing the card game (e.g., eye-to-eye gaze, the monitoring or inhibition of deception cues like arousal responses or facial expressions and of the game-related information exchange). The structure of our trials and the turn-response sequence of behaviour is likely to mediate this pattern, because each trial had an average duration of 4.9 seconds. It is worth noting that our preliminary results did not survive the FDR correction. This could be related to the need for a larger samples size (N>23) when having naturalistic experiments as higher variability in the data might be expected. Given that our participants could interact freely with no particular experimental restraints and dyads were created with no particular criteria besides gender balancing, the brain-to-brain synchrony might be modulated by the type of interaction that was established between the two experimental partners (e.g. some dyads were more engaged and laughed together more than others), which can translate in differences in the social signals exchanged between participants. Hence, further studies will involve more detailed behavioural recordings of e.g. the eye contact and facial expressions of the participants or extensive de-briefs and questionnaires that could help us determine exactly what leads or not lead to brain-to-brain coupling in this adversarial communicative context.

However, our study has some limitations. The first and probably most important limitation is the challenge of precisely determining the onset timings of events in ecological experiments. Considering the freedom that participants had in terms of timing of responses, event timings had to be determined retrospectively through video scoring. This way of determining events' timings and durations is very time consuming and operator-dependent. Therefore, in our future studies, we aim to incorporate additional monitors of participants' behaviour, such as motion capture systems, face expression tracking, eye-tracking, to develop an automatic procedure that identifies events of interest

from participants' movements and behaviour. The use of a multimodal approach will also be crucial to investigate the neurophysiology underlying dynamic social interactions. In this work, we explored a novel analysis method (Stephens et al., 2010) to look at brain-to-brain synchrony but further studies are needed to understand which social behaviours and cognitive processes are at the basis of such synchronization. This may require a multimodal platform that senses the brain, the body, and the behaviour. For instance, fNIRS is particularly suitable for multimodal measures and can be easily interfaced with other neuroimaging modalities such as EEG, to look at brain-to-brain synchrony at different time scales; monitors of systemic physiology such as heart rate and breathing rate, to gather information on arousal and emotional responses; behavioural signals by means of e.g. eye-tracking, video cameras, motion capture, to track changes in eye-gaze patterns, facial expressions and body movements, respectively. The combination of these measures can help in understanding what element or what combination of stimuli drives the dyadic synchrony during the fast exchange of cues in live face-to-face interactions. Another limitation lies in the inability to fully recreate the complexity of a real-world situation. Our task attempted to capture the type of lying found in card games such as Poker or Cheat, which motivate participants to lie and succeed in doing so through rewards and punishments. However, this might not reflect all types of lying, such as social white lies or lies a criminal investigation. These different social encounters could lead to very different patterns of brain activity. In addition, our experimental design did not include a non-social control condition that could help in further understanding if the increased activation of anterior PFC that we found is related to the social interactivity of our task. This was due to time constraints imposed by the fNIRS cap becoming uncomfortable if worn for long periods (>30-40 minutes) and to the need to include enough trials for both participants. Hence, in our future studies we will investigate if similar patterns of activity in anterior PFC can be observed during a non-social version of our task. In order to improve the reliability of fNIRS data, it is also recommended to monitor physiological variables alongside fNIRS recordings. This can help in further denoising the fNIRS signals from systemic contamination (Tachtsidis and Scholkmann, 2016) taking into account emotional effects that might be expected in

experiments involving lie production that can elicit changes in heart rate, respiration and autonomic activity (Furedy, 1986). Here, we used the CBSI method that, by incorporating both HbO$_2$ and HbR, has the potential to reduce false positives associated to physiological interferences, but we aim to include physiological measures in our next experiments to better account for such systemic confounding factors. Other methods to minimize the impact of physiological changes such as the global mean removal (Zhang et al., 2016) and the use of short-separation channels (Gagnon et al., 2014) will be explored as well. fNIRS also limits the investigation to the more superficial layers of the cortex. However, there are other brain regions which have a key role in deception that cannot be reached by the NIR light, such as the ACC (e.g., involved in the estimation of rewards and risks) and the amygdala (e.g., involved in the emotional responses to being deceived) (Sip et al., 2008).

5. Conclusion

In this study, we explored the brain activation patterns in the prefrontal cortex during deception in live face-to-face interactions. To increase the ecological validity of the experimental setting, we used fNIRS and monitored pairs of participants playing an interactive card game face-to-face. Our results are consistent with the view that face-to-face lying requires cognitive resources additional to merely telling the truth, and is associated to higher activity in anterior PFC, in agreement with previous fMRI and fNIRS results performed in less ecological settings. Beyond past research, our task design allowed us to gather new knowledge on the neural mechanisms underlying interpersonal deception, not only by looking at the production of lies but also the detection of lies. In particular, we suggest that the naturalistic face-to-face interaction between two people is a key factor for the involvement of anterior PFC in deception, meaning that in such contexts people use their BA 10-supported (or anterior PFC) stimulus-oriented attending to a large degree in watching and monitoring the other person's behaviours whilst avoiding the leakage of deception cues. In addition, we found that the successful detection of deception is accompanied by increased activation in anterior

PFC, especially on the left. This is unlikely to be reward-related and might be linked to a release of 'goal tension' or to particular signals in the social interaction. Further experimentation using additional measures like e.g. motion tracking or eye-tracking could help to understand the source of such PFC activations.

As an initial hyperscanning fNIRS study of lying and lie detection in a naturalistic context, our results open up many directions for future research. First, it would be valuable to conduct further studies with more detailed behavioural recordings and physiological recordings to obtain better experimental control and understand the factors driving brain-to-brain coupling. Second, it would be interesting to explore the neural mechanisms of lie detection in more detail, to understand what makes a person a good lie detector (Wright et al., 2012) and what behaviour and brain systems help us detect the lies told by others.

References

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.

Abe, N., Suzuki, M., Mori, E., Itoh, M., & Fujii, T. (2007). Deceiving others: distinct neural responses of the prefrontal cortex and amygdala in simple fabrication and deception with social interactions. *Journal of Cognitive Neuroscience*, *19*(2), 287-295.

Abe, N., Okuda, J., Suzuki, M., Sasaki, H., Matsuda, T., Mori, E., … Fujii, T. (2008). Neural correlates of true memory, false memory, and deception. *Cerebral Cortex*, 18(12), 2811–2819.

Abe, N., Fujii, T., Ito, A., Ueno, A., Koseki, Y., Hashimoto, R., Hayashi, A., Mugikura, S., Shoki, T & Mori, E. (2014). The neural basis of dishonest decisions that serve to harm or help the target. *Brain and Cognition*, 90, 41-49.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. In Discovering the Social Mind (pp. 183-207). Psychology Press.

Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *Cortex*, 49(5), 1195-1205.

Benoit, R. G., Gilbert, S. J., Frith, C. D., & Burgess, P. W. (2012). Rostral prefrontal cortex and the focus of attention in prospective memory. *Cerebral Cortex*, 22(8), 1876-1886.

Bhatt, S., Mbwana, J., Adeyemo, A., Sawyer, A., Hailu, A., & Vanmeter, J. (2009). Lying about facial recognition: an fMRI study. *Brain and Cognition*, *69*(2), 382-390.

Bhutta, M. R., Hong, M. J., Kim, Y. H., & Hong, K. S. (2015). Single-trial lie detection using a combined fNIRS-polygraph system. *Frontiers in psychology*, *6*, 709.

Bunce, S. C., Devaraj, A., Izzetoglu, M., Onaral, B., & Pourrezaei, K. (2005). Detecting deception in the brain: A functional near-infrared spectroscopy study of neural correlates of intentional deception. In *Nondestructive detection and measurement for homeland security III*, 5769, 24-32. International Society for Optics and Photonics.

Burgess, P. W., Dumontheil, I., & Gilbert, S. J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in cognitive sciences*, 11(7), 290-298.

Burgess PW, Gonen-Yaacovi G, Volle E. (2011). Functional neuroimaging studies of prospective memory: What have we learnt so far? *Neuropsychologia*, 49(8): 2246-2257.

Burgess, P. W. & Wu, H. (2013). Rostral prefrontal cortex (Brodmann area 10): metacognition in the brain. In: Stuss, D. T. and Knight, R. (Eds). In *Principles of frontal lobe function*, 524-544. Oxford University Press, New York.

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: Evidence from activation likelihood estimate meta-analyses. *Cerebral Cortex*, 19(7), 1557–1566.

Cui, X., Bray, S., & Reiss, A. L. (2010). Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage*, *49*(4), 3039-3046.

Cui, X., Bryant, D. M., & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage*, 59(3), 2430-2437.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, 129(1), 74.

Ding, X. P., Gao, X., Fu, G., & Lee, K. (2013). Neural correlates of spontaneous deception: a functional near-infrared spectroscopy (fNIRS) study. *Neuropsychologia*, *51*(4), 704-712.

Ding, X. P., Sai, L., Fu, G., Liu, J., & Lee, K. (2014). Neural correlates of second-order verbal deception: A functional near-infrared spectroscopy (fNIRS) study. *Neuroimage*, *87*, 505-514.

Fellows, L. K. (2011). Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage. *Annals of the New York Academy of Sciences*, *1239*(1), 51-58.

Fett, A. K. J., Gromann, P. M., Giampietro, V., Shergill, S. S., & Krabbendam, L. (2012). Default distrust? An fMRI investigation of the neural development of trust and cooperation. *Social cognitive and affective neuroscience*, *9*(4), 395-402.

Fishburn, F. A., Murty, V. P., Hlutkowsky, C. O., MacGillivray, C. E., Bemis, L. M., Murphy, M. E., ... & Perlman, S. B. (2018). Putting our heads together: interpersonal neural synchronization as a biological mechanism for shared intentionality. Social cognitive and affective neuroscience, 13(8), 841-849.

Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. Human brain mapping, 32(12), 2207-2216.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D. & Frackowiak, R.S.J. (1994) Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2, 189-210.

Furedy, J. J. (1986). Lie detection as psychophysiological differentiation: Some fine lines. *Psychophysiology: Systems, processes, and applications*, 683-701.

Gagnon, L., Yücel, M. A., Boas, D. A., & Cooper, R. J. (2014). Further improvement in reducing superficial contamination in NIRS using double short separation measurements. *Neuroimage*, *85*, 127-135.

Gamer, M., Bauermann, T., Stoeter, P., & Vossel, G. (2007). Covariations among fMRI, skin conductance, and behavioral data during processing of concealed information. *Human brain mapping*, *28*(12), 1287-1301.

Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., & Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: an fMRI investigation. *Cerebral cortex*, *13*(8), 830-836.

Ganis, G., Morris, R. R., & Kosslyn, S. M. (2009). Neural processes underlying self-and other-related lies: an individual difference approach using fMRI. *Social Neuroscience*, *4*(6), 539-553.

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., & Burgess, P. W. (2006). Functional specialization within rostral prefrontal cortex (area 10): a meta-analysis. *Journal of cognitive neuroscience*, *18*(6), 932-948.

Gilbert, S. J., Williamson, I, Dumontheil, I,. Simons, J., Frith, C. D., & Burgess, P. W. (2007). Distinct regions of medial rostral prefrontal cortex supporting social and nonsocial functions. Social cognitive and affective neuroscience, 2.

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2008). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral cortex*, *19*(2), 483-495.

Gvirts, H. Z., & Perlmutter, R. (2020). What guides us to neurally and behaviorally align with anyone specific? A neurobiological model based on fNIRS hyperscanning studies. The Neuroscientist, 26(2), 108-116.

Hartwright, C. E., Apperly, I. A., & Hansen, P. C. (2014). Representation, control, or reasoning? Distinct functions for theory of mind within the medial prefrontal cortex. Journal of cognitive neuroscience, 26(4), 683-698.

Henseler, I., Krüger, S., Dechent, P., & Gruber, O. (2011). A gateway system in rostral PFC? Evidence from biasing attention to perceptual information and internal representations. *NeuroImage*. 56. 1666-76.

Hikosaka, O., Nakamura, K., & Nakahara, H. (2006). Basal ganglia orient eyes to reward. Journal of neurophysiology, 95(2), 567-584.

Hirsch, J., Zhang, X., Noah, J. A., & Ono, Y. (2017). Frontal temporal and parietal systems synchronize within and across brains during live eye-to-eye contact. *Neuroimage*, 157, 314-330.

Hirsch, J., Adam Noah, J., Zhang, X., Dravida, S., & Ono, Y. (2018). A cross-brain neural mechanism for human-to-human verbal communication. *Social cognitive and affective neuroscience*, *13*(9), 907-920.

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied optics*, 48(10), D280-D298.

Karim, A. A., Schneider, M., Lotze, M., Veit, R., Sauseng, P., Braun, C., & Birbaumer, N. (2010). The truth about lying: Inhibition of the anterior prefrontal cortex improves deceptive behavior. *Cerebral Cortex*, 20(1), 205–213.

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

Kingsbury, L., Huang, S., Wang, J., Gu, K., Golshani, P., Wu, Y. E., & Hong, W. (2019). Correlated neural activity and encoding of behavior across brains of socially interacting animals. *Cell*, *178*(2), 429-446.

Kireev, M., Korotkov, A., Medvedeva, N., & Medvedev, S. (2013). Possible role of an error detection mechanism in brain processing of deception: PET-fMRI study. *International Journal of Psychophysiology*, 90(3), 291–299.

Kozel, F. A., Revell, L. J., Lorberbaum, J. P., Shastri, A., Elhai, J. D., Horner, M. D., ... & George, M. S. (2004a). A pilot study of functional magnetic resonance imaging brain correlates of deception in healthy young men. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *16*(3), 295-305.

Kozel, F. A., Padgett, T. M., & George, M. S. (2004b). A replication study of the neural correlates of deception. *Behavioral neuroscience*, *118*(4), 852.

Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J., & George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biological psychiatry*, *58*(8), 605-613.

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., Mok, C., Witt, E. E., Pradhan, A. H., Chen, J. E., & Reiss, A. L. (2016). NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative Jenga game with face-to-face communication. *Frontiers in human neuroscience*, 10, 82.

Lisofsky, N., Kazzer, P., Heekeren, H.R., & Prehn, K. (2014). Investigating socio-cognitive processes in deception: A quantitative meta-analysis of neuroimaging studies. *Neuropsychologia*, 61, 113-122.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2005). Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage*, *26*(1), 251-257.

Molavi, B., & Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological measurement*, 33(2), 259.

Pinti, P., Aichelburg, C., Lind, F., Power, S., Swingler, E., Merla, A., ... & Tachtsidis, I. (2015). Using fiberless, wearable fNIRS to monitor brain activity in real-world cognitive tasks. *JoVE (Journal of Visualized Experiments)*, (106), e53336.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2018a). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*.

Pinti, P., Aichelburg, C., Gilbert, S., Hamilton, A., Hirsch, J., Burgess, P., & Tachtsidis, I. (2018b). A review on the use of wearable functional near-infrared spectroscopy in naturalistic environments. *Japanese Psychological Research*, *60*(4), 347-373.

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., & Tachtsidis, I. (2018c). Current status and issues regarding pre-processing of fNIRS neuroimaging data: An investigation of diverse signal filtering methods within a General Linear Model framework. *Frontiers in human neuroscience*, 12, 505.

Piva, M., Zhang, X., Noah, J. A., Chang, S. W., & Hirsch, J. (2017). Distributed neural activity patterns during human-to-human competition. *Frontiers in human neuroscience*, *11*, 571.

Quaresima, V., & Ferrari, M. (2019) A Mini-Review on Functional Near-Infrared Spectroscopy (fNIRS): Where Do We Stand, and Where Should We Go? *Photonics*, 6(3), 87.s

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and brain sciences*, *36*(4), 393-414.

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U., & Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage*, 85, 6-27.

Sip, K. E., Roepstorff, A., McGregor, W., & Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends in cognitive sciences*, *12*(2), 48-53.

Sip, K. E., Lynge, M., Wallentin, M., McGregor, W. B., Frith, C. D., & Roepstorff, A. (2010). The production and detection of deception in an interactive game. *Neuropsychologia*, 48(12), 3619-3626.

Southgate, V., Begus, K., Lloyd-Fox, S., di Gangi, V., & Hamilton, A. (2014). Goal representation in the infant brain. *NeuroImage*, 85, 294-301.

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 of London. Series B, Biological sciences*, *359*(1451), 1755–1762.

Spence, S. A., Kaylor-Hughes, C., Farrow, T. F., & Wilkinson, I. D. (2008). Speaking of secrets and lies: the contribution of ventrolateral prefrontal cortex to vocal deception. *Neuroimage*, *40*(3), 1411-1418.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425-14430.

Tachtsidis, I., & Scholkmann, F. (2016). False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward. *Neurophotonics*, 3(3), 031405.

Tian, F., Sharma, V., Kozel, F. A., & Liu, H. (2009). Functional near-infrared spectroscopy to investigate hemodynamic responses to deception in the prefrontal cortex. *Brain research*, *1303*, 120-130.

Wallis, J. D. (2012). Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature neuroscience*, *15*(1), 13.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. *In 5th International Conference on Language Resources and Evaluation (LREC 2006),* pp. 1556-1559.

Wright, G. R., Berry, C. J., & Bird, G. (2012). "You can't kid a kidder": association between production and detection of deception in an interactive deception task. *Frontiers in Human Neuroscience*, *6*, 87.

Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *Neuroimage*, *44*(2), 428-447.

Yin, L., Reuter, M., & Weber, B. (2016). Let the man choose what to do: Neural correlates of spontaneous lying and truth-telling. *Brain and Cognition*, 102, 13–25.

Yücel, M. A., Selb, J., Aasted, C. M., Lin, P. Y., Borsook, D., Becerra, L., et al. (2016). Mayer waves reduce the accuracy of estimated hemodynamic response functions in functional near-infrared spectroscopy. *Biomedical Optics Express*, 7, 3078–3088. doi: 10.1364/BOE.7.003078

Yu, J., Tao, Q., Zhang, R., Chetwyn C.C.H., & Tatia M.C.L. (2019). Can fMRI discriminate between deception and false memory? A meta-analytic comparison between deception and false memory studies. *Neuroscience & Biobehavioral Reviews,* 104, 43-55.

Zhang, X., Noah, J. A., & Hirsch, J. (2016). Separation of the global and local components in functional near-infrared spectroscopy signals using principal component spatial filtering. *Neurophotonics*, *3*(1), 015004.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In *Advances in experimental social psychology*,14, 1-59. Academic Press.