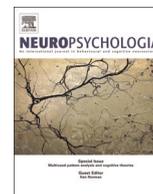




ELSEVIER

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Neural correlates of outcome processing post dishonest choice: An fMRI and ERP study



Delin Sun ^{a,b,1}, Chetwyn C.H. Chan ^{c,1}, Yang Hu ^d, Zhaoxin Wang ^{d,**}, Tatia M.C. Lee ^{a,b,e,*}

^a Laboratory of Neuropsychology, The University of Hong Kong, Hong Kong

^b Laboratory of Cognitive Affective Neuroscience, The University of Hong Kong, Hong Kong

^c Applied Cognitive Neuroscience Laboratory, Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong

^d Key Laboratory of Brain Functional Genomics, Institute of Cognitive Neuroscience, East China Normal University, Shanghai, China

^e The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong,

ARTICLE INFO

Article history:

Received 26 July 2014

Received in revised form

19 December 2014

Accepted 9 January 2015

Available online 10 January 2015

Keywords:

Dishonesty

fMRI

ERP

Outcome

Reward

Attention

ABSTRACT

A dishonest person often utilizes another person's obliviousness to appropriate the property that belongs to the other person. Previous researchers have studied the making of a dishonest choice and the manipulation of truthful information. Here, we have investigated the neural correlates of processing the outcomes of dishonest decisions. Participants in this study were asked to interact with counterparts in an economic game. They could accept the counterparts' proposals on how to divide the profits (honest choice) or choose the alternative plan that was advantageous to themselves (dishonest choice), playing to the ignorance of their counterparts who had a 50% chance of detecting the situation. Successful dishonest choices (not being detected) would bring large rewards, whereas honest choices would lead to less of a reward, and failed dishonest choices (being caught) would result in no reward. Participants' neural responses during the outcome presentations were recorded by functional magnetic resonance imaging (fMRI) and event-related potential (ERP) methods in different sessions. We found that the outcomes of successful dishonest (vs. honest) choices elicited stronger activations in the ventral striatum and posterior cingulate cortex and a smaller ERP component called feedback-related negativity (FRN), which suggests that positive outcome evaluation and attention processing were aroused by successful dishonest choices. Moreover, the outcomes of failed dishonest (relative to honest) choices were associated with different neural response patterns in the medial orbitofrontal cortex and P3b ERP component between human and computer counterparts, suggesting that processing the output of social decision making (playing human) is different from that of risk taking (playing computer). The findings advanced our understanding about the neural processing of outcome presentation after a dishonest choice has been made.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Dishonesty is a common social phenomenon. A dishonest person appropriates the property of another person by deliberately utilizing his or her ignorance. The dishonest individual may either fabricate non-factual information or take advantage of

* Corresponding author at: Rm 656, Laboratory of Neuropsychology, The University of Hong Kong, Pokfulam Road, Hong Kong. Fax: +852 2819 0978.

** Corresponding author at: Key Laboratory of Brain Functional Genomics, Ministry of Education, Shanghai Key Laboratory of Brain Functional Genomics, Institute of Cognitive Neuroscience, School of Psychology and Cognitive Science, East China Normal University, Shanghai, China. Fax: +86 021 6223 3352.

E-mail addresses: zxwang@nbic.ecnu.edu.cn (Z. Wang),

tmclee@hku.hk (T.M.C. Lee).

¹ Both authors contributed equally to this work.

another person's lack of knowledge without manipulating the truth. For instance, a salesman either exaggerates the efficiency of the product or conceals the defects of it to his or her customers in a trade show. Similarly, a government officer may embezzle public funds and run away when supervision is lacking. Given its frequency across different social contexts, dishonesty and its inner mechanisms have aroused intense interest in both public and academic researchers.

The neural correlates of dishonest behaviors have been investigated for more than a decade (for a recent review, see Abe, 2011). The majority of such studies cite the manipulation of truthful information, in another words, deception (see Vrij, 2004) as their focus (Abe et al., 2008; Langleben et al., 2002; Lee et al., 2002; Spence et al., 2001; Sun et al., 2013). Recently, researchers have begun to investigate the neural circuits that underpin a dishonest decision (Abe et al., 2014; Baumgartner et al., 2009; Greene

and Paxton, 2009; Sip et al., 2010, 2012). However, to the best of our knowledge, few researches have investigated neural processing after a dishonest action. Compared with honesty in real life, a successful dishonest action (not being caught) often brings large rewards while failed dishonest responses (being caught and punished) bring large losses. In line with this thought, the neural correlates of outcome processing may reflect whether a dishonest action has been made, even after the actual action. Our current findings could provide important insights to future investigations on using neural patterns to identify honest/dishonest behaviors.

In this study, we aimed to delineate neural signals during the outcome presentation after a dishonest choice had been made. Participants in this study were asked to interact with counterparts in an economic game. They could behave honestly by accepting the counterpart's offer on how to divide a reward, or they could make a dishonest choice that was advantageous to them but harmed their counterparts. The probability of the situation (both honesty and dishonesty) being detected was 50%. A successful dishonest choice (not being detected) would lead to a large reward, the honest choice would bring less of a reward, and a failed dishonest action (being caught) would result in no reward. This task paradigm mimics dishonest behaviors in real life, such as occupying another person's property, taking advantage of the victim's ignorance. Furthermore, it models the risk of being caught and punished for dishonesty. Our previous behavioral studies showed that this task could robustly evoke dishonest decisions (Zhang et al., 2012). Here, we were particularly interested in the neural processing for the outcome of dishonest choices, which could consist of two processes, including the processing of rewards associated with the decision" (i.e., whether the monetary goal of the dishonest choice is achieved) and "success of the action itself" (i.e., whether the dishonest action is detected), and we did not aim to differentiate them in this study. Therefore, a "successful dishonest action" denoted that a dishonest choice was not caught, which led to large reward. In contrast, a "failed dishonest action" meant that a dishonest choice was detected, which resulted in no reward. We did not ask participants to manipulate truth in this task because we were interested in the neural mechanism of processing the outcome of a dishonest decision and hence did not want it to be confused by that of fabricating information. Future studies on telling lies with bad intentions could employ more complex task paradigms based on our current findings. In the present study, methods of high spatial resolution (fMRI) and of high temporal resolution (ERP) were employed (in different sessions) to record participants' neural responses during outcome presentation.

Previous fMRI studies have shown that monetary rewards elicit stronger blood oxygen level dependent (BOLD) activations than do no rewards or negative consequences in the ventral striatum (vStr) (Breiter et al., 2001; Coricelli et al., 2005; Xue et al., 2011), medial orbitofrontal cortex (mOFC) (Knutson et al., 2003; Liu et al., 2007; Ursu and Carter, 2005), and posterior cingulate cortex (PCC) (Fujiwara et al., 2009; Izuma et al., 2008; Nieuwenhuis et al., 2005c). A recent meta-analysis by Liu et al. (2011) investigated 142 neuroimaging studies and showed that vStr, mOFC, and PCC all respond more strongly to positive than negative outcomes. The vStr and mOFC are two major projection sites of the midbrain dopamine system, which is widely hypothesized to serve as a common-reward metric for evaluating rewards delivered in different modalities (for a recent review, see Haber and Knutson, 2010). On the other hand, the PCC activity was shown to increase during attentional biasing to targets with high motivational value, i.e., food vs. tools (Mohanty et al., 2008). In an electrophysiological study on monkeys while they shifted their gaze to visual targets for liquid rewards, McCoy et al. (2003) found that the PCC neurons responded following saccades, as well as following reward delivery,

and these responses were related to the magnitude of the reward. These findings suggest that the brain responses in PCC encode the neural processing of attention to stimulus with reward. In line with these previous findings, considering that successful/failed dishonest action (relative to honest action) in this study means larger/smaller material reward, the stronger/weaker brain activations in vStr, mOFC, and PCC may be utilized to reflect more positive/negative outcome evaluation and enhanced/reduced attention allocation elicited by successful/failed dishonest action than by an honest choice. We thus hypothesized to find stronger BOLD activations in vStr, mOFC, and PCC for dishonest (relative to honest) choices when undetected than detected.

The neuroimaging method locates the regions of interest with high spatial resolution (a few millimeters) in the brain, whereas the neurophysiological technique provides neural information with high temporal resolution (about a millisecond). Previous studies using event-related potential (ERP) found that the presentation of monetary loss (relative to gain) was followed by negative-going amplitudes at medial frontal scalp sites about 250 ms post-feedback. The waveform is termed feedback related negativity (FRN) and is supposed to be associated with quick neural responses to the valence of outcome at the early stage of information processing (Gehring and Willoughby, 2002; Hewig et al., 2007; Osinsky et al., 2013). A previous study by Holroyd et al. (2004) showed that the FRN to an outcome depended on the relative value of the outcome in possible outcomes. They found that the worst and medium outcomes were related to larger FRN than the best outcomes, regardless of whether the possible outcomes were all no-positive outputs (i.e., losing 0, 2.5, or 5 cents), no-negative outputs (i.e., gaining 0, 2.5, or 5 cents) or outputs with both positive and negative values (i.e., getting 0 cents, winning 10 cents, or losing 10 cents). Therefore, the FRN may be used to represent the relative value of the outcome among possible outcomes. We therefore hypothesized that smaller FRN would be found when dishonest (relative to honest) choices were not detected compared to that of being detected.

Another ERP component, P3, was also found to reflect the information of outcome processing. The P3 is a positive potential with peak latency after 300 ms post stimulus presentation and was found to be sensitive to the motivational significance of the eliciting stimulus (Nieuwenhuis et al., 2005a). It consists of two sub-components, i.e., P3a and P3b. The P3a peaks 60–80 ms earlier than the P3b. Further, the P3a exhibits a frontocentral distribution while the P3b is prominent at parietocentral sites. The P3a is scarcely found to be associated with outcome processing. In contrast, Yeung and Sanfey (2004) found that the P3b is sensitive to the magnitude of outcome regardless of whether it concerns a gain or a loss of money. In line with this finding, the outcome of large (successful dishonest action), less (honest choice), and no reward (failed dishonest behavior) should be represented by large, medium, and small P3b amplitudes, respectively. We thus also hypothesized that larger P3b would be found when dishonest choices (relative to honest) were not detected compared to the case where they are detected.

2. Material and methods

2.1. Participants

Twenty-five female undergraduate/postgraduate Chinese students (20–25 years old) in the East China Normal University, China, were recruited for both fMRI and ERP sessions. Only females were recruited for this study in order to control for the sex-related effect on decision making (Lee et al., 2009; Zhang et al., 2012). They were all right-handed (Oldfield, 1971), and they had no metal

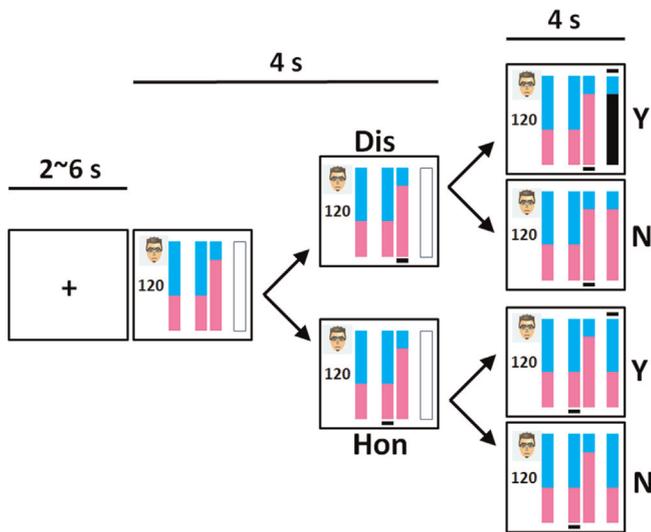


Fig. 1. Task paradigm. In this example, the counterpart is a human being (represented by an icon of a human face). The cyan and purple areas in the vertical stacked bars represented the proportions of reward assigned to the counterpart and the participant, respectively. The total amount to be divided and the proposal bar by the counterpart were shown on the left-hand side. The participant could choose between the two bars in the middle of the screen. One bar was consistent with the proposal by the counterpart (i.e., Hon bar); the other bar (i.e., Dis bar) indicated a plan advantageous to the participant herself. The result was represented by the outcome bar on the right-hand side. Each trial was preceded by an interval randomly varying between 2 s and 6 s and consisted of two 4-s phases. During the first phase, the participant had to make a choice by pressing one of the two buttons with the right index or middle finger. Once the response was made, a black line appeared under the chosen bar. During the second phase, a black line appeared above the outcome bar if the real situation was detected. On the contrary, no line was shown if the detection did not occur. When a Dis action was detected, the participant gained nothing in this trial and her area in the outcome bar became black. Under the other conditions, the participant kept the amount (as well as the area) for herself. If there was no response or the response exceeded the first 4-s phase, all of the reward would be sent to the counterpart. Dis—dishonest, Hon—honest, Y—being detected, and N—not being detected.

or medical device implants or histories of neurological or cardiovascular disorders. Participants all gave informed consent for a protocol approved by the local ethics committee. Among these 25 participants, five of them were excluded from fMRI analysis because of data recording errors. Seven participants were excluded from ERP analysis because of large noises and/or small (< 12) trial numbers under some task condition. Fifteen participants were included in both fMRI and ERP analyses.

2.2. Experimental task

The participant in this task was instructed to play as a trustee who received a monetary investment from a counterpart (investor), repaid a proportion of the increased investment in each trial, and held the rest (Fig. 1). The participant also received a proposal from the counterpart on how to divide the reward. She could choose to repay as much as (honest, Hon) or less than (dishonest, Dis) the proportions proposed. The participant was instructed beforehand that, after the choice was made, the possibility of the real situation being detected (Y) vs. not being detected (N) was 50%. The participant won nothing in the trial if a Dis choice was caught but gained the amount kept for herself in the other conditions. The participant was asked to treat each trial as a single-shot interaction since no players (investors and trustees) could recognize whether the counterparts had met in previous trials. The participant was told that half of the counterparts were human beings (Hm), while the others were computer programs (PC). Either a photo of human face representing a human

counterpart (there were a total of 2 males and 2 females, of front view with neutral expression, randomly shown) or a laptop denoting a computer counterpart was presented during each trial. The participant was told that the photo was only used to show the type of counterpart (Hm or PC) but not other information. She did not know that, in fact, computer programs mimicked all of the responses of her counterparts. The participant was asked to make her own decision in the task and would ultimately receive a real monetary bonus proportional to the amount earned during the experimental session. Because of the ethical regulations, each participant was only given 250 Chinese Yuan as compensation. She was told that the bonus gained in the task was included in this amount.

In each trial, the amount of the increased investment was shown on the screen for 4 s during which the participant had to make her choice by pressing one of two buttons. One of the buttons signified the Dis choice and the other signified the Hon choice (the order of buttons was counterbalanced across trials). In the following 4 s, the participant was informed of whether the real situation was detected and how much she gained in this trial. The proportions of reward assigned to the counterpart and the participant were represented by the cyan and purple areas in the vertical stacked bars, respectively. The total amount to be divided and the proposal bar by the counterpart were shown on the left-hand side. The participant could choose between the two bars in the middle of the screen. One of them was consistent with the proposal by the counterpart (i.e. Hon bar); the other bar (i.e. Dis bar) indicated a plan advantageous to the participant herself. The result was represented by the outcome bar on the right-hand side. Each trial was preceded by an inter-trial interval (ITI) randomly varying between 2 s and 6 s and consisted of two 4-s phases. During the first phase, the participant had to make a choice by pressing one of the two buttons with the right index or middle finger (the order of buttons was counterbalanced across trials). Once the response was made, a black line appeared under the chosen bar. During the second phase, a black line appeared above the outcome bar if the real situation was detected. On the contrary, no line was shown if the detection did not occur. When a Dis action was detected, the participant gained nothing in this trial and her area in the outcome bar became black. Under the other conditions, the participant kept the amount (as well as the area) for herself. If there was no response, or should the response exceed the first 4-s phase, all reward would be sent to the counterpart. The subjective utilities (reward \times possibility) were the same for the two choices. The permutation of the total amount, i.e. the amount to be divided (a number randomly generated between 80 and 150), the proposed portion of repayment to the counterpart (60%, 65%, 70%), and the location (left or right) of the Dis/Hon bars were randomized for each participant.

2.3. Procedure

Participants were required to participate in an fMRI session and an ERP session in two separate time slots within two weeks. The order of the two sessions was counterbalanced across participants. Prior to data collection, participants practiced about 10 trials using a set of symbols different from those used during data collection. In the formal task, there were 5 runs of 24 trials in the fMRI session and 12 runs of 24 trials in the ERP session. More trials were included in the ERP session to compensate for the relatively low SNR in ERP data recording. Participants were debriefed after completing the two sessions and were paid.

2.4. Data acquisition

2.4.1. fMRI data

Scans were conducted on a 3-Tesla Siemens Trio MR scanner. Thirty-five axial slices covering the whole brain were obtained using a T2*-weighted echo planar imaging (EPI) sequence (TR=2000 ms, TE=30 ms, flip angle=90°, matrix=64 × 64, Field of View (FOV)=240 × 240 mm², slice thickness=4 mm without gap) for functional images. The axial slices were adjusted to be parallel to the AC-PC plane. A high-resolution structural image for each participant was also acquired using 3D MRI sequences (TR=1900 ms, TE=3.43 ms, flip angle=7°, matrix=256 × 256, FOV=210 × 210 mm², slice thickness=1 mm). The visual stimuli presentations and response collections were performed through the integrated functional imaging system (IFIS).

2.4.2. ERP data

Scalp electrical potentials were recorded through an elastic electroencephalogram (EEG) cap (Brain Products Company, Germany) embedded with 64 tin scalp electrodes according to the extended international 10–20 system. All channel impedances were kept below 10 kΩ. The EEG signals were amplified using a 0.05–100 Hz band-pass filter and continuously sampled at 500 Hz. Vertical and horizontal eye movements were recorded by two electrodes at the temporal and lower sides of the left eye, respectively. The visual stimuli presentations and response collections were performed through E-Prime software (Psychology Software Tools, Inc.).

2.5. Statistical analysis

2.5.1. Behavioral data

The trials without responses were less than 5% of the total trials and were thus excluded from analyses. The behavioral performance consisted of the frequency of Dis choice and the respective reaction time before the actualization of Dis and Hon choices. The frequency of Dis choice was calculated through dividing the number of trials of Dis choice by the number of total trials having responses (i.e., Dis+Hon). Both the frequency of choice and the reaction time were calculated for Hm and PC counterparts, respectively. The behavioral performances in both sessions were collected from all participants, although some of them were excluded in later fMRI or ERP analyses. To test whether participants' responses were consistent between the two sessions, we conducted the Pearson correlation analyses on the behavioral performances between sessions using SPSS version 20.0 software (IBM Corp.).

2.5.2. fMRI data

Neuroimaging data were analyzed using the SPM8 software (Wellcome Department of Imaging Neuroscience, UK). The functional scans of each participant were spatially realigned to adjust for head movement and corrected for slice-acquisition timing. Anatomical images were then co-registered to the mean functional image and were segmented into gray/white matter according to an anatomical template of Eastern Asian brains. After that, the functional images were normalized to the Montreal Neurological Institute (MNI) brain template and smoothed with an 8-mm full-width half-maximum (FWHM) Gaussian filter.

In this study, we were only interested in the processing of outcome. The findings (both fMRI and ERP) during decision making (the first 4 s phase involving pressing the button) are reported elsewhere. The general linear model (GLM) was used to examine the experimental effects across task events. Eight regressors were employed to model the onset of outcome presentation. They were combinations of choice (Dis and Hon), counterpart (Hm and PC), and detection (Y and N). A regressor modeling the onset of button

pressing and six extra regressors modeling residual head motions were also included. These regressors were convolved with the SPM canonical hemodynamic response function. High-pass temporal filtering with a cut-off of 128 s was recruited to remove low-frequency drifts. Beta-weight images of regressors were employed to form within-subject contrasts between Dis and Hon choices under each condition, i.e., Hm&Y, Hm&N, PC&Y, and PC&N.

The contrasting images from all participants were then group-level analyzed through a two-way ANOVA model with 2 within-subject factors, i.e., Counterpart and Detection. Results of whole brain analyzes were voxel-level height thresholded at $p < 0.001$ and survived cluster- or peak-level Family Wise Error (FWE) correction ($p < 0.025$). The xjView toolbox (<http://www.alivelearn.net/xjview>) was used for the anatomical definition. The MarsBaR toolbox (<http://marsbar.sourceforge.net/>) was used to extract from each participant the % signal change of each event (there were totally 8 events, i.e. 2 [Choice: Dis vs. Hon] by 2 [counterpart: Hm vs. PC] by 2 [Detection: Y vs. N] combinations). We then calculated the % signal change of Dis–Hon in 4 conditions (i.e. Hm&Y, Hm&N, PC&Y, and PC&N). For example, the % signal change of the event Hm&Dis&N was contrasted with that of the event Hm&Hon&N to form the % signal change of Dis–Hon in the condition of Hm&N.

2.5.3. ERP data

The scalp electrical potentials were preprocessed by the SPM8 software. The continuous recordings were filtered (0.1–30 Hz), corrected for eye-movement (Berg and Scherg, 1994), cut into epochs (–200–2300 ms after outcome onset), and corrected for baseline signals (–200–0 ms). Epochs containing amplitudes exceeding $\pm 100 \mu\text{V}$ were removed. Epochs were averaged (robust averaging) (Wager et al., 2005) in each channel for each participant. An additional low-pass filter (< 30 Hz) was employed to remove the high-frequency noises elicited by the robust averaging. The ERP data were then re-referenced to a computed average of the whole-scalp EEG channels.

Statistical analysis on the mean amplitudes of ERP data was conducted using SPSS software. Consistent with the published guidelines for studies using electroencephalography (Keil et al., 2014), the representative channels and time windows of the three ERP components of interest that we selected were based on previous research cited in Introduction section, and on visual inspection of our own data: (I) FRN, channel Fz, 250–280 ms; (II) P3a, channel Fz, 370–420 ms; and (III) P3b, channel Pz, 450–490 ms. Mean amplitudes of the contrast Dis–Hon under each condition were extracted from the representative channel within the time window for each component and entered into a two-way ANOVA model with 2 within-subject factors, i.e., Counterpart and Detection.

3. Results

3.1. Behavioral findings

Results showed that, regardless of the type of the counterpart, the frequencies of Dis choice across the sessions were positively correlated, and the reaction times of the difference between Dis and Hon choices between sessions were positively correlated (for correlation coefficients and statistical significance, please see Fig. 2). These results suggested that participants behaved consistently in the fMRI session and the ERP session.

3.2. fMRI findings

An average effect of condition comparing the Dis and Hon choices across conditions showed that Dis (vs. Hon) choices

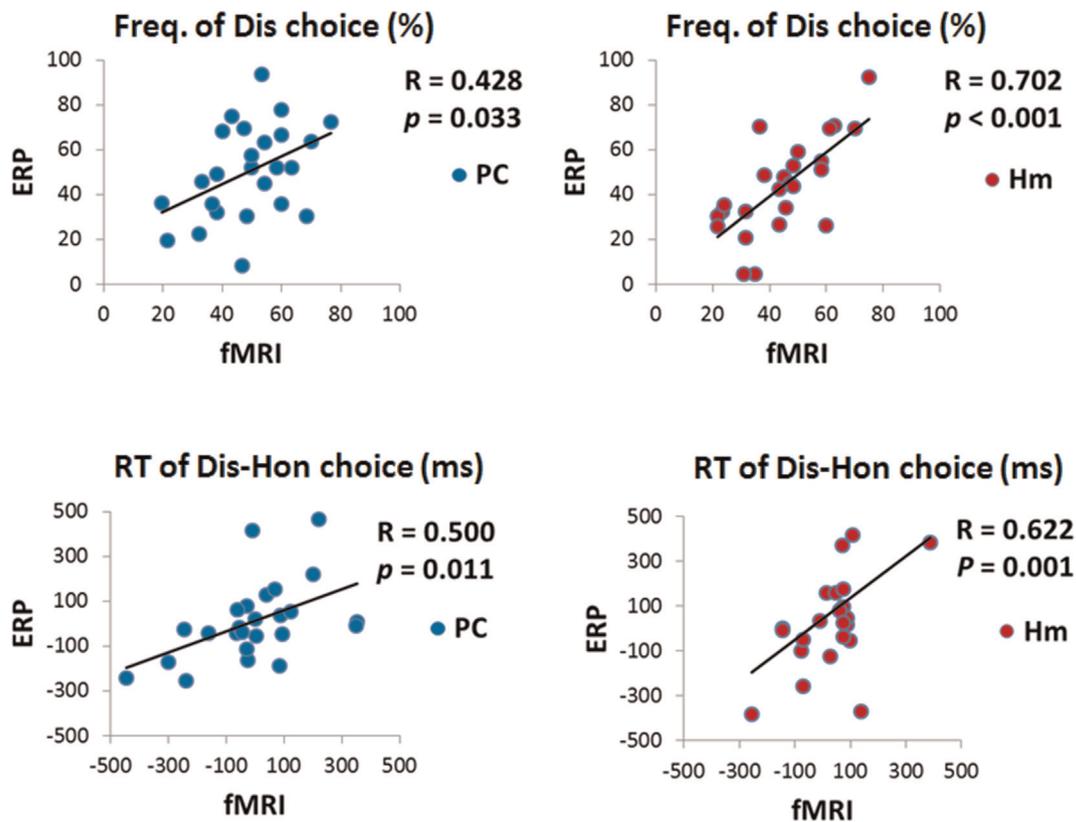


Fig. 2. Behavioral performance is highly correlated between the fMRI and the ERP session. The Pearson correlation coefficient and statistical significance were shown. Dishonest and Hon–honest.

elicited stronger activations (positive effect) in bilateral inferior frontal gyrus/insula, medial frontal gyrus (including anterior cingulate gyrus and thalamus) and middle cingulate gyrus and the right middle frontal gyrus, whereas weaker activations (negative effect) in right supramarginal gyrus and superior temporal gyrus. The main effect of Detection (Table 1 and Fig. 3) showed that dishonesty, relative to honest choices (i.e., the contrast Dis–Hon), elicited stronger BOLD activations in bilateral vStr, mOFC, and PCC when not being detected (N) than being detected (Y). Whole brain analysis did not show significant results for the main effect of Counterpart or the interaction between Counterpart and Detection. However, the non-significance may be attributed to the too strict statistical threshold (i.e. FWE correction within the whole brain). Therefore, to investigate whether the brain activations during Dis choice are different from those for the Hon choice under different conditions, we extracted the % signal changes of the contrast of Dis–Hon from the above significant clusters (spheres with radius=8 mm, the centers located at the MNI coordinates of peak significance listed in Table 1) for each of the four conditions (Hm&Y, Hm&N, PC&Y, and PC&N) and compared them with 0 (using one sample *t* test). Results (Table 3 and Fig. 3) showed that the N condition elicited stronger activations in bilateral vStr and PCC, regardless of the counterpart type, whereas no significant results were found for the Y condition in these areas. On the other hand, the Y condition elicited weaker activations in mOFC for the Hm but not the PC counterpart whereas there was no significant result for the N condition in mOFC.

3.3. ERP findings

The significant main effect of Detection (Table 2 and Fig. 4) showed that the contrasting Dis–Hon elicited smaller FRN and P3a when not being detected (N) than being detected (Y). The

significant main effect of Counterpart showed that the contrast Dis–Hon elicited smaller FRN and P3a when the counterpart was Hm rather than PC. A significant interaction effect between Counterpart and Detection was detected in P3b. Post hoc analysis with Bonferroni correction showed that the contrast Dis–Hon elicited smaller P3b in condition N than Y only when the counterpart was PC ($t(17)=2.813$, $p=0.024$) but not Hm ($t < 1$).

To investigate whether the ERP responses during the Dis choice are different from those for the Hon choice under different conditions, we extracted the amplitudes of the contrast Dis–Hon from the interested ERP components for each of the four conditions (Hm&Y, Hm&N, PC&Y and PC&N) and compared them with 0 (using one sample *t* test). Results (Table 3 and Fig. 4) showed that the N condition was associated with smaller FRN (more positive-going amplitudes) regardless of the counterpart type whereas there was no significant FRN result for the Y condition. On the contrary, the Y condition related to larger P3a (more positive-going amplitudes) regardless of the counterpart type whereas there was no significant P3a result for the N condition. Analysis on the P3b showed that the Y condition was associated with larger P3b (more positive-going amplitudes) for PC but not Hm counterpart whereas the N condition was related with larger P3b for both Hm and PC (a trend of significance) counterparts. The average effect of condition for the contrast Dis–Hon in fMRI data and ERP waveforms of all conditions in channel Fz and Pz could be seen in Fig. 5.

4. Discussion

A dishonest choice made in real life is often to gain a larger reward than an honest action. However, the violation of social norms or laws may cause punishment and brings dishonesty a

Table 1
fMRI results of the 2 (Counterpart, Hm vs. PC) by 2 (Detection, Y vs. N) ANOVA analysis on the contrast of outcomes between dishonest and honest choices.

Brain Area	Cluster	MNI coordinates			Z
		x	y	z	
Average effect of condition					
<i>Positive effect</i>					
R IFG/INS (BA13/47)	625	39	23	7	6.663
L/R Med FG (BA6/32)	2243	9	35	19	6.245
L IFG/INS (BA13/47)	445	−33	23	−5	5.335
R Mid FG (BA6)	112	45	−1	46	5.062
L/R MCC (BA31)	176	9	−28	43	4.637
<i>Negative effect</i>					
R Supr (BA39)	128	45	−55	31	4.343
R STG (BA22)	137	63	−4	22	4.195
Main effect of Counterpart					
NS					
Main effect of Detection					
Y > N					
NS					
Y < N					
R vStr	102	15	2	−8	5.280
L vStr	169	−12	5	−8	4.999
L/R PCC (BA23)	164	0	−22	25	4.603
L/R mOFC (BA10)	240	−3	53	1	3.913
Interaction between Counterpart and Detection					
NS					

Note: Sample size=20. All results were height thresholded at $p < 0.001$ and survived cluster- or peak-level FWE correction ($p < 0.025$). BA—Brodmann's area; Cluster—number of voxels within the cluster; Z—Z value; L—Left; R—right; NS—no significance; Hm—human counterpart, PC—computer counterpart, Y—being detected; N—not being detected; MCC—middle cingulate cortex; Med FG—medial frontal gyrus; Mid FG—middle frontal gyrus; IFG—inferior frontal gyrus; INS—inula; mOFC—medial orbitofrontal cortex; PCC—posterior cingulate cortex; STG—superior temporal gyrus; Supr—Supramarginal Gyrus; and vStr—ventral striatum.

smaller reward (or even more loss) than honesty. Therefore, dishonesty in reality often shows this pattern of outcome: a successful dishonest action (not being caught) > an honest decision > a failed dishonest response (being caught and punished). In this study, we mimicked this pattern in an economic game and investigated the neural signals during the resultant presentation post a dishonest (or an honest) choice. Participants in this study could increase self-benefits through reducing the repayment to the counterparts. By utilizing the counterparts' ignorance (i.e., no detection), they had chances to gain greater rewards relative to the outcomes of honest actions. However, they received nothing (smaller rewards than those of honesty) if the dishonesty was caught. Both fMRI and ERP methods were employed to record neural signals in the same group of participants (in different sessions). These findings suggested that the neural signals could be used to uncover the dishonest action, even after it has been made. Importantly, the dishonest action that is behaviorally undetectable (i.e., successful dishonest behavior) may still be uncovered through the neural responses to the outcome of such an action. The findings advanced our understanding of neurocognitive functions in dishonesty.

4.1. Neural circuits of processing the outcomes of dishonest choices

As consistent with our first *a priori* hypothesis, stronger BOLD activations in vStr, mOFC, and PCC were found during dishonest (relative to honest) choices when not being detected than being

detected. The vStr and mOFC have been widely accepted as key brain regions for outcome evaluation (Haber and Knutson, 2010). Stronger activations in these two brain areas were consistently found to be related with relatively positive outcomes (Knutson et al., 2003; Liu et al., 2007; Ursu and Carter, 2005; Xue et al., 2009, 2011). On the other hand, the PCC has been found to play a crucial role in regulating attention (Hahn et al., 2007; Hampson et al., 2006), and its increased activations have been detected for those being rewarded compared to those without (Mohanty et al., 2008). In line with these previous findings, the stronger activations in the three brain areas may reflect that, relative to the outcomes of honest choices (medium reward whether being detected or not), successful dishonest behavior (large reward) compared to failed dishonest action (no reward) elicited more positive reward evaluation and enhanced attention processing.

Further analyses showed that, relative to the outcome of the honest choice, the consequence of the dishonest choice elicited stronger activations in bilateral vStr and PCC only when not detected but not when caught. In other words, a large reward elicited stronger activations than did the medium-amount reward, whereas there was no difference detected across a medium-amount and no reward. A previous study by (Nieuwenhuis et al., 2005b) found that part of the vStr (i.e., caudate) and the PCC exhibited comparable BOLD activations for intermediate outcomes and worst outcomes, whereas the best outcomes showed the largest BOLD responses. Their findings suggested that these brain areas classify the outcomes into two categories: the best outcomes are classified as good and all other outcomes are classified as bad. The similarity between our results and Nieuwenhuis et al.'s findings support the validity of our study. More importantly, our findings suggest that the brain activations in bilateral vStr and PCC reflect dishonest actions only when the participants perceived that they had escaped from detection and punishment. Our finding has the potential to be utilized in real life to detect dishonest behaviors, especially when traditional methods (such as behavioral performance test) seem to fail. This in turn evokes confidence in the dishonest person regarding his/her success. However, caution should be taken when applying the findings of this study to reality. Participants in this study gained material rewards, whereas more complex rewards, including monetary gain, the joy of escaping from punishment, and the sense of achievement in fooling an intellectual agent, serve as the incentives for dishonest people in real life. These different processes are interesting but could not be discriminated in this study and should be investigated in future research.

It is interesting that the mOFC showed different activation patterns compared with the vStr and PCC. The differences between dishonest and honest outcomes were found only when the real situation was detected by the human but not computer counterpart. McCabe et al. (2001) in an earlier study found that cooperators in a Trust Game exhibited stronger activation in mOFC when playing a human rather than computer counterpart, whereas there was no significant difference in non-cooperators. The mOFC may attribute extra rewards to the cooperative (vs. non-cooperative) behaviors when playing human counterparts. It is possible that, when being detected during the final presentation, compared with playing a computer honestly, playing a human honestly was regarded as more rewarding and thus elicited stronger activations in mOFC. That explains why we found a significant negative-going activation pattern when dishonest choices were made compared to honest choices when the counterpart was a human being. More importantly, the activations in mOFC may reflect the different response strategies to human and to computer counterparts, respectively, and may thus be utilized to differentiate dishonesty (i.e., risky+antisocial decision) from simply taking risk (i.e., risky decision). Methodologies may be used to

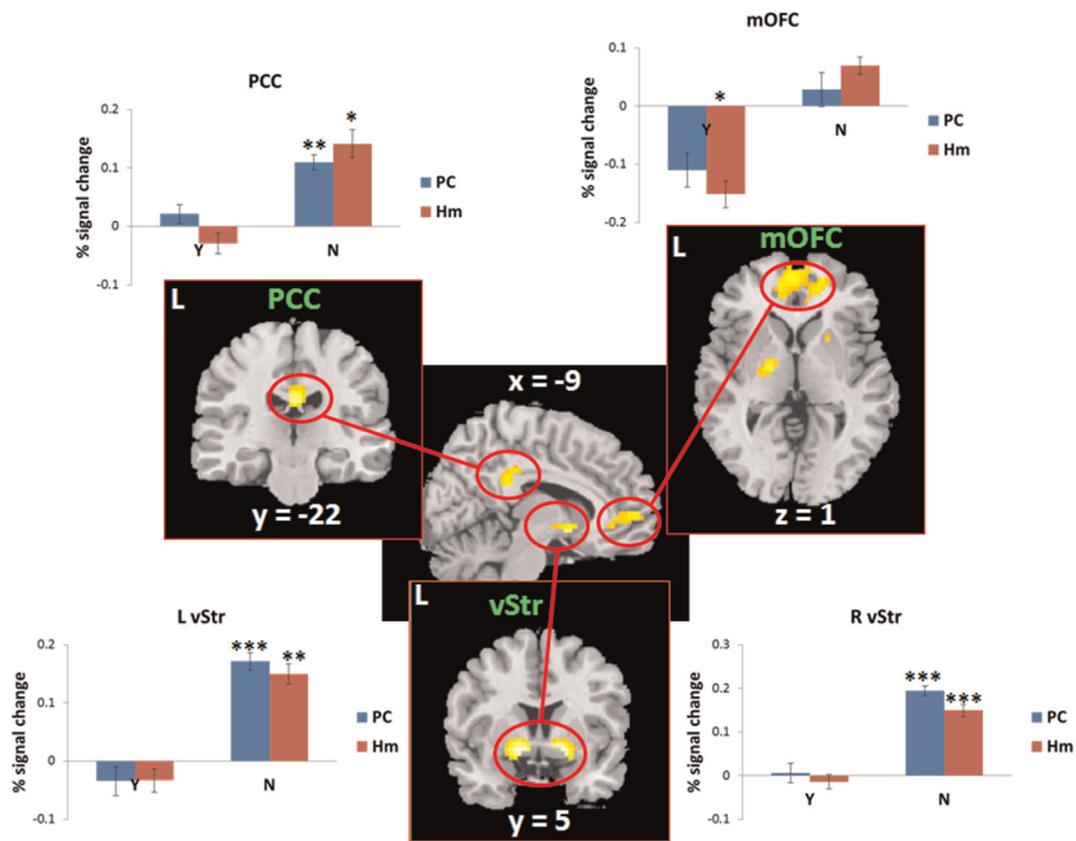


Fig. 3. fMRI findings. The outcomes of dishonest (vs. honest) choices elicited stronger BOLD activations in bilateral mOFC (medial orbitofrontal cortex), PCC (posterior cingulate cortex), and vStr in the N condition (not being detected) than in the Y condition (being detected). This effect was similar between PC (computer) and Hm (human) counterparts. The SPM T map is voxel-level thresholded $p < 0.001$ with cluster size > 100 voxels for observation purpose only. The y-axis indicated the % signal change of the contrast Dis–Hon. L–left; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. Error bar denotes sem.

Table 2
ERP results of the 2 (Counterpart, Hm vs. PC) by 2 (Detection, Y vs. N) ANOVA analysis on the contrast of outcomes between dishonest and honest choices.

	FRN		P3a		P3b	
	mean	sem	mean	sem	mean	sem
Mean amplitudes (μV)						
PC&Y	0.815	0.609	3.204	0.785	3.472	0.753
PC&N	2.960	0.679	1.133	0.671	1.452	0.527
Hm&Y	0.702	0.467	1.996	0.615	2.066	0.851
Hm&N	1.481	0.446	0.265	0.381	1.417	0.369
Statistical values						
	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Counterpart	5.245	0.035	5.807	0.028	3.385	0.083
Detection	7.378	0.015	10.732	0.004	4.258	0.055
Counterpart \times Detection	2.649	0.122	0.211	0.652	4.727	0.044

Note: Sample size=18; sem—standard error of the mean.

specifically reduce dishonest actions but not risk-taking behaviors in future studies. For example, oxytocin has been found to be able to raise the level of trust, but not risk-taking behaviors (Kosfeld et al., 2005).

We found that Dis (vs. Hon) choices were related with stronger BOLD activations in bilateral inferior frontal gyrus/insula and medial frontal gyrus, including the anterior cingulate cortex (ACC). A plenty of previous studies showed joint fMRI activations of these

Table 3
One sample *t*-tests (two-tailed) of the neural responses (% signal changes for fMRI data and mean amplitudes for ERP data) for the contrast of outcomes between dishonest and honest choices in different ROIs/ERP components under different conditions.

	R vStr		L vStr		PCC		mOFC	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
fMRI (n=20)								
PC&Y	0.138	3.568	-0.687	2.001	0.641	2.117	-1.902	0.290
PC&N	8.119	0.000	5.922	0.000	4.254	0.002	0.489	2.522
Hm&Y	-0.428	2.694	-0.825	1.678	-0.836	1.654	-3.252	0.017
Hm&N	5.301	0.000	4.356	0.001	2.972	0.031	2.304	0.131
ERP (n=18)								
	FRN		P3a		P3b			
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>		
PC&Y	1.339	0.793	4.084	0.003	4.610	0.001		
PC&N	4.361	0.002	1.689	0.438	2.757	0.054		
Hm&Y	1.505	0.603	3.247	0.019	2.428	0.106		
Hm&N	3.322	0.016	0.695	1.985	3.843	0.005		

Note: Upper panel, one sample *T* tests (two-tailed) on the % signal changes of the contrast Dis–Hon in areas of interest (i.e. L/R vStr, PCC and mOFC) were conducted under each of the four conditions. Lower panel, one sample *T* tests (two-tailed) on the mean amplitudes of the contrast Dis–Hon in ERP components of interest (i.e. FRN, P3a and P3b) were conducted under each of the four conditions. All *p* values were adjusted by Bonferroni correction. Hm—human counterpart, PC—computer counterpart, Y—being detected, and N—being not detected.

regions in subjects experiencing emotional feelings of a large variety of emotions, including fear, anger, sadness and happiness (Craig, 2009). They were then supposed to serve as a salience

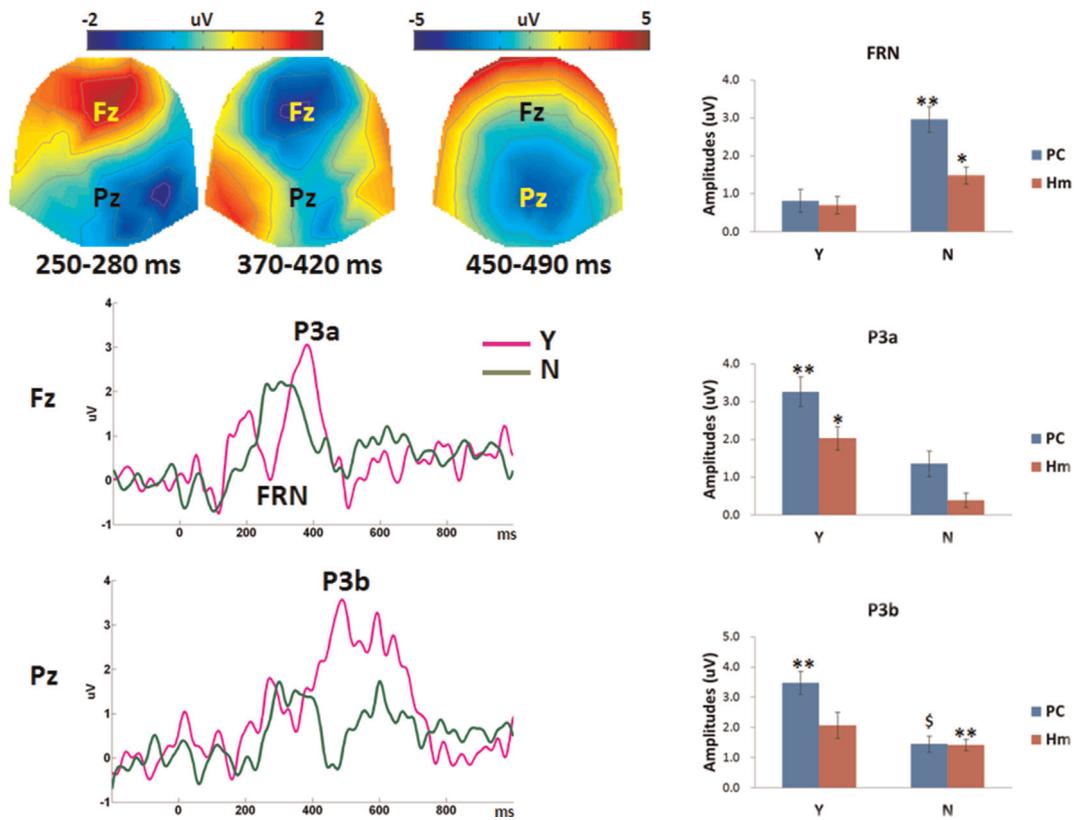


Fig. 4. ERP findings. The outcomes of dishonest (vs. honest) choices elicited smaller FRN (250–280 ms, most prominent at Fz), P3a (370–420 ms, most prominent at Fz), and P3b (450–490 ms, most prominent at Pz; only for the PC [computer] but not Hm [human] counterpart) in the N condition (not being detected) than in the Y condition (being detected). Scalp topographies and waveforms of (dishonest–honest) choices are shown. The y-axis indicated the amplitudes (μV) of the contrast Dis–Hon. $^{\$}p < 0.06$, $^*p < 0.05$, and $^{**}p < 0.01$. Error bar denotes sem.

system which responds to the degree of subjective salience (Me-
 non and Uddin, 2010). Considering that the Dis (vs. Hon) choices
 were related with larger positive (gain because of not detected) or
 negative (loss because of being caught) feelings, the fMRI activa-
 tions in bilateral inferior frontal gyrus/insula and ACC might reflect
 the intensity of feelings to the outcome.

4.2. Time course of processing the outcomes of dishonest choices

Consistent with our second *a priori* hypothesis, we found that
 there is smaller FRN (more positive-going amplitudes) when dis-
 honest (relative to honest) choices were made but were not de-
 tected compared to detected. The FRN was found to be able to

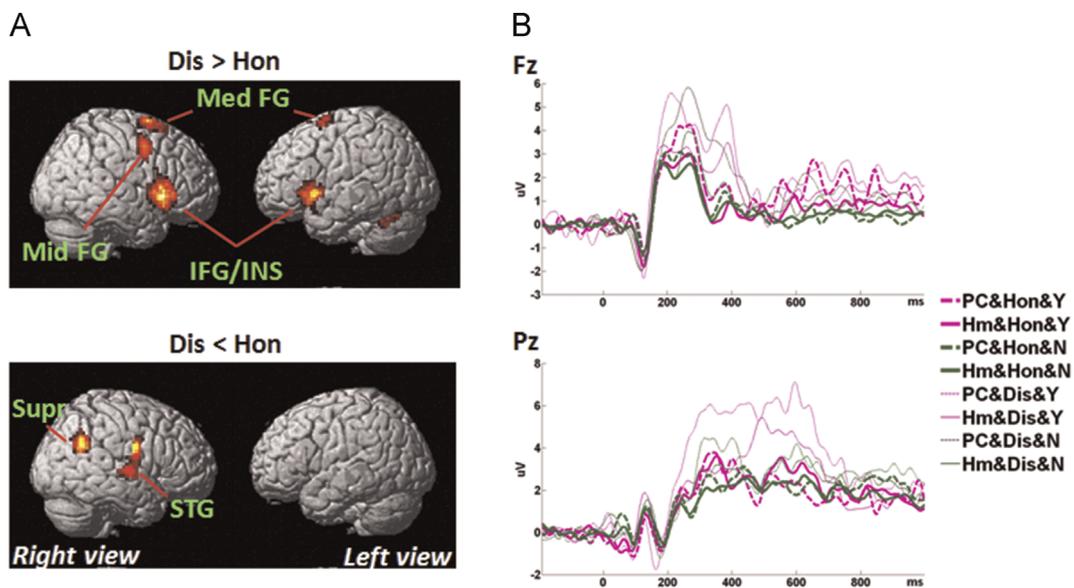


Fig. 5. (A) Average effect of condition for the contrast Dis–Hon in fMRI data. IFG/INS—inferior frontal gyrus/insula, Med FG—medial frontal gyrus, Mid FG—middle frontal gyrus, STG—superior temporal gyrus, and Supr—supramarginal gyrus. (B) ERP waveforms of all conditions in channel Fz and Pz. Thick (thin) lines represent Hon (Dis) choices, solid (dashed) lines represent human (computer) counterpart, and purple (green) lines represent being (not) detected.

quickly represent the relative value of an outcome in possible outcomes, with larger FRN corresponding to the other outputs than the best feedback (Holroyd, et al., 2004). In line with this thought, our result suggests that the successful dishonest action was recognized as the best output compared with the consequences of the other conditions in the very early neural processing of feedback information.

Further analyses showed that the difference between the dishonest outcome and honest outcome was found only when not being detected and suggested that the FRN responds to the outcome of successful (but not failed) dishonest choice. Holroyd et al. (2004) found that both the medium and the worst outcomes elicited larger FRN than the best outcomes, whereas there was no difference between the medium and the worst outcomes. The similar FRN patterns between Holroyd et al.'s and our findings support the validity of this study and suggest using this early ERP component to uncover the dishonest persons when they know that they have escaped detection.

This FRN response pattern is similar to that of the BOLD activations in bilateral vStr and PCC. The convergent findings from both fMRI and ERP datasets again support the validity of our task design. Furthermore, they also suggested that the vStr and PCC are the sources generating the FRN. Consistent with this idea, some previous studies using source reconstruction analysis also showed that the vStr (Foti et al., 2011; Martin et al., 2009) and PCC (Donamayor et al., 2011; Nieuwenhuis et al., 2005c) are the possible origins of FRN.

Surprisingly, larger P3a was elicited by a failed dishonest action (no reward) relative to an honest choice (medium reward), while there was no difference of P3a when comparing successful dishonest action (large reward) with honest choice. P3a is often aroused by novel and highly deviant or salient task-irrelevant stimuli (for a review, see Nieuwenhuis et al., 2005a) and is suggested to reflect the early attention allocation to stimuli (for a review, see Polich and Criado (2006)). Therefore, our results may reflect that enhanced attention was invested into a failed dishonest action (relative to honest choice), whereas a comparable amount of attention given to a successful dishonest behavior and an honest choice. Interestingly, the P3a response pattern was a mirrored reverse to the FRN response pattern. It is possible that the dishonest persons firstly processed the feedback of successful action (reflected by enhanced FRN) and then processed the outcome of failed action (reflected by increased P3a). However, we did not find the similar activation pattern in fMRI data. Previous studies did not get a convergent conclusion about the neural sources of P3. One opinion is that P3-like neural responses were generated by multiple, relatively independent sources while the other opinion is that the P3 activity reflects the influence of a broadly distributed neural system that synchronously impacts several brain areas (for a review, see Nieuwenhuis et al., 2005a). Future studies should clarify what neural source(s) generated the P3a for outcomes of failed dishonest actions.

Contrary to our third *a priori* hypothesis, we found smaller P3b (more positive-going amplitudes) when dishonest (relative to honest) choices were made but not detected than detected dishonest choices when the counterpart was a computer (but not human). This result was different from the Yeung and Sanfey's finding (2004) that larger P3b was elicited by larger reward magnitude, suggesting that the P3b amplitudes in our study did not simply reflect the magnitude of rewards. The P3b was found to be larger for unexpected than expected events, regardless of the valence (for a review, see Martin, 2012). Participants made dishonest choices in this task to gain a larger reward. The outcome of a failed dishonest choice was contrary to their predictions and may thus elicit larger P3b. Moreover, we found that dishonest (vs. honest) outcomes were associated with larger P3b in almost all

conditions (except for the condition of the human counterpart and being detected). The P3b amplitudes were modulated by the amount of attention paid to the stimulus (Nieuwenhuis et al., 2005a). Compared to the outcome of an honest action (a default choice in social interaction), participants in this study might have attributed larger attention to the dishonest choice, which is risky and often violates social norms. More importantly, we found that failed dishonest action elicited larger P3b with the computer but not human counterpart. This is the only difference between counterpart types we found in ERP data. Considering that the only difference between counterpart types in fMRI results was found in mOFC, we speculate that a similar neural processing is reflected by both datasets. That is to say, when being detected, compared with the output of a safe action (i.e., playing computer honestly), the outcome of an honest choice (i.e., playing human honestly) was attributed with an extra reward, which was represented by stronger activations in mOFC and larger P3b amplitudes for an honest outcome. Therefore, when comparing the dishonest outcome with the honest outcome, relatively negative activations in mOFC were detected, whereas no significant P3b difference was found. This speculation is interesting but requires further study to validate it.

4.3. Limitation and future studies

There are a few shortcomings in this study. First, we only studied females of a narrow age range in order to control for the sex- and age-related cognitive effects on the experimental task. However, the trade-off is the limitation of the generalization of our findings to other populations, e.g. to females older than the participants in this study and to a male population. Previous studies have shown significant sex-related effects on risky decision making. For example, Dreber and Johannesson (2008) have shown that males are more likely than females to lie to gain monetary benefits. A recent fMRI study found that, when instructed to tell lies about personal information, males elicited stronger activations in the left middle frontal gyrus than did females (Marchewka et al., 2012). Also, difference in gender was found to have influenced emotion processing; for example, in a task requiring cognitive control of emotions, females showed increased activations in more emotion-associated areas (e.g., the amygdala and the orbitofrontal cortex), while males showed stronger activations in more cognition-associated regions (e.g., prefrontal and superior parietal areas) (Koch et al., 2007). In line with these findings, males and females may respond differently while being dishonest as well as when encountering the consequences of their choices. Future studies should directly compare the neural correlates between males and females to fully understand the neural processing of dishonesty. Second, the small sample size could have been lowered the power of the analyses and henceforth may increase the risk for false negative results. Future studies should consider conducting a priori sample size estimation and having better control on the attrition of the subjects throughout the experiment. Also, a large sample size would add stability of the findings and increase the power of the analyses. Third, a laboratory experimental task paradigm was recruited in this study to control the variables of interest or no interest. However, participants in this environment may behave differently compared to dishonest persons in real life. Further studies should test whether the findings in the laboratory could be generalized to real life contexts. Last but not the least, participants in this study were all healthy volunteers whereas some of the dishonest persons in real life are driven by psychological and/or neurological deficits (Dike et al., 2005; Poletti et al., 2011). It is also interesting to investigate the neural processing of dishonest behaviors in people with mental disorders.

5. Conclusion

We investigated the neural correlates of processing the outcomes of dishonest decisions. Both fMRI and ERP methodologies were recruited to record neural responses in participants when they were playing an economic exchanging game task. We found that the outcomes of successful dishonest (vs. honest) choices elicited both stronger BOLD activations in vStr and PCC and smaller FRN. These results suggest that successful dishonest choices elicited more positive outcome evaluation and attention processing. Moreover, the outcomes of failed dishonest (relative to honest) choices were associated with different BOLD activations in mOFC and different P3b amplitudes between human and computer counterparts. These findings suggest that processing the output of social decision making (playing human) requires different neural mechanisms than that of risk taking (playing computer). The findings advanced our understanding about the neural processing of outcome presentation post a dishonest choice.

Acknowledgments

This work was supported by the May Endowed Professorship of The University of Hong Kong, the Research Grants Council Humanities and Social Sciences Prestigious Fellowship Scheme (Ref. no. HKU703-HSS-13) and the National Natural Science Foundation of China (Ref. no. 31070986). The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Abe, N., 2011. How the brain shapes deception: an integrated review of the literature. *Neuroscientist* 17, 560–574.
- Abe, N., Fujii, T., Ito, A., Ueno, A., Koseki, Y., Hashimoto, R., Hayashi, A., Mugikura, S., Takahashi, S., Mori, E., 2014. The neural basis of dishonest decisions that serve to harm or help the target. *Brain Cognit.* 90C, 41–49.
- Abe, N., Okuda, J., Suzuki, M., Sasaki, H., Matsuda, T., Mori, E., Tsukada, M., Fujii, T., 2008. Neural correlates of true memory, false memory, and deception. *Cereb. Cortex* 18, 2811–2819.
- Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., Fehr, E., 2009. The neural circuitry of a broken promise. *Neuron* 64, 756–770.
- Berg, P., Scherg, M., 1994. A multiple source approach to the correction of eye artifacts. *Electroencephalogr. Clin. Neurophysiol.* 90, 229–241.
- Breiter, H.C., Aharon, I., Kahneman, D., Dale, A., Shizgal, P., 2001. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30, 619–639.
- Coricelli, G., Critchley, H.D., Joffily, M., O'Doherty, J.P., Sirigu, A., Dolan, R.J., 2005. Regret and its avoidance: a neuroimaging study of choice behavior. *Nat. Neurosci.* 8, 1255–1262.
- Craig, A.D., 2009. How do you feel–now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70.
- Dike, C.C., Baranoski, M., Griffith, E.E.H., 2005. Pathological lying revisited. *J. Am. Acad. Psychiatry Law* 33, 342–349.
- Donamayor, N., Marco-Pallares, J., Heldmann, M., Schoenfeld, M.A., Munte, T.F., 2011. Temporal dynamics of reward processing revealed by magnetoencephalography. *Hum. Brain Mapp.* 32, 2228–2240.
- Dreber, A., Johannesson, M., 2008. Gender differences in deception. *Econ. Lett.* 99, 197–199.
- 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. *Hum. Brain Mapp.* 32, 2207–2216.
- Fujiwara, J., Tobler, P.N., Taira, M., Iijima, T., Tsutsui, K., 2009. Segregated and integrated coding of reward and punishment in the cingulate cortex. *J. Neurophysiol.* 101, 3284–3293.
- Gehring, W.J., Willoughby, A.R., 2002. The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295, 2279–2282.
- Greene, J.D., Paxton, J.M., 2009. Patterns of neural activity associated with honest and dishonest moral decisions. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12506–12511.
- Haber, S.N., Knutson, B., 2010. The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35, 4–26.
- Hahn, B., Ross, T.J., Stein, E.A., 2007. Cingulate activation increases dynamically with response speed under stimulus unpredictability. *Cereb. Cortex* 17, 1664–1671.
- Hampson, M., Driesen, N.R., Skudlarski, P., Gore, J.C., Constable, R.T., 2006. Brain connectivity related to working memory performance. *J. Neurosci.* 26, 13338–13343.
- Hewig, J., Trippel, R., Hecht, H., Coles, M.G., Holroyd, C.B., Miltner, W.H., 2007. Decision-making in blackjack: an electrophysiological analysis. *Cereb. Cortex* 17, 865–877.
- Holroyd, C.B., Larsen, J.T., Cohen, J.D., 2004. Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology* 41, 245–253.
- Izuma, K., Saito, D.N., Sadato, N., 2008. Processing of social and monetary rewards in the human striatum. *Neuron* 58, 284–294.
- Keil, A., Debener, S., Gratton, G., Junghofer, M., Kappenman, E.S., Luck, S.J., Luu, P., Miller, G.A., Yee, C.M., 2014. Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology* 51, 1–21.
- Knutson, B., Fong, G.W., Bennett, S.M., Adams, C.M., Homme, D., 2003. A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *Neuroimage* 18, 263–272.
- Koch, K., Pauly, K., Kellermann, T., Seiferth, N.Y., Reske, M., Backes, V., Stocker, T., Shah, N. J., Amunts, K., Kircher, T., Schneider, F., Habel, U., 2007. Gender differences in the cognitive control of emotion: an fMRI study. *Neuropsychologia* 45, 2744–2754.
- Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U., Fehr, E., 2005. Oxytocin increases trust in humans. *Nature* 435, 673–676.
- Langleben, D.D., Schroeder, L., Maldjian, J.A., Gur, R.C., McDonald, S., Ragland, J.D., O'Brien, C.P., Childress, A.R., 2002. Brain activity during simulated deception: an event-related functional magnetic resonance study. *Neuroimage* 15, 727–732.
- Lee, T.M., Chan, C.C., Leung, A.W., Fox, P.T., Gao, J.H., 2009. Sex-related differences in neural activity during risk taking: an fMRI study. *Cereb. Cortex* 19, 1303–1312.
- Lee, T.M.C., Liu, H.L., Tan, L.H., Chan, C.H., Mahankali, S., Feng, C.M., Hou, J.W., Fox, P.T., Gao, J.H., 2002. Lie detection by functional magnetic resonance imaging. *Hum. Brain Mapp.* 15, 157–164.
- Liu, X., Hairston, J., Schrier, M., Fan, J., 2011. Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neurosci. Biobehav. Rev.* 35, 1219–1236.
- Liu, X., Powell, D.K., Wang, H.B., Gold, B.T., Corby, C.R., Joseph, J.E., 2007. Functional dissociation in frontal and striatal areas for processing of positive and negative reward information. *J. Neurosci.* 27, 4587–4597.
- Marchewka, A., Jednorog, K., Falkiewicz, M., Szeszkowski, W., Grabowska, A., Szatkowska, I., 2012. Sex, lies and fMRI–gender differences in neural basis of deception. *PLoS ONE*, 7.
- Martin, L.E., Potts, G.F., Burton, P.C., Montague, P.R., 2009. Electrophysiological and hemodynamic responses to reward prediction violation. *Neuroreport* 20, 1140–1143.
- Martin, R.S., 2012. Event-related potential studies of outcome processing and feedback-guided learning. *Front. Hum. Neurosci.* 6.
- McCabe, K., Houser, D., Ryan, L., Smith, V., Trouard, T., 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci.* 98, 11832–11835.
- McCoy, A.N., Crowley, J.C., Haghghian, G., Dean, H.L., Platt, M.L., 2003. Saccade reward signals in posterior cingulate cortex. *Neuron* 40, 1031–1040.
- Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. *Brain Struct. Funct.* 214, 655–667.
- Mohanty, A., Gitelman, D.R., Small, D.M., Mesulam, M.M., 2008. The spatial attention network interacts with limbic and monoaminergic systems to modulate motivation-induced attention shifts. *Cereb. Cortex* 18, 2604–2613.
- Nieuwenhuis, S., Aston-Jones, G., Cohen, J.D., 2005a. Decision making, the p3, and the locus coeruleus–norepinephrine system. *Psychol. Bull.* 131, 510–532.
- Nieuwenhuis, S., Heslenfeld, D.J., von Geusau, N.J., Mars, R.B., Holroyd, C.B., Yeung, N., 2005b. Activity in human reward-sensitive brain areas is strongly context dependent. *Neuroimage* 25, 1302–1309.
- Nieuwenhuis, S., Slagter, H.A., von Geusau, N.J., Heslenfeld, D.J., Holroyd, C.B., 2005c. Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. *Eur. J. Neurosci.* 21, 3161–3168.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Osinsky, R., Mussel, P., Ohrlin, L., Hewig, J., 2013. A neural signature of the creation of social evaluation. *Soc. Cognit. Affect. Neurosci.* 9, 731–736.
- Poletti, M., Borelli, P., Bonucelli, U., 2011. The neuropsychological correlates of pathological lying: evidence from behavioral variant frontotemporal dementia. *J. Neurol.* 258, 2009–2013.
- Polich, J., Criado, J.R., 2006. Neuropsychology and neuropharmacology of P3a and P3b. *Int. J. Psychophysiol.* 60, 172–185.
- Sip, K.E., Lyng, M., Wallentin, M., McGregor, W.B., Frith, C.D., Roepstorff, A., 2010. The production and detection of deception in an interactive game. *Neuropsychologia* 48, 3619–3626.
- Sip, K.E., Skewes, J.C., Marchant, J.L., McGregor, W.B., Roepstorff, A., Frith, C.D., 2012. What if I Get Busted? Deception, choice, and decision-making in social interaction. *Front. Neurosci.* 6, 58.
- Spence, S.A., Farrow, T.F.D., Herford, A.E., Wilkinson, I.D., Zheng, Y., Woodruff, P.W.R., 2001. Behavioural and functional anatomical correlates of deception in humans. *Neuroreport* 12, 2849–2853.
- Sun, D., Lee, T.M., Chan, C.C., 2013. Unfolding the spatial and temporal neural processing of lying about face familiarity. *Cereb. Cortex*.
- Ursu, S., Carter, C.S., 2005. Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: implications for neuroimaging studies of decision-making. *Brain Res. Cognit. Brain Res.* 23, 51–60.
- Vrij, A., 2004. Guidelines to catch a liar, *The Detection of Deception in Forensic Contexts*. Cambridge University Press, New York.
- Wager, T.D., Keller, M.C., Lacey, S.C., Jonides, J., 2005. Increased sensitivity in neuroimaging analyses using robust regression. *Neuroimage* 26, 99–113.
- Xue, G., Lu, Z., Levin, I.P., Weller, J.A., Li, X., Bechara, A., 2009. Functional dissociations of risk and reward processing in the medial prefrontal cortex. *Cereb. Cortex* 19, 1019–1027.
- Xue, G., Lu, Z.L., Levin, I.P., Bechara, A., 2011. An fMRI study of risk-taking following wins and losses: implications for the Gambler's fallacy? *Hum. Brain Mapp.* 32, 271–281.
- Yeung, N., Sanfey, A.G., 2004. Independent coding of reward magnitude and valence in the human brain. *J. Neurosci.* 24, 6258–6264.
- Zhang, H.J., Sun, D., Lee, T.M., 2012. Impaired social decision making in patients with major depressive disorder. *Brain Behav.* 2, 415–423.