Brain and Cognition 125 (2018) 149–156

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Telling a truth to deceive: Examining executive control and reward-related processes underlying interpersonal deception

Liyang Sai[a], Haiyan Wu[b], Xiaoqing Hu[c,*], Genyue Fu[a,*]

[a] Institute of Psychological Science, Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Center for Cognition and Brain Disorders, Hangzhou Normal University, Hangzhou, China

[b] Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Science, Beijing, China

[c] Department of Psychology, The State Key Laboratory of Brain and Cognitive Science, The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:
ERP
Executive control
Outcome evaluation
Deception
Interpersonal deception

ABSTRACT

Does deception necessarily involve false statements that are incompatible with the truth? In some cases, people choose truthful statements in order to mislead others. This type of deception has been investigated less. The current study employed event-related brain potentials (ERPs) to investigate the neurocognitive processes when both truthful and false statements were used to deceive others. We focused our ERP analysis on two stages: a decision making stage during which participants decided whether to tell a false or a truthful statement, and an outcome evaluation stage during which participants evaluated whether their deception had succeeded or not. Results showed that in the decision making stage, intentions to deceive elicited larger N200s and smaller P300s than an honest control condition. During the outcome evaluation stage, success/failure feedback in the deception condition elicited larger Reward positivity (RewP) and feedback-P300 than feedback after honest responses. Importantly, whether participants chose to tell false or true statements, did not further modulate executive control or reward-related processes. Taken together, these results suggest that during interpersonal deception, having deceptive intentions engages executive control and reward-related processes regardless of the veracity of statements.

1. Introduction

To deceive others, people may spontaneously tell a falsified statement that is inconsistent with the truth. However, this strategy may not be optimal when potential recipients are already aware of the senders' deceptive intentions and therefore may not believe the senders' messages. In this scenario, the senders could strategically choose a truthful statement so that the recipient would take the truth as false. To date, the neurocognitive processes underlying such strategic deception involving truthful statements remain unclear. The present study employed an interpersonal deception game in which people deceived their opponents using both true and false statements. Compared with previous deception studies that compared truthful vs. false responses, we were able to use this manipulation to compare deceptive (regardless of the veracity of statements) with honest responses.

Previous studies have examined neurocognitive processes underlying both instructed and voluntary deception. In instructed deception, participants are instructed to lie and give false statements (e.g., deny their involvement of previous acts, see Abe et al., 2006; Lee et al., 2002; Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003; Spence et al., 2001). In voluntary deception, participants choose whether to make honest or dishonest decisions and they can over-report their performance for incentives (see Abe & Greene, 2014; Cui et al., 2018; Greene & Paxton, 2009; Hu, Pornpattananangkul, & Nusslock, 2015; Sip et al., 2010; Yin, Reuter, & Weber, 2016). Although instructed and voluntary deception differ along important dimensions such as social and motivational processes (e.g., perspective taking and reward processing, see Lisofsky, Kazzer, Heekeren, & Prehn, 2014), both deceptions require participants to execute a falsified response that is inconsistent with the truth. Specifically, the execution of truth-inconsistent responses requires the detection of conflict between two competing responses and then the inhibition of the goal-irrelevant truthful response (Abe et al., 2006; Greene & Paxton, 2009; Hu et al., 2015; Hu, Wu, & Fu, 2011; Johnson, Henkell, Simon, & Zhu, 2008; Johnson, Barnhardt, & Zhu, 2004). Based on these results, researchers hypothesize that truth-telling is the default response tendency (Christ, Van Essen, Watson, Brubaker, & McDermott, 2009; Farah, Hutchinson, Phelps & Wagner, 2014; Vrij, 2008; Vrij, Fisher, Mann, & Leal, 2006; but see Bereby-Meyer & Shalvi,

[*] Corresponding authors at: Department of Psychology, The University of Hong Kong, Pokfulam, Hong Kong, China (X. Hu). Institute of Psychological Science, Hangzhou Normal University, Hangzhou 311121, China (G. Fu).
E-mail addresses: xiaoqinghu@hku.hk (X. Hu), fugenyue@hznu.edu.cn (G. Fu).

https://doi.org/10.1016/j.bandc.2018.06.009
Received 11 January 2018; Received in revised form 26 June 2018; Accepted 27 June 2018

2015).

However, deception can be achieved by truthful statements as well. Critically, the defining feature of deception is that the message sender has an intention to mislead the recipient (Vrij, 2008). According to this definition, a deceiver could intentionally tell a truthful statement in order to mislead the recipient to believe its opposite. This type of deception can be adaptive especially when the recipient is already aware of the sender's deceptive intention, for example, in highly competitive scenarios such as negotiation (Rogers, Zeckhauser, Gino, Norton, & Schweitzer, 2017).

To date, very few studies have examined this type of deception (Carrión, Keenan, & Sebanz, 2010; Ding, Sai, Fu, Liu, & Lee, 2014; Sip et al., 2010; Volz, Vogeley, Tittgemeyer, von Cramon, & Sutter, 2015). Employing different methodologies such as ERP, fMRI, fNIRS, researchers consistently found that when the participants' goal was to mislead their opponents, telling a true statement would still engage similar executive control processes as telling a false statement. For example, in Sip et al. (2010), participants played the zero-sum dice game, Meyer, with a confederate. The participants' goal was to deceive the confederate about a dice combination. Sometimes participants chose to tell the truth about the dice combination, however, this was done with the expectation that the opponent would not trust them and thus believe the opposite to be true. Sip et al. (2010) found that both true and false claims about the dice combination were associated with higher activities in the fronto-polar cortex than that in a non-competitive control condition. Moreover, relative to truthful claims, false claims were associated with greater activity in the premotor and parietal cortices, which was taken as evidence that choosing a false claim additionally engaged response selection processes. Employing ERPs, Carrión et al. (2010) showed that both truthful and false claims with a deceptive intention elicited larger executive control-related ERPs, the medial frontal negativity (MFN), than truthful responses without deceptive intentions. These findings provide initial evidence that when truthful statements are used to deceive others, it involves similar executive control processes as when telling false statements to deceive.

Furthermore, because deception involves both information management (e.g. decision making) as well as risk management (e.g. outcome evaluation, see Sip, Roepstorff, McGregor, & Frith, 2008), the present study aims to extend previous research by examining both decision making and outcome evaluation processes in truth-telling deception. Critically, during interpersonal deception, a deceiver may not only decide whether or when to tell a false or a truthful statement, but the deceiver also needs to evaluate whether the deception has succeeded or not. This latter outcome evaluation stage may tap into reward-related processes (Hu et al., 2015; Luo, Sun, Mai, Gu, & Zhang, 2011; Sun, Chan, Hu, Wang, & Lee, 2015). To capture these two essential aspects of interpersonal deception, we leverage ERPs' high temporal resolution in a zero-sum, interpersonal deception game. Examining both stages of deception allows us to provide a more complete picture regarding interpersonal deception and its underlying neurocognitive mechanisms.

In the decision making stage, we focused on the fronto-central N200 and the centro-parietal P300, both of which are implicated in executive control processes. Specifically, it has been suggested that the N200 is a sensitive neural marker of response conflict (Bartholow et al., 2005; for a review, see Folstein & Van Petten, 2008); while the later P300 has been associated with cognitive resource allocation and conflict resolution (Johnson, 1988; Johnson, Barnhardt, & Zhu, 2003). Examining these two ERP components would also be consistent with previous ERP studies on both instructed and voluntary deception (Hu et al., 2015; Hu et al., 2011; Johnson et al., 2004; Johnson et al., 2008; Suchotzi, Crombez, Smulders, Meijer, & Verschuere, 2015; Wu, Hu, & Fu, 2009). Regarding how deception may modulate N200-P300, we hypothesize that telling a truth to deceive would engage similar executive control processes as when telling a lie to deceive (see Carrión et al., 2010). Moreover, deceptive responses, regardless of whether they were true or false, would elicit a larger N200 and smaller P300 than honest responses without deceptive intentions.

Regarding the outcome evaluation stage, we focused on two ERP components that have been intensively studied in the outcome evaluation literature: the Reward Positivity (RewP) and the feedback-P300. The RewP is typically observed during the 200–300 ms time window after the onset of the performance feedback, which indicates whether participants' behavior has led to good or bad outcomes (for a review, see Proudfit, 2015). Specifically, positive feedback would enhance this RewP while negative feedback would attenuate this RewP (Gehring & Willoughby, 2002; Miltner, Braun, & Coles, 1997; for reviews, see Proudfit, 2015; Walsh & Anderson, 2012).

The feedback-P300 is another ERP component that occurs later than RewP and is also intensively studied in the outcome evaluation research. Compared to the RewP, the results of feedback-P300 are less consistent across studies: Some studies have found that the feedback-P300 is sensitive to reward magnitude but not to valence (Sato et al., 2005; Yeung & Sanfey, 2004); while other studies have found that the feedback-P300 does encodes reward valence but may also implicate more cognitive processing of the feedback (Hajcak, Holroyd, Moser, & Simons, 2005; Hajcak, Moser, Holroyd, & Simons, 2007).

In relation to deception, Luo et al. (2011) employed an instructed deception paradigm and reported that the outcomes after instructed deception had elicited a larger RewP and feedback-P300 than the outcomes after honest responses (Luo et al., 2011). This finding suggests that instructed deception versus honesty modulates outcome evaluation processes. Based on this study, we also predicted that feedback after deception would elicit larger RewP and feedback-P300 than feedback after honest responses. Moreover, since both truthful and false statements serve the same goal to deceive others, we predict that there are no significant differences for RewP and feedback-P300 between truth-deceive and false-deceive responses.

2. Methods

2.1. Participants

Twenty-one undergraduates from Zhejiang Normal University were recruited in the study (9 males, Mean age = 24.05 years, age range 21–29 years). Two participants were excluded from behavioral and ERP analyses due to excessive eye blinks in the experiment; one additional participant was excluded from analyses in the outcome evaluation stage because he or she had insufficient outcome evaluation trials ($n < 20$) in the four conditions of the experiment. Thus, the final sample for the decision-making stage was 19 (7 males, Mean age = 24.16 years, age range 21–29 years), and the final sample for the outcome evaluation stage was 18 (7 males, Mean age = 24.17 years, age range 21–29 years). This sample size was consistent with previous studies on the topic ($n = 11$ in Carrión et al., 2010, $n = 25$ in Ding et al., 2014; $n = 14$ in Sip et al., 2010). All participants were right-handed with normal or corrected-to-normal vision. The study was approved by the Ethics Committee of Zhejiang Normal University.

2.2. Procedure

Participants completed two task sessions: an honest control session in which no deceptive intentions were involved; and an interpersonal deceptive game session during which participants were asked to mislead their opponents. These two sessions were presented in a fixed order, with the honest control session always coming first, followed by the interpersonal deception session. We chose this fixed order because if the honest session followed the interpersonal deception session, participants' honest behavior might be influenced by their previous deceptive intentions even when no deception was required (for a similar task order and rationale, see Carrión et al., 2010). In the interpersonal deceptive game, participants were told that they were about to play a

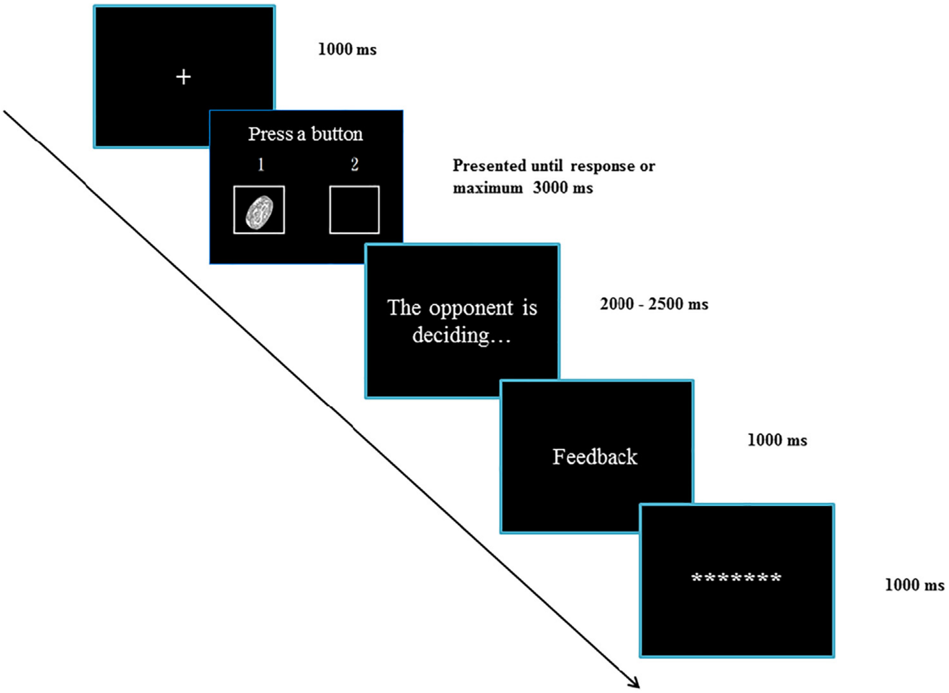
+
1000 ms
Press a button
1
2
Presented until response or maximum 3000 ms
The opponent is deciding…
2000 - 2500 ms
Feedback
1000 ms
****
1000 ms

Fig. 1. a sample trial structure in the interpersonal deception game.

"coin guessing" game with an opponent (i.e., a confederate). Participants were then introduced to their opponents before the experiment, and were told that they would sit in two different rooms and interact with each other through the intranet. In fact, participants played the game by themselves.

During the game, participants were presented with two boxes on a computer monitor. Only one of the boxes contained a coin. Participants were told that only they could see which box contained the coin, and that they needed to mislead their opponents about the location of the coin in order to win. Participants were also informed that the payoff of the game would be zero-sum: they would lose the coin to their opponent if their opponent correctly guessed its location. Importantly, the word "deception" was never mentioned in these instructions. Participants were further told that their goal was to win as many coins as possible, and their task performance would determine their final payment.

Each trial began with a fixation point that lasted for 1000 ms, signaling for the participants to fix their gaze on the center of the screen (for details of a trial structure, see Fig. 1). Two boxes were then presented on the screen for a maximum of 3000 ms, one of which contained a coin. Participants were asked to show their opponents which box (e.g., left or right) contained the coin by pressing one of the two buttons as soon as possible. The period from the onset of the box stimuli to the moment when the participants made a response was referred to as the *decision making stage*. After participants entered their response, the phrase "your opponent is deciding" was presented for 2000–2500 ms, which indicated that the opponent was deciding whether or not to believe the message the participants just delivered. Subsequently, participants were notified with feedback that they had either won (a number "2") or lost (a number "0"). The feedback was presented for 1000 ms on the monitor, which was referred to as the *outcome evaluation stage*. Note that all feedback was pre-determined and was presented in a random order, i.e., it was not contingent upon participants' actual responses. A debriefing session after the study showed that all participants believed that they had played the game with another person.

There were 240 trials in the interpersonal deceptive game session. The coins were presented equally in the left- and right-side boxes. Participants could take a short break every 6 trials to minimize fatigue. Participants were also instructed to minimize body movements during the experiment. Before the experimental session, participants received 10 practice trials to familiarize them with the task procedure.

In the honest control session, the procedure was identical to the interpersonal deceptive game session except that participants were told that the other participant (i.e., a confederate) was their partner and participants' goal is to make the partner believe that they were his/her teammates. Thus, participants needed to be honest about the location of the coins. The session will end if the partner believed what participants had said in 5 consecutive trials. Because participants always delivered the truthful message and they needed to ensure that their partners trusted them, no reward was involved in this honest session. Therefore, unlike the interpersonal deceptive session, participants in the honest session did not have deceptive intentions. Our pilot study has indicated that in the experimental deception condition, the average amount of trials that participants told the truth in order to deceive was around 80. Thus, we had programed 75 trials in the honest control condition to match the trial numbers of these two conditions: tell a truth to deceive vs. tell a truth to be honest.

2.3. EEG recordings and analyses

Continuous EEGs were recorded from 32 scalp sites using Ag/AgCl electrodes (FP1/2; F7/8; F3/4; Fz; FT7/8; FC3/4; FCz; T3/4; C3/4; Cz; TP7/8; CP3/4;CPz;T5/6; P3/4; Pz;O1/2;Oz) mounted in an elastic cap (Neuroscan Inc., USA) according to the international 10–20 system, with references on linked mastoids. Electrode impedances were kept below 5 kΩ. The vertical electro-oculograms (EOGs) were recorded supra-orbitally and infra-orbitally from the right eye; the horizontal EOG was recorded from electrodes placed at the outer canthus of the left eye and right eye. The sampling rate was 500 Hz with 0.1–100 Hz online band-pass filtering.

For offline analyses, the data was first filtered using a band-pass filter of 0.1–30 Hz. For the decision making stage and outcome evaluation stage, EEG epochs were extracted −200 to 1000 ms post stimulus onset (the "box" stimuli and the feedback stimuli, respectively). Epochs exceeding ± 100 μV in amplitude were excluded from ERP averaging. For the decision making stage, segmented EEG epochs were averaged into three conditions: tell a truth to deceive, tell a lie to deceive (from the interpersonal deception session) and tell a truth to be honest (from the honest control session). During the outcome evaluation stage, the segmented EEG epochs were averaged into six conditions: truth-deceive-success vs. failure, false-deceive-success vs. failure, honest-success vs. failure. All segmented EEG epochs were baseline corrected using the mean amplitude of the 200 ms pre-stimulus interval. At least 30 clean trials were included to generate ERPs for each condition.

During the decision making stage, we focused on the frontal-central N200 and parietal P300. For N200 analyses, we measured the mean amplitude between 270 and 330 ms after the "boxes stimulus" collapsing across Fz and FCz electrodes; For P300 analyses, we measured the mean amplitude between 340 and 400 ms after the "boxes stimulus" collapsing across CPz and Pz electrodes. The time windows of N200 and P300 are similar to previous deception studies (Hu et al., 2011; Suchotzi et al., 2015; Wu et al., 2009). During the outcome evaluation stage, we focused on RewP and feedback-P300. For RewP analyses, we measured the mean amplitude between 250 and 350 ms collapsing across Fz and FCz electrodes (Bress & Hajcak, 2013; Foti & Hajcak, 2009). For the feedback-P300, we calculated the mean amplitude in the time window of 350–450 ms collapsing across CPz and Pz electrodes. Analyses of variance (ANOVA) were performed on SPSS 20.0. Greenhouse-Geisser correction was applied whenever the assumption of sphericity was violated. Post-hoc comparisons were computed with Fisher's Protected Least Significant Difference.

3. Results

3.1. Behavioral results

During the interpersonal deception game, on average participants chose false statements 55% of the time (Mean ± S.E., 132.32 ± 2.33 times), and truthful statements 45% of the time (107.42 ± 2.31). A paired sample *t*-test showed that participants were more likely to choose false statements to deceive their opponents (t (18) = 5.37, $p < .001$, Cohen's $d = 1.23$).

A repeated measure ANOVA with response type (truth-deceive, false-deceive, honest) as a within-subject factor was conducted on reaction time (RT) as the dependent variable. Results showed a significant main effect of response type, F (2, 36) = 19.92, $p < .01$, $\eta_p^2 = 0.53$. Post-hoc tests showed that both choosing false and truthful statements to deceive took a significantly longer time than choosing honest responses: false-deceive 1725 ± 239 ms, truth-deceive 1646 ± 239 ms vs. vs. honest 613 ± 21 ms, $ps < .01$, Cohen's $d = 1.00$, 1.08 respectively. However, there was no significant difference between the RTs for false-deceive and truth-deceive, t (18) = 1.33, $p = .20$, Cohen's $d = 0.31$ (see Fig. 2).

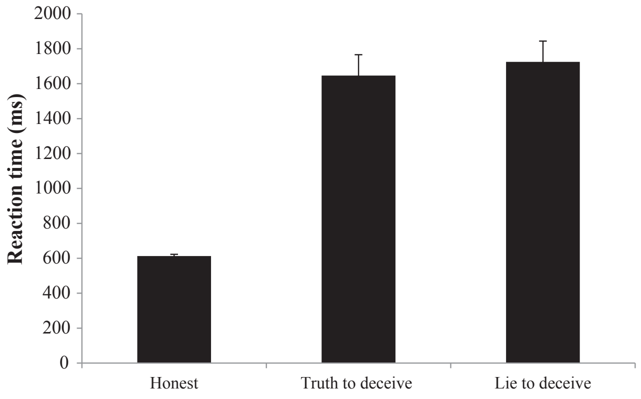

Fig. 2. RTs of each response condition. Error bars indicate standard error of means.

3.2. ERP results

3.2.1. Decision making stage: N200

A 3-level (response type: truth-deceive, false-deceive, honest) one-way repeated measure ANOVA on N200 amplitude revealed a significant main effect of response type, F(2, 36) = 8.25, $p = .004$, $\eta_p^2 = 0.31$. Post-hoc analyses found that both the false-deceive (−3.78 ± 0.86 μV) and truth-deceive (−3.13 ± 1.14 μV) elicited a larger N200 than honest responses (−0.81 ± 1.21 μV), t (18) = 3.33p = .004, Cohen's $d = 0.76$; t (18) = 2.63, $p = .017$, Cohen's $d = 0.60$, respectively. But there was no significant difference for N200 between truth-deceive and false-deceive, t (18) = 1.46, $p = .16$, Cohen's $d = 0.33$. To further corroborate this null result, we employed Bayesian analyses to calculate the probability that the present data support the null hypothesis, i.e., no differences between false- and truth-deceive N200s. The result shows that the BF_{01} factor (the data supporting H0 over H1) was 1.67, thus providing moderate support to the null hypothesis.

3.2.2. Decision making stage: P300

The same repeated measure ANOVA was conducted on P300 amplitude. Results showed that there was a significant main effect of response type, F(2, 36) = 37.41, $p < .001$, $\eta_p^2 = 0.68$. Post-hoc analyses found that both the false-deceive (6.67 ± 0.81 μV) and truth-deceive (5.95 ± 0.95 μV) elicited smaller P300 than honest responses (12.13 ± 1.19 μV), t (18) = −6.37, $p < .001$, Cohen's $d = 1.46$; t (18) = −6.76, $p < .001$, Cohen's $d = 1.55$ respectively. Again, there was no significant P300 differences between truth-deceive and false-deceive responses, t (18) = 1.42, $p = .17$, Cohen's $d = 0.32$ (see Fig. 3 a & b for grand-average ERPs in this decision making stage). To further confirm this null result, the Bayesian analysis revealed the BF_{01} factor (the data supporting H0 over H1) to be 1.75.

3.2.3. Outcome evaluation stage: RewP

To investigate the RewP effect of outcome evaluation, a 3 (response type: truth-deceive, false-deceive, honest) by 2 (feedback value: success vs. failure) repeated measure ANOVA was conducted with RewP amplitude. First, replicating previous findings regarding outcome evaluation, we found a significant main effect of outcome valence, F (1, 17) = 38.72, $p < .001$, $\eta_p^2 = 0.70$, such that success feedback elicited a larger RewP than failure feedback: 10.68 ± 1.08 μV vs. 6.52 ± 0.80 μV, t (17) = 6.22, $p < .001$, Cohen's $d = 1.46$. Second, we found a significant main effect of response type, F (2, 34) = 30.42, $p < .001$, $\eta_p^2 = 0.64$. Post-hoc analyses found that feedback after both truth-deceive (11.16 ± 1.31 μV) and false-deceive (10.62 ± 1.19 μV) responses elicited larger RewP than feedback after honest responses (4.02 ± 0.53 μV), t (17) = 5.63, $p < .001$, Cohen's $d = 1.33$; t (17) = 5.73, $p < .001$, Cohen's $d = 1.35$, respectively. Again, there was no significant RewP difference between feedback after truth-deceive and false-deceive responses, t (17) = 1.27, $p = .22$, Cohen's $d = 0.30$. The Bayesian analysis revealed the BF_{01} factor to be 2.14. Lastly, there was a significant interaction between response type and outcome valence, F (2, 34) = 13.97, $p < .001$, $\eta_p^2 = 0.45$. Further analyses found that this interaction was driven by significant success vs. failure RewP differences after both truth- and false-deceive responses ($ps < .001$), and non-significant success vs. failure RewP differences after honest responses ($p = .36$) (for detailed descriptive and statistical results, see Table 1).

3.2.4. Outcome evaluation stage: Feedback-P300

The same repeated-measure ANOVA was conducted on feedback-P300 amplitudes. First, we found a significant main effect of outcome valence, F(1, 17) = 67.04, $p < .001$, $\eta_p^2 = 0.80$, suggesting that success feedback elicited a larger feedback-P300 than failure feedback:

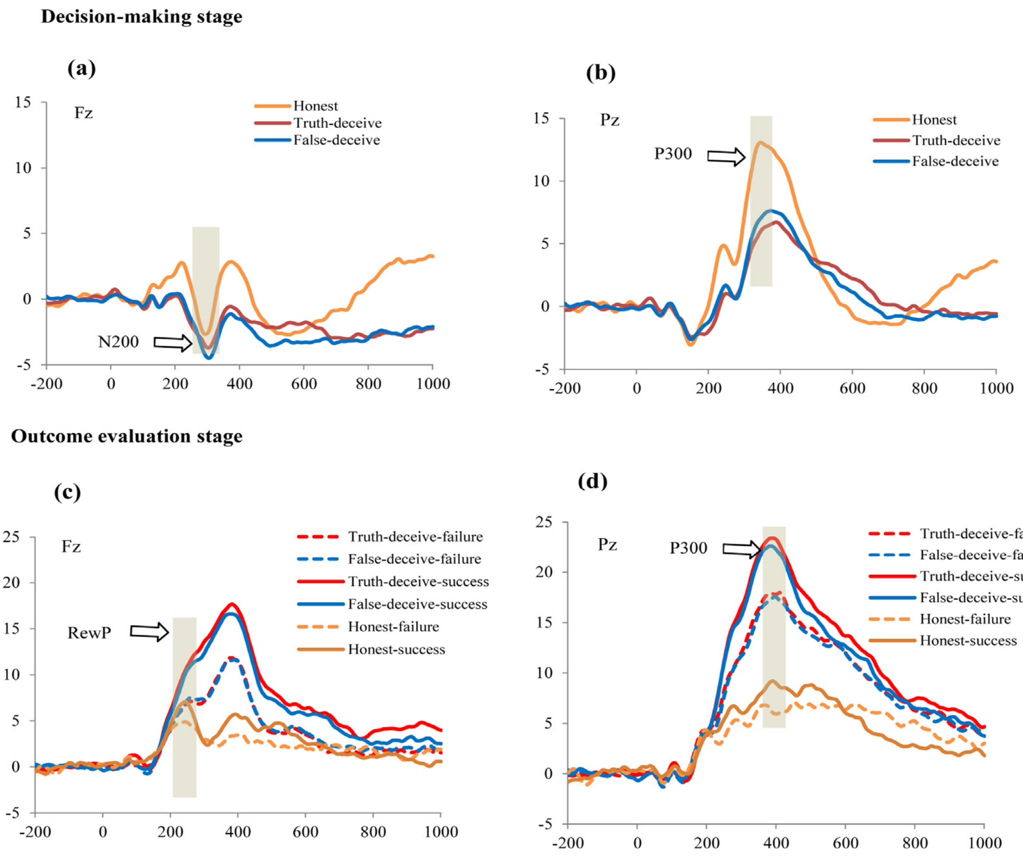

Fig. 3. The grand average ERP for each condition in the decision making stage (a and b) and in the outcome evaluation stage (c and d).

Table 1
Statistics of interactions between response type and outcomes.

	RewP	Feedback-P300
Truth-deceive-success	14.28 (1.53)	22.61(1.27)
Truth-deceive-failure	8.04 (1.26)	16.85 (1.26)
	t (17) = 6.16, *p* < .001, Cohen's *d* = 1.45	*t* (17) = 8.02, *p* < .001, Cohen's *d* = 1.89
False-deceive-success	13.38(1.46)	21.76(1.31)
False-deceive-failure	7.85(1.10)	16.29(1.17)
	t (17) = 5.46, *p* < .001, Cohen's *d* = 1.28	*t* (17) = 6.28, *p* < .001, Cohen's *d* = 1.48
Honest-success	4.38(0.67)	8.05(0.71)
Honest -failure	3.66(0.63)	6.05(0.89)
	t (17) = .95, *p* = .36, Cohen's *d* = 0.22	*t* (17) = 3.32, *p* = .004, Cohen's *d* = 0.78

17.48 ± 0.94 μV vs. 13.06 ± 0.95 μV, t (17) = 8.19, $p < .001$, Cohen's $d = 1.54$. Second, we found a significant main effect of response type, F (2, 34) = 107.53, $p < .001$, $\eta_p^2 = 0.86$. Post-hoc analyses found that feedback after both truth-deceive (19.73 ± 1.22 μV) and false-deceive responses (19.03 ± 1.16 μV) elicited a larger feedback-P300 than feedback after honest responses (7.05 ± 0.75 μV), t (17) = 10.46, $p < .001$, Cohen's $d = 2.47$; t (17) = 10.91, $p < .001$, Cohen's $d = 2.57$. Again, no significant difference was found for feedback-P300 between truth-deceive and false-deceive responses, t (17) = 1.73, $p = .10$, Cohen's $d = 0.41$. A Bayesian analysis reveals that the corresponding BF_{01} factor being 1.19. Lastly, we found a significant interaction between outcome valence and response type, F (2, 34) = 11.40, $p < .001$, $\eta_p^2 = 0.40$, which is driven by significantly larger success vs. failure differences in deceptive conditions than found in the honest condition (see Fig. 4).

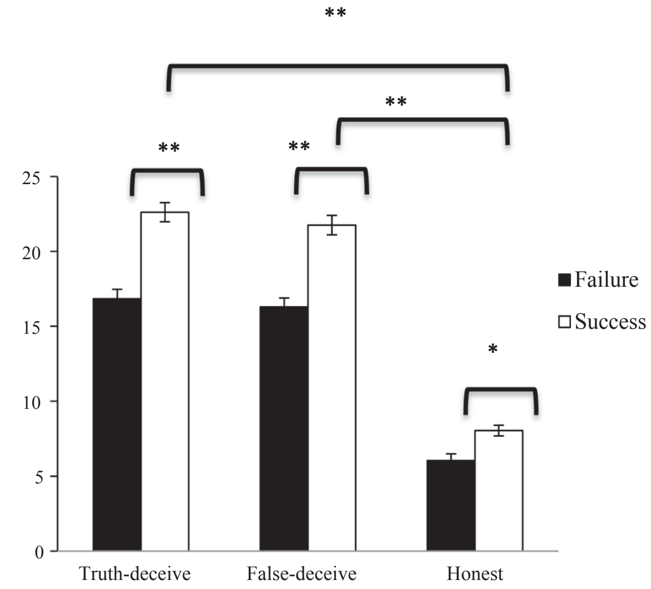

Fig. 4. The interaction between response type and feedback valence for feedback-P300. Error bars indicate standard error of mean. Although success elicited larger feedback-P300 than failure feedback across all conditions, the differences were significantly larger following deceptive responses than following honest responses. $^{**}p < 0.01$, $^{***}p < 0.001$.

4. Discussion

Employing a zero-sum, competitive interpersonal deception game that involved monetary incentives, we investigated the neurocognitive processes underlying interpersonal deception during both the decision making and outcome evaluation stages. To maximize one's payoff, participants not only had to decide which message to deliver so as to mislead their opponents, but they also needed to monitor whether their deception was successful. Replicating previous deception studies, we found that participants engaged in executive control processes when they needed to deceive others compared to when they gave honest responses without deceptive intentions. Furthermore, deception modulated reward-related processes were implicated in the outcome evaluation stage. Most importantly, extending previous studies, we found that when participants tried to deceive others, the veracity of their statements did not modulate neural signals associated with either executive control or reward processing.

4.1. *Executive control processes during decision making*

During the interpersonal deception game, participants had to decide for each trial whether they needed to make a truthful or a false statement regarding the location of the coins. Since participants' primary goal was to maximize their payoff by misleading their opponents, their responses could be considered deceptive regardless of the veracity of their message. Compared with the honest condition in which participants were asked to be honest and to gain their opponents' trust, engaging in interpersonal deception elicited larger frontal-central N200s and reduced parietal P300s. This frontal-central N200 has been implicated in executive control tasks that involve response conflict and resolution, such as the Go/Nogo and Flanker task (Bruin, Wijers, & Van Staveren, 2001; Enriquez-Geppert, Konrad, Pantev, & Huster, 2010; Huster, Westerhausen, Pantev, & Konrad, 2010; Kropotov, Ponomarev, Hollup, & Mueller, 2011; for a review, see Folstein & Van Petten, 2008). Moreover, this frontal-central N200 has been repeatedly found in both instructed and voluntary deception ERP studies, when participants are instructed to deceive or to conceal memories, and when participants over-report their performance for monetary incentives (Gamer & Berti, 2010; Hu et al., 2011; Hu, Pornpattananangkul, & Rosenfeld, 2013; Hu et al., 2015; Suchotzi et al., 2015; Wu et al., 2009; see also Ganis, Bridges, Hsu, & Schendan, 2016). In these deception tasks, people always give a false response to deceive (e.g., denying recognition of their own names), therefore the enhanced frontal-central N200 could be attributed to either falsifying a response that is incompatible with the truth or to participants' deceptive intentions (either being instructed or spontaneous). Supporting the first hypothesis, research shows that telling lies that are incompatible with the truth also elicit response-locked, error-related negativity, suggesting that giving a false statement to deceive may be processed as an error (Johnson, Barnhardt, & Zhu, 2005).

Extending these previous studies, our design allows us to investigate whether the observed brain activity is due to the choice of truthful vs. false statements or honest vs. deceptive intentions. Specifically, even when participants strategically chose a truthful statement to deceive others, this choice elicited enhanced frontal-central N200s than when participants made the same statement without deceptive intentions. Furthermore, during the deception game session, we did not find any N200 differences between truthful or false statements. Thus, this conflict-sensitive frontal-central N200 result suggests that in this interpersonal deception situation, it is the deceptive intention, rather than choosing an incompatible response, that elicits response conflict.

Because response conflict is aversive (Dreisbach & Fischer, 2015), people need to devote cognitive resources to resolve such conflict. The P300s that follow N200s are typically associated with conflict resolution and the amplitude of P300 is inversely related with task demand and the amount of executive control needed (Johnson, 1986). Again, reduced P300s are repeatedly reported in both instructed and voluntary deception research (Johnson et al., 2003, 2005; Hu et al., 2011; Hu et al., 2015; Suchotzi et al., 2015; Wu et al., 2009). Similar to the N200, the P300 is only modulated by deceptive vs. honest intentions instead of the veracity of the statements. Thus, this N200/P300 pattern supports the claim that when people make a decision to deceive others, even conveying a truthful statement will engage executive control processes. It should also be noted that the P300 is involved in a range of cognitive processes other than executive control. In relation to memory concealment, P300 has been widely employed as a neural signal for recognition/familiarity or personal significance in memory detection studies where participants attempt to conceal their recognition of certain stimuli (for a review, see Rosenfeld, Hu, Labkovsky, Meixner, & Winograd, 2013). In this concealed information paradigm, P300s are typically enhanced in response to the to-be-concealed probes compared with irrelevant stimuli, because of their signal value and personal significance in the test, e.g., a guilty examinee would recognize a crime-relevant item such as a murdering weapon.

To date, there are four studies that have investigated the neurocognitive basis of sophisticated deception during interpersonal interaction games that are similar to the present study. Our results are highly consistent with Carrión et al. (2010), which they also reported that telling a lie or a truth to deceive did not elicit significant ERP differences in the conflict-sensitive N450 signals. On the other hand, the other three studies employing fMRI and fNIRS have reported both similarities and dissimilarities between telling a lie and telling a truth to deceive. For example, Volz et al. (2015) reported that the temporo-parietal junction (TPJ) was engaged in both sophisticated deception and plain deception, when compared with plain honest responses. However, the activity of right TPJ could still distinguish between sophisticated deception and plain deceptions. Moreover, Ding et al. (2014) and Sip et al. (2010) reported that while engaging in deception (with both false and truthful messages) recruited cortical regions (e.g., the right superior frontal gyrus, the frontal-polar cortex) that are associated with executive control, telling false to deceive would additionally elicit activities in premotor cortex, middle frontal gyrus, etc. Thus, across different studies employing different technologies, data strongly suggest that telling a truth to deceive would engage in executive control processes compared with honest responses or plain truth; however, whether telling a lie and telling a truth to deceive would engage in different levels of executive control processes may depend on specific experimental design and neuroimaging techniques.

4.2. *Reward-related processes in outcome evaluation*

Evaluating whether one's deception succeeded or not is an indispensable component of deception (Sip et al., 2010; Sun et al., 2015). Because deception is typically driven by self-interest to increase ones' gains, deceivers need to monitor the outcome of deception and thus engage in reward-related processes. Reward positivity (RewP) is an ERP component that has been associated with reward processing, especially when people evaluate their performance outcome in a binary manner such as good or bad, win or lose, success or failure (Proudfit, 2015). Note that many previous studies have regarded this ERP component as a negative-going waveform that is enhanced by negative feedback (i.e., feedback-related negativity, medial frontal negativity or feedback-negativity, Gehring & Willoughby, 2002; Miltner et al., 1997). However, recent studies show that this ERP component may actually be elicited by positive feedback or reward, and is attenuated by the omission of reward or by punishment (Proudfit, 2015). Our recent studies that focused on reward processing in memory concealment also support this reward-positivity hypothesis. Specifically, by employing principal component analysis to decompose overlapping ERP activities, we showed that feedback indicating success in concealing memories elicited a stronger RewP. In contrast, this positive ERP was attenuated and became negligible when participants received negative feedback (Sai et al., 2016, 2014). In the current study, we have successfully replicated our own and others' findings, such that successful deception that led to monetary gains elicited a more positive RewP than failed deception that did not lead to any gains. In addition to RewP, the later feedback-P300 is also sensitive to the outcome valence: success or gains would elicit a larger feedback-P300 than failure or losses (Hajcak et al., 2005; Hajcak et al., 2007; Long, Jiang, & Zhou, 2012; Wu & Zhou, 2009).

Beyond the binary encoding of positive vs. negative feedback, we provide novel evidence that both RewP and feedback-P300 can be significantly influenced by deceptive and honest responses. Given that RewP tracks the motivational significance of on-going events (Gehring & Willoughby, 2002; Yeung, Holroyd & Cohen, 2005), our results suggest that participants are more motivated to evaluate whether or not they succeed when they try to deceive their opponents (Luo et al., 2011; Sun et al., 2015). Similar to ERPs in the earlier decision making stage, whether participants told a false or a truthful statement did not modulate RewP or feedback-P300. Therefore, for ERPs across both the decision making stage and the subsequent outcome evaluation stage, our data provides consistent evidence that it is the deceptive intention that modulates executive control and reward-related brain activities.

Being deceptive inherently elicits conflict between two competing response tendencies: to tell a false statement or to tell a truthful statement. Thus, participants in the interpersonal deception condition were confronted with response uncertainties, as for each trial they needed to decide whether to make a true or false statement. In contrast, participants in the honest control condition simply sent a truthful message without any intention to mislead others. Therefore, observed ERP differences between the deceptive and honest conditions likely reflect domain-general executive control processes that arise when people are confronted with response competition or decision ambiguities. This is also consistent with recent meta-analyses of neuroimaging findings that observed that the brain activations involved in deception are associated with domain-general executive control processes such as working memory, response inhibition and task switching (Christ et al., 2009; Farah et al., 2014).

The current study has general implications for future research into deception and moral psychology. The data reported here suggests that the core characteristic of deception is the intention to mislead others regardless of the veracity of statements. It can be predicted that during highly competitive social interactions such as negotiations, people would employ both true and false accounts to mislead their opponents (see Rogers et al., 2017). Thus, this study will help people understand how intentions and social contexts may modulate the neurocognitive processes underlying interpersonal deception. Furthermore, when people convey truthful messages to mislead others, they may not experience strong negative emotions such as guilt that are typically associated with deception. Thus, future studies could also investigate the emotional consequences associated with each sub-type of deception.

Despite these promising findings, the current study also has limitations that should be addressed in future studies. First, although our sample size is similar to previous studies (e.g., Carrión et al., 2010; Ding et al., 2014; Sip et al., 2010), it is desirable to further investigate this question with larger sample sizes. Second, although we intentionally placed the honest session before the deception session so that participants' honest responses were not influenced by prior deceptive attempts (for a similar task order, see Carrión et al., 2010), this fixed task order might also cause fatigue and might have unexpected influences on ERPs and/or behavior in the later deception session. Thus, a counterbalanced task order combined with a larger sample size is warranted in future studies. Third, although participants were motivated by monetary incentives to deceive their opponents, we did not explicitly assess participants' deception intentions on a trial-by-trial basis. Thus, we cannot guarantee that all truthful responses in the deception session were serving to mislead the opponents, despite the significant different ERPs between truthful responses in the deception and in the honest conditions. To address this concern, future studies can directly ask participants about their deception intentions even when they respond truthfully (see Volz et al., 2015).

In conclusion, the present study examined the neurocognitive processes underlying interpersonal deception. Critically, to maximize ones' payments, participants needed to employ both truthful and false statements to mislead their opponents. During such strategic deceptions, telling truthful and false statements engage in similar executive control and reward-related processes. Thus, it is the deceptive intention, instead of the veracity of statement, which is driving the observed ERP and behavioral differences between honest and deceptive behavior.

Funding

The work was supported by National Natural Science Foundation of China (grant numbers 31070894, 31371041 to G. Fu and 31600875 to L. Sai, and 31700953 to X. Hu and U1736125 to H. Wu) and Hong Kong Research Grants Council Early Career Scheme No. 27610617 to X. Hu.

Declaration of interest

None.

References

Abe, N., & Greene, J. D. (2014). Response to anticipated reward in the nucleus accumbens predicts behavior in an independent test of honesty. *Journal of Neuroscience, 34*(32), 10564–10572.

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.

Bartholow, B. D., Pearson, M. A., Dickter, C. L., Sher, K. J., Fabiani, M., & Gratton, G. (2005). Strategic control and medial frontal negativity: Beyond errors and response conflict. *Psychophysiology, 42*(1), 33–42.

Bereby-Meyer, Y., & Shalvi, S. (2015). Deliberate honesty. *Current Opinion in Psychology, 6*, 195–198.

Bress, J. N., & Hajcak, G. (2013). Self-report and behavioral measures of reward sensitivity predict the feedback negativity. *Psychophysiology, 50*(7), 610–616.

Bruin, K., Wijers, A., & Van Staveren, A. (2001). Response priming in a go/nogo task: Do we have to explain the go/nogo N2 effect in terms of response activation instead of inhibition? *Clinical Neurophysiology, 112*(9), 1660–1671.

Carrión, R. E., Keenan, J. P., & Sebanz, N. (2010). A truth that's told with bad intent: An ERP study of deception. *Cognition, 114*(1), 105–110.

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, F., Wu, S., Wu, H., Wang, C., Jiao, C., & Luo, Y. (2018). Altruistic and self-serving goals modulate behavioral and neural responses in deception. *Social Cognitive and Affective Neuroscience, 13*(1), 63–71.

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.

Dreisbach, G., & Fischer, R. (2015). Conflicts as aversive signals for control adaptation. *Current Directions in Psychological Science, 24*(4), 255–260.

Enriquez-Geppert, S., Konrad, C., Pantev, C., & Huster, R. J. (2010). Conflict and inhibition differentially affect the N200/P300 complex in a combined go/nogo and stop-signal task. *Neuroimage, 51*(2), 877–887.

Farah, M. J., Hutchinson, J. B., Phelps, E. A., & Wagner, A. D. (2014). Functional MRI-based lie detection: Scientific and societal challenges. *Nature Reviews Neuroscience, 15*(2), 123–131.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152–170.

Foti, D., & Hajcak, G. (2009). Depression and reduced sensitivity to non-rewards versus rewards: Evidence from event-related potentials. *Biological Psychology, 81*(1), 1–8.

Ganis, G., Bridges, D., Hsu, C. W., & Schendan, H. E. (2016). Is anterior N2 enhancement a reliable electrophysiological index of concealed information? *NeuroImage, 143*, 152–165.

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

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295*(5563), 2279–2282.

Greene, J. D., & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences of the United States of America, 106*(30), 12506–12511.

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology, 42*(2), 161–170.

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology, 44*(6), 905–912.

Hu, X., Pornpattananangkul, N., & Nusslock, R. (2015). Executive control-and reward-related neural processes associated with the opportunity to engage in voluntary dishonest moral decision making. *Cognitive, Affective, & Behavioral Neuroscience, 15*(2), 475–491.

Hu, X., Wu, H., & Fu, G. (2011). Temporal course of executive control when lying about self-and other-referential information: An ERP study. *Brain Research, 1369*, 149–157.

Hu, X., Pornpattananangkul, N., & Rosenfeld, J. P. (2013). N200 and P300 as orthogonal and integrable indicators of distinct awareness and recognition processes in memory detection. *Psychophysiology, 50*(5), 454–464.

Huster, R., Westerhausen, R., Pantev, C., & Konrad, C. (2010). The role of the cingulate cortex as neural generator of the N200 and P300 in a tactile response inhibition task. *Human Brain Mapping, 31*(8), 1260–1271.

Johnson, R. (1988). The amplitude of the P300 component of the event-related potential: Review and synthesis. *Advances in psychophysiology, 3*, 69–137.

Johnson, R., Henkell, H., Simon, E., & Zhu, J. (2008). The self in conflict: The role of executive processes during truthful and deceptive responses about attitudes. *Neuroimage, 39*(1), 469–482.

Johnson, R., Barnhardt, J., & Zhu, J. (2003). The deceptive response: Effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. *Biological Psychology, 64*(3), 217–253.

Johnson, R., Barnhardt, J., & Zhu, J. (2004). The contribution of executive processes to deceptive responding. *Neuropsychologia, 42*(7), 878–901.

Johnson, R., Barnhardt, J., & Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: An event-related brain potential study. *Cognitive Brain Research, 24*(3), 386–404.

Kropotov, J. D., Ponomarev, V. A., Hollup, S., & Mueller, A. (2011). Dissociating action inhibition, conflict monitoring and sensory mismatch into independent components of event related potentials in GO/NOGO task. *Neuroimage, 57*(2), 565–575.

Lee, T. M. C., Liu, H. L., Tan, L. H., Chan, C. C. H., Mahankali, S., Feng, C. M., ... Gao, J. H. (2002). Lie detection by functional magnetic resonance imaging. *Human Brain Mapping, 15*(3), 157–164.

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.

Long, Y., Jiang, X., & Zhou, X. (2012). To believe or not to believe: Trust choice modulates brain responses in outcome evaluation. *Neuroscience, 200*, 50–58.

Luo, Y. J., Sun, S. Y., Mai, X. Q., Gu, R. L., & Zhang, H. J. (2011). Outcome evaluation in decision making: ERP studies. In S. Han, & E. Pöppel (Eds.). *Culture and neural frames of cognition and communication* (pp. 249–285). Berlin, Heidelberg: Springer Berlin Heidelberg.

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience, 9*(6), 788–798.

Proudfit, G. H. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology, 52*(4), 449–459.

Rogers, T., Zeckhauser, R., Gino, F., Norton, M. I., & Schweitzer, M. E. (2017). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of Personality and Social Psychology, 112*(3), 456–473.

Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J., & Winograd, M. R. (2013). Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology, 90*(2), 118–134.

Sai, L., Lin, X., Hu, X., & Fu, G. (2014). Detecting concealed information using feedback related event-related brain potentials. *Brain & Cognition, 90*(4), 142–150.

Sai, L., Lin, X., Rosenfeld, J. P., Sang, B., Hu, X., & Fu, G. (2016). Novel, ERP-based, concealed Information detection: Combining recognition-based and feedback-evoked ERPs. *Biological Psychology, 114*, 13–22.

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., & Kuboki, T. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport, 16*(4), 407–411.

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.

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.

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.

Suchotzi, K., Crombez, G., Smulders, F. T. Y., Meijer, E., & Verschuere, B. (2015). The cognitive mechanisms underlying deception: An event-related potential study. *International Journal of Psychophysiology, 95*(3), 395–405.

Sun, D., Chan, C. C., Hu, Y., Wang, Z., & Lee, T. M. (2015). Neural correlates of outcome processing post dishonest choice: An fMRI and ERP study. *Neuropsychologia, 68*, 148–157.

Volz, K. G., Vogeley, K., Tittgemeyer, M., von Cramon, D. Y., & Sutter, M. (2015). The neural basis of deception in strategic interactions. *Frontiers in Behavioral Neuroscience, 9*(27).

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities.* Wiley.com.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences, 10*(4), 141–142.

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews, 36*(8), 1870–1884.

Wu, H., Hu, X., & Fu, G. (2009). Does willingness affect the N2–P3 effect of deceptive and honest responses? *Neuroscience Letters, 467*(2), 63–66.

Wu, Y., & Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Research, 1286*, 114–122.

Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex, 15*(5), 535–544.

Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *The Journal of Neuroscience, 24*(28), 6258–6264.

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.