

Altruistic and self-serving goals modulate behavioral and neural responses in deception

Fang Cui,^{1,2} Song Wu,¹ Haiyan Wu,^{3,4} Chengyao Wang,¹ Can Jiao,¹ and Yuejia Luo^{1,2,5}

¹College of Psychology and Sociology, ²Shenzhen Key Laboratory of Affective and Social Cognitive Science, Shenzhen University, Shenzhen, China, ³CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, ⁴Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, and ⁵Shenzhen Institute of Neuroscience, Shenzhen, China

Correspondence should be addressed to Fang Cui, College of Psychology and Sociology, Shenzhen University, Nanhai Ave 3688, Shenzhen, Guangdong 518060, China. E-mail: cuifang0826@gmail.com

Abstract

People tell lies not only for their own self-interests but sometimes also to help others. Little is known about the ways in which different types of goals modulate behaviors and neural responses in deception. The present study investigated the neural processes associated with spontaneous deception that occurs with altruistic reasons (i.e. the money would be donated to charity), self-serving reasons (i.e. the participant receives all of the money) and mixed goals (i.e. the money would be equally split between the participant and the charity). Altruistic motivation for deception reduced the intensity of moral conflict and the subsequent mental cost of resolving this conflict, reflected by a smaller N2–P3 effect in the purely altruistic condition. When making decisions about whether to lie, self-interest was a stronger motivator than others' interests, and the participants tended to lie more for themselves than for others. When the lie could be mutually beneficial for both of the self and others, the participants tended to lie even when they knew that they could be easily caught, but they actually lied for their own self-interest rather than for altruistic reasons. These findings shed light on the neural basis of 'good lies' and decision-making in mutually beneficial situations.

Key words: altruistic; self-serving; deception; moral conflict; ERPs

Introduction

Deception occurs when one person attempts to convince another person to accept as true what the prevaricator knows is untrue. Genuinely dishonest/honest decisions usually involve a trade-off between the expected benefits and costs of the honest/dishonest behavior (Becker, 1968; Allingham and andmo, 1972). Previous functional magnetic resonance imaging (fMRI) studies have consistently reported that deception elicits the activation of a network in the prefrontal cortex (PFC), including the dorsolateral PFC, the ventrolateral PFC, the medial frontal cortex and the anterior cingulate cortex (ACC) (for a meta-

analysis, see Christ *et al.*, 2009). These brain regions (i.e. the PFC and surrounding regions) have been heavily implicated in executive control-related tasks (Chudasama, 2011; Kesner and Churchwell, 2011; Koechlin, 2016). Event-related potential (ERP) studies have found that both instructed and voluntary deception elicited an increase in the N2 component and a decrease in the P3 component. Specifically, the enhanced N2 component reflected conflict monitoring and the decreased P3 component reflected the involvement of executive control demands (Christ *et al.*, 2009; Greene and Paxton, 2009; Wu *et al.*, 2009; Baker and Holroyd, 2011; Ito *et al.*, 2012; Hu *et al.*, 2015; Sun *et al.*, 2015,

Received: 20 May 2017; Revised: 6 November 2017; Accepted: 11 November 2017

© The Author (2017). Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

2016). These findings convergently indicate that the automatic and default response is honesty. To be deceptive, one needs to suppress the automatic, prepotent honest response and generate a dishonest one. Therefore, more cognitive resources and more effort need to be expended to resolve the moral conflict in deceptive responding (Greene and Paxton, 2009).

In most studies about deception, the experimenter designs a situation in which the participants can decide to lie for monetary rewards for themselves (i.e. self-serving deception) (Greene and Paxton, 2009; Hu et al., 2011, 2015). However, people tell lies not only for their own self-interests but sometimes also to help others. For example, a previous study found that ~25% of lies in daily life are told for the benefits of others (DePaulo et al., 1996). Thus, deception appears to have at least two basic types of motivations: selfish motivation (i.e. to gain a benefit or avoid punishment for the self) and altruistic motivation (i.e. to lie for another's benefits) (Caspi and Gorsky, 2006; Wu et al., 2011). Self-serving motivation increases the immoral nature of deception and triggers intense moral conflict (Hayashi et al., 2010; Parkinson et al., 2011; Abe et al., 2014). In contrast, altruistically motivated deceptive behaviors are correlated with less moral conflicts because they are more acceptable both socially and morally (Lewis et al., 2012). For example, previous studies found that deception that was oriented toward the benefit of others was more likely to be perceived as moral rather than immoral (Wu et al., 2016). The so-called 'Robin Hood effect' (i.e. steal from the rich and give to the poor) refers to individuals use the money they gain through immoral ways, such as deception, to help people who are in need (Poddar et al., 2012). Studies have found that the objective of deception modulates the evaluation and moral judgment of these deceptive behaviors. Subjects with other-oriented and well-intended behaviors are typically rated less negatively and considered more acceptable, compared with deceptive behaviors with a selfish objective (Brody, 1982; Lindskold and Han, 1986; Rigoulot et al., 2014).

However, little is known about the ways in which different types of goals can modulate behaviors and neural responses in deception. A previous behavioral study reported that when individuals were instructed to win money in a 'die-under-the-cup' game, they over-reported their accuracy to gain more money when the money would be donated to charity compared with direct personal gain (Lewis et al., 2012). A very recent fMRI study reported similar behavioral results, in which participants lied more when the goal was altruistic compared with a self-serving goal. The modulation of goals was found in the anterior insula with reduced activity under altruistic goal comparing to self-serving goal (Yin et al., 2017). These results suggest that a 'good' objective may justify dishonest behavior, reduce the moral conflict and promote deceptive behaviors accordingly. Notably, only psychology and economics students participated in the study by Lewis et al. (2012). Students who major in psychology may more easily infer the true intention of the study, and students who major in economics may be more sensitive to mathematics and probability theory. In fact, these authors found that economics students were more likely to lie than psychology students (Lewis et al., 2012). Therefore, these findings are difficult to generalize to the general population. In the study by Yin et al. (2017), the participants were shown two payoff options and one of the options was selected by the computer. The participants' task was to send this message (i.e. the option that the computer chose) to a receiver. If the receiver believed this message then the option that was sent by the participant would be implemented. In some of the trials, when the computer chose the option with a lower payoff to the participant/charity the

participant would choose to send a false message in order to receive a higher payoff (Yin et al., 2017). However, the participants knew that both the true information and their choices were recorded by the computer and observed by the experimenter, which may have strongly biased the participants' behavior by making them consider their own self-image and reputation in another person's eyes. Therefore, more valid studies need to be conducted to determine the ways in which motivation influences deceptive behavior when the participants believe that their lies will not be detected by others. Additionally, the studies by Yin et al. (2017) and Lewis et al. (2012) examined only purely altruistic (PA) and purely selfish goals. In real life, mutually beneficial goals that benefit both the liars and others also exist. This intermediate situation has not been investigated yet.

Therefore, in the present study, we recruited students from various college majors and employed a paradigm that enabled the participants to engage in deceptive behaviors privately and voluntarily. The participants were instructed to perform a coin-flip guessing task in which they could win money if they correctly predicted the outcome of a coin-flip. In half of the trials, the participants recorded their prediction by a button-pressing before the coin-flip (they did not have the opportunity to lie without being caught; NoOp condition) while in the other half of the trials; they only needed to make a mental prediction (they did have the opportunity to lie without being caught; Op condition). The Op condition allowed the participants to freely engage in voluntary deception by over-reporting the accuracy of their prediction without being detected. Importantly, the participants were informed that there were three ways of distributing the money they won in each trial: (1) the money would be given to the participant him/herself [*purely self-serving (PS)*]; (2) donated to a charity project (PA) or (3) equally split between the participant and the charity (*mixed*). Electroencephalograms (EEGs) were recorded during the task, and ERPs were locked to the phase of the trial in which the participants first saw the outcome of the coin-flip. This was the moment at which the participant would know whether they won according to their prior prediction. Meanwhile, this outcome phase was also the moment at which the participants evaluated their outcomes and made a decision about whether to lie.

Behaviorally, the present study tested two hypotheses. First, the reported accuracy would be higher in the trials in which the participants made their prediction privately (i.e. mental prediction) than in the trials in which they made their prediction publicly (i.e. computer-recorded prediction). According to the results of previous studies that used similar paradigms, the participants were expected to be honest in the NoOp trials when they knew that they would be caught if they lied (the reported accuracy rate would be ~50%, which would reflect a random guessing). Conversely, the participants would be more deceptive in the Op trials in which their lies would not be detected (the reported accuracy rate would be significantly greater than 50%, indicating the existence of dishonest responses) (Hu et al., 2015). Second, we predicted that the reported accuracy would be highest in the trials in which the charity would receive all of the money, which would be consistent with the previous findings (Lewis et al., 2012; Yin et al., 2017).

At the neural level, we mainly focused on two ERP components: frontal N2 and parietal P3. The frontal N2 is a negative component that peaks around 200–400 ms after stimulus onset and has been suggested to represent detection and monitoring of conflict (Folstein and Van, 2008). N2 has been reported to be enhanced when participants are deceptive than when they are

honest (Hu et al., 2011; Ji et al., 2012). An ERP study by Hu et al. (2015) used a similar paradigm as in the present study, in which Op trials elicited a larger N2 component than NoOp trials, suggesting that the opportunity to deceive recruited more executive control processes, such as conflict monitoring (Hu et al., 2015). We hypothesized that altruistic motivation would reduce the moral conflict between honest and dishonest responses, and the difference in N2 between the two conditions (Op vs NoOp) would decrease when the charity received the gains. Specifically, the difference in N2 (Op minus NoOp) would be smallest when all of the money went to charity and largest when the participants received all of the money. The parietal P3 component has been suggested to be attenuated when experimental manipulations increase executive control demands (Johnson, Jr. et al., 2003, 2004, 2008). In the study by Hu et al. (2015), the participants had a smaller P3 amplitude in Op trials than in NoOp trials (Hu et al., 2015). A stronger moral conflict that was evoked in the time window of N2 was associated with greater executive functions resources that were required to resolve this conflict during the time window of P3. Therefore, we predicted a similar trend in P3 as in N2.

Materials and methods

Ethics statement

All research procedures were approved by the Medical Ethical Committee of College of Psychology and Sociology in Shenzhen University according to the Declaration of Helsinki. All participants provided written informed consent after they fully understand the study.

Participants

Thirty-five right-handed participants recruited at Shenzhen University participated in the study. Participants were screened for history of neurological disorders, brain injury or developmental disabilities. All of them had a normal or corrected-to-normal vision. All participants' data were included in the analysis of behavioral performance [11 females, age: 22.0 ± 0.33 years (mean \pm s.e.)]. One participants' EEG data were rejected due to a technical problem (recording failed). Therefore, 34 participants' data were included in the analysis of EEG data (11 females, age: 21.98 ± 0.33 years).

Experimental procedure

The stimulus display and behavioral data acquisition were performed using E-Prime 2.0 software (Psychology Software Tools, Inc; Pittsburgh, USA). During the task, the participants sat comfortably in an electrically shielded room approximately 90 cm from a 15-inch color computer screen. After the electrode application, participants completed a coin-guess task [adapted from Greene and Paxton (2009) and Hu et al. (2015)].

Participants were informed that they were to predict the outcomes of the coin flips and they would earn 5 RMB for each correct prediction and earn 0 RMB for an incorrect one. However, the distribution of the money they won was different in each block. The task consisted of 3 blocks and each block contains 108 trials. In one of the three blocks, the participants would gain all of the money they won (PS); in the second of the three blocks, the money would be donated to a specific charity project chosen by the participant him/herself before the task (PA) (see [Supplementary Material S1](#)). In the last one block, the money would be split equally between the participant and the charity

project (mixed). The order of the blocks was counterbalanced among participants. In each block, half of the trials were no-opportunity-to-deceive-without-being-caught trials (NoOp). In these NoOp trials, participants recorded their predictions by button-pressing. In each NoOp trials, after a 1000 ms fixation, the Chinese Words 'Head' and 'Tail' would appear on each side of the screen for 3000 ms. The participants were instructed to record their prediction about the upcoming coin flip when the word 'Press' appeared on the screen by pressing a button labeled 'H' if they predicted 'head' and a button labeled 'T' if they predicted 'tail' for that particular trial. The buttons labeled were the 'F' and 'J' keys on a keyboard and the labels were counterbalanced among participants. They have maximum 2000 ms to response. After pressing, the outcome of the coin flip would appear for 2000 ms. Then, the question 'Correct?' appeared on the screen next, the participants were instructed to indicate whether their prediction was correct by pressing the Y key for 'Yes' or the N key for 'No'. After the response, a screen would show how much money the participant earned in this trial for 1000 ms. The other half of the trials was opportunity-to-deceive-without-being-caught trials (Op). In these Op trials, participants made their predictions privately. The structure of the Op trials was the same as the NoOp trials. The only difference was instead of the 'Head' and 'Tail' in the NoOp trials, the word 'Random' would appear on both side of the screen for 3000 ms. When the word 'Random' appeared, the participants would make a prediction mentally and press the 'R' key to indicate they finished their prediction (Figure 1). Therefore, in the NoOp trials, if the participant lied they could be detected explicitly since their predictions were recorded by the computer; while in the Op trials, the participants could lie without being caught since their predictions were only recorded in their own minds. Given our two factorial within-subject 3 (Beneficiaries: PS, PA and mixed) \times 2 (Opportunity: Op and NoOp) design, six conditions were generated accordingly: PS_Op, PS_NoOp, PA_Op, PA_NoOp, mixed_Op and mixed_NoOp. The whole task consisted of 324 trials in total and lasted for approximately 1 hour.

Data recording and analysis

Accuracies and reaction times. The accuracies were calculated based on the participants' reported results. Notable, since the accuracy of the NoOp condition was self-reported participants still can choose to deceive even they can be detected explicitly. Therefore, we also calculated their real accuracy in the NoOp trials based on their true responses (Real_NoOp). Then, we run a repeated measures ANOVA [3 (Beneficiaries: PS, PA and mixed) \times 2(Opportunity: Op and NoOp)] and three paired t-test to compare the reported accuracies and real accuracies in the NoOp condition with different beneficiaries. For the reaction times, we run a repeated measures ANOVA [(3(Beneficiaries: PS, PA and mixed) \times 2(Opportunity: Op and NoOp)]. Degrees of freedom for F-ratios were corrected according to the Sphericity assumed method. Statistical differences were considered significant at $P < 0.05$.

ERP recordings and analyses. EEG data were recorded from a 64-electrode scalp cap using the 10–20 system with a sampling frequency of 1000 Hz (Brain Products, Munich, Germany). The electrode at the right mastoid was used as the reference during recording while the electrode on the medial-frontal aspect was used as a ground electrode. Two electrodes were used to measure the electrooculogram (EOG). EEG and EOG activity was amplified at 0.01–100 Hz band-passes and downsampled at

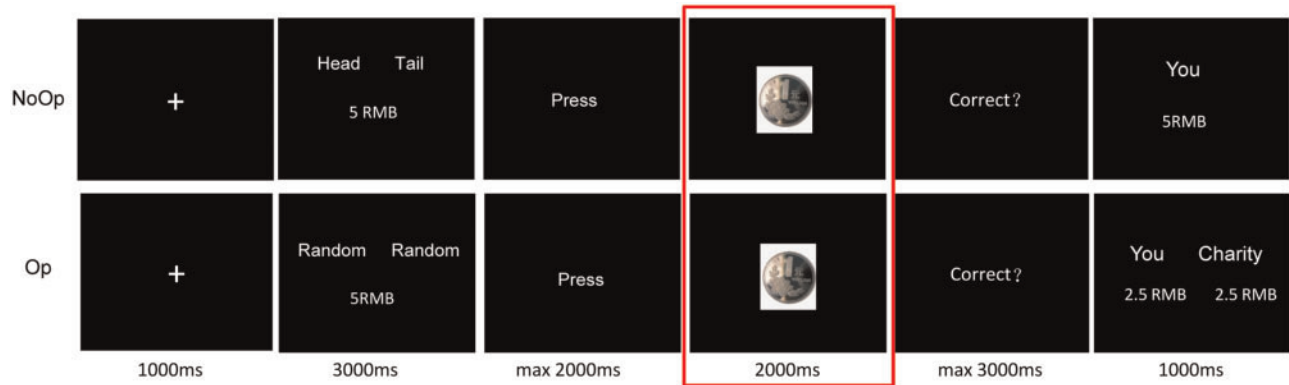


Fig. 1. Task sequence. Participants do a coin-flip guessing task. In the NoOp trials, participants were instructed to record their prediction about the upcoming coin flip when the word 'Press' appeared on the screen by pressing a button labeled 'H' if they predicted 'head' and a button labeled 'T' if they predicted 'tail'. In the Op trials, participants made their predictions privately: instead of the 'Head' and 'Tail' in the NoOp trials, the word 'Random' would appear and the participants would press the 'R' key to indicate they finished their prediction. The ERPs were locked to the outcome cue screen (marked with a red box).

500 Hz. All electrode impedances were maintained below 5 k Ω . EEG data were preprocessed and analyzed using BrainVision Analyzer 2.0.1 (Brain Products GmbH, Germany). EEG data were rereferenced to the common average. The signal passed through a 0.01–30 Hz band-pass filter. EOG artifacts were corrected using an independent component analysis (Jung et al., 2001). Segmented EEG data were stimulus-locked to the onset of the outcome of the coin flip. The ERP epochs were trimmed (from –200 ms to 1000 ms) and the prestimulus baseline (–200 ms to 0 ms) were corrected. Epochs with amplitude values exceeding +75 μ V at any electrode were excluded from the average.

Further statistical analysis was conducted in IBM SPSS Statistics 22 (IBM Corp., Armonk, NY, USA). Based on the literature and visual inspection of the grand-averaged ERPs, the N2 component between 200 and 400 ms and the P3 component between 300 and 600 ms were measured. Different sets of electrodes for N2 and P3 were chosen. The N2 was distributed in the frontal region; thus, F4, Fz, F3, FC3, FCz and FC4 were selected. P3 was distributed in the central-parietal region; thus, CP3, CPz, CP4, P3, Pz and P4 were selected. Mean amplitudes were obtained from waveform averaged for all selected electrode sites. Repeated measures ANOVA [3 (Beneficiaries: PS vs PA vs mixed) \times 2 (Opportunity: Op vs NoOp)] were performed for each component. Degrees of freedom for F-ratios were corrected according to the Greenhouse–Geisser method. Statistical differences were considered significant at $P < 0.05$; *post hoc* comparisons were Bonferroni-corrected at $P < 0.05$.

It was noticed that there is an imbalance of gender in our sample (11 females vs 23 males). Given that there are studies mentioned the gender differences in dishonest behaviors (Marchewka et al., 2012), we did analyses for both of the behavioral and ERP data with 'gender' as a between-subject factor to determine the effect of gender. The results of these analyses suggested that this imbalance in gender does not influence our main findings (for details of the results, see [Supplementary Material S2](#)).

Results

Behaviors

The repeated measure ANOVA revealed that the main effect of Beneficiary was significant [$F(2, 68) = 13.154, P < 0.001, \eta_p^2 = 0.279$] such that participants reported highest accuracy in the PS condition and lowest accuracy in the PA condition (PS:

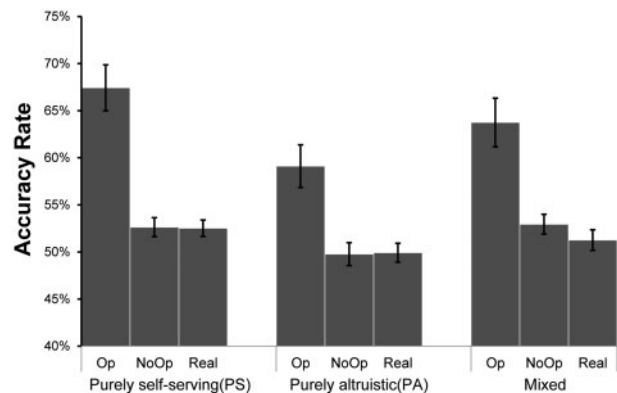


Fig. 2. Accuracy rate per group per condition (Real: real accuracy in the NoOp condition; NoOp: reported accuracy in the NoOp condition; Op: reported accuracy in the Op condition) (mean \pm s.e.).

60.5 \pm 1.4%; PA: 54.5 \pm 1.4%; mixed: 58.4 \pm 1.5%; PS vs PA, $P < 0.001$, Cohen's $d = 0.493$; mixed vs PA, $P = 0.026$, Cohen's $d = 0.317$; PS vs mixed: $P = 0.068$, Cohen's $d = 0.169$). The main effect of Opportunity was significant [$F(1, 34) = 29.723, P < 0.001, \eta_p^2 = 0.466$] such that participants reported higher accuracy in the Op trials than in the NoOp trials (Op: 63.4 \pm 2.2%; NoOp: 52.1 \pm 0.7%). The interaction of Beneficiary \times Opportunity was insignificant [$F(2, 68) = 2.009, P = 0.153, \eta_p^2 = 0.056$].

The paired t-tests revealed that in the mixed conditions, participants reported significantly higher accuracy than the real accuracy in the NoOp trials [$t(34) = 2.230, P = 0.032$, Cohen's $d = 0.266$, NoOp: 52.93 \pm 6.3%; Real_NoOp: 51.26 \pm 6.3%]. In the PS and PA conditions, the reported accuracy and the real accuracy were comparable [PS: $t(34) = 1.705, P = 0.097$, Cohen's $d = 0.198$, NoOp: 53.57 \pm 5.9%; Real_NoOp: 52.46 \pm 5.1%; PA: $t(34) = -0.039, P = 0.969$, Cohen's $d = 0.022$, NoOp: 49.90 \pm 7.2%; Real_NoOp: 49.95 \pm 5.9%] (Figure 2).

For the RTs, the main effect of Beneficiary [$F(2, 68) = 0.017, P = 0.975, \eta_p^2 < 0.001$], Opportunity [$F(1, 34) = 0.220, P = 0.642, \eta_p^2 = 0.006$] and the interaction of Beneficiary \times Opportunity [$F(2, 68) = 0.086, P = 0.910, \eta_p^2 = 0.003$] was insignificant.

Event-related potentials

N2 (200–400 ms). The main effect of Beneficiary was significant [$F(2, 66) = 6.553, P = 0.003, \eta_p^2 = 0.166$] such that PS and mixed

conditions elicited significantly more positive amplitudes than PA condition (PS: $1.124 \pm 0.411 \mu\text{V}$; PA: $0.093 \pm 0.484 \mu\text{V}$; mixed: $0.944 \pm 0.512 \mu\text{V}$; PS vs PA, $P = 0.005$; mixed vs PA, $P = 0.016$; PS vs mixed: $P > 0.999$). The main effect of Opportunity was significant [$F(1, 33) = 6.744$, $P = 0.014$, $\eta_p^2 = 0.170$] such that Op trials elicited significantly more negative N2 amplitudes than NoOp trials (Op: $0.352 \pm 0.450 \mu\text{V}$; NoOp: $1.088 \pm 0.468 \mu\text{V}$). A significant interaction of Beneficiary \times Opportunity [$F(2, 66) = 4.210$, $P = 0.022$, $\eta_p^2 = 0.113$] was observed. Pairwise comparison revealed that in the PS condition, Op trials elicited a larger N2 than NoOp trials (Op: $0.494 \pm 0.493 \mu\text{V}$; NoOp: $1.753 \pm 0.462 \mu\text{V}$, $P = 0.002$; Cohen's $d = 0.479$); in the mixed condition the Op trials elicited a more negative N2 than NoOp trials (Op: $0.516 \pm 0.551 \mu\text{V}$; NoOp: $1.371 \pm 0.545 \mu\text{V}$, $P = 0.037$; Cohen's $d = 0.267$); in the PA condition the difference between the Op and NoOp trials was insignificant (Op: $0.045 \pm 0.510 \mu\text{V}$; NoOp: $0.141 \pm 0.516 \mu\text{V}$, $P = 0.779$; Cohen's $d = 0.032$) (Figure 3A and B).

P3 (300–600 ms). The main effect of Beneficiary was significant [$F(2, 66) = 6.816$, $P = 0.002$, $\eta_p^2 = 0.171$] such that PS and mixed conditions elicited significantly larger amplitudes than PA condition (PS: $6.065 \pm 0.543 \mu\text{V}$; PA: $5.034 \pm 0.467 \mu\text{V}$; mixed: $5.994 \pm 0.485 \mu\text{V}$; PS vs PA, $P = 0.005$; mixed vs PA, $P = 0.009$; PS vs mixed: $P > 0.999$). The main effect of Opportunity was significant [$F(1, 33) = 5.001$, $P = 0.032$, $\eta_p^2 = 0.132$] such the Op trials elicited that significantly smaller amplitudes than the NoOp trials (Op: $5.435 \pm 0.458 \mu\text{V}$; NoOp: $5.961 \pm 0.502 \mu\text{V}$). A marginally significant interaction of Beneficiary \times Opportunity [$F(2, 66) = 3.269$, $P = 0.050$, $\eta_p^2 = 0.090$] was observed. Pairwise comparison revealed that in the PS condition the Op trials elicited a smaller

P3 than the NoOp trials (Op: $5.627 \pm 0.546 \mu\text{V}$; NoOp: $6.503 \pm 0.586 \mu\text{V}$, $P = 0.010$; Cohen's $d = 0.265$); in the mixed condition the Op trials elicited a smaller P3 than the NoOp trials (Op: $5.623 \pm 0.495 \mu\text{V}$; NoOp: $6.365 \pm 0.516 \mu\text{V}$, $P = 0.013$; Cohen's $d = 0.252$); in the PA condition the P3 amplitude difference between the Op and NoOp trials was insignificant (Op: $5.055 \pm 0.500 \mu\text{V}$; NoOp: $5.014 \pm 0.503 \mu\text{V}$, $P = 0.912$; Cohen's $d = 0.013$) (Figure 3C and D).

Correlation between ERP and behavioral data. We did Pearson's correlation analysis between the difference of ERP components and the difference of accuracy of the Op minus NoOp for each kind of beneficiary. On N2 component ($N2_{Op-NoOp}$), we reported a marginally significant negative correlation between the mean amplitude and accuracy on PS_Op > PS_NoOp ($r = -0.334$, $P = 0.053$) and a negative correlation between mixed_Op > mixed_NoOp ($r = -0.452$, $P = 0.007$). On P3 component ($P3_{Op-NoOp}$), we found a significant negative correlation between the mean amplitude and accuracy on PS_Op > PS_NoOp ($r = -0.446$, $P = 0.008$), as well as a significant negative correlation between mixed_Op > mixed_NoOp ($r = -0.582$, $P < 0.001$) (Figure 4).

Discussion

In the present study, we investigated the neural processes that are associated with spontaneous deception with altruistic, self-serving and mixed goals. In the task, the participants were free to make their own honest or dishonest choices to win money for different beneficiaries. Three goals were manipulated: win money for themselves (PS); win money for a pre-selected charity

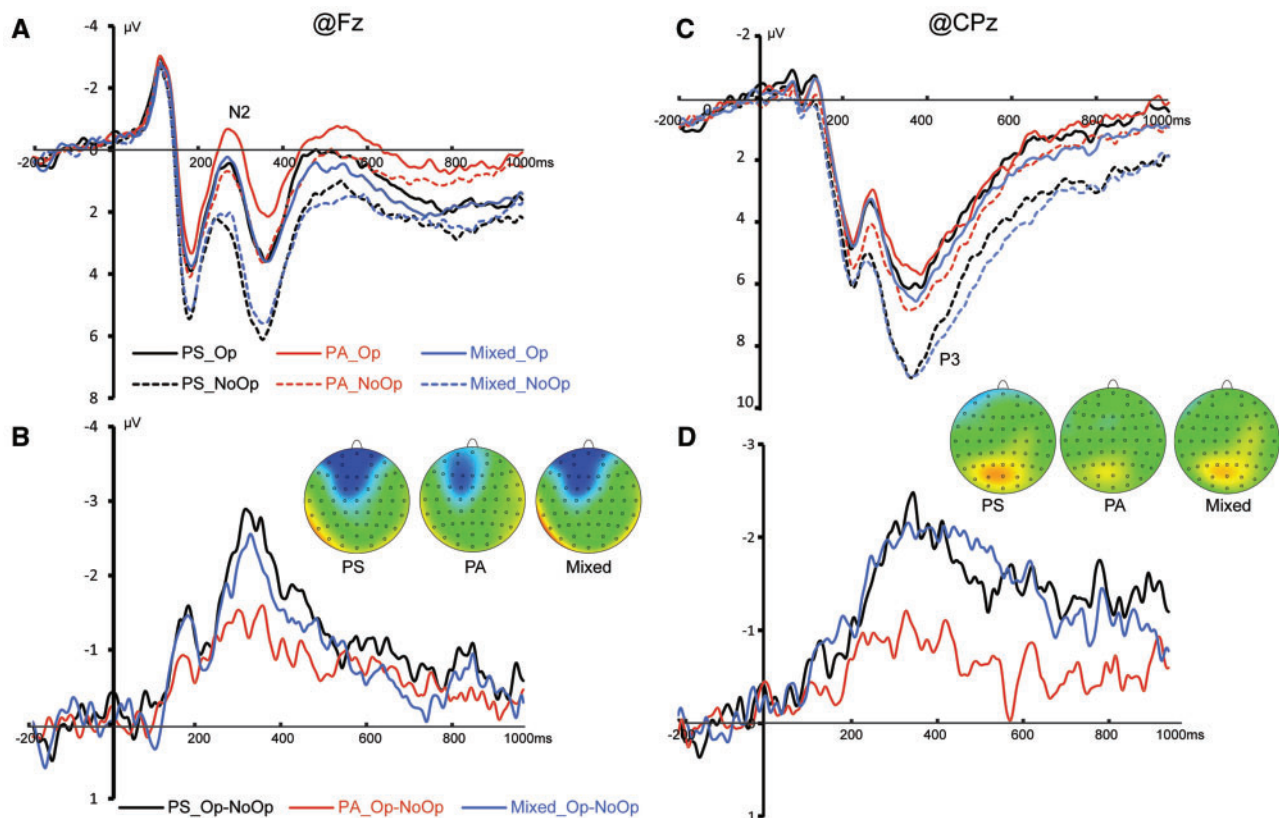


Fig. 3. (A) Grand average of N2 at Fz. (B) The difference wave of Op–NoOp and the topographies of the difference wave between 250 and 300 ms. (C) Grand average of P3 at CPz. (D) The difference wave of Op–NoOp and the topographies of the difference wave between 400 and 450 ms.

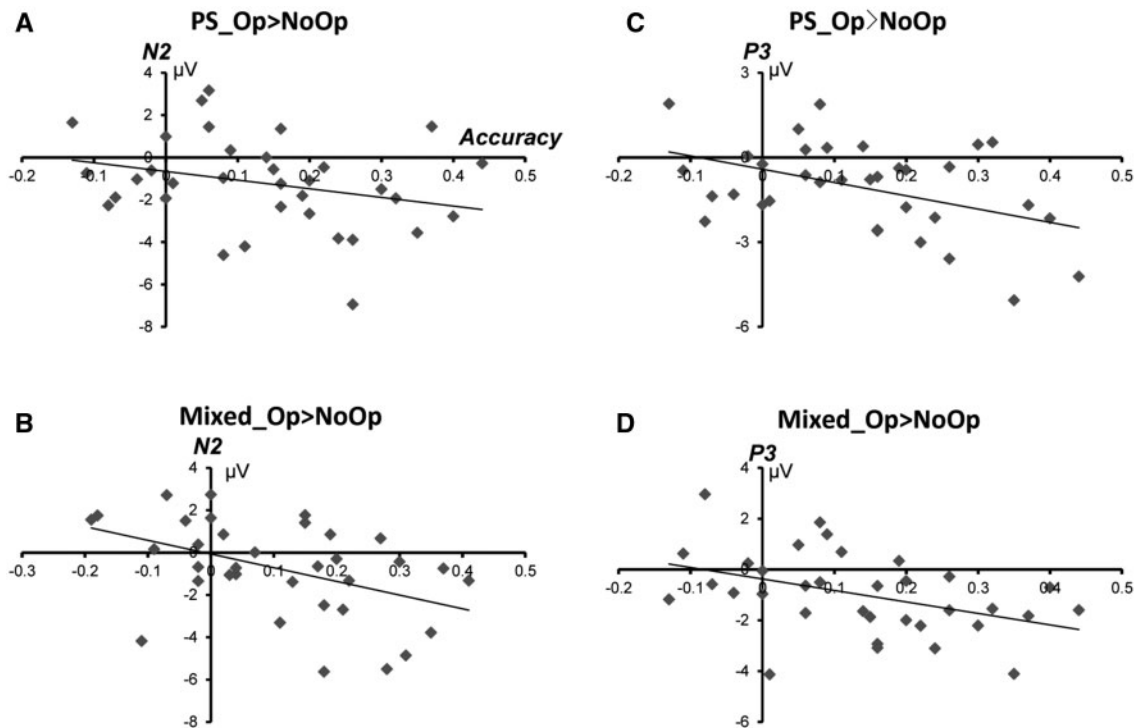


Fig. 4. Correlation between ERP components and accuracy rates. The horizontal axis is the reported accuracy of Op-NoOp(0-1) and the longitudinal axis is the mean amplitude of Op-NoOp (μV). (A) N2_(PS_Op > PS_NoOp) and accuracy_(PS_Op > PS_NoOp); (B) N2_(mixed_Op > mixed_NoOp) and accuracy_(mixed_Op > mixed_NoOp); (C) P3_(PS_Op > PS_NoOp) and accuracy_(PS_Op > PS_NoOp); (D) P3_(mixed_Op > mixed_NoOp) and accuracy_(mixed_Op > mixed_NoOp).

(PA) or win money for both themselves and the charity (mixed: mutually beneficial). Notably, in this paradigm, the participants could make dishonest responses without being explicitly detected or viewed by the experimenter. The authenticity of self-report in each single Op trials could not be directly determined, and we can only use probability statistics to estimate the extent of lying. This paradigm prevented the influence of social factors such as positive reputation or self-image on the participants' behavior. In the context that was created by the paradigm, the pros and cons of being dishonest were simplified. The motivation to lie was to gain more money for oneself, a charity, or both oneself and a charity. Thus, the motivation could be totally self-serving, altruistic, or mixed. Lying is resisted in an effort to avoid negative feelings that are associated with moral conflict and the expenditure of cognitive resources to resolve this conflict. To be dishonest, one needs to inhibit the default, honest response, experience the moral conflict, and execute a deceptive response (Christ et al., 2009; Hu et al., 2011; Debey et al., 2012). The interplay between drive and resistance leads an individual to make honest or dishonest choices. In the present study, when the individuals had the opportunity to deceive without being detected in the Op trials, they experienced elevated conflict monitoring (with a concomitant increase in N2) and expended greater cognitive resources to resolve this conflict (with a concomitant decrease in P3) compared with the NoOp trials in which they did not have the opportunity to deceive. Such an effect on N2 and P3 was significant in the PS and mixed conditions, but not in the PA condition. The participants also reported significantly higher accuracy in the Op trials in the PS condition than in the mixed condition and significantly higher accuracy in the mixed condition than in the PA condition. In the mixed condition, the participants reported significantly higher accuracy in the NoOp trials

than their actual accuracy and there was no significant difference between reported accuracy and actual accuracy in the PS and PA conditions.

Altruistic motivation reduces the moral conflict in deception

The frontal N2 component had been consistently reported to be enhanced when participants need to override one prepotent response in order to executing an alternative response to achieve a goal, reflected the conflict monitoring, cognitive control, and/or response inhibition (Baker and Holroyd, 2011). The N2 component is generated in the ACC, a brain region that signals the occurrence of conflicts in information processing, thereby triggering compensatory adjustments in cognitive control (Bioulac et al., 2005; Botvinick, 2007; Sohn et al., 2007; Braem et al., 2017). ERP and fMRI studies have found that both instructed and voluntary deception elicited enhanced N2/activations of the ACC (Christ et al., 2009; Greene and Paxton, 2009; Ito et al., 2012; Hu et al., 2015; Sun et al., 2015; Sun et al., 2016). To be deceptive, one needs to suppress the automatic, prepotent honest response and generate a dishonest response. Therefore, more cognitive resources and more effort need to be expended to resolve this conflict (Greene and Paxton, 2009). In the Op trials, the participants reported higher accuracy (i.e. significantly higher than the random probability of 50%) than in the NoOp trials, indicating that they lied more in the Op trials. In the PS and mixed conditions, Op trials were associated with a more negative N2, reflecting the relative higher conflict processing than the NoOp trials. Previous studies found that the amplitude of P3 is attenuated when experimental tasks involved executive control demands (Chen et al., 2008; Hu et al., 2012). Notably, the amplitude of P3 was reported to decrease for a deceptive

response than for a truthful response in both of the instructed deception and voluntary deception (Johnson, Jr. et al., 2003, 2005, 2008). Deception is an executive control intensive task which is associated with a substantially higher mental-load. Participants need to recruit higher level executive control processes to resolve this conflict and implement behavioral responses that are incompatible with truthful responses. We observed a significant decrease in P3 in the Op trials compared with the NoOp trials, indicating that more executive resources were engaged in the PS and mixed conditions.

In the PA condition, no significant differences in the N2 or P3 components were found between the Op and the NoOp trials. Is that because the participants did not consider lying in the PA condition? We suggest that the answer is no. Behavioral data suggest that the participants in the PA condition reported significantly higher accuracy in the Op trials than in the NoOp trials. Previous studies compared the neural activities between a helpful honest decision and a helpful dishonest decision and found no significant differences, indicating that altruistic goals might help reduce the moral conflict associated with lying and overcome the psychological barrier of lying (Abe et al., 2014; Rigoulot et al., 2014; Yin et al., 2017). Our results indicated that altruistic goals can weaken the intensity of the moral conflict in deception and reduce the psychological cost of lying. Importantly, the Op trials elicited a significantly larger N2 and smaller P3 than the NoOp trials in both the PS and mixed conditions. Besides, in the correlation analysis we found that for both N2 and P3, the larger the difference of Op–NoOp in the ERP amplitudes, the larger the difference of the reported accuracies. This result may indicate that the more intensive moral conflict (enlarged N2), the more mental efforts expended to resolve this conflict (decreased P3), the more deceptive responses were generated. The similarity in the pattern of N2–P3 effect and correlations may indicate that the influence of altruistic goals on the moral conflict works in an all-or-none fashion, and only the PA goals can reduce moral conflict. With regard to self-interest, the moral conflict triggered by lying cannot be negated.

Self-serving motivation was stronger than altruistic motivation

We found that the rank order of reported accuracy was PS > Mixed > PA. Greater money that the participants received was associated with higher reported accuracy in the Op trials. This finding conflicts with previous studies, in which participants were more willing to lie for charity than for themselves (Lewis et al., 2012; Yin et al., 2017). We propose that this discrepant result might be attributable to the different paradigms and participants samples. In the study by Yin et al. (2017), the participants' choices could be observed and recorded by the experimenter, meaning that they might take in to consideration their own reputation and other's evaluations of them. Therefore, the participants' self-image could be negatively impacted if they lied for their own self-interest. In contrast, if they lied for charity, this altruistic behavior may gain them social rewards, such as a better reputation. Therefore, the altruistic goal may promote deceptive behavior and self-serving goals may suppress deceptive behavior. In the present study, when the goal was PS, the moral conflict appeared to be strongest because of the tradeoff between obtaining the maximum amount of money for self-interest and resistance to doing so due to social pressures. Therefore, the moral conflict was more intense in PS condition compared with the mixed and PA conditions. However, the participants still reported the highest accuracy in the PS condition.

This result indicates that the self-serving motivation is a more powerful driver than altruistic motivation. Although the participants would experience stronger moral conflict and expend more cognitive resources to resolve the conflict, they still chose to lie more for their own self-interest. The motivation to lie was reduced when the self-benefitting gains were reduced (PS > mixed > PA), resulting in a decrease in the reported accuracy in the Op trials. In the PA condition, the moral conflict was minimized. Completely altruistic motivation was not sufficiently enough to drive them to lie as much as they did in the PS condition. The experimental design employed here sought to control the confounding social factors, in which the participants' choices between being honest and dishonest were undetected by the experimenter. This setting may prevent the participants' tendency of 'being altruistic but not being selfish in public' and show the real power of self-interest drive.

Self-serving motivation rather than altruistic motivation promoted dishonest choice in the mixed condition

In the Op trials, the accuracy was calculated based on self-reported results because the true predictions were only known to the participants. In the NoOp trials, we calculated two types of accuracy: reported accuracy and real accuracy based on button-pressing. In the NoOp trials, the participants were expected to not engage in deception because such deception would be explicitly detected by the experimenter. Surprisingly, we found that the participants significantly over-reported their accuracy in the mixed_NoOp condition. Why would they lie in a circumstance in which their lies would be detected? In the mixed condition, deceptive behavior was driven by both self-serving and altruistic motivation. We propose that the altruistic aspect of motivation may compromise the immoral nature of deception. If the participants were caught lying for their own self-interest, then this would be judged as immoral. The participants were faced with this conflict and possible negative social judgments if they were caught. However, in the mixed condition, the participants were not the only one who could benefit from the lying. Half of the money would be donated to charity. Previous studies suggested that when the goal is altruistic, the evaluation and judgment of deceptive behavior is more positive (Talwar et al., 2007; Talwar and Crossman, 2011; Rigoulot et al., 2014; Wu et al., 2016). For the participants, being caught lying for charity was more acceptable because they had a noble reason for doing so. However, if they actually lied because they were being altruistic, then we would expect them to lie more in the PA_NoOp condition. Because in this condition, they could be deceptive without any moral burden and win more money to donate. Out of our prediction, they did not lie in the PA_NoOp condition. Therefore, we propose that the main driver of the deceptive behavior in the mixed_NoOp condition was self-oriented. The participants were driven by the desire to win more money for themselves, and they happened to have a good reason for this moral compromise, which released them from being judged as immoral. Moreover, the ERP results in the PS and mixed conditions indicated similar patterns of effects in N2 and P3. In the PA condition, no such pattern was observed; implying that the main driver of deceptive behavior in the PS and mixed conditions was self-oriented. Otherwise, if deceptive behavior is mainly driven by altruistic motivation, then we would have expected to observe similar patterns in mixed and PA conditions.

In conclusion, the present study examined the ways in which different types of goals can modulate behaviors and neural

responses of deceptive behavior. Altruistic goals reduced the intensity of moral conflict between honest and dishonest responses and lowered the mental cost of resolving this conflict, reflected by a diminished N2-P3 effect in the PA condition. When making decisions about whether to lie self-interest is a stronger motivator than other's interest, in which people exhibit more of a tendency to lie for themselves than for charity. When the lie could be mutually beneficial to both the self and others, people tended to lie even when they could be easily caught, but they actually lied for their own self-interest rather than for altruistic reasons. Our findings shed light on the neural basis of 'good lies' and moral decision-making in mutually beneficial situations.

Funding

This study was funded the National Natural Science Foundation of China no. 31500877 to F.C. and no. 31400963 to H.Y.W.; Humanity and Social Science Youth Foundation of Ministry of Education of China (16YJC190021) to S.W. and the Outstanding Young Faculty Award of Guangdong province (YQ2014149) to C.J.

Supplementary data

Supplementary data are available at SCAN online.

Conflict of interest. None declared.

References

- Abe, N., Fujii, T., Ito, A., et al. (2014). The neural basis of dishonest decisions that serve to harm or help the target. *Brain and Cognition*, *90*, 41–9.
- Allingham, M.G., andmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, *1*(3–4), 323–38.
- Baker, T.E., Holroyd, C.B. (2011). Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biological Psychology*, *87*(1), 25–34.
- Becker, G.S. (1968). Crime and punishment: an economic approach. *Journal of Political Economy*, *76*(2), 169–217.
- Bioulac, B., Michelet, T., Guehl, D., Aouizerate, B., Burbaud, P. (2005). The anterior cingulate cortex in error detection and conflict monitoring. Unitary neuronal activity in monkeys. *Bulletin de L'Académie Nationale de Médecine*, *189*, 1529–38.
- Botvinick, M.M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(4), 356–66.
- Braem, S., King, J.A., Korb, F.M., Krebs, R.M., Notebaert, W., Egner, T. (2017). The role of anterior cingulate cortex in the affective evaluation of conflict. *Journal of Cognitive Neuroscience*, *29*(1), 137–49.
- Brody, H. (1982). The lie that heals: the ethics of giving placebos. *Annals of Internal Medicine*, *97*(1), 112–8.
- Caspi, A., Gorsky, P. (2006). Online deception: prevalence, motivation, and emotion. *Cyberpsychology & Behavior*, *9*(1), 54–9.
- Chen, A., Xu, P., Wang, Q., et al. (2008). The timing of cognitive control in partially incongruent categorization. *Human Brain Mapping*, *29*(9), 1028–39.
- 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), 1557–66.
- Chudasama, Y. (2011). Animal models of prefrontal-executive function. *Behavioral Neuroscience*, *125*, 327–43.
- Debey, E., Verschuere, B., Crombez, G. (2012). Lying and executive control: an experimental investigation using ego depletion and goal neglect. *Acta Psychol. (Amst)*, *140*, 133–41.
- DePaulo, B.M., Kashy, D.A., Kirkendol, S.E., Wyer, M.M., Epstein, J.A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979–95.
- Folstein, J.R., Van, P.C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*, *45*(1), 152–70.
- 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), 12506–11.
- Hayashi, A., Abe, N., Ueno, A., et al. (2010). Neural correlates of forgiveness for moral transgressions involving deception. *Brain Research*, *1332*, 90–9.
- Hu, X., Hegeman, D., Landry, E., Rosenfeld, J.P. (2012). Increasing the number of irrelevant stimuli increases ability to detect countermeasures to the P300-based Complex Trial Protocol for concealed information detection. *Psychophysiology*, *49*(1), 85–95.
- Hu, X., Pompattananangkul, 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), 475–91.
- 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*, 149–57.
- Ito, A., Abe, N., Fujii, T., et al. (2012). The contribution of the dorsolateral prefrontal cortex to the preparation for deception and truth-telling. *Brain Research*, *1464*, 43–52.
- Ji, S., Liu, P., Shen, H., Li, W., Bian, Z. (2012). [Studies on the ERP underlying premeditated and temporary deception]. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, *29*, 272–6.
- Johnson, R., Jr., 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), 217–53.
- Johnson, R., Jr., Barnhardt, J., Zhu, J. (2004). The contribution of executive processes to deceptive responding. *Neuropsychologia*, *42*, 878–901.
- Johnson, R., Jr., 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. *Brain Research. Cognitive Brain Research*, *24*(3), 386–404.
- Johnson, R., Jr., 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), 469–82.
- Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J. (2001). Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping*, *14*(3), 166–85.
- Kesner, R.P., Churchwell, J.C. (2011). An analysis of rat prefrontal cortex in mediating executive function. *Neurobiology of Learning and Memory*, *96*, 417–31.
- Koechlin, E. (2016). Prefrontal executive function and adaptive behavior in complex environments. *Current Opinion in Neurobiology*, *37*, 1–6.
- Lewis, A., Bardis, A., Flint, C., et al. (2012). Drawing the line somewhere: an experimental study of moral compromise. *Journal of Economic Psychology*, *33*(4), 718–25.
- Lindskold, S., Han, G. (1986). Intent and the judgment of lies. *Journal of Social Psychology*, *126*(1), 129–30.

- 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*(8), e43076.
- Parkinson, C., Sinnott-Armstrong, W., Koralus, P.E., Mendelovici, A., McGeer, V., Wheatley, T. (2011). Is morality unified? Evidence that distinct neural systems underlie moral judgments of harm, dishonesty, and disgust. *Journal of Cognitive Neuroscience*, *23*(10), 3162–80.
- Poddar, A., Foreman, J., Banerjee, S.(S.), Ellen, P.S. (2012). Exploring the Robin Hood effect: moral profiteering motives for purchasing counterfeit products. *Journal of Business Research*, *65*(10), 1500–6.
- Rigoulot, S., Fish, K., Pell, M.D. (2014). Neural correlates of inferring speaker sincerity from white lies: an event-related potential source localization study. *Brain Research*, *1565*, 48–62.
- Sohn, M.H., Albert, M.V., Jung, K., Carter, C.S., Anderson, J.R. (2007). Anticipation of conflict monitoring in the anterior cingulate cortex and the prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(25), 10330–4.
- Sun, D., Lee, T.M., Chan, C.C. (2015). Unfolding the spatial and temporal neural processing of lying about face familiarity. *Cerebral Cortex*, *25*(4), 927–36.
- Sun, D., Lee, T.M., Wang, Z., Chan, C.C. (2016). Unfolding the spatial and temporal neural processing of making dishonest choices. *PLoS One*, *11*(4), e0153660.
- Talwar, V., Crossman, A. (2011). From little white lies to filthy liars: the evolution of honesty and deception in young children. *Advances in Child Development and Behavior*, *40*, 139–79.
- Talwar, V., Murphy, S.M., Lee, K. (2007). White lie-telling in children for politeness purposes. *International Journal of Behavioral Development*, *31*(1), 1–11.
- Wu, D., Loke, I.C., Xu, F., Lee, K. (2011). Neural correlates of evaluations of lying and truth-telling in different social contexts. *Brain Research*, *1389*, 115–24.
- Wu, H., Hu, X., Fu, G. (2009). Does willingness affect the N2-P3 effect of deceptive and honest responses? *Neuroscience Letters*, *467*(2), 63–6.
- Wu, S., Cai, W., Jin, S. (2016). Gain or non-loss: the message matching effect of regulatory focus on moral judgements of other-orientation lies. *International Journal of Psychology*, DOI: 10.1002/ijop.12286 (2 June 2016).
- Yin, L., Hu, Y., Dynowski, D., Li, J., Weber, B. (2017). The good lies: altruistic goals modulate processing of deception in the anterior insula. *Human Brain Mapping*, *38*(7), 3675–90.