Published in final edited form as:
Cogn Affect Behav Neurosci. 2015 June ; 15(2): 475–491. doi:10.3758/s13415-015-0336-9.

Executive Control- and Reward-Related Neural Processes Associated with the Opportunity to Engage in Voluntary Dishonest Moral Decision Making

Xiaoqing Hu[*], **Narun Pornpattananangkul**[*], and **Robin Nusslock**
Department of Psychology, Northwestern University

Abstract

Research has begun to examine the neurocognitive processes underlying voluntary moral decision making, which involves engaging in honest or dishonest behavior in a setting where the individual is free to make his or her own moral decisions. Employing event-related potentials (ERPs), we measured executive control and reward-related neural processes during an incentivized coin-guessing task where participants had the opportunity to voluntarily engage in dishonest behavior by over-reporting their wins to maximize earnings. We report four primary findings: First, the opportunity to deceive recruited executive control processes involving conflict monitoring and conflict resolution, as evidenced by a higher N2 and a smaller P3; Second, processing the outcome of the coin-flips engaged reward-related processes, as evidenced by a larger medial feedback-negativity (MFN) for incorrect (loss) than correct (win) guesses, reflecting a reward prediction error signal. Third, elevated executive control-related neural activity reflecting conflict resolution (i.e., attenuated executive control-P3) predicted a greater likelihood of engaging in overall deceptive behavior. Finally, whereas elevated reward-related neural activity (reward-P3) was as associated with a greater likelihood of engaging in overall deceptive behavior, an elevated reward prediction error signal (MFN difference score) predicted increased trial-by-trial moral behavioral adjustment (i.e., a greater likelihood to over-report wins following a previous honest loss than a previous honest win trial). Collectively, these findings suggest that both executive control- and reward-related neural processes are implicated in moral decision making.

Keywords

executive control; medial frontal negativity; moral decision making; reward process; reward prediction error

Real-world deception or cheating behavior, such as a Ponzi Scheme, can cause significant harm to society, organizations and individuals. Examining the neural processes underlying honest and dishonest decision making has important implications for understanding the foundation of both moral and immoral behavior, as well as for the study of ethics, psychology, neuroscience and law. Previous research suggests that individuals recruit

Address correspondences to: Xiaoqing Hu, Department of Psychology, Northwestern University, 2029 Sheridan Rd. Evanston, IL 60208 USA, xhu@u.northwestern.edu.

[*]The first two authors contributed equally.

distinct neural circuits when they are instructed by an experimenter to deceive others (i.e. instructed deception) compared to when they engage in truthful behavior (Abe 2011; Abe et al. 2006, 2007; Ganis et al. 2003; Langleben et al. 2002; Lee et al. 2002; Priori et al. 2008; Spence et al. 2001; for a meta-analysis, see Christ et al. 2009). Although these studies have contributed to our understanding of the neurocognitive processes associated with deception, real-life deception often involves a voluntary intention to deceive rather than explicit instructions to do so. Thus, examining the neural processes associated with voluntary moral decision making has important implications for understanding im/moral behavior as it frequently occurs in real-life settings (Sip et al. 2008).

Researchers have recently begun to investigate the neural processes underlying voluntary deception, which involves engaging in dishonest behavior in a setting in which the individual is free to make their own honest or dishonest choices (Abe & Greene, 2014; Baumgartner et al., 2009; Ding et al., 2013; Greene & Paxton, 2009; Sip et al. 2010, 2012). This initial work suggests both commonalities and distinctions in the neural mechanisms underlying voluntary versus instructed deception. Specifically, both instructed and voluntary deception recruit neural networks involved in executive control, including the dorsolateral prefrontal cortex (DLPFC) and the anterior cingulate cortex (ACC). Voluntary deception, however, additionally recruits neural regions implicated in reward-related processing, specifically the ventral striatum (e.g., Abe & Greene, 2014; Baumgartner et al., 2009). Thus, voluntary deception appears to involve both executive control and reward-related neural processes.

Understanding the neurocognitive processes associated with the opportunity to engage in voluntary dishonest choices not only involves identifying the implicated neural regions, but also when, and in what sequence, these neurocognitive processes occur along the temporal scale. The present study employed scalp-recorded event-related potentials (ERPs), which provide millisecond temporal resolution, to examine the neural temporal dynamics underlying the opportunity to engage in voluntary deception. Specifically, we recorded ERPs during a coin-guessing task in which participants could win rewards if they correctly predicted the outcome of a coin-flip (adapted from Greene & Paxton, 2009). Importantly, during certain trials, participants could freely engage in voluntary deception by over-reporting the accuracy of their prediction in order to maximize their monetary winnings. ERPs were stimulus locked to the outcome phase of the trial where participants first learn and evaluate the outcome of the coin-flip (heads versus tails). This outcome serves as a predictive cue for the participant as to whether or not they will win based on their prior prediction. The outcome phase is also the moment participants evaluate their outcomes (e.g., loss or win), and make a decision whether or not to engage in voluntary deception in order to increase their earnings by claiming a correct prediction for trials in which they actually made an incorrect prediction. The present study focuses on three questions. First, we examine the neural temporal dynamics of executive control processes associated with the opportunity to engage in honest or dishonest behavior. Second, we examine the extent to which reward-related neural activity, as indexed by the reward prediction error signal to the outcome cue, is implicated in voluntary dishonest behavior. Lastly, we examine whether executive control and/or reward-related neural activity modulate a) the overall likelihood of engaging in voluntary deceptive behavior and, b) moral behavioral adjustment on a trial-by-trial basis (i.e. a greater likelihood of engaging in voluntary deception on a subsequent trial following a loss on a previous trial).

We predict that compared with no opportunity to deceive, having an opportunity to deceive will recruit executive control processes involving elevated conflict monitoring and subsequent conflict resolution. Here, conflict monitoring involves a rapid evaluation about whether response conflict is involved. The detection of conflict then initiates conflict resolution, which is exerted by a resource-limited control system to resolve conflict (Botvinick et al., 2001). For the present study, analyses of conflict monitoring focused on the fronto-central N2 and analyses of conflict resolution focused on the parietal P3 (we define the parietal P3 that occurs in response to an opportunity to deceive as the “executive control-P3”). Elevated N2 has consistently been observed during tasks requiring conflict monitoring, such as the Flanker task and the Go/Nogo task. The augmented N2 likely reflects the activity of the dorsal ACC associated with the detection of conflict and response monitoring processes (Nieuwenhuis et al. 2003; Ridderinkhof et al. 2004; van Veen and Carter 2002; Yeung and Cohen 2006). Most importantly, enhanced N2 has been found in both instructed deception and information concealment (Carrion et al. 2010; Gamer and Berti 2010; Hu et al. 2011, 2013). These findings suggest that either being deceptive or concealing information requires participants to actively monitor response conflict because two competing response tendencies are activated (honest vs. dishonest response). Based on these findings, we predict that when compared with having no opportunity to deceive, having the opportunity to deceive (i.e., over-report one's performance) will trigger two competing response tendencies: to honestly report one's actual performance versus to dishonestly over-report one's actual performance in an attempt to maximize earnings. Regardless of whether the ultimate behavior is honest or dishonest, we predict these competing response tendencies between making an honest versus dishonest choice will elicit an elevated N2.

The amplitude of the parietal executive control-P3 can be modulated by a variety of factors, such as a stimulus's subjective probability, participants' devoted attentional resources and stimuli categorization uncertainty, etc (Donchin and Coles 1988; Johnson 1986, 1993). In particular, it has been found that the parietal executive control-P3 is attenuated when experimental manipulations increase the executive control demands involved in the task. These manipulations include perceptual/memory load, dual-task, categorization ambiguity, stimulus-response incompatibility, among others (Chen et al. 2008; Garcia-Larrea & Cezanne-Bert 1998; Hu et al., 2012; Kok 2001; Lorist et al. 1996; Wickens et al. 1983). Moreover, research has consistently demonstrated that instructed deception is associated with attenuated parietal executive control-P3, which was taken as evidence that deception involves executive control processes (e.g., Johnson et al., 2003). More specifically, it has been argued that deception forces people to manage two competing response tendencies (lie versus truth) in working memory and that engaging in deception necessitates suppressing the truth in order to give the deceptive responses (Hu et al. 2011; Johnson et al. 2003, 2008). We predict that having the opportunity to deceive will activate conflict resolution processes to resolve the conflict between competing response tendencies and that these processes will require cognitive effort. We further predict that this increase in executive control processes will attenuate the executive control-P3 when people have the opportunity to deceive relative to when they do not have the opportunity to deceive. Examining both the fronto-central conflict-sensitive N2 and the parietal executive control-P3 will allow us to assess multiple psychological processes underlying the opportunity to engage in dishonest moral choices along the temporal scale. Specifically, we propose that the N2 will inform our understanding of the initial conflict associated with the opportunity to engage in voluntary deception, and that the subsequent executive control-P3 will inform our understanding of the higher-level regulatory processes to resolve this conflict. Lastly, a goal of the present study is to examine the extent to which individual differences in executive control-related neural processes modulate one's likelihood of engaging in voluntary deceptive behavior. We predict that individuals who exhibit increased executive control-related neural activity (i.e., a larger N2 and/or a smaller executive control-P3) during trials in which they have an opportunity to deceive will be more likely to engage in voluntary deceptive behavior. This prediction is based on previous studies showing that being dishonest engages executive control-related neural networks (Baumgartner et al., 2009; Greene & Paxton, 2009).

Relative to executive control processes such as conflict monitoring and conflict resolution, less is known about the role that reward-related processes play in voluntary deception. Given deception can serve the goal of maximizing gains, we predict that reward-related neural activity will play an important role in voluntary deception. To investigate this, we examined the medial frontal negativity (MFN, Gehring & Willoughby, 2002) locked to the outcome phase of the coin task where participants learn whether they accurately predicted the coin flip (e.g. heads vs. tails). The MFN (also known as the feedback negativity, FN, or feedback error-related negativity, fERN, Miltner et al. 1997; Gehring & Willoughby, 2002) is a negative-going, fronto-central distributed waveform that peaks approximately 200-400 ms after the presentation of negative outcome/feedback compared to positive outcome/feedback. Research on reward processing demonstrates that midbrain dopamine neurons encode prediction errors. When a reward prediction is violated such that the outcome is worse than desired or does not meet one's goal, the firing rate of midbrain dopamine neurons will be temporarily dropped, generating a negative reward prediction error (Schultz 2002). Such reward prediction error signals from the midbrain dopamine system subsequently modulate the activity of the anterior cinguate cortex (ACC), reflected by a scalp-recorded MFN (Holroyd & Coles 2002; Holroyd & Yeung, 2012). Indeed, the MFN has been proposed as a neural proxy of prediction error and has been studied intensively in the context of decision making or reinforcement-learning tasks that involve gains/losses and trial-by-error learning (for reviews, see Holroyd & Coles 2002; Nieuwenhuis et al. 2004; Walsh & Anderson 2012). Source localization analyses suggest that the MFN is likely produced in the ACC (Miltner et al. 1997; Gehring & Willoughby 2002; but see Carlson, Foti et al. 2012), consistent with the computation model of the MFN (Holroyd & Coles 2002). The MFN has been demonstrated to update the contingency between behavioral choice and reward outcome, and to guide one's subsequent behavior to pursue wins and/or avoid losses on a trial-by-trial basis (Cohen & Ranganath 2007; van der Helden et al. 2010, Walsh & Anderson, 2011; for a review see Walsh & Anderson 2012).

These characteristics of the MFN allow us to examine the role that reward-related neural activity, and specifically the reward prediction error signal, plays in voluntary deception choices involving possible gains. We hypothesize that an incorrect prediction of the outcome of the coin-flip (e.g., the outcome is heads when the prediction was tails) will generate an elevated reward prediction error signal reflected in a larger MFN to the incorrect outcome cue relative to the correct outcome cue. Based on the logic that individuals with elevated reward-related neural activity (i.e., an elevated reward prediction error signal) may be more willing to deceive for personal gain, we further predict that participants with a particularly large MFN (i.e., a larger reward prediction error) to the incorrect outcome cue will be more likely to engage in voluntary deception to maximize gains. Lastly, given the role that reward prediction error plays in adjusting one's behavior to maximize performance-based gains, we predict that a larger MFN during trials in which participants have no opportunity to engage in voluntary deception will be associated with a greater likelihood of engaging in voluntary deception on *subsequent* trials in which there is an opportunity to deceive (i.e. behavioral adjustment).

In addition to the MFN, recent work indicates that the P3 in response to gain/loss cues (what we refer to as the reward-P3) is also implicated in reward processing. For example, the reward-P3, which follows the MFN, is larger to both unexpected outcomes (Hajcak et al., 2005; von Borries et al. 2013) and rewards of greater magnitudes (Sanfey & Yeung 2004). Moreover, the reward-P3 is larger during tasks that involve active choices rather than mere observations, which may reflect a higher level of attentional engagement associated with making active choices (Yeung et al. 2005). Because the reward-P3 may reflect one's attentional engagement to outcome cues, we predict that a larger reward-P3 response to gain than to loss cues (i.e., higher reward processing) will predict a greater likelihood of engaging in voluntary deception. Moreover, recent studies have also linked the reward-P3 with behavioral adjustment strategies in reinforcement learning tasks (Chase et al. 2011; Martin et al. 2013; von Borries et al. 2013). Based on these findings, we also tested in the present study whether the MFN or the P3 can predict moral behavioral adjustment strategies to maximize one's gains. Collectively, these analyses have important implications for understanding the extent to which reward-related neural activity influences one's likelihood of engaging in both overall voluntary deception and moral behavioral adjustment on a trial-by-trial basis (i.e. switching from honest to dishonest behavior to maximize earnings).

Materials and Methods

Participants

Twenty-six right-handed participants (nine males, age: 18-22) at Northwestern University received partial course credit for their participation. This sample size is consistent with previous studies that employed a similar voluntary dis/honest choice paradigm (e.g., Ding et al., 2013, *N*=18; Abe & Greene, 2014, *N*=28; Shavli & De Deru, 2014, *N*=30). Six additional participants were excluded given excessive blinks and artifacts. Participants were screened for neurological history and had normal or corrected vision. The study was approved by the Northwestern Institutional Review Board and participants provided written consents prior to the experiment.

Procedure

Participants were seated in an electromagnetically shielded, sound-attenuated booth in front of a computer monitor. Following electrode application, participants completed an incentivebased-coin-guess task (adapted from Greene and Paxton, 2009) to examine neurocognitive profiles of voluntary moral decision makings. Following the coin-guess task, participants were debriefed and dismissed.

Coin-guess Task

Participants were instructed that they were to predict the outcome of each coin-flip and that they would win raffle tickets for correct guesses (Figure 1). Specifically, participants were informed they could earn either 3 or 5 raffle tickets for each correct guess on a given trial, and that they would not win any raffle tickets for a given trial if their guess was incorrect. The cumulative tickets each participant gained over the course of the task were then placed into a lottery from which they could win one of three $25 Amazon gift cards (i.e., the better their performance, the higher the probability they would win one of the gift-cards). The task consisted of 200 trials and lasted approximately 40 minutes. Half of the trials (i.e., 100 trials) were no-opportunity-to-deceive trials (NoOp). In these NoOp trials, the word “RECORD” was presented for 3000 ms at the beginning of the trial. For 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 predict “heads” and a button labeled T if they predict “tails” for that particular trial. Requiring participants to record their prediction on NoOp trials prevented them from engaging in voluntary deception on these trials. The remaining trials (i.e., 100 trials) were opportunity-to-deceive trials (Op). In these Op trials, the word “RANDOM” was presented for 3000 ms at the beginning of the trial. When the word “RANDOM” appeared on the screen participants were instructed to make a prediction in their mind about the upcoming coin flip, but they did not have to record their prediction via external button-press. To justify this manipulation, participants were informed that previous research suggests that people's ability to predict the future (i.e. coin-flip) might be better if they made the predictions privately to themselves (see also Greene & Paxton, 2009). During this coin-guessing task, Op trials and NoOp trials were intermixed, and were presented to participants in a randomized order. To balance motor output across Op and NoOp trials, participants were instructed to randomly press one of two buttons on the button box labeled R (random) for the Op trials when the word “Press” appeared on the screen. For all trials, the outcome of the coin-flip (visual depiction of head versus tail) was next presented for 2000 ms. Epoched EEG data was stimulus locked to the outcome of the coin-flip for all analyses. The question “Correct?” next appeared on the monitor, prompting participants to indicate whether or not their prediction was accurate. For NoOp (i.e., RECORD) trials, participants were instructed to press either a “YES” button (i.e., correct prediction) or a “NO” button (i.e., incorrect prediction) based on their previously recorded responses. For Op trials (i.e., RANDOM trials), participants were instructed to press either the “YES” button (i.e. correct prediction) or the “NO” button (i.e. incorrect prediction) based on their prior, non-recorded predictions. The fact that participants did not record their predictions during Op trials afforded them the opportunity to over-report their accuracy rate in order to increase their possible winnings (i.e., voluntary deception). During debriefing, all participants reported that they were aware that they could cheat during the Op trials.

Operationalizing Overall Deception—Because participants were instructed to predict the outcome of a coin-flip in the coin guessing task, the expected reported accuracy for honest participants should be comparable across Op and NoOp trials (i.e., ~50%). Accordingly, a higher reported accuracy for Op than NoOp trials suggests a higher likelihood of dishonesty. Thus, by comparing claimed wins between the Op and the NoOp trials (Reported Wins in the Op trials minus Reported Wins in the NoOp trials), we can infer the likelihood that a participant engaged in overall voluntary deception during the coin-guess task. A higher difference score suggests a higher likelihood of voluntary deception. One can also use 50% rather than the actual accuracy in the NoOp trials as the baseline because people's prediction in the NoOp trials should be 50%. However, examination of the data suggested that there was a wide range of variance across individuals for their prediction accuracy in the NoOp trials (38% to 60%, see Figure 2). We therefore calculated the Op-minus-NoOp performance difference to account for individual differences in prediction accuracy during the NoOp trials.

Operationalizing Trial-by-Trial Moral Behavioral Adjustment—To examine one's behavioral adjustment on a trial-by-trial basis, we measured the probability of reported wins during Op trials that were preceded by a NoOp_Loss trial. To ensure this behavioral adjustment was due to a previous loss rather than a general craving for wins, we also calculated the probability of reported wins during Op trials that were preceded by NoOp_Win trials. The behavioral adjustment score was calculated as the difference between the proportion of reported wins during Op trials following NoOp_Loss trials and the proportion of reported wins during Op trials following NoOp_Win trials ($_{\text{NoOp_Loss}}$OpWin –minus- $_{\text{NoOp_Win}}$OpWin). A higher behavioral adjustment score indicates that a participant was more likely to make a dishonest decision following an honest loss than following an honest win. That is, participants with a high behavioral adjustment score were more likely to engage in voluntary deception on a *subsequent* trial given a previous honest loss.

EEG recordings and analyses

EEG data were collected from 19 electrodes (FP1/2, Fz, F3/4, F7/8, FCz, Cz, C3/4, C7/8, CPz, Pz, P4/5, T6/8) grounded at AFz. The on-line reference was the left mastoid and data were recorded from the right mastoid enabling computation of an off-line linked mastoid reference (impedances<5kΩ). Data were filtered (DC-100 Hz), amplified, and digitized (500Hz).

During offline analyses, eye blinks were first corrected with PCA algorithms implemented in NeuroScan EDIT software (Neuroscan Inc.). Saccades and movement-related artifacts were removed manually. EEG data were then high-pass filtered (.1 Hz, 24 db). For all analyses, epoched EEG data were stimulus locked to the onset of the outcome of the coin-flips (visual depiction of head versus tails). A linear detrend algorithm on a large epoch (from -1,100 to 2000 ms) was used to remove leftover drifts in data. ERP epochs were then trimmed (from -200 to 1000 ms) and the pre-stimulus baseline (-200-0 ms) corrected. Epochs containing artifacts (±75 μV) were rejected, and the remaining clean trials were low-pass filtered (30 Hz, 12 db).

We stimulus locked EEG data to the onset of the outcome of the coin-flips for two reasons. First, the outcome of the coin-flip occurs right before the question “Correct?’ appears on the monitor which is the moment participants need to decide whether to be honest or dishonest in reporting the accuracy of their prediction about outcome of the coin toss (i.e., whether or not to engage in voluntary deception). Thus, this time window is relevant to the N2 and P3 executive control analyses. Second, it is the point where participants first evaluate the outcome of the coin flip [incorrect prediction (loss) versus correct prediction (win)], and thus relevant to the MFN and P3 reward processing analyses.

ERP measurements

Artifact-free EEG epochs were averaged into four categories: Op_Loss, Op_Win, NoOp_Loss and NoOp_Win given our within-subject 2 (Opportunity: Op vs. NoOp) by 2 (Outcome: Win vs. Loss) design. Again, whereas the distribution of Op vs. NoOp trials were established a priori by the experimenters, whether a trial was a Win or Loss trial was based on participants' “YES” or “NO” responses to the “Correct?” slide in the coin-guessing task. Given participants had the opportunity to cheat during the Op trials, an individual's tendency to engage in dishonest behavior would influence the number of Op_Win trials and Op_Loss trials available for analyses. However, because participants in the present study were largely honest (see below for behavioral results), the number of trials for Op_Loss and Op_Win were comparable: on average, there were 39, 40, 38 and 42 artifact-free trials available for averaging for NoOp_Loss, NoOp_Win, Op_Loss, and Op_Win trials, respectively[1].

Based on visual inspection of the grand averaged ERPs, we measured the mean ERP amplitude between 200 and 450 ms as the N2/MFN, and the mean ERP amplitude between 450 and 650 ms as the P3. We termed the ERPs that were responsive to Op vs. NoOp as executive control-related ERPs, including executive control-N2 and executive control-P3. We termed the ERPs that were responsive to Loss vs. Win as reward-related ERPs, including reward-MFN and reward-P3. These labels are based on recent theory and empirical evidence suggesting that, although N2 and MFN may temporally overlap with each other, the N2 is more sensitive to conflict monitoring whereas the MFN is more sensitive to the reward processing (Baker & Holroyd 2011; Warren & Holroyd, 2012).

Results

All within-subject ANOVA results are reported with Greenhouse-Geisser corrected *p* value when the assumption of sphericity was violated. Partial eta squared values (η_p^2) are used to estimate effect size in repeated measure ANOVAs, with 0.01, 0.06 and 0.14 considered as small, medium and large effect sizes, respectively. For individual difference correlational analyses, correlation coefficients of 0.1, 0.3 and 0.5 are considered as small, medium and large effect sizes, respectively. All tests were two-tailed.

[1] Although participants were largely honest in the present study, there were six “dishonest” participants identified using binominal tests who reported significantly more wins than losses (55 vs. 31, $t(5)=3.996$, $p<0.02$) during the Op trials. Although the unequal number of Op_Win and Op_Loss trials from these six “dishonest” individuals could introduce confounds into quantifying ERP amplitude, this concern is mitigated by the fact that we used mean amplitude (as opposed to peak amplitude) to quantify ERP amplitude. Mean amplitude is less susceptible to bias than peak amplitude for studies/conditions with unequal trial numbers (Luck, 2014).

Behavioral Assessment of Overall Dis/honest Behavior

Figure 2 depicts the distributions of reported wins in percentages for Op and No-Op trials, and the difference between the Op and NoOp trials. A paired-sample t-test was conducted to compare the percentage of reported wins in the Op and the NoOp trials. Results indicated that participants did report winning slightly more often in the Op trials than in the NoOp trials, though the differences were not significant (Op vs. NoOp, Mean ± S.E.: 52.61 ± 1.62 % vs. 49.50 ± 0.97 %, $t(25)=1.69$, $p >.10$).

We further analyzed participants' accuracy and RTs to the “Correct?” slides in the coin task. For accuracy, because participants made a Yes (Win) or No (Loss) decision based on predictions made in their mind in the Op trials, we can only analyze accuracy in the NoOp trials. Results showed that participants were highly accurate in the NoOp trials, (Mean ± S.E.: 98 ± 0.54%), suggesting that they were following instruction. A 2 (Opportunity Op vs. NoOp) × 2 (Outcome: Loss vs.Win) within-subjects ANOVA was conducted on RTs. This ANOVA yielded marginally significant main effects for both Opportunity and Outcome. With respect to the main effect of Opportunity, participants had faster RTs during Op trials than NoOp trials: $F(1,25)=3.95$, $p=.058$, $\eta_p^2=0.14$, Mean ± S.E. Op vs. NoOp: 334.13 ± 17.18 ms vs. 346.07 ± 19.21 ms. With respect to the main effect of outcome, participants were slower in reporting a Loss outcome than a Win outcome: $F(1,25)=3.54$, $p=.072$, $\eta_p^2=0.12$, Mean ± S.E. Loss vs. Win: 350.04 ± 21.96 ms vs. 330.16 ± 14.82 ms. Importantly, these marginally significant main effects were qualified by a significant Opportunity × Outcome interaction: $F(1,25)=6.14$, $p<.02$, $\eta_p^2=0.20$. Follow-up paired t-tests showed that during NoOp trials, it took participants longer to report a Loss than to report a Win (Mean ± S.E.: 363.35 ± 22.90 ms vs. 328.79 ± 16.54 ms, $t(25)=3.15$, $p<.01$). In contrast, there were no RT differences between for Loss and Win in Op trials (336.73 ± 21.76 ms vs. 331.54 ± 14.26 ms, $t(25)=0.40$, $p>.70$, see Figure 3). This suggests that when participants did not have an opportunity to deceive, they found it more difficult to report a loss outcome than a win outcome, which may be due to loss aversions. However, when participants had an opportunity to deceive and when they could voluntarily avoid a loss by cheating, this difficulty in reporting loss disappeared.

To better understand rates of dishonest behavior in the present study, we conducted binominal tests on each participant's accuracy for Op and NoOp trials. A significantly higher accuracy rate in the Op than NoOp trials indicates that a particular participant likely engaged in dishonest behavior. Results showed that six of 26 participants (23%) in the present study can be classified as having engaged in extreme dishonest behavior, *p*s<.01. This rate of dishonest behavior is comparable with previous studies that used either the same or a similar voluntary honest/dishonest moral choice paradigm (Abe & Greene, 2014; Gino & Ariely, 2012; Shalvi & De Dreu, 2014). This rate of dishonest behavior is smaller than that observed in Greene and Paxton (2009) given the authors of this study intentionally selected participants who were highly likely to cheat based on a pilot behavioral testing.

Behavioral Assessment of Trial-by-Trial Moral Behavioral Adjustment

Across all participants, there were 663 Op trials that were preceded by NoOp_Loss, and 671 Op trials there were preceded by NoOp_Win. To estimate one's trial-by-trial moral behavioral adjustment, we calculated 1) the percentage of reported wins during Op trials following a previous NoOp loss trial ($_{\text{NoOp_Loss}}$OpWin, Mean ± S.E., 53.69 ± 0.02 %), and 2) the percentage of reported wins during Op trials following a previous NoOp win trial ($_{\text{NoOp_Win}}$OpWin, Mean ± S.E., 52.59 ± 0.02 %). A paired t-test showed that there was no significant difference between these two measures ($t(25)=0.39$, $p>.70$), suggesting that whether a previous NoOp trial was a loss or win trial did not modulate participants' decision making on subsequent Op trials. We then calculated each individual's moral behavioral adjustment score, defined as $_{\text{NoOp_Loss}}$OpWin –minus- $_{\text{NoOp_Win}}$OpWin in terms of percentage of reported wins.

There was no relationship between one's moral behavioral adjustment score and one's overall deception (as estimated by the difference between claimed wins for Op and NoOp trials), $r(26)=-0.064$, $p>.76$. This suggests that distinct processes may underlie the likelihood to engage in trial-by-trial moral behavioral adjustment versus overall voluntary deceptive behavior.

ERP Results

Within-Subject ANOVA on N2/MFN (200-450 ms) and P3 (450-650 ms)—200-450 ms N2/MFN: A 2 (Opportunity: Op vs. NoOp) × 2 (Outcome: Loss vs. Win) × 5 (Electrode: Fz, FCz,Cz,CPz,Pz) within-subject ANOVA on N2/MFN amplitude yielded a significant main effect of Opportunity: $F(1,25)=15.10$, $p<.001$, $\eta_p^2=0.38$, such that the N2 for the Op trials was significantly greater than the N2 for the NoOp trials (Mean ± S.E., 0.04 ± 1.02 vs. 2.16 ± 1.23 μV). There was also a significant main effect of Outcome, $F(1,25)=18.98$, $p<.001$, $\eta_p^2=0.43$, such that Loss cues elicited a larger MFN than Win cues (Mean ± S.E., 0.32 ± 1.00 vs. 1.88 ± 1.10 μV). Lastly, there was a significant main effect of Electrode, $F(2.14, 53.61)=96.85$, $p<.001$, $\eta_p^2=0.80$. Pairwise comparisons showed that the N2/MFN at Fz (-2.71 ± 0.99 μV) and FCz (-1.28 ± 1.18 μV) were more negative than the N2/MFN at Cz (1.29 ± 1.07 μV), CPz (3.12 ± 1.08 μV) and Pz (5.08 ± 1.05 μV, all ps<. 001). This indicates a fronto-central distribution to the N2/MFN.

None of the 2-way or 3-way interactions were significant for the N2/MFN. [Opportunity × Outcome, $F(1,25)=0.66$, $p>.42$, $\eta_p^2=0.03$; Opportunity × Electrode, $F(1.53, 38.31)=0.57$, $p>.53$, $\eta_p^2=0.02$; Outcome × Electrodes, $F(2.02, 50.43)=0.33$, $p>.72$, $\eta_p^2=0.01$; and Opportunity × Outcome × Electrode, $F(1.66, 41.44)=1.78$, $p>.18$, $\eta_p^2=0.07$].

450-650 ms P3: A 2 (Opportunity: Op vs. NoOp) × 2 (Outcome: Loss vs. Win) × 5 (Electrode: Fz, FCz,Cz,CPz,Pz) within-subject ANOVA on P3 amplitude yielded a significant main effect of Opportunity: $F(1,25)=11.71$, $p<.005$, $\eta_p^2=0.32$, such that the P3 for the Op trials was significantly smaller than the P3 for the NoOp trials (Mean ± S.E., 3.95 ± 0.84 vs. 5.75 ± 0.97 μV). There was also a significant main effect of Electrode, $F(2.11, 52.71)=93.79$, $p<.001$, $\eta_p^2=0.79$. Pairwise comparisons showed that P3 amplitude was greater at Pz (8.67 ± 0.96 μV) compared to P3 amplitude at anterior electrodes [CPz (7.55 ± 0.97 μV), Cz (5.65 ± 0.92 μV), FCz (2.39 ± 1.01 μV) and Fz (-0.02 ± 0.78 μV)] (all ps<. 001). There was no main effect of Outcome, $F(1,25)=1.98$, $p>.17$, $\eta_p^2=0.07$, such that Loss and Win cues generated comparable P3 amplitudes (Mean ± S.E., Loss: 5.18 ± 0.89 vs. Win: 4.52 ± 0.91 μV).

There was a significant Outcome × Electrode interaction for P3, $F(1.77, 44.27)=3.78$, $p<.05$, $\eta_p^2=0.13$. Follow-up tests indicated that Loss cues elicited a significantly larger P3 than Win cues at Fz, (0.59 ± 0.84 vs. -0.62 ± 0.79 μV, $t(25)=2.65$, $p<.02$). However, the Loss-P3 and Win-P3 were comparable at FCz, (2.87 ± 1.02 vs. 1.92 ± 1.07 μV, $t(25)=1.88$, $p>.07$), Cz (5.95 ± 0.96 vs. 5.34 ± 0.96 μV, $t(25)=1.23$, $p>.20$), CPz (7.61 ± 0.97 vs. 7.49 ± 1.03 μV, $t(25)=0.22$, $p>.80$) and Pz (8.88 ± 0.93 vs. 8.46 ± 1.07 μV, $t(25)=0.75$, $p>.46$). None of the other 2-way or 3-way interactions were significant for P3 [Opportunity × Outcome, $F(1,25)=0.18$, $p>.60$, $\eta_p^2=0.01$; Opportunity × Electrode, $F(2.53, 63.26)=1.18$, $p>.32$, $\eta_p^2=0.05$; Opportunity × Outcome × Electrode, $F(1.72, 43.30)=0.44$, $p>.60$, $\eta_p^2=0.02$].

Individual Differences Analyses: Executive control- and reward-related neural activity as predictors of voluntary deception and moral behavioral adjustment

The lack of a significant Opportunity × Outcome interaction for N2, MFN and P3 suggests that having or not having the opportunity to cheat (i.e., Opportunity) and processing wins and losses (i.e., Outcome) influenced ERPs independently. This finding is in line with our a priori theorization that the Opportunity condition (Op vs. NoOp) primarily engages executive control processes and that the Outcome condition (Loss vs. Win) primarily engages reward processes. Accordingly, the subsequent individual differences analyses focus on the two significant main effects [Opportunity (Op vs. NoOp) and Outcome (Win vs. Loss)]. To isolate ERPs of interests, we computed two separate difference waves: one for Op-minus-NoOp to reflect executive control processes and one for Loss-minus-Win to reflect reward processes (for similar difference wave approaches, see Bress & Hajcak 2013; Cohen & Ranganath, 2007; Yeung et al. 2005).

To quantify the executive control-N2, we calculated the mean amplitude for the 200-350 ms time window for the Op-minus-NoOp difference wave (collapsing across Win and Loss trials) at FCz (Hu et al., 2011; Johnson et al., 2008). An elevated (i.e., more negative) N2 during Op relative to NoOp trials (as reflected by a larger N2 Op-minus-NoOp difference wave) suggests that a participant experienced greater conflict between honest and dishonest response tendencies during Op trials.

The executive control-related P3 was calculated as the mean amplitude for the 450-650 ms time window for the Op-minus-NoOp difference wave (collapsing across Win and Loss trials) at Pz (Hu et al., 2011). An attenuated executive control-P3 during Op relative to NoOp trials (as reflected by a smaller executive control-P3 Op-minus-NoOp difference wave) suggests that a participant devoted more efforts to resolve conflict between honest and dishonest responses tendencies during Op trials (for executive control-related ERPs, see Fig. 4a).

We calculated the MFN as the mean amplitude for the 300-450 ms time window for the Loss-minus-Win difference wave at FCz (collapsing across Op and NoOp trials). We chose FCz to be consistent with previous MFN literature (e.g., Cohen & Ranganath, 2007). The loss-minus-win MFN difference score served as an indicator of reward prediction error, with a larger MFN difference score (i.e., more negative) reflecting a larger reward prediction error.

We calculated the loss-minus-win P3 as the mean amplitude for the 450-650 ms time window for the Loss-minus-Win difference wave (collapsing across Op and NoOp trials) at Pz (Yeung & Sanfey, 2004). A larger value for this difference score indicates a larger P3 (i.e., a higher level of attention engagement) for loss cues than for win cues (for reward-related ERPs, see Fig. 5a).

Analyses of overall voluntary honest and dishonest decision makings—Prior to proceeding with individual difference analyses, we first examined whether the distribution of participants' overall dishonest behavioral tendencies was normal using the Shapiro-Wilk test. A normal distribution of dishonest tendencies would justify the use of parametric tests, i.e., the Pearson correlation. By contrast, a non-normal distribution of participants' dishonest behavioral tendencies would indicate that non-parametric tests would be more appropriate, i.e., the Spearman rank tests. Results indicated that participants' overall dishonest behavioral tendency was not normally distributed (W=0.91, df=26, p=.022). Therefore, we report results from Spearman rank-order tests for correlational analyses involving one's overall dishonest behavioral tendency.

In line with prediction, the Op-minus-NoOp executive control-P3 at Pz was negatively correlated with one's overall dishonest behavioral tendency, r_s (26)=0.404, p<.045. This indicates that individuals with a smaller P3 amplitude during Op trials relative to NoOp trials were more likely to over-report their wins (i.e., engage in overall deception) (see Figure 4b). There was no relationship between individual differences in the Op-minus-NoOp executive control-N2 at FCz and one's likelihood to engage in overall deception (r_s (26)=0.059, p > .70).

Regarding reward-related neural activity (i.e., the Loss-minus-Win ERPs), and in line with prediction, the loss-minus-win reward-P3 at Pz was negatively associated with one's overall tendency to engage in voluntary deception (r_s (26)=-0.47, p<.02). Thus, a larger P3 to win relative to loss was associated with a greater likelihood of engaging in voluntary deception (see Figure 5b). There was no significant relationship between the loss-minus-win MFN difference score at FCz and one's overall tendency to engage in voluntary deception r_s(26)=0.14, p>.50.

Analyses of moral behavioral adjustment—We next examined the relationship between trial-by-trial moral behavioral adjustment and reward-related neural activity (loss-minus-win ERPs) exclusively during the NoOp trials. This analysis is restricted to NoOp trials because, by definition, moral behavioral adjustment involves being dishonest following a NoOp loss trial. We did not examine the relationship between executive control-related neural activity and behavioral adjustment as this analysis would require use of both Op and NoOp trials.

The Shapiro-Wilk normality test indicated that the distribution of moral adjustment was normal (W=0.97, df=26, p>.60), justifying the employment of parametric tests. Pearson correlational analyses suggested that neither the loss-minus-win MFN at FCz (r(26)=-0.195, p>.30) nor the loss-minus-win P3 at Pz (r(26)=-0.017, p>.90) predicted behavioral adjustment. We next conducted exploratory correlational analyses using all five midline electrodes. To reduce the chance of a false positive, we interpolated Fz/FCz/Cz as a fronto-central cluster, and CPz/Pz as a centro-parietal cluster. Analyses indicated that a larger centro-parietal loss-minus-win MFN during NoOp trials uniquely predicted a greater likelihood of trial-by-trial moral behavioral adjustment in order to maximize gains (r(26)=-0.390, p<.05, see Figure 6a for grand average ERPs and Figure 6b for the scatter-plot). This correlation was not significant when the fronto-central MFN was used: r(26)=-0.229, p>.25. Lastly, there was no relationship between the reward-P3 and trial-by-trial moral behavioral adjustment (for the fronto-central P3, r(26)=0.002, p>.99, for the centro-parietal P3, r(26)=-0.062, p>.76).

Discussion

The present study investigated the neurocognitive processes associated with having the opportunity to engage in voluntary deception. In this context, an individual is free to make his or her own honest or dishonest moral choices. We report four primary findings. First, when individuals had the opportunity to deceive, they experienced elevated conflict monitoring (a more negative N2) and devoted increased cognitive resources to resolve this conflict (attenuated executive control-P3) compared to when they did not have the option to deceive. Second, evaluating the outcome cue of the coin toss (win versus loss) elicited reward-related neural activity. Specifically, an incorrect prediction of the outcome of the coin toss (e.g., the outcome was heads when the prediction was tails) generated an elevated reward prediction error signal reflected in a larger MFN to the incorrect outcome cue relative to the correct outcome cue. Third, elevated executive control-related neural activity reflecting conflict resolution (attenuated executive control-P3) was associated with a greater likelihood of engaging in overall deceptive behavior. Finally, whereas elevated reward-related neural activity (reward-P3) was as associated with a greater likelihood of engaging in overall deceptive behavior, an elevated reward prediction error signal (MFN difference score) predicted increased trial-by-trial moral behavioral adjustment.

The Opportunity to Engage in Voluntary Deception Recruits Executive Control-Related Neural Activity

The opportunity-to-deceive (Op) versus no-opportunity-to-deceive (NoOp) contrast in the present study mimics real-life scenarios in which individuals navigate the temptation to deceive or engage in dishonest behavior for personal gains. We report that having the opportunity to engage in voluntary deception is associated with a more negative N2 than not having the opportunity to engage in such deception. The N2 has been observed during tasks involving conflict monitoring and response uncertainties, such as the Go/Nogo or Flanker tasks, in which participants need to override one response tendency to execute an alternative, goal-directed response (for a review, see Folstein and Van Petten, 2008). Increased N2 in the present study suggests that having the opportunity to engage in voluntary deception triggers two competing response tendencies: to honestly report one's actual performance versus to dishonestly over-report one's actual performance in order to maximize earnings. A larger (i.e., more negative) N2 has also been documented in previous studies of instructed deception reflecting a conflict between participants' automatic tendency to engage in honest behavior and instructions from an experimenter to engage in deceptive behavior (Hu et al., 2011). Considering that the N2 typically reflects activity in the ACC (Van Veen and Carter, 2002), results from the present study concur with fMRI studies that found elevated ACC activation during both instructed and voluntary deception (Abe & Greene, 2014; Baumgartner et al., 2009; Christ et al., 2009; Greene and Paxton, 2009). Collectively, these data suggest that both voluntary and instructed deception are associated with elevated conflict monitoring related neural activity.

An advantage of the temporal resolution afforded by ERP is that it allowed us to assess multiple executive control processes associated with having the opportunity to deceive along the temporal scale. In line with prediction, having the opportunity to engage in voluntary deception was associated with a more attenuated executive control-related P3 than not having the opportunity to deceive. A similar attenuation of executive control-P3 has been observed during studies of instructed deception (Hu et al., 2011; Johnson et al., 2003; Johnson et al., 2008). Drawing on this previous research on instructed deception, we argue that an attenuated executive control-related P3 reflects the engagement of regulatory processes to resolve the detected conflict (i.e., N2) between the tendencies to honestly report one's actual performance versus engage in deceptive behavior to maximize earnings. In line with this argument, theories of cognitive control state that when conflict is detected within the ACC it alerts higher cognitive control systems (such as the dorsolateral PFC) to expend resources in attempt to resolve such conflict (Botvinick et al. 2001). Reduced executive control-related P3 in this context likely reflects the engagement of these higher-level executive control processes to resolve the detected conflict and implement the behavioral response. The present study is the first to report attenuated executive control-related P3 when people have the opportunity to engage in voluntarily deception, suggesting that both instructed and voluntary deception engage conflict resolution related neural activity aimed at resolving the tension between honest and deceptive behavior.

The Opportunity to Engage in Voluntary Deception Recruits Reward-Related Neural Activity

Previous research highlights the involvement of motivational/reward-related processes in voluntary moral decision making (Abe & Greene, 2014; Baumgartner et al 2009; Ding et al., 2013). By investigating ERPs locked to the outcome of the coin flip, we provide the first electrophysiological evidence regarding how reward processing, and in particular the reward prediction error signal, contributes to voluntary moral decision making.

In line with prediction, cues that were indicative of loss (incorrect prediction of the outcome of the coin flip) elicited a larger MFN (i.e., more negative) than cues indicative of win (correct prediction of the outcome of the coin flip). The reinforcement learning model of the MFN proposes that the MFN reflects the impact of reward prediction error signals from the midbrain dopamine system on the ACC. The ACC will then use this signal to improve one's behavior to obtain a desired goal and to maximize one's personal gains (Holroyd & Coles 2002; Walsh & Anderson 2012). The MFN has been observed in numerous previous studies of reward processing and reinforcement learning in non-moral domains ((Bress & Hajcak 2013; Foti et al. 2011; Gehring & Willoughby 2002; Hajcak et al. 2005; Holroyd et al. 2003; Miltner et al. 1997; Nieuwenhuis et al. 2004; Yeung et al. 2005). The present study is the first to report a reward prediction error signal in the context of moral decision making when such (im)moral behavior is associated with gains/losses. Collectively these findings suggest that subsequent research on voluntary moral decision making should assess both executive control and reward-related neural activity as reward processes play an important role in modulating voluntary deceptive behavior.

It should be noted that recent empirical evidence questions the assumption that the MFN encodes reward prediction errors. In particular, the MFN is enhanced when an anticipated pain was omitted (i.e., a rewarding event, see Talmi et al., 2013). The MFN thus appears to encode salience prediction error (Talmi et al., 2013). The results in our study, however, are consistent with both the reward and salience perspective of the MFN: an experienced loss can be perceived as more salient than an experienced win, which is reflected by an enhanced MFN.

Examining the P3 that follows the MFN to win versus loss cues (what we refer to as the reward-P3) provides a window into subsequent stages of reward processing in the context of voluntary moral choices. Consistent with previous research (von Borries et al., 2013; Yeung & Sanfey, 2004; Yeung et al., 2005), we did not find that the outcome valence modulated the reward-P3. Despite the lack of main effect of valence, individual differences in the reward-P3 may reflect participants' different levels of attentional engagement during processing gain and loss outcomes (Yeung et al., 2005). In line with this perspective, the reward-P3 predicted one's overall dishonest tendency, as we report below.

Individual Differences in Overall Dishonest Tendency and Moral Behavioral Adjustment

Analyses of individual differences allow us examine the extent to which executive control-related processes and/or reward-related processes modulate one's propensity to engage in voluntary deception. In addition to mechanistic implications, examining this topic has important practical implications, as it can provide insight into which specific processes should be targeted in the promotion of honest and ethical behavior.

Here we operationalized voluntary deception at two different levels of analysis. First, we obtained an overall measure of deception for each individual, defined as the likelihood of over-claiming one's gains throughout the task. Second, we obtained a trial-by-trial moral behavioral adjustment estimate for each individual. This was defined as the likelihood of one making a dishonest or deceptive decision on the present trial if that trial followed an honest loss trial. Thus, participants with a high behavioral adjustment score were more likely to engage in voluntary deception on a *subsequent* trial given a previous honest loss.

With respect to overall dishonest tendencies, we report for the first time that both executive control-related and reward-related P3 modulate one's likelihood of engaging in voluntary deceptive behavior. Reduced executive control-related P3 in this context likely reflects the engagement of conflict resolution during Op trials to resolve the conflict between honest and dishonest response tendencies. The present study reports that individuals with a smaller P3 amplitude during Op trials relative to NoOp trials were more likely to over-report their wins (i.e., engage in overall deception). This finding suggests that the more people deliberate and resolve conflict between honest and dishonest responses, the higher likelihood they will violate moral norms and exhibit morally questionable behavior. This result is consistent with previous neuroimaging research in showing that activity in the executive control neural network (e.g., DLPFC, parietal lobe) is positively correlated with actual cheating behavior (Baumgartner et al., 2009; Greene & Paxton, 2009). Collectively, these findings support the hypothesis that 1) being honest or adhering to moral norms is a default behavioral tendency, and 2) cheating or violating moral norms to pursue self-interests actually requires active cognitive control (Baumgartner et al., 2009; Greene & Paxton, 2009; Rand et al., 2012; Sip et al., 2010).

Another neural indicator that predicted one's overall dishonest tendency is the reward-P3. Here, a larger reward-P3 to win cues than to loss cues was associated with a greater likelihood of engaging in overall voluntary deception. Given P3's close relationship with attention allocation (Donchin and Coles, 1988; Johnson, 1986), and the fact that P3 is associated with reward magnitude in gambling tasks (Yeung & Sanfey, 2004), this finding suggests that individuals who were more sensitive to win cues were more likely to engage in voluntary deception to maximize personal gain.

Regarding one's trial-by-trial moral behavioral adjustment, we report that the larger the magnitude of the loss-minus-win MFN during previous NoOp trials (i.e., larger reward prediction errors) the more likely participants were to report a win on the subsequent Op trial. Because participants' responses on the NoOp trials were always honest, higher reported wins on the subsequent Op trials suggests that participants switched their response tendencies to be more dishonest once they were given the opportunity. As argued in the reinforcement learning theory of the MFN, the reward prediction error signal from the midbrain dopamine system alerts the ACC and the PFC to update the stimulus/response-reinforcement contingency and to adjust behavior for optimal outcomes such as reward (Holroyd & Coles 2002). In accordance with this perspective, the MFN predicts a range of feedback-based choices and learning efficiencies on a trial-by-trial basis in non-moral tasks (Cohen & Ranganath 2007; van der Helden et al. 2010). The present study is the first to illustrate that the reinforcement learning signal MFN is involved in trial-by-trial moral behavioral adjustment to maximize personal gains.

Collectively, these results reveal an interesting dissociation. Whereas the early (300-450 ms) loss-minus-win MFN during NoOp trials predicts one's moral behavioral adjustment on a trial-by-trial basis, the later (450-650 ms) executive control-P3 and reward-P3 are associated with one's overall dishonest tendency. It is possible that because the MFN provides rapid and initial evaluations of on-going performance (Hajcak et al., 2006), it is more likely to drive *trial-by-trial* deceptive behavior instead of modulating *overall* dishonest response tendencies. In contrast, given that maintaining a general dishonest behavioral tendency requires increased cognitive control, it seems appropriate that this overall dishonest tendency was uniquely predicted by late P3 activity and not by earlier MFN or N2 activity. Future research is needed to more fully examine these hypotheses and the possible dissociation in neural processes underlying *trial-by-trial* versus *overall* deceptive behavior.

Given the novelty and complexity of the coin-guess task, it is not surprising that the loss-minus-win MFN observed in the present study shows some differences from the MFN reported in previous literature (e.g., Gehring & Willoughby 2002). First, the MFN in the present study occurs relatively late (~300 ms) compared to previous studies in which the MFN typically peaks between 200-300 ms (Gehring & Willoughby, 2002). The delayed MFN observed in the present study may reflect the nature of the coin-guess task. Upon the receipt of the outcome cue of the coin-toss, participants need to maintain their predictions in working memory and to choose between a subsequent honest or dishonest response. The prolonged latency of the MFN may reflect the involvement of such complex cognitive processes (Baker & Holroyd 2012).

Second, whereas we found that the centro-parietal MFN predicted moral behavioral adjustment, most previous studies found that a fronto-central MFN modulated non-moral behavioral adjustment (for a review, see Walsh & Anderson, 2012). This discrepancy may reflect the nature of the coin-task used in the present study which assesses moral behavioral adjustment, as opposed to tasks used in previous MFN research that assess more basic reward-based processes outside the scope of moral behavioral adjustment (Walsh & Anderson, 2012). Specifically, moral decisions likely involve elevated conflict detection and control-related processes as compared to non-moral choices (Greene & Paxton, 2009; Sip et al. 2010), and our posterior MFN may reflect such processes that are critical for moral behavioral adjustment (see Folstein & Van Petten, 2008). Future studies are needed to replicate the current findings and to investigate similarities and differences between reward-related neural processes within, and outside, the scope of moral behavioral adjustment.

A potential limitation of the present study is that we cannot know for certain at which point during the trial of the coin-guess task a participant decided to make an honest or dishonest decision. For example, participants could decide to cheat as soon as they learned the next trial is an Op trial. In addition, a participant may initially decide to make an honest decision but then change his or her mind to be dishonest during the “Correct?” slide. Similarly, a participant may even flip between honest and dishonest decisions as the trial proceeds. These possibilities would make it precarious to focus analyses on the earlier stages of a trial given the ambivalence that participants may be experiencing at such points during the trial. In contrast, the “Outcome” stage of the trial clearly involves Loss vs. Win outcome evaluations and these Outcome-locked ERPs can serve as ideal neural signals for reward processing based on previous literature (Walsh & Anderson, 2012). Thus, despite uncertainty as to when a participant precisely decides to make an honest or dishonest decision, we believe the Outcome stage of the trial is the ideal time period to focus analyses for the present study.

Another limitation of the present study is the relatively small sample size. Although our sample size is consistent with previous research that used a similar paradigm (*Ns*=18-30, Abe & Greene, 2014; Ding et al., 2013; Shalvi & De Deru, 2014), studies with this sample size may have difficulty detecting smaller effect sizes (i.e., Type II errors). This problem is rather common in cognitive, affective, and behavioral neuroscience (Lieberman & Cunningham, 2009). One practical solution to obtain a balance between Type I and Type II errors is to aggregate research findings by meta-analyses (Lieberman & Cunningham, 2009). Thus, it will be important for future research to investigate whether pattern of relationships observed in the present study can be replicated across studies and laboratories, and meta-analytic research should be used to provide more precise estimates of effect sizes.

In conclusion, the present study examined the neurocognitive processes and neural temporal dynamics involved in voluntary honest and dishonest moral decision making. This work has implications for a broad range of topics including ethics, philosophy, neuroscience and forensic science. We found that having the opportunity to cheat recruited both executive control (i.e., N2, executive control-P3) and reward (i.e., MFN, reward-P3) related neural activity. Moreover, early/late ERPs differentially predicted one's trial-by-trial moral behavioral adjustment and overall dishonest tendency. This work sheds new light on the neurocognitive processes underlying both voluntary deception and moral decision making.

References

Abe N. How the brain shapes deception: An integrated review of the literature. Neuroscientist. 2011; 17(5):560–574. [PubMed: 21454323]

Abe N, Greene JD. Response to anticipated reward in the nucleus accumbens predicts behavior in an independent test of honesty. Journal of Neuroscience. 2014; 34(32):10564–10572. [PubMed: 25100590]

Abe N, Suzuki M, Mori E, Itoh M, Fujii T. Deceiving others: Distinct neural responses of the prefrontal cortex and amygdala in simple fabrication and deception with social interactions. J Cogn Neurosci. 2007; 19(2):287–295. [PubMed: 17280517]

Abe N, Suzuki M, Tsukiura T, Mori E, Yamaguchi K, Itoh M, Fujii T. Dissociable Roles of Prefrontal and *Anterior* Cingulate Cortices in Deception. Cereb Cortex. 2006; 16(2):192–199. [PubMed: 15858160]

Baker TE, Holroyd CB. Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by feedback error-related negativity and N200. Bio Psy. 2011; 87:25–34.

Baumgartner T, Fischbacher U, Feierabend A, Lutz K, Fehr E. The neural circuitry of a broken promise. Neuron. 2009; 64(5):756–70. [PubMed: 20005830]

Bizzi, E.; Greely, HT. Using imaging to identify deceit: Scientific and ethical questions. American Academy of Arts & Science; 2009.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and cognitive control. Psychol Rev. 2001; 108(3):624–52. [PubMed: 11488380]

Bress JN, Hajcak G. Self-report and behavioral measures of reward sensitivity predict the feedback negativity. Psychophysiology. 2013; 50(7):610–616. [PubMed: 23656631]

Carlson JM, Foti D, Mujica-Parodi LR, Harmon-Jones E, Hajcak G. Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: a combined ERP and fMRI study. Neuroimage. 2011; 57(4):1608–1616. [PubMed: 21624476]

Carrion RE, Keenan JP, Sebanz N. A truth that's told with bad intent: an ERP study of deception. Cognition. 2010; 114(1):105–10. [PubMed: 19836013]

Chase HW, Swainson R, Durham L, Benham L, Cools R. Feedback-related Negativity Codes Prediction Error but Not Behavioral Adjustment during Probabilistic Reversal Learning. J Cog Neurosci. 2010; 23:936–946.

Chen A, Xu P, Wang Q, Luo Y, Yuan J, Yao D, Li H. The timing of cognitive control in partially incongruent categorization. Hum Brain Mapp. 2008; 29(9):1028–39. [PubMed: 17894393]

Christ SE, Van Essen DC, Watson JM, Brubaker LE, McDermott KB. The contributions of prefrontal cortex and executive control to deception: Evidence from activation likelihood estimate meta analyses. Cereb Cortex. 2009; 19(7):1557–1566. [PubMed: 18980948]

Cohen J, Polich J. On the number of trials needed for P300. International Journal of Psychophysiology. 1997; 25(3):249–255. [PubMed: 9105949]

Cohen MX, Ranganath C. Reinforcement learning signals predict future decisions. J Neurosci. 2007; 27(2):371–378. [PubMed: 17215398]

Ding XP, Gao X, Fu G, Lee K. Neural correlates of spontaneous deception: a functional near-infrared spectroscopy (fNIRS) study. Neuropsychologia. 2013; 51(4):704–712. [PubMed: 23340482]

Donchin E, Coles MG. Is the P300 component a manifestation of context updating? Behavioral and Brain Sciences. 1988; 11(3):357–427.

Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods. 2007; 39:175–191. [PubMed: 17695343]

Folstein JR, Van Petten C. Influence of cognitive control and mismatch on the N2 component of the ERP: A review. Psychophysiology. 2008; 45:152–170. [PubMed: 17850238]

Foti D, Weinberg A, Dien J, Hajcak G. Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. Hum Brain Mapp. 2011; 32(12):2207–2216. [PubMed: 21305664]

Gamer M, Berti S. Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials. Psychophysiology. 2010; 47(2):355–64. [PubMed: 20003148]

Ganis G, Kosslyn SM, Stose S, Thompson WL, Yurgelun-Todd DA. Neural correlates of different types of deception: an fMRI investigation. Cereb Cortex. 2003; 13(8):830–6. [PubMed: 12853369]

Garcia-Larrea L, Cezanne-Bert G. P3, positive slow wave and working memory load: a study on the functional correlates of slow wave activity. Electroencephalogr Clin Neurophysiol. 1998; 108(3):260–73. [PubMed: 9607515]

Gehring WJ, Willoughby AR. The medial frontal cortex and the rapid processing of monetary gains and losses. Science. 2002; 295(5563):2279–2282. [PubMed: 11910116]

Gino F, Ariely D. The dark side of creativity: Original thinkers can be more dishonest. Journal of Personality and Social Psychology. 2012; 102:445–459. [PubMed: 22121888]

Greene JD, Paxton JM. Patterns of neural activity associated with honest and dishonest moral decisions. Proc Natl Acad Sci U S A. 2009; 106(30):12506–12511. [PubMed: 19622733]

Hajcak G, Holroyd CB, Moser JS, Simons RF. Brain potentials associated with expected and unexpected good and bad outcomes. Psychophysiology. 2005; 42(2):161–170. [PubMed: 15787853]

Hajcak G, Moser JS, Holroyd CB, Simons RF. The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. Biological Psychology. 2006; 71(2):148–154. [PubMed: 16005561]

Holroyd CB, Coles MG. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Rev. 2002; 109(4):679–709. [PubMed: 12374324]

Holroyd CB, Nieuwenhuis S, Yeung N, Cohen JD. Errors in reward prediction are reflected in the event-related brain potential. Neuroreport. 2003; 14(18):2481–4. [PubMed: 14663214]

Holroyd CB, Yeung N. Motivation of extended behaviors by anterior cingulate cortex. 2012. Trend Cog Sci. 2012:122–128.

Hu X, Hegeman D, Landry E, Rosenfeld JP. Increasing the number of irrelevant stimuli increases ability to detect countermeasures to the P300- based Complex Trial Protocol for concealed information detection. Psychophysiology. 2012; 49(1):85–95. [PubMed: 22091554]

Hu X, Pornpattananangkul N, Rosenfeld JP. N200 and P300 as orthogonal and integrable indicators of distinct awareness and recognition processes in memory detection. Psychophysiology. 2013; 50(5):454–464. [PubMed: 23317115]

Hu X, Wu H, Fu G. Temporal course of executive control when lying about self- and other-referential information: An ERP study. Brain Res. 2011; 1369:149–157. [PubMed: 21059343]

Johnson R. A Triarchic Model of P300 amplitude. Psychophysiology. 1986; 23(4):367–384. [PubMed: 3774922]

Johnson R. On the neural generators of the P300 component of the event-related potential. Psychophysiology. 1993; 30(1):90–97. [PubMed: 8416066]

Johnson R Jr, Barnhardt J, Zhu J. The deceptive response: effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. Biol Psychol. 2003; 64(3):217–53. [PubMed: 14630405]

Johnson R Jr, Henkell H, Simon E, Zhu J. The self in conflict: the role of executive processes during truthful and deceptive responses about attitudes. Neuroimage. 2008; 39(1):469–82. [PubMed: 17919934]

Keil A, Debener S, Gratton G, Junghöfer M, Kappenman ES, Luck SJ, et al. Yee CM. Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. Psychophysiology. 2014; 51(1):1–21. [PubMed: 24147581]

Kok A. On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology. 2001; 38(3):557–77. [PubMed: 11352145]

Langleben DD, Schroeder L, Maldjian JA, Gur RC, McDonald S, Ragland JD, O'Brien CP, Childress AR. Brain activity during simulated deception: an event-related functional magnetic resonance study. Neuroimage. 2002; 15(3):727–32. [PubMed: 11848716]

Lee TMC, Liu HL, Tan LH, Chan CCH, Mahankali S, Feng CM, Hou J, Fox PT, Gao JH. Lie detection by functional magnetic resonance imaging. Hum Brain Mapp. 2002; 15(3):157–164. [PubMed: 11835606]

Lorist MM, Snel J, Kok A, Mulder G. Acute effects of caffeine on selective attention and visual search processes. Psychophysiology. 1996; 33(4):354–61. [PubMed: 8753934]

Luck, SJ. An introduction to the event-related potential technique. MIT press; 2014.

Marco-Pallares J, Cucurell D, Münte TF, Strien N, Rodriguez-Fornells A. On the number of trials needed for a stable feedback-related negativity. Psychophysiology. 2011; 48(6):852–860. [PubMed: 21091959]

Martin RS, Appelbaum LG, Pearson JM, Huettel SA, Woldorff MG. Rapid Brain Responses Independently Predict Gain Maximization and Loss Minimization during Economic Decision Making. J Neurosci. 2013; 33:7011–7019. [PubMed: 23595758]

Miltner WHR, Braun CH, Coles MGH. Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a ‘generic’ neural system for error detection. J Cog Neurosci. 1997; 9(6):788–798.

Nieuwenhuis S, Holroyd CB, Mol N, Coles MGH. Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. Neurosci Biobehav Rev. 2004; 28:441–448. [PubMed: 15289008]

Nieuwenhuis S, Yeung N, Holroyd CB, Schurger A, Cohen JD. Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. Cereb Cortex. 2004; 14(7):741–747. [PubMed: 15054053]

Nieuwenhuis S, Yeung N, van den Wildenberg W, Ridderinkhof KR. Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. Cogn Affect Behav Neurosci. 2003; 3(1):17–26. [PubMed: 12822595]

Priori A, Mameli F, Cogiamanian F, Marceglia S, Tiriticco M, Mrakic-Sposta S, Ferrucci R, Zago S, Polezzi D, Sartori G. Lie-specific involvement of dorsolateral prefrontal cortex in deception. Cereb Cortex. 2008; 18(2):451–5. [PubMed: 17584853]

Rand DG, Greene JD, Nowak MA. Spontaneous giving and calculated greed. Nature. 2012; 489(7416): 427–430. [PubMed: 22996558]

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S. The role of the medial frontal cortex in cognitive control. Science. 2004; 306(5695):443–7. [PubMed: 15486290]

Schultz W. Getting formal with Dopamine and Reward. Neuron. 2002; 36:241–263. [PubMed: 12383780]

Shalvi S, De Dreu CKW. Oxytocin promotes group-serving dishonesty. Proceedings of the National Academy of Sciences. 2014; 111(15):5503–5507.

Sip KE, Lynge M, Wallentin M, McGregor WB, Frith CD, Roepstorff A. The production and detection of deception in an interactive game. Neuropsychologia. 2010; 48(12):3619–26. [PubMed: 20727906]

Sip KE, Roepstorff A, McGregor W, Frith CD. Detecting deception: the scope and limits. Trends Cogn Sci. 2008; 12(2):48–53. [PubMed: 18178516]

Sip KE, Skewes JC, Marchant JL, McGregor WB, Roepstorff A, Frith CD. What if I Get Busted? Deception, Choice, and Decision-Making in Social Interaction. Front Neurosci. 2012; 6:58. [PubMed: 22529772]

Spence SA, Farrow TFD, Herford AE, Wilkinson ID, Zheng Y, Woodruff PWR. Behavioural and functional anatomical correlates of deception in humans. Neuroreport. 2001; 12(13):2849–2853. [PubMed: 11588589]

Talmi D, Atkinson R, El-Deredy W. The feedback-related negativity signals salience prediction errors, not reward prediction errors. The Journal of Neuroscience. 2013; 33(19):8264–8269. [PubMed: 23658166]

van der Helden J, Boksem MAS, Blom JHG. The importance of failure: Feedback-related negativity predicts motor learning efficiency. Cereb Cortex. 2010; 20(7):1596–1603. [PubMed: 19840974]

Van Veen V, Carter CS. The timing of action-monitoring processes in the anterior cingulate cortex. J Cogn Neurosci. 2002; 14(4):593–602. [PubMed: 12126500]

von Borries AK, Verkes RJ, Bulten BH, Cools R, de Bruijn ER. Feedback-related negativity codes outcome valence, but not outcome expectancy, during reversal learning. Cogn Affect Behav Neurosci. 2013; 13:737–746. [PubMed: 24146314]

Walsh MM, Anderson JR. Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. Neurosci Biobehav Rev. 2012; 36:1870–1894. [PubMed: 22683741]

Walsh MM, Anderson JR. Modulation of the feedback-related negativity by instruction and experience. Proc Natl Acad Sci U S A. 2011; 108(47):19048–19053. [PubMed: 22065792]

Warren CM, Holroyd CB. The impact of deliberative strategy dissociates ERP components related to speeded responding vs. reinforcement learning. Front Decisi Neurosci. 2012; 6:43.

Wickens C, Kramer A, Vanasse L, Donchin E. Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. Science. 1983; 221(4615):1080–1082. [PubMed: 6879207]

Yeung N, Cohen JD. The impact of cognitive deficits on conflict monitoring. Predictable dissociations between the error-related negativity and N2. Psychol Sci. 2006; 17(2):164–71. [PubMed: 16466425]

Yeung N, Holroyd CB, Cohen JD. ERP correlates of feedback and reward processing in the presence and absence of response choice. Cereb Cortex. 2005; 15(5):535–44. [PubMed: 15319308]

Yeung N, Sanfey A. Independent Coding of Reward Magnitude and Valence in the Human Brain. J Neurosci. 2004; 24:6258–6264. [PubMed: 15254080]

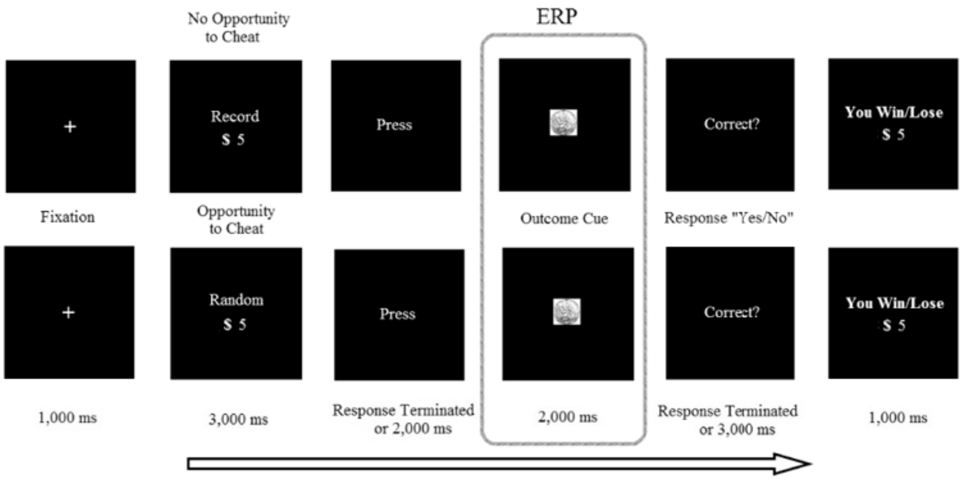


Figure 1.

The task structure of the coin-guess task. Participants made predictions for either 3 or 5 raffle tickets (only the $5 trial was shown here as an exemplar). Note for the “Press” screen participants were instructed during NoOp trials to enter their prediction as either “Heads” or “Tails” by pressing either the “H” or “T” key. During Op trials, participants were instructed to randomly press one of the two “R” keys to control for motor activity. ERPs were locked to the Outcome Cue Slide for both Op/NoOp trials (for executive control analyses) and Win/Loss trials that were coded based on their later report on the Response Slide (correct/win vs. incorrect/loss, for reward process analyses).

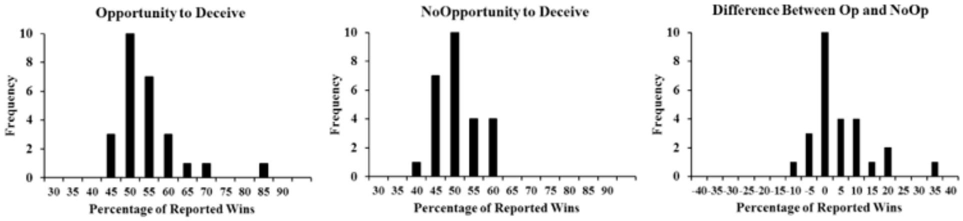


Figure 2.

Distribution of reported wins in percentages for a) Opportunity-to-deceive trials Op trials; b) No-Opportunity-to-deceive trials the NoOp trials; and c) the differences between Op and NoOp trials.

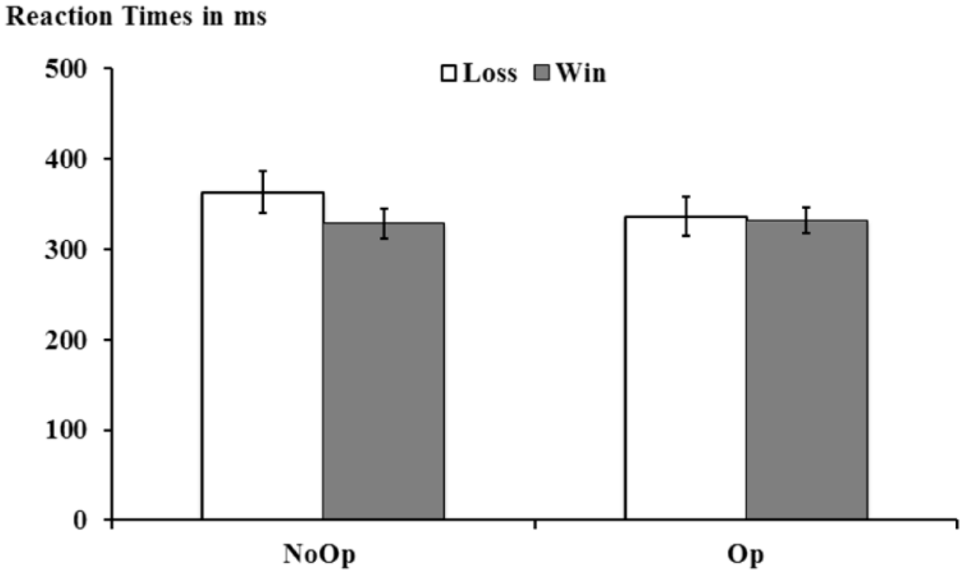
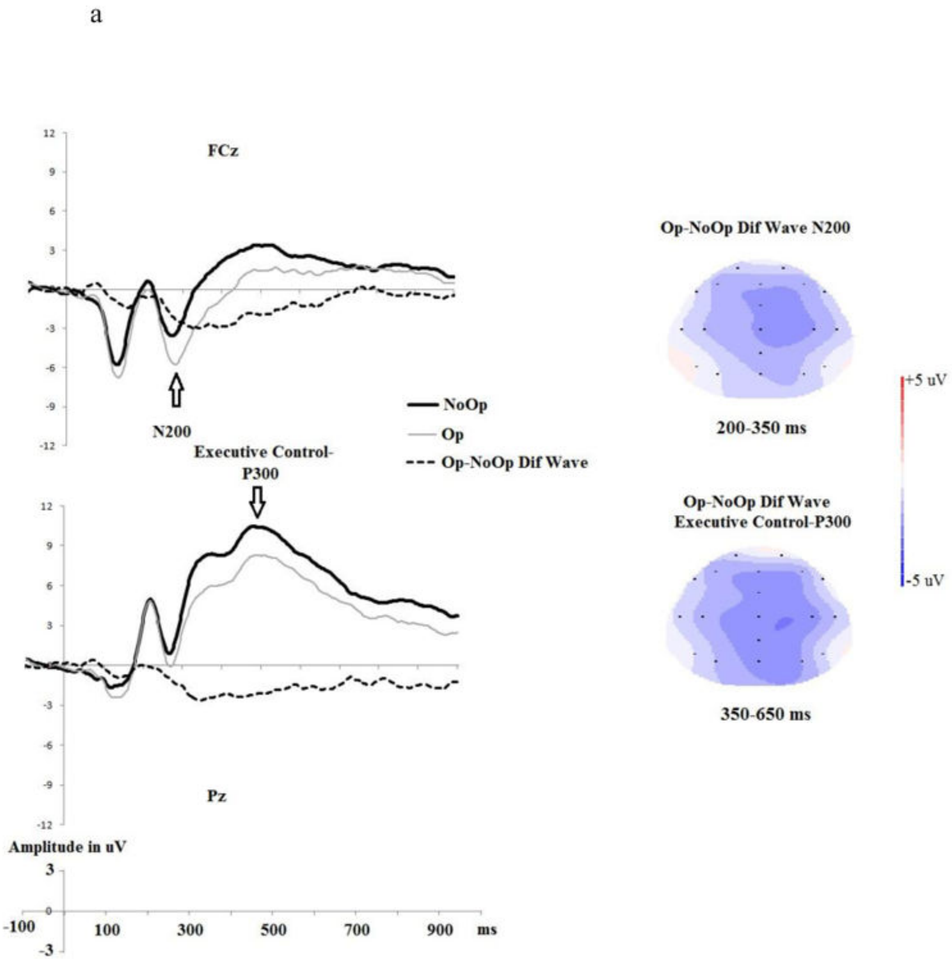


Figure 3.

RTs to “Correct?” Slide during the coin-guessing task. Error bar indicates 1 standard error.


a
FCz
12
9
6
3
-3
-6
-9
-12
N200
NoOp
Op
Op-NoOp Dif Wave
Executive Control-
P300
12
9
6
3
-3
-6
-9
-12
Pz
Amplitude in uV
3
0
-3
-100
100
300
500
700
900
ms
Op-NoOp Dif Wave N200
200-350 ms
Op-NoOp Dif Wave
Executive Control-P300
350-650 ms
+5 uV
-5 uV

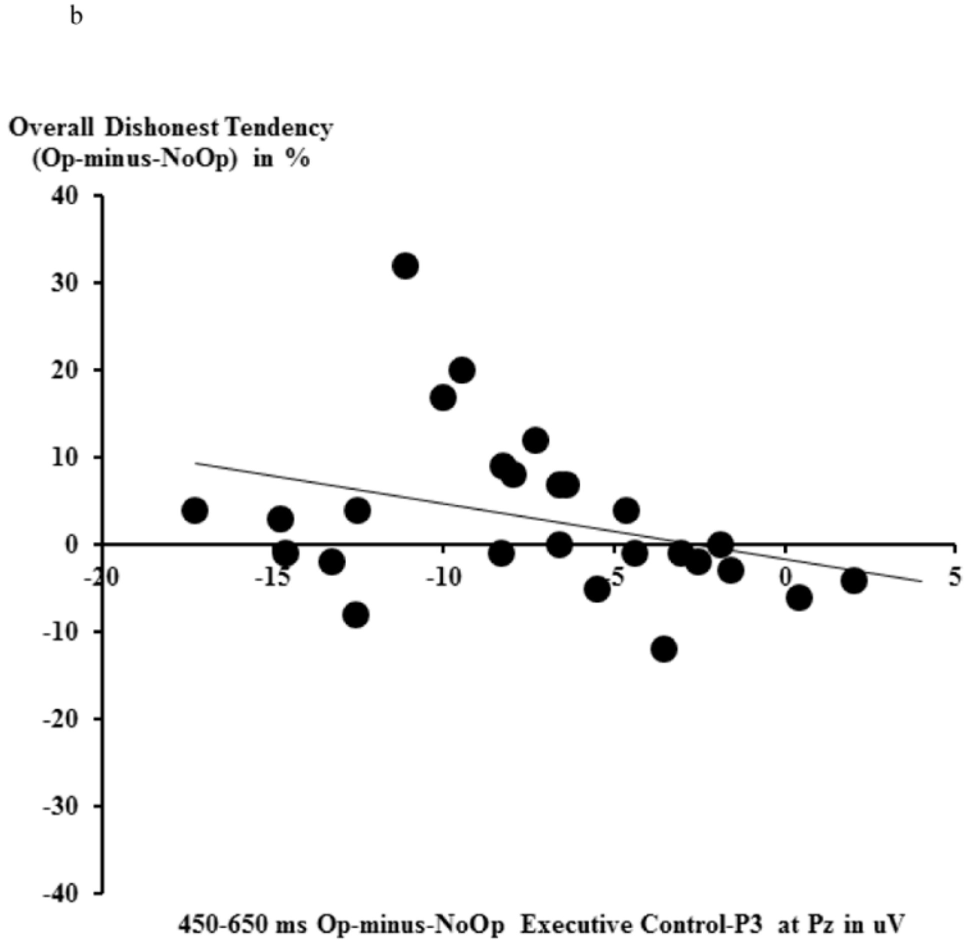
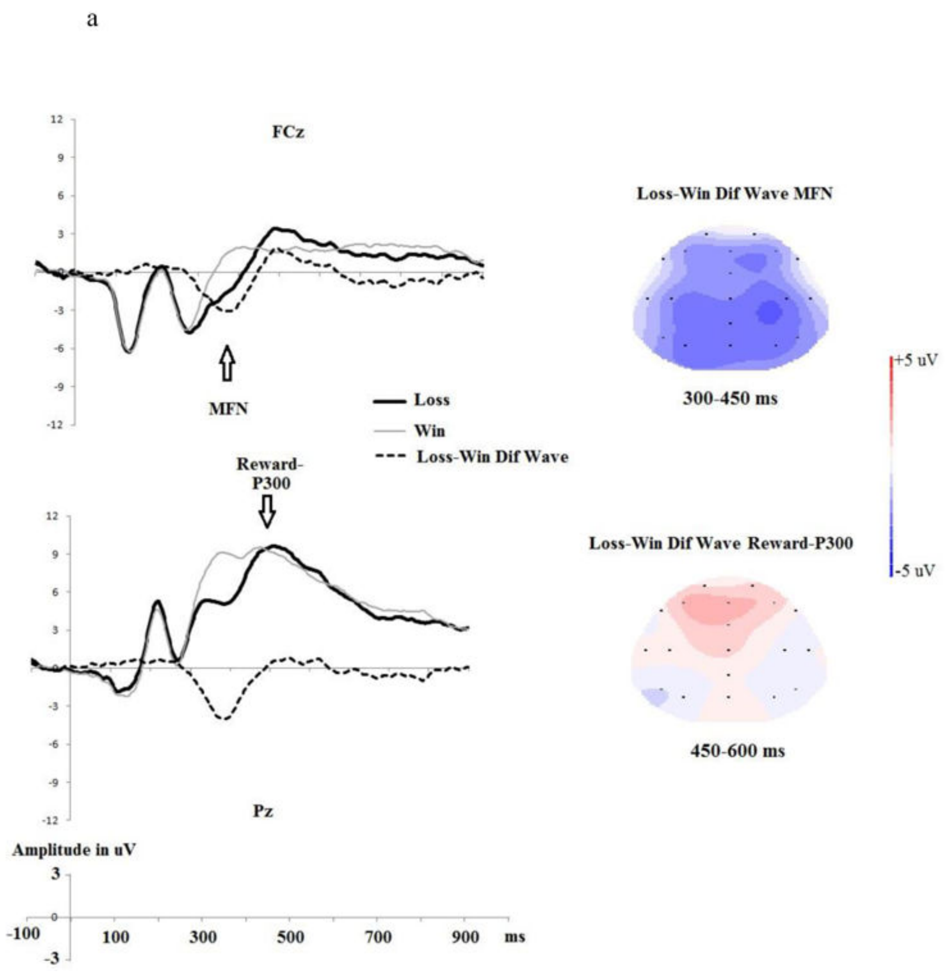


Figure 4.

a) No-Opportunity-to-deceive trials (NoOp) vs. Opportunity-to-deceive trials (Op) and the Op-minus-NoOp difference wave ERPs at fronto-central (FCz) and parietal (Pz) electrodes. The topographical map depicts the mean amplitude of the N2 and the executive control- P3 during each time window based on the grand average waveforms. b) The relationship between the Op-minus-NoOp executive control-P3 and one's overall voluntary dishonest tendencies (percentage differences in reported wins between Op and NoOp trials).


a
FCz
12
9
6
3
-3
-6
-9
-12
MFN
Loss
Win
Loss-Win Dif Wave
Loss-Win Dif Wave MFN
300-450 ms
+5 uV
-5 uV
Reward-
P300
Loss-Win Dif Wave Reward-P300
450-600 ms
Pz
Amplitude in uV
3
0
-3
-100
100
300
500
700
900
ms

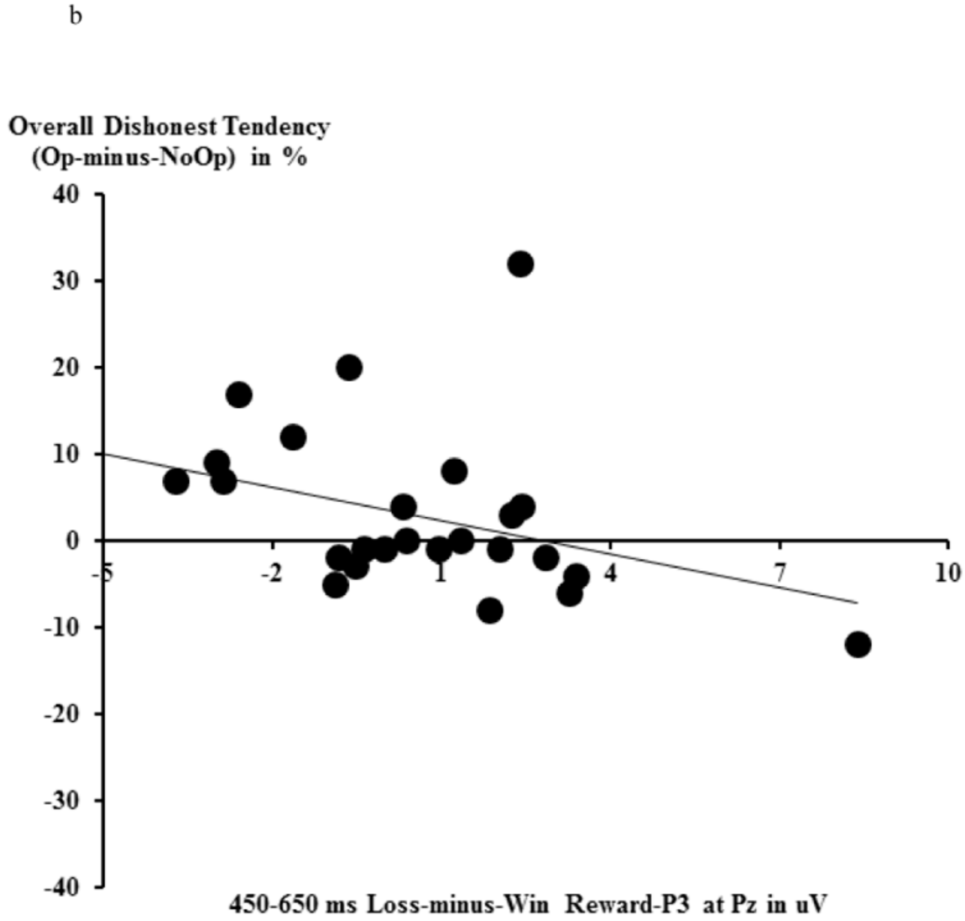
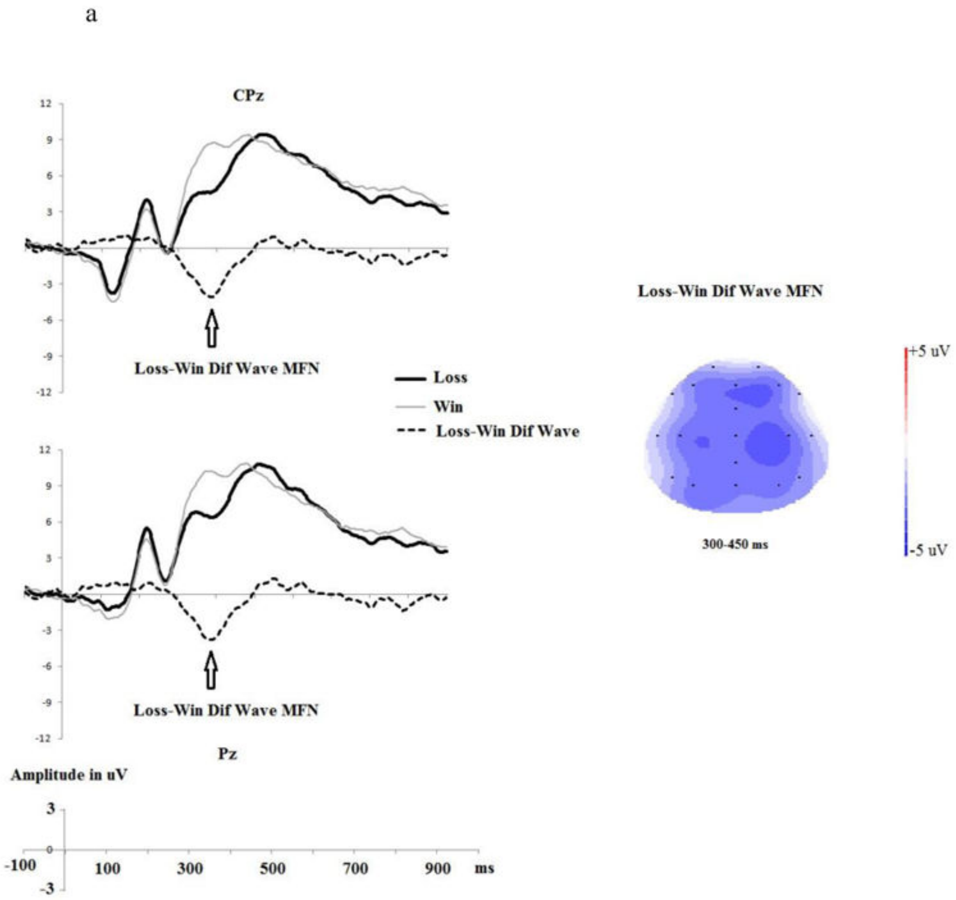


Figure 5.

a) Loss trials (incorrect prediction of the coin toss) vs. Win trials (correct prediction of the coin toss) and the Loss-minus-Win difference wave ERPs at fronto-central (FCz) and parietal (Pz) electrodes. The topographical map depicts the mean amplitude of MFN and reward P300 during each time window based on the grand average waveforms. b) The relationship between the loss-minus-win reward-P3 and one's overall voluntary dishonest tendencies (percentage differences in reported wins between Op and NoOp trials).


a
CPz
12
9
6
3
-3
-6
-9
-12
Loss-Win Dif Wave MFN
Loss
Win
Loss-Win Dif Wave
12
9
6
3
-3
-6
-9
-12
Loss-Win Dif Wave MFN
Pz
Amplitude in uV
3
0
-3
-100
100
300
500
700
900
ms
Loss-Win Dif Wave MFN
+5 uV
-5 uV
300-450 ms

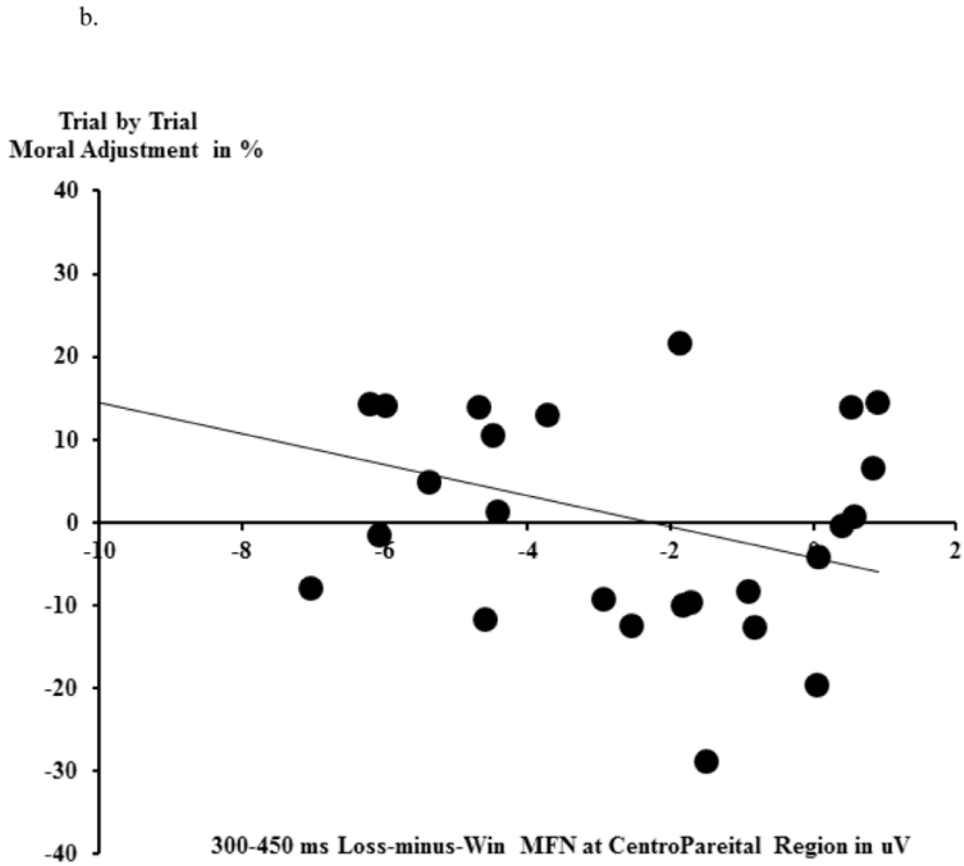


Figure 6.

Moral behavioral adjustment analyses. a): Within NoOp trials, Loss trials (incorrect prediction of the coin toss) vs. Win trials (correct prediction of the coin toss) and the Loss-minus-Win difference wave ERPs at centro-parietal (CPz and Pz) electrodes. The topographical map depicts the mean amplitude of MFN during the 300-450 ms time window based on grand average waveforms. b) the relationship between Loss-minus-Win MFN (within NoOp trials) and one's moral behavioral adjustment (in %). A larger Loss-minus-Win MFN reflects a larger reward prediction error.