Event-Related Potential Components Associated with the Preparation and Execution of Self-Motivated Deception within a Morally Accountable Context

Nolan B. O'Hara

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Psychology

from the University of Michigan

2013

Advisor: Dr. William Gehring

Abstract

Recent neuroimaging research has attempted to deconstruct the cognitive components of dishonest behavior, but it often confines itself within the arena of instructed deceit and thereby fails to observe deceptive activity that bears genuine moral relevance. We report on an ERP analogue of an fMRI study in which participants are not told to lie, but rather choose to lie of their own volition after realizing that our experimental structure can be exploited for dishonest monetary gain. Subjects who were willing to act deceptively in this morally accountable atmosphere showed distinctive ERP responses preceding potential lies. Specifically, stimuli about which a dishonest participant was able lie elicited more negative Feedback-related negativities and less positive P3 waveforms. The observed patterns of activity may point to the importance and detectability of affective and motivational factors that precede real-world deception.

Keywords: deception, lie detection, moral cognition, moral culpability, temptation, feedback-related negativity, late positive component, contingent negative variation, error-related negativity

Event-Related Potential Components Associated with the Preparation and Execution of Self-Motivated Deception within a Morally Accountable Context

The proverbial fork in the road that is reached during opportunities for dishonest gain offers widely diverging paths: honest and dishonest behaviors are fundamentally built upon different goals, and much recent neuroimaging research has shown that they also exhibit distinguishable patters of brain activity (for reviews, see Abe, 2011; Christ, Van Essen, Watson, Brubaker, & McDermott, 2009). Explanations for such differences are often rooted in the idea that being truthful is less demanding than behaving deceptively, and that honesty is the default, or pre-potent, behavior during simple response tasks (Greene and Paxton, 2009; Walczyk, Roper Seemann, & Humphrey, 2003). In asking what precisely qualifies behavior as deceptive, however, the field shows less consistency. Though most neuroimaging research on the subject seems to assume semantically similar definitions of deception, there is a critical discrepancy between these definitions and the experimental contexts in which research participants produce lies. In a way, the nature of a deception-based experimental paradigm defines any given study's meaning of "deception" much more than the explicit definition it provides. Though the environment of natural deceptive behavior must be somewhat simplified for the sake of experimental control, extreme care must be taken to ensure that it is not getting simplified along dimensions that are directly relevant to the idea under study. The topic of deception is certainly a moral one, and consequently deception research must comprehensively ensure that behavior under study still embodies the morally culpable actions with which our ethical and legal systems are concerned. Revisiting previous neuroimaging studies of deceptive behavior with this concept in mind will provide an insightful foundation from which to approach analysis of the present study.

On a very broad level, nearly all studies of deception investigate the production of a response that is known to be counterfactual. This discrepancy between what a liar understands as truth and what a liar communicates as truth constitutes the most basic condition necessary for defining deception. The neural activity of this basic deception has been studied using neuroimaging techniques such as electroencephalography (EEG) (Johnson, Barnhardt, & Zhu, 2003; Johnson, Barnhardt, & Zhu, 2004; Hu, Haiyan, & Fu, 2011) and functional magnetic resonance imaging (fMRI) (Spence et al., 2001; Lee et al., 2002; Ito et al., 2012), and these data suggest that deception is associated with increased activity in processes of executive control and working memory. The goal of such research and the analysis of its data are relatively straightforward: they seek to characterize neural signals that the mind necessarily generates when managing differences between what is known and what is communicated (Ganis & Keenan, 2009). Although the processing and management of lie-relevant information implicated here is certainly a large part of deception, these counterfactual productions are orphaned from a liar's willful decision to produce a lie. In order to draw relevant conclusions about deception, it is essential to know if the volition to lie alters the cognitive control activity associated with a deceptive act.

In this vein, some recent neuroimaging research has studied deceptive behavior in which a participant must not only execute a deceptive response but also generate the circumstantial intent to do so. Maxim Kireev (2007, 2012, 2013), Shi-Yue Sun (2011), and their respective colleagues have been utilizing paradigms in which participants deliberately attempt to "trick" a computer program that judges the truth of their response. Participants in these studies are presented with a stimulus, and then subsequently choose to report what they saw with a strategically truthful or strategically deceitful response. The neural activity that is recorded when

a participant attempts to deceive the computer, then, represents the combined cognitive efforts of purposely deciding to lie and also executing a deceptive response. In these conditions, data from EEG (Kireev et al., 2007; Sun et al., 2011), fMRI (Kireev et al., 2012), and positron emission topography (PET) (Kireev et al., 2013) all suggest that the activity of various executive processes can distinguish a participant's honest and dishonest behavior. Importantly, studying the volition to lie broadens the analytical focus to look beyond the exact time of response execution and consider the potentially deception-motivating events that precede it.

However, while the incorporation of this relevant data does help paint a fuller picture of deceptive cognition, the decision to lie in this context lacks any element of moral accountability. A participant playing "trick-the-computer" types of games understands that lying is a sanctioned and even encouraged type of performance for which they cannot be held culpable. This is quite unlike the natural world, in which legal jurors, social peers, and believers in free will alike acknowledge that individuals are accountable for their deliberate actions to some extent. Insofar as deception research aims to develop technologies and broaden understanding within these ethically oriented fields, it is paramount to understand how executing a lie not only with intent, but also with morally accountable intent, affects the cognition surrounding deception.

Motivated by this sentiment, Joshua Greene and Joseph Paxton (2009) utilized a novel paradigm to study self-motivated and morally relevant deceptive behavior. Rather than observing activity that was explicitly directed to be honest or dishonest, Greene and Paxton simply had participants participate in a financially incentivized heads-or-tails guessing game, and at times they made participants responsible for reporting their own success or failure. This responsibility was explained to the participant as a way to investigate private prediction abilities, but in truth its purpose was to make dishonest gain possible. They found that different participants took

advantage of this cheatable system to varying degrees, with some reporting correct guesses far more often than was statistically feasible. The proportion of self-reported successful guesses was used as an index of probable deception, and participants were categorized as those who were likely to have behaved honestly and those who were likely to have lied at times throughout the task. Since they do not realize that deception is the focus of the study, Dishonest participants must break the rules of the game in order to dishonestly report a successful guess. In so doing they forgo the experimentally dictated mindset, which means that their deception transcends the limits of experimental sanction and is a deliberate, naturally motivated, morally culpable choice.

The ability to experimentally observe this natural deception is entirely facilitated by the novel, albeit counterintuitive, acceptance of uncertainty regarding the identity of any specific response as honest or dishonest. Data cannot be labeled as honest or dishonest, but rather are linked to a participant who is probabilistically categorized as "Dishonest" if they show a willingness to deceive the researchers. Caution should therefore be exercised when comparing these results to previous neuroimaging studies of deception that can identify specific deceptive actions. Nevertheless, the data still permit very meaningful and morally oriented analysis. Greene and Paxton offer insight by speaking of honesty as a routine behavior rather than an isolated action: those willing to break the rules exhibit fundamentally different set of motivations than honest group, who knowingly "leave money on the table" (Greene & Paxton, 2009). Moreover, Greene and Paxton show differences in fMRI activity between "Honest" and "Dishonest" participants that do, in fact, replicate many of the neural activation patterns observed in previous studies of instructed deceit. This lends credence to the idea that their probabilistic classification of participants at the individual level can still capture the general physiological effects associated with a single deceptive action. More importantly, it suggests that

unique neural activity exhibited by Honest or Dishonest participants during this paradigm could be attributed to distinct, morally relevant processes, or to considerations of moral accountability. Here, we report on an event-related potential (ERP) analog of Greene and Paxton's study, which demonstrates the EEG response during this paradigm and all its opportunities for self-motivated, morally accountable deception. The high temporal resolution of EEG data, along with the large amount of research that has characterized the following ERP components during both deceptive tasks and basic cognitive tasks, should allow a very thorough cognitive analysis of natural deception and its associated moral decisions.

Response Conflict and The ERN

Greene and Paxton's results showed that, compared to Honest participants with statistically feasible self-reported win rates, those who were categorized as Dishonest had increased activation of the left dorsolateral prefrontal cortex (DLPFC), dorsomedial prefrontal cortex (DMPFC) and anterior cingulate cortex (ACC) when self-reporting their own success or loss. The ACC was specifically implicated when dishonest participants willingly reported an incorrect guess, or in other words, when they offered a truthful admission in a context where they are willing to tell a lie. ACC activity is often associated with response conflict and the detection of errors (Gehring, Liu, Orr, & Carp, 2012; C. Kim, Kroger, & J. Kim, 2011), and past research shows evidence for the involvement of these processes during a deceptive response. This cognitive commonality seems intuitively reasonable, considering the similarities of erroneously doing one thing while intending to do another and deceptively doing one thing while acknowledging the truth of another (Kireev et al., 2012). However, there is fMRI evidence that suggests slight differences in brain activity during unintentionally erroneous responses and instructed deceptive responses (Lee et al., 2009), so the involvement of a cognitive error-detector

during deception may not be so simple. One prominent theory of the ACC posits that a cognitive error-detector does not sense some abstract quality of "incorrectness," but rather is sensitive to the experience of conflict between an intended correct response and a committed erroneous response (Gehring et al., 2012). This theory can provide further reasonable explanations of ACC involvement in deception, as conflict may be detected when a self-interested dishonest response competes for representation with a pre-potent truthful one (Greene & Paxton, 2009; Johnson et al., 2003; Walczyk et al., 2003). Alternatively, the conflict may be more emotional in nature, and activity in the ACC could represent conflict between mutually exclusive feelings towards the execution of a deceptive response (Baumgartner, Fischbacher, Feierabend, Lutz, & Fehr, 2009).

Focusing on ERP data, the activity of this ACC error detector is typically indexed by the error-related negativity (ERN), a negative-going deflection of EEG activity near the center of the scalp that is observed in trials where a participant gives an erroneous response (for a review, see Gehring et al., 2012). Despite the previously mentioned differences between fMRI activity during instructed deceit and fMRI activity during unintentional error commission (Lee et al., 2009), ERP studies of deception have shown ERN-like activity time-locked to the moment that participants execute instructed counterfactual responses (Johnson et al., 2003, Johnson et al., 2004). Moreover, Kireev et al. (2007) have shown that ERN-like activity is also at work during the self-motivated deceptions that participants execute during purposeful attempts to trick a computer. Taken together, these findings suggest that cognitive response conflict occurs during deceptive responses regardless of whether the lie is superficially motivated by instruction or willfully motivated by the participant's volition. The present study hopes to determine if the addition of moral accountability to the deceptive environment has any affect on the conflict of a deceptive response or on ERN activity.

Lie Preparation Activity and the Advantage of ERP Components in its Analysis

The study of self-motivated, morally accountable behavior focuses not only on deceptive action, but also on the complete deceptive process, and as a result the neural activity surrounding any event capable of influencing deception-related decisions is relevant. Data from past fMRI research, unfortunately, lack the temporal resolution to finely distinguish stimulus-provoked lie preparation activity and response-associated lie execution. In order to understand the decision to lie in a morally relevant context, Greene and Paxton's findings of DLPFC, DMPFC, and ACC activity must be attributed specific events during the deceptive process. One recent fMRI study (Ito et al., 2012) circumvented this problem by cueing participants to lie about a stimulus either prior to or along with the presentation of the lie-provoking stimulus, and found an increased response of the ACC and DLPFC in the former, lie-preparation condition. Furthermore, a recent study (Ding, Gao, Fu, & Lee, 2013) utilized functional near-infrared spectroscopy (fNIRS), a technique that measures the same hemodynamic response as fMRI with better temporal resolution, to replicate Greene and Paxton's paradigm and more finely parse the cognitive timeframe of morally relevant deception. Their results, like the results of Ito et al., illustrated the significant contributions of DLPFC activity following the presentation of a stimulus that might later be lied about. These findings suggest a significant involvement of cognitive control preceding the execution of deception, and also illustrate the merit of studying the response to lie-preceding stimuli, but it is evident that ERP methods could be better suited to this task. EEG data are sensitive to the activity of deception-relevant cortices such as the ACC that are too deep for fNIRS to fully detect, and EEG data exhibit the fine temporal resolution needed to distinguish activity of the following stimulus-locked ERP components.

Feedback Processing and the FRN. The feedback-related negativity (FRN) is a centrally located, negative-going deflection of EEG activity that occurs when a participant receives negative feedback, and is thought to reflect an ACC process related to the one that generates the ERN (Gehring et al., 2012). It is prominent not only when feedback explicitly indicates a loss but also when an unexpected outcome is presented (Gehring, Gratton, Coles, & Donchin, 1992), though most choices imply an expected affirmation of correctness that renders these two situations nearly equivalent. In the context of Greene and Paxton's paradigm, this means that the FRN may show how expected a certain result is to a participant before they report on it. Regardless of how those willing to deceive actually respond, they necessarily appreciate the unexpectedness of a lie-prompting outcome. In this line of thought, previous research of instructed deceit has shown an enhanced FRN response to stimuli when participants are expected to respond to it dishonestly (Johnson et al., 2004; Hu, Haiyan, & Fu, 2011). In tasks where a deceptive action is volition-driven but not morally relevant, however, the FRN response to lie-preceding stimuli has not been detected (Sun et al., 2011, Kireev et al., 2007).

Despite this finding of non-significance, there is some reason to expect FRN responses to lie-preceding stimuli during morally relevant deception, because FRN amplitude is also sensitive to more affective conditions than outcome valence. Individuals who show a more negative trait affect tend to show a larger FRN (Sato et al., 2005), and when there is more at stake during a choice, larger FRNs are likely to result (San Martín, Manes, Hurtado, Isla, & Ibañez, 2010). Given the motivational and emotional factors that contribute to the FRN, receiving feedback that motivates a morally accountable intent to deceive may inherently produce a distinctive ERP: a strongly negative affective response seems capable of driving a choice to forgo virtue and tell a lie.

Context Updating and the P3. Participants who lie during the Greene and Paxton paradigm are also expected to show distinctive cognition following the FRN-related recognition that an outcome has been incorrectly predicted. They must next appreciate a difference between the perceived state of their loss and the preferred state of their success, and they must acknowledge that deception is now the necessary course if they want to reach their preferred successful state. The P3, a centro-pariatally maximal ERP component that peaks around 300 milliseconds after a stimulus (Polich, 2012), may be useful in identifying this circumstance. It, like the FRN, exists in a temporal space where cognitive processing beyond basic sensory input can begin, and it is classically attributed to an updating of a mental representation or context (Polich, 2012). More precisely, the deflection is thought to reflect a neural inhibition that facilitates this updating process by focusing attention on the stimulus (Polich, 2012).

When someone interprets a stimulus as an indication that they must deceive, they would presumably engage their attention in a distinctive way in order to arrange their deceptive plans. Attributing their result to such an engagement of executive processes, Johnson et al. (2003) found a decreased P3 response following a stimulus that participants were instructed to subsequently lie about. They speculated that attentional and processing resources are taken away from the default honest response when lying, which caused the decrease in P3 amplitude. Interestingly, participants who were executing volition-driven lies in order to trick a computer program during the studies conducted by Kireev et al. (2007) and Sun et al. (2011) did not show P3-like activity in response to lie-preceding stimuli. The experimental sanction of deceptive behavior differentiates these paradigms from Greene and Paxton's, however, and the present study will investigate if the relevance of moral accountability when making a deceptive response affects P3-generating activity.

Response Preparation and the CNV. Deception requires that liars mentally construct a fictitious state of affairs, and then further requires that they maintain awareness of both their mental invention and the world as they truly know it. With this in mind, many studies of deception discuss the cognitive strain brought about by telling a lie (Abe, 2011; Christ et al., 2009), and it seems reasonable to expect that some sort of cognitive preparatory activity helps to preserve this strenuous cognitive maintenance throughout a deceptive response. The contingent negative variation (CNV) is a slow-going negative deflection at frontal electrode sites that is generally observed in preparation of a response: after a participant is warned of an upcoming need to respond, the CNV emerges and is sustained until the actual presentation of a response-warranting stimulus (Luck, 2005). The CNV is partially related to motor preparation (Luck, 2005), but research suggests a more motor independent, cognitive nature to the component (Damen and Bruna, 1987).

Moreover, the CNV has shown sensitivity to motivation (Rebert, McAdam, Knott, & Irwin, 1967), the need for additional feedback (Picton & Lowe, 1971), and other clues that might be telling of the response being prepared. Previous studies of instructed deceit have taken notice of these cognitive sensitivities, and found that heightened CNV-like activity similarly preceded counterfactual responses when participants were instructed to lie (Fang, Liu, & Shen, 2003). Even in a game-like setting, where there may be more variation in the timing and strategy of any participant's self-motivated deceptive process, Sun et al. (2011) showed distinct CNV activity that preceded dishonest responses. Despite this finding, it is important to note that participants in such studies are still aware of the focus on their deception, and CNV amplitude seems very dependent on attention and common motivation between participant and researcher. Considering

this, it will be interesting to see if a significantly different CNV precedes the self-assessments of participants who decide to deceive in our morally relevant context and those who do not.

Behavioral Expectations

Many studies of deception have found slightly increased reaction time when comparing deceptive responses to truthful ones (Abe, 2011), and in explaining such results they often appeal to the same previously described cognitive strain used to justify CNV activity. Such results are not perfectly consistent throughout deception literature, however, and the significance of response time differences have shown sensitivity to the content of a lie (Mameli et al., 2010), and studies of self-motivated deceit in the context of a computer game have found non-significant differences in dishonest and honest reaction times (Kireev et al., 2007; Sun et al., 2011). Greene and Paxton (2009) found that, on average, participants who were executing self-motivated and morally culpable deceptive behavior by deliberately over-reporting their own success did not show longer reaction times when doing so. However, they did take significantly longer to self-report a failed prediction than Honest-categorized participants did. In other words, Greene and Paxton found that admitting loss took a much longer time if one was already willing to lie in the experimental context. Given the extreme similarity between Greene and Paxton's paradigm and the paradigm of the present study, we expect to replicate their findings of reaction time.

Method

Participants

The participants of this study were 30 undergraduate college students (11 men, 19 women) recruited by flyers posted on the University of Michigan campus. They ranged in age from 18 to 26 years old, had normal or corrected-to-normal vision, were all right-handed, and had no history of psychiatric disorder. On the day of participation, they confirmed that they were

not on any type of psychiatric medicine and had not recently ingested any possibly confounding stimulants such as caffeine. Participants were compensated twenty dollars for about two hours of participation, and were able to earn up to thirty additional dollars depending on their task performance. The Institutional Review Board at University of Michigan approved the study (IRB Number HUM00051463).

Apparatus, Stimuli, and Procedure

Participants were in a soundproof and dimly lit booth sitting approximately 1 meter from a 14-inch computer monitor on which the stimuli were presented. They responded to the stimuli by using the index finger of their left and right hands to press one of two buttons on a low-profile Apple keyboard that correspond to respective left and right response choices within the stimuli. Specifically, the "Z" key was used to represent leftmost response choices and the forward slash ("/") key was used to represent rightmost response choices. The exact timing of these responses is recorded along with the exact timing of stimulus presentation.

Participants completed 180 trials where they predicted the result of a computerized coin flip (see Figure 1). Prior to each prediction, a value that ranged from $0.25 to $2.00 was presented as a wager. Participants were credited this amount if their prediction was correct and were penalized by this amount if their prediction was incorrect. After the wager was displayed for 1.5 seconds, a stimulus requesting a prediction of heads or tails was displayed until a participant responded properly. 90 of the 180 trials were categorized as "No-Opportunity" trials, where the participant was directed to explicitly select heads or tails by pressing either the left or right button that corresponded to their choice. The other 90 of the 180 trials were categorized as "Opportunity" trials, in which a participant was still instructed to make an explicit prediction for the coin flip outcome in their head, but instead of publically recording their choice they pressed

both left and right buttons to advance and kept their prediction private. After a two second fixation cross, the randomized coin flip outcome was presented. This outcome remained on the screen for 1.5 seconds before a new stimulus appeared requesting a self-report of predictive success. After a participant confirmed or denied that their prediction was correct, a fixation cross again appeared for 2.0 seconds. Afterwards, the words "correct" or "incorrect" appeared on the screen for 2.0 seconds. Then, the reward or penalty of the wager was applied and summarized for the participant as a 1.0 second presentation of a "running total." These additional winnings, which could not fall below $0.00 or exceed $30.00, were given to the participant at the end of the experiment in addition to their compensation for participating. The left-right orientation of responses ("HEADS" vs. "TAILS" and "YES" vs. "NO") was randomized for each trial. All stimuli were programmed and presented using Presentation ® software (Version 14.9, www.neurobs.com).

Participant Instructions and Cover Story

Participants were falsely led to believe that they were taking part in a study of neurally "hard wired" social awareness and memory formation. The researcher explained that, according to recent research, very short-term memories are maintained in neurologically distinct ways depending on whether the matter being remembered was public ("I know what the others know") or private ("I alone possess this information"). The researcher then speculated that the origins of this neural distinction likely coincided with the social evolution of animal herds, and that they granted individuals the survival benefits of group member specialization before sophisticated communication methods. With this neural difference, he explained, individuals were able to determine if a problem could be solved by anyone in their group or if it must be solved by them and their particular knowledge set. The researcher continued by saying that they were interested

in how this neural difference appears in people today, and how it varies with personal resolution and decision making. Consequently, the experiment at hand would require the participant to make a series of public and private decisions while EEG data was collected. This cover story differs from the "ESP study" ruse used by Greene and Paxton (2009), and it was used because it seemed less likely to arouse participant suspicion concerning the focus of the study. Importantly, our social evolution cover story still makes it clear to the participant that their cognitive cooperation is critical to the success of our laboratory's intentions.

Participant Categorization and Debriefing

Following the example of Greene and Paxton, participants were categorized as a function of their Opportunity-trial-specific win rate. A one-tailed binomial test compared each participant's win rate during opportunity trials to the 45 out of 90 wins expected from a random-guess strategy. Individuals reporting improbably high levels of accuracy ($p < 0.005$) displayed a willingness to deceive the researchers that presently classifies them as "Dishonest," even though we cannot confidently say how often they actually lied. A debriefing questionnaire (see Appendix A) verified that participants were aware of the possibility to cheat and thus ensured that any exhibition of honesty was deliberate and not due to ignorance. The questionnaire also asked participants how often they felt tempted to falsely report a successful guess, and solicited the participant's estimate of how often they actually exhibited dishonest behavior. All participants in the Dishonest group necessarily admitted to falsely reporting wins to some extent or another, but there were several participants with disproportionately high win rates in the Opportunity trials that also reported relatively low instances of cheating. Some may have not had a good memory of how often they actually cheated throughout the hour-long task, or they may have underreported the amount that they cheated to avoid embarrassment.

Physiological (EEG/EOG) Method

EEG data were collected from 64 Ag/AgCl electrodes embedded in an elastic cap using the ActiveTwo BioSemi system and two additional electrodes on each mastoid. Electrooculographic (EOG) data from eye movements were also collected from 6 electrodes on the face: two about 1 centimeter above and below each of the participant's eyes and two about 1 centimeter lateral to each eye. A feedback loop was formed using a Common Mode Sense (CMS) active electrode and a Driven Right Leg (DRL) passive electrode to drive up the average potential of the participant. All data were sampled at 512 Hz and then resampled at 256 Hz after recording.

After recording was complete, the continuous EEG data from each participant were inspected for artifacts. Sections of the recording that were obviously affected by task-irrelevant movement or contained excessive alpha waves (10 Hz oscillations) were removed. EOG data were used to correct for eye-movement artifacts in the EEG data using a version of the eye movement correction procedure developed by Gratton, Coles, and Donchin (1983). Data from all electrode sites were referenced to averaged right and left mastoid activity and then band-pass filtered using 0.1 Hz and 50 Hz cut-offs. Recordings from different trials were separated and categorized as Opportunity Win trials, No-Opportunity Win trials, Opportunity Loss trials, or No-Opportunity Loss trials.

To produce stimulus-locked waveforms, epochs of EEG data were isolated beginning 500 milliseconds before the presentation of the coin flip outcome ("HEADS" or "TAILS") and ending 2.0 seconds afterwards. For response-locked waveforms epochs of EEG data were isolated beginning 500 milliseconds before the button-press that indicates a correct or incorrect response and ending 1.0 seconds afterwards. After the trials have been epoched, these data were

again inspected by eye to ensure that the data processing methods described above were successful. At any given electrode site, activity during epochs of a certain trial type are then averaged for the participant. These data are graphed using a combination of MATLAB and EEGLAB (Delorme & Makeig, 2004) software to create the waveforms seen in Figures 2 and 5. The topographic scalp plots seen in Figures 3 and 4 were also generated using a combination of MATLAB and EEGLAB (Delorme & Makeig, 2004), and represent average electrode activity over the course of the time range indicated below each plot.

The 200 millisecond period preceding the time-locking event was used as a baseline to compare EEG data from trial to trial. For the present study, the amplitudes of various ERP components in each trial were defined as the mean amplitude of a participant's EEG data over a characteristic latency period. The FRN was defined as the mean amplitude from 200 milliseconds to 300 milliseconds after the presentation of the coin flip outcome. The P3 immediately follows, and was defined as the mean amplitude from 300 to 400 milliseconds post-stimulus. The CNV was defined as the mean amplitude during the 500 milliseconds that preceded the presentation of the stimulus requesting a self-report. Finally, the ERN was calculated as the mean amplitude during the 100 milliseconds that followed a button press signifying a self-reported win or loss.

Results

Survey Data and General Response Tendencies

One participant was excluded from all of the following analysis and behavioral analysis because he suspected that our study intended to focus on dishonesty. He admitted during debriefing that this realization influenced his behavior, and therefore his responses do not reflect the type of unsanctioned deception we intend to study. After this exclusion, there were 16 participants categorized as "honest" and 13 participants categorized as "dishonest." For the most part, Dishonest participants seemed to be more or less aware of how much they over-reported success throughout the task, because there was a relatively strong correlation between the Dishonest participants' win rate during opportunity trials and their estimation of how often they cheated on a scale of one to ten (r = 0.755, n= 13, p = .002) during these trials. The correlation between the Opportunity trial win rate for all participants and their self-reported frequency of temptation to cheat was also significant (r = 0.647, n = 29, p <.001), and showed that Dishonest participants tended to report a much higher temptation to cheat than the Honest participants did.

Reaction Time Data

The mean reaction times taken to confirm a prediction as correct or incorrect are summarized in Table 1. Since the deceptive behavior of interest can only exist during Opportunity trials and should only be exhibited by the Dishonest group, the relationship between Group and Condition was assessed. Reaction times were subjected to a 2 (group: Honest vs. Dishonest) × 2 (condition: Opportunity vs. No-Opportunity) ANOVA. Across all participants, there was a main effect of Group (F(1, 28) = 5.220, p = .030), showing that the Dishonest group took less time to confirm and deny the success of their predictions than the Honest group did. No main effect of trial condition (F(1, 28) = 0.967, p = .334) was found, but there was a significant

interaction between Group and Condition, $F(1, 28) = 4.404$, $p = .045$. This suggests that reaction times for the Dishonest group and Honest group were affected differently by the Opportunity and No-Opportunity conditions.

Taking this interaction into account, we specifically compared the reaction times of Opportunity Win trials to No-Opportunity Win trials within each group using one-way ANOVA. Comparing the Honest group's reaction times found no significant difference, $F(1, 14) = 1.801$, $p = .201$. The same data from the Dishonest group, however, did show a significant 40ms difference in reaction time ($F(1,11) = 5.420$, $p = .040$), with self-reported successes taking longer during Opportunity trials. During Loss trials, there was no significant difference in how long the Honest group took to report a loss during the Opportunity condition as compared to the No-Opportunity condition, $F(1, 14) = 1.063$, $p = .320$. We found that the Dishonest group, in contrast to the findings of Greene and Paxton, also showed no significant difference between reaction time during Opportunity Losses and No-Opportunity Losses, $F(1, 706) = 1.502$, $p = .246$.

In an attempt to understand what moderates these differences in opportunity-trial reaction time, reaction times from the dishonest group's Opportunity Win trials and reaction times from the honest group's opportunity loss trials were further analyzed with respect to the kind of trial that preceded it. These data are presented in Table 2. The reaction times were subjected to a 2 (preceding condition: Opportunity vs. No-Opportunity) × 2 (preceding outcome: Win vs. Loss) between-subjects ANOVA. There was a significant interaction between the preceding trial condition and preceding trial outcome which affected the time that Dishonest participants took, on average, to self-report a win $F = 5.464$, $p = .020$. This showed that Dishonest participants reported wins on Opportunity trials more slowly if the previous Opportunity trial was reported as

a win rather than a loss. Moreover, this effect was not seen in instances of self-reported success that were preceded by a No-Opportunity trial. For the ANOVA that was run on the Honest group's No-Opportunity loss trials, there was a significant main effect of previous trial condition (Opportunity or No-Opportunity), $F = 4.144$, $p = .043$. This showed that if Opportunity trials were preceded by Opportunity conditions, Honest participants tended to report their losses more quickly than if the trial was preceded by a No-Opportunity condition.

ERP Data

Two of additional participants from the early stages of the experiment were excluded from ERP analysis due to a hardware issue with a mastoid electrode, which rendered the EEG data of these participants unusable. After removing these two participants, 12 remained in the Dishonest group and 15 remained in the Honest group. Mean amplitude measurements were analyzed separately for each component of interest. Data from win trials and loss trials were also analyzed independently. In order to examine any conditional differences in the topographic distribution of EEG activity, the data were subjected to a 2 (condition: Opportunity vs. No-Opportunity) × 3 (electrode anteriority) × 5 (electrode laterality) within-subjects ANOVA with group (Honest vs. Dishonest) as a between-subjects factor. Of particular interest to this research are interactions between condition and group, which would indicate that ERP response to the opportunity for dishonest gain showed particular differences in Dishonest subjects, and consequently suggest that the ERP component under study is somehow implicated in deceptive cognition. When such an interaction was found, the recordings from electrodes that best represented the ERP component under study were sometimes analyzed further in planned comparisons, which are summarized in Tables 3 and 4.

ERN Amplitude. Subjecting averaged EEG data over the ERN time frame to the four-way ANOVA test showed a significant interaction between electrode anteriority and laterality, $F(2, 38) = 0.037$, $p = .037$. Across all participants, this interaction manifested as a maximally negative peak at central electrode Cz. However, the interaction effect of Opportunity/No-Opportunity trial condition and Honest-Dishonest group classification on ERN amplitude was non-significant, $F(1, 19) = 3.402$, $p = .081$. Consequently, the waveforms of different Honest-categorized and Dishonest-categorized participants did not show any notable differences during trials with or without the opportunity for dishonest gain.

FRN Amplitude. During the FRN time window, EEG measurements from Opportunity Win trials and No-Opportunity Win trials appeared to be most negative at central electrode Cz, characteristic of FRN activity. Consequently, data from this electrode are graphed in Figure 2, in which the FRN is represented by differences between the two waveforms during the 200ms-300ms range. The overall topography of this difference is shown in Figure 3. Subjecting data from this FRN time window to the four-way ANOVA revealed a significant interaction between electrode anteriority and laterality ($F(8, 192) = 8.406$, $p < 0.001$), and supported the idea that the data under study truly did reflect the FRN component. The four-way ANOVA also revealed a significant interaction between Opportunity/No-Opportunity trial differences and Honest-Dishonest group categorization ($F(1, 192) = 4.258$, $p < 0.05$). So in response to "correctly" guessed coin flip outcomes, the FRN activity of Dishonest participants, more so than the FRN activity of Honest participants, changed depending on whether the trial offered the potential for dishonest gain. Further analysis of the voltage differences at this maximal electrode showed that the Dishonest group had a significantly larger FRN response during Opportunity win trials than

they did during No-Opportunity win trials ($F(1, 21) = 4.569$, $p < 0.05$). The same comparison within the Honest group showed no significant results ($F(1, 29) = 0.981$, $p < 0.05$).

In order to further investigate this significant difference and its relationship to the preceding-trial reaction time differences mentioned earlier, FRN amplitude during Opportunity Win trials was investigated with respect to what type of trial came before the FRN-prompting stimulus. Data from Opportunity Win trials were submitted to a variation of the normal four-way-ANOVA, in which preceding trial opportunity conditions, rather than current trial opportunity conditions, were considered. FRN data from Opportunity Win Trials was subjected to this 4 (preceding condition: Opportunity Win vs. Opportunity Loss vs. No-Opportunity Win vs. No-Opportunity Loss) × 3 (electrode anteriority) × 5 (electrode laterality) within-subjects ANOVA with group (Honest vs. Dishonest) as a between-subjects factor. The data show a significant interaction of electrode anteriority and laterality ($F(8, 128) = 4.797$, $p = .002$), and the addition of Preceding-Trial-Type to this interaction renders it non-significant $F(24, 384) = 1.193$, $p = .244$. Such a finding lends credence to the idea that these different waveforms are comparable FRN responses, and not a result of different underlying combinations of ERP components. Further comparisons of these measurements were therefore made, and are shown in Table 4. Analysis of the Dishonest participant data showed that Opportunity Win trials following another Opportunity Win had significantly higher FRNs than Opportunity Win trials following an Opportunity Loss, $F(1, 15) = 4.861$, $p = .045$. The corresponding comparison for Honest participants was non-significant, $F(1, 19) = 0.971$, $p = .338$. Comparisons of Opportunity Win trials that were preceded by either an Opportunity win or a No-Opportunity win found non-significant FRN differences for both Honest ($F(1, 19) = 0.014$, $p = .907$) and Dishonest ($F(1, 15) = 0.724$, $p = .409$) participants.

P3 Amplitude. During the P3 time window, the difference between Opportunity Win trials and No-Opportunity Win trials appeared to be maximally negative at both central electrode Cz and parietal electrode Pz. As a result of this centro-parietal distribution, P3 activity is apparent in the 300-400ms range of Figure 2, and the overall topography of the difference between the two waveforms is shown in Figure 4. Consistent with the appearance of this scalp plot, the four-way ANOVA showed significant effects of electrode anteriority ($F(2, 48) = 138.847$, $p < .001$), electrode laterality ($F(4, 96) = 196.007$, $p < .001$), and the interaction of the two ($F(8, 192) = 8.216$, $p < .001$) on P3 amplitude. Comparing P3 data from Win trials in both Opportunity and No-Opportunity conditions using the four-way ANOVA also showed that the interaction between Opportunity/No-Opportunity trial differences and Honest-Dishonest group categorization was significant, $F(1, 24) = 5.300$, $p = .030$. Much like the FRN measurements, the P3 response of Dishonest participants to a coin flip outcome was more sensitive than the response of Honest participants to whether the present trial offered the potential for dishonest gain. Further planned comparisons, shown in Table 3, revealed that the Dishonest group's P3 amplitude in response to Opportunity Win coin flip outcomes was significantly smaller than their P3 amplitude in response to No-Opportunity Win coin flip outcomes, $F(1, 21) = 5.062$, $p = .036$. The equivalent comparison within the Honest group yielded non-significant results, $F(1, 29) = 0.192$, $p = .665$.

To probe at what might be moderating this significant difference in the Dishonest participants, ERP amplitudes from Opportunity Win trials were again analyzed with respect to a given waveform's preceding trial type. The results of this analysis are shown in Table 4. Sorting the Opportunity trials in this way and subjecting them to a 4 (preceding condition: Opportunity Win vs. Opportunity Loss vs. No-Opportunity Win vs. No-Opportunity Loss) × 3 (electrode

anteriority) × 5 (electrode laterality) within-subjects ANOVA with group (Honest vs. Dishonest) as a between-subjects factor revealed a significant interaction between electrode anteriority and laterality, $F(8, 120) = 6.347$, $p < .001$. As it was with the FRN data, the addition of Preceding-Trial-Type to this interaction resulted in non-significance ($F(24,360) = 0.909$, $p = .590$), and suggests that these EEG are comparable variations of the same P3 component. Unlike the FRN data, however, Dishonest participants did not show differences in P3 amplitude depending on preceding trial type: Opportunity Win trials preceded by Opportunity Win trials did not show significantly different P3 responses than Opportunity Win trials preceded by Opportunity Loss trials ($F(1, 15) = 1.111$, $p = .310$), or than Opportunity Win trials preceded by No-Opportunity Win trials, $F(1, 15) = 0.931$, $p = .350$. The corresponding comparisons for Honest participants ($F(1, 18) = 0.036$, $p = .853$ and $F(1, 18) = 0.014$, $p = .908$, respectively) were also non-significant.

CNV Amplitude. Data preceding a response in the CNV time window appeared to be maximal at frontal electrode site Fz. Subjecting amplitudes within this time range to the four-way ANOVA showed a significant interaction between electrode anteriority and laterality ($F(8, 168) = 2.853$, $p < .001$) which is consistent with the localized EEG peak at its maximal Fz site. However, no significant difference in CNV amplitude was apparent between different Honest-Dishonest group categorization and differences in Opportunity/No-Opportunity trial conditions, and there was no discernable waveform differences between these data in the CNV time window.

Discussion

Collectively, these results emphasize the relative cognitive importance, as well as the superior physiological detectability, of behaviors that plan and motivate morally relevant deception compared to the behaviors of performing morally relevant deception. Differences

between the response-associated ERN and CNV components, associated with the self-report of successful or unsuccessful prediction of the previously presented coin-flip outcome, were non-significant across Honest and Dishonest participants. Significant differences in the ERP response of Honest participants and Dishonest participants were only found for the stimuli-associated FRN and P3 components, in response to the outcome of a coin flip that a participant was attempting to guess. Within Dishonest participants, trials that afforded the opportunity for dishonest gain by falsely reporting a correct prediction showed a marked increase in FRN negativity and a reduction in the positivity of the P3, as compared to trials that did not afford such an opportunity. Corresponding comparisons within the Honest participants were non-significant.

Lack of significance between response-associate ERP differences across Honest and Dishonest groups is somewhat surprising, since previous studies of both instructed deceit (Johnson et al., 2003; Johnson et al., 2004) and volition-driven deceit in a game context (Kireev et al., 2007; Sun et al., 2011) have found such differences in ERP activity immediately preceding and throughout following a deceptive response. Explanations for why this trend does not persist in our study of morally accountable and volition-driven deception will need to differentiate the mindsets of Dishonest participants in either context. Participants in previous studies of deceit are very aware of honesty's experimental relevance, and as such, their efforts as conscientious participants will make them cooperate with researchers in experiencing the conflict of dishonesty with full attention. Removing the sanction of dishonesty from the experimental paradigm does not simply hold our participants to different expectations of how honest their behavior must be; Our participants are placed in a situation where the amount of attention that they must dedicate to the concept of honesty can be entirely different. Participants are free to lie about their coin flip

prediction without paying attention to their lie in the same way that they might pay attention in an experimental setup that implicitly suggests a focus on dishonesty. Such an inhibition of attention is consistent with the small P3 response to Opportunity Win coin flip outcomes observed in Dishonest participants, and this process could allow participants to personally distance themselves from their actions. Studies have found that distancing self from action in this way, or by vaguely perceiving self as "cheating" rather than truly considering the details of one "being a cheater," can lead to a marked increase dishonesty (Bryan, Adams, & Monin, 2013). Therefore, participants in our experimental paradigm can employ deceptive strategies that direct attention away from their own deceptions, resulting in a dishonest categorization and a significance-reducing lack of physiologically detectable conflict during response execution. Morally accountable, volition-driven instances of deceptive response, then, seem to show a greater variety of ERP patterns that are not so easily detectable as previously studied instances of deceptive response.

In explaining the Dishonest group's enhanced FRN during opportunity trials, several non-mutually exclusive reasons for the discrepancy can be addressed. In the most simple sense, one could consider that heightened FRN amplitudes are elicited by unexpected outcomes, and realize that the some of the outcomes subsequently reported as correctly guessed by the Dishonest group were actually incorrectly guessed. The pool of Dishonest group ERP data in response to "correctly guessed" outcomes, then, are actually polluted by ERP responses to unpredicted stimuli, and as a result these data are also polluted by heightened FRN amplitudes. Such a heterogeneous mixture of expected and unexpected FRN responses would, by nature of the average, lead to a higher FRN response. Assuming that the Dishonest participants were properly engaged in the task and attentively viewing outcome stimuli with their prediction in mind, then

this sort of effect could very well have contributed to their heightened FRN response during Opportunity Win trials.

However, the results showing that FRN amplitude varied depending on the context of preceding trials suggests that the above explanation may be overly simplistic, and that explanation of the Dishonest group FRN response requires a discussion of not only *what* they are perceiving during the coin flip outcome, but also *how* they are perceiving during the coin flip outcome. Since amplitude of the FRN has shown sensitivity to not only outcome expectedness but also to a participant's stake in the outcome (San Martin et al., 2010), it is possible that FRN differences can be partly explained by differences in the Honest and Dishonest groups' perceived importance of an outcome stimulus. Regardless of whether a coin flip prediction is confirmed as correct or incorrect upon outcome presentation, the outcome stimulus conveys a certain kind of information to only the members of the Dishonest group: whether or not this upcoming response will require a lie. This consideration need not trouble the members of the Honest group, but it could possibly explain perception of the outcome stimulus in a way that is unique to the Dishonest. The stimulus might be considered high stake in that in provokes unwanted reflection on how often a participant has been lying, and how often they can expect to get away with lying in the experimental context. Alternatively, the high FRN response may simply reflect a motivational or emotional trait of a participant (Sato et al., 2005), which indicates a strong response to unexpected outcomes that may ultimately act as a stronger motivator for deceptive behavior.

The idea that an outcome stimulus may provoke consideration of deceptive strategy within the Dishonest group also speaks to group differences in P3 amplitude. The involvement of the P3 response in facilitating an update of context is understandable, given the Dishonest

group's need to keep tabs on prediction correctness, the relative need of dishonesty, the recent amount of dishonesty, and other factors. The Honest group, on the other hand, would not seem to entertain these thoughts, and the additional attentional resources they can dedicate to the predication task may account for their larger P3 response. Beyond the need to allocate attention to their deceptive strategy, the Dishonest group may directly benefit from allocating their attention away from the outcome stimulus. Just as depletion of task-related cognitive control resources brought about by sleepiness can facilitate dishonest behavior (Mead, Baumeister, Gino, Schweitzer, & Ariely, 2009), the under-involvement of cognitive control resources in processing outcome stimuli can circumvent conflict and ease the act of deception. If deception requires a pre-potent truthful response to be inhibited, allocation of attention to the coin flip outcome may reinforce tendencies toward the honest response and must be inhibited. Limited attention to a lie-promoting stimulus in this way may facilitate a less-conflicted response when it comes to actually executing a deceptive behavior.

These significant FRN and P3 responses, however, again contrast the findings of some previous ERP studies of deception. In particular, ERP studies that seem to observe deceptive decision in a game context (Kireev et al., 2007; Sun et al., 2011) do not show significant differences between Honesty-preceding and Dishonesty-preceding outcome stimulus evaluation during the FRN and P3 time ranges. Investigating this discrepancy between findings by looking to differences in experimental paradigm lends insight into the electrophysiological effects of considering moral accountability. A notable difference between falsely reporting an incorrect coin flip prediction and intentionally giving a misleading statement to a computer is that, in predicting a coin flip, the need to lie is not known until after the lie-preceding stimulus is presented. The significant amplitudes of the FRN and P3 following outcome presentation in

Dishonest participants suggest that this deception-related decision occurs consistently upon presentation of the relevant coin flip outcome – even if a participant is fully willing to lie about an upcoming coin flip outcome, the decision to lie is not likely pre-determined, and a dishonest participant will still choose to tell the truth if they do happen to predict the coin flip correctly. Instances of dishonesty within Sun et al. or Kireev et al.'s paradigm of deliberate computer misdirection, on the other hand, come about due to an outcome-independent, strategic decision to lie. With this in mind, it is not surprising that ERP evidence of a decision to lie is not apparent upon presentation of the lie-preceding stimulus.

The roles of the FRN and the P3 in this process are further delineated by looking to the amplitude effects of preceding trial type. The significant effect of previous trial type on Dishonest group FRN amplitude in Opportunity Win trials is evidence that the Dishonest group may be processing outcome stimuli with consideration of contextual factors that the Honest group is not considering. Dishonest group FRN measures during Opportunity Win trials are significantly higher if they are preceded by another Opportunity Win, which seems to suggest that the Dishonest group perceives their "correctly predicted" coin flip outcomes as more unexpected or involving higher stakes than usual. Understandably, the report of two potentially deceptive successes in a row is a noteworthy occurrence for any participant maintaining awareness of a deceptive strategy. This makes such an outcome, and the information it carries about the possible need to lie, of particular interest and motivational worth to a Dishonest participant, which may help explain the heightened FRN. Interestingly, P3 amplitudes do not show the same sensitivity to contextual differences in preceding trial type. In this respect, a heightened P3 seems like a deception-facilitating condition that is common to all Dishonest participants, but insensitive to the actual extent or range of deceptive behavior. The FRN on the

other hand, seems to be a more direct reflection of deceptive motivation. It is difficult in the present research to tease apart the two components given their relatively similar distribution and temporal window, but future deception research ought to consider the relationship between these two underlying processes.

When appreciating the implications of these findings, a tempered point of critique might again emphasize the study's inability to ever really know what behavior is being studied during Opportunity Win trials. With this in mind, there is a limitation on the comparability of the present study to previous studies of deception, which investigate basic and straightforward contrasts of superficial deceptive acts. Current stratagems of advanced lie detection technology also rely on these basic contrasts (Ganis, Rosenfeld, Meixner, Kievit, & Schendan, 2011), and data collected using the paradigm introduced by Greene and Paxton does not easily fit into the puzzle. Imagining a machine that can reliably detect the lie-facilitating FRN and P3 that precede potentially deceptive responses, however, illustrates an important point: Morally relevant cognitive research possesses a particular ability to not only help describe moral behavior, but also to help define moral behavior and actively shape its perception. For example, the present association between morally relevant deceptive activity and heightened FRN and P3 amplitudes preceding a response not only furthers understanding of how to detect natural deception, but has the potential to reshape opinions about what the morally culpable part of deception really is. FRN and P3 amplitude differences can, in this way, lend some credence to the opinion that moral fault arises well before a lie is even executed, when information is processed in a self-serving way that can later facilitate deception.

Detailed consideration of moral deception in this fashion is of critical importance as an industry of lie detection continues to move forward. Misunderstanding the results of such

technology could potentially lead to ethically disastrous decisions in legal and social arenas (Wolpe, Foster, & Langleben, 2005). Recent survey work has shown that fMRI lie-detection evidence would significantly influence potential jurors' tendency to arrive at guilty verdicts (Ganis & Keenan, 2009), despite the fact that these data are imperfect (albeit improving) evidence, that they are vulnerable to countermeasures of disruptive thought (Ganis et al., 2011) and that they are based on a developing science. The goal of understanding the underlying cognition of deception is thus critical to achieve across all participants. Such work will need to rely on clever experimental paradigms such as the one developed by Greene and Paxton, and analysis of data will need to constantly stay aware of discrepancies between natural deception seen in the real world and artificial deception seen in an experimental context.

References

Abe, N. (2011). How the Brain Shapes Deception: An Integrated Review of the Literature. *The Neuroscientist* 17(5), 560-574.

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., and Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron* 64, 756-770.

Bryan, C. J., Adams, G. S., Monin, B. (2013). When cheating would make you a cheater: implicating the self prevents unethical behavior. *Journal of Experimental Psychology: General* 142(4), 1001-1005.

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. *Cereb Cortex* 19(7), 1557-1566.

Damen, E. J. P. and Brunia, C. H. M. (1987). Changes in heart rate and slow brain potentials related to motor preparation and stimulus anticipation in a time estimation task. *Psychophysiology* 24(6), 700-713.

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods* 134, 9-21.

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

Fang, F., Liu, Y., Shen, Z. (2003). Lie detection with contingent negative variation. *International Journal of Psychophysiology* 50(3), 247-255.

Ganis, G., Keenan, J. P. (2009). The cognitive neuroscience of deception. *Social Neuroscience* 4(6), 465-472.

Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A., Schendan, H. E. (2011). Lying in the scanner: covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *NeuroImage* 55(1), 312-319.

Gehring, W. J., Gratton, G., Coles, M. G. H., Donchin, E. (1992). Probability effects on stimulus evaluation and response processes. *Journal of Experimental Psychology: Human Perception and Performance* 18(1), 198-216.

Gehring, W. J., Liu, Y., Orr, J. M., Carp, J. (2012) The error-related negativity (ERN/Ne). In S. J. Luck & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 231-294). New York: Oxford University Press.

Graton, G., Coles, M. G., Donchin, E. (1983) A new method for the off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology* 55(4), 468-484.

Greene, J. D. and Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *PNAS*, 106(30), 12506-12501.

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

Ito, A., Abe, N., Fujii, T., Ueno, A., Koseki, Y., Hashimoto, R., Mugikura, S., Takahashi, S., Mori, E. (2011). The role of the dorsolateral prefrontal cortex in deception when remembering neutral and emotional events. *Neuroscience Research* 69(2), 121-128.

Ito, A., Abe, N., Fujii, T., Hayashi, A., Ueno, A., Mugikura, S., Takahashi, S., Mori, E. (2012). The contribution of the dorsolateral prefrontal cortex to the preparation for deception and truth telling. *Brain Research* 1464, 43-52.

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

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

Kim, C., Kroger, J. K., Kim, J. (2011). A functional dissociation of conflict processing within anterior cingulate cortex. *Human Brain Mapping* 32(2), 304-312.

Kireev, M. V., Starchenko, M. G., Pakhomov, S. V., Medvedev, S. V. (2007). Stages of the cerebral mechanisms of deceptive responses. *Human Physiology* 33(6), 659-666.

Kireev, M. V., Korotkov, A. D., Medvedev, S. V. (2012). Functional magnetic resonance study of deliberate deception. *Human Physiology* 38(1), 32-39.

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

Lee, T. M, Liu, H., Tan, L., Chan, C. C., Mahankali, S., Feng, C., Hou, J., Fox, P., Gao, J. (2002). Lie Detection by Functional Magnetic Resonance Imaging. *Human Brain Mapping* 15. 157-164.

Lee, T. M., Au, R. K., Liu, H. L., Ting, K. H., Huang, C. M., Chan, C. C. (2009). Are errors differentiable from deceptive responses when feigning memory impairment? An fMRI study. *Brain Cognition* 69(2), 406-412.

Luck, S. J. (2005). *An introduction to the event-related potential technique* (4th ed). Cambridge: MIT Press.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288(5472), 1835-1838.

Mameli, F., Mrakic-Sposta, S., Vergari, M., Fumagalli, M., Macis, M., Ferrucci, R., Nordio, F., Consonni, D., Sartori, G., Priori, A. (2010). Dorsolateral prefrontal cortex specifically processes general – but not personal – knowledge deception: Multiple brain networks for lying. *Behavioural Brain Research* 211, 164-168.

McCabe, D. P., Castel, A. D., Rhodes, M. G. (2011). The influence of fMRI lie detection evidence on juror decision-making. *Behavioral Sciences and the Law* 29, 566-577

Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., and Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology* 45, 594-597.

Picton, R. W. and Low, M. D. (1971). The CNV and semantic content of stimuli in the experimental paradigm: effects of feedback. *Electroencephalography and Clinical Neurophysiology* 31(5), 451-456.

Polich, J. (2012). Neuropsychology of P300. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 159-188). New York: Oxford University Press.

Rebert, C. S., McAdam, D. W., Knott, J. R., Irwin, D. A. (1967). Slow potential change in human brain related to level of motivation. *Journal of Comparative and Physiological Psychology* 63(1), 20-23.

San Martín, R., Manes, F., Hurtado, E., Isla, P., Ibañez, A. (2010) Size and probability of rewards modulate the feedback error-related negativity associated with wins but not losses in a monetarily rewarded gambling task. *Neuroimage* 51(3), 1194-1204.

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., Kuboki, R. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Cognitive Neuroscience and Neuropsychology* 16(4), 407-411.

Sun, S., Mai, X., Liu, C., Liu, J., Luo, Y. (2011). The processes leading to deception: ERP spatiotemporal principal component analysis and source analysis. *Social Neuroscience* 6(4), 348-359.

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Brain Imaging* 12(13), 2849-2853.

Spence, S. A., Hunter, M. D., Farrow, T. F. D., Green, R. D., Leung, D. H., Hughes, C. J., Gansean, V. (2004). A cognitive neurobiological account of deception: evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359(1451), 1755-1762.

Van Overwalle, F. and Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *Neuroimage* 48(3), 564-584.

Walczyk, J. J., Roper, K. S., Seemann, E., Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Applied Cognitive Psychology* 17(7), 755-774.

Wolpe, P. R., Foster, K., Langleben, D. D. (2005) Emerging neurotechnologies for lie-detection: promises and perils. *American Journal of Bioethics* 5(2), 39-49.

Author Note

Nolan B. O'Hara, Department of Psychology, University of Michigan, Ann Arbor

I would like to thank Dr. Bill Gehring, whose willingness to share his guidance, resources, and expertise has entirely made this project possible. For the facilitating the process of data collection, I owe many thanks to the University of Michigan and to the family of Wilson P. Tanner: expenses of participant payment were covered in part by funds from Rackham Graduate School and in part by funds from the Tanner Memorial Award, received in April of 2012.

Table 1

Reaction Time Data

Group	Condition	Outcome	Reaction Time (ms)	Std. Deviation (ms)
Honest	Opportunity	Win	619	178
		Loss	635	196
	No-Opportunity	Win	632	185
		Loss	673	219
Dishonest	Opportunity	Win	598	220
		Loss	555	238
	No-Opportunity	Win	558	207
		Loss	599	254

Note. ms = milliseconds. Reaction times represent the average time taken to report a prediction as correct or incorrect after being presented with the stimulus requesting a report. A main effect of Group on reaction time was found, with the Dishonest group responding more quickly than the Honest group. Within the Dishonest group, the Opportunity Win reaction time was significantly longer than the No-Opportunity Win reaction time. Within the Honest group, the No-Opportunity Loss reaction time was significantly longer than the Opportunity Loss reaction time.

Table 2

Opportunity Win Reaction Times

Group	Previous Trial Type	Current Trial Type	Reaction Time (ms)	Std. Deviation (ms)
Honest	Opportunity Win	Opportunity Win	610	162
		No-Opportunity Win	638	176
	No-Opportunity Win	Opportunity Win	630	181
		No-Opportunity Win	626	186
Dishonest	Opportunity Win	Opportunity Win	623	226
		No-Opportunity Win	529	214
	No-Opportunity Win	Opportunity Win	590	221
		No-Opportunity Win	571	206

Note. ms = milliseconds. Reaction times represent the average time taken to report a prediction as correct or incorrect after being presented with the stimulus requesting a report.

Table 3

Results of Planned ERP Amplitude Contrasts

Contrast	df_{bet}	df_{within}	*F*	*p*
FRN Response to Coin Flip Outcome				
Dishonest Group Win Trials				
OpW vs. No-OpW at Cz	1	21	4.569	.045
Honest Group Win Trials				
OpW vs. No-OpW at Cz	1	29	0.981	.330
P3 Response to Coin Flip Outcome				
Dishonest Group Win Trials				
OpW vs. No-OpW at Pz	1	21	5.062	.036
Honest Group Win Trials				
OpW vs. No-OpW at Pz	1	29	0.192	.665

Note. Op = Opportunity, No-Op = No-Opportunity, W = Win, L = Loss.

Table 4

Results of Planned ERP Amplitude Contrasts based on Preceding Trial Type

Contrast	df_{bet}	df_{within}	*F*	*p*
OpW FRN Response to Coin Flip Outcome				
Dishonest Group				
Following OpW vs. Following No-OpW	1	15	0.724	.409
Following OpW vs. Following OpL	1	15	4.861	.045
Honest Group				
Following OpW vs. Following No-OpW	1	19	0.014	.907
Following OpW vs. Following OpL	1	19	0.971	.338
OpW P3 Response to Coin Flip Outcome				
Dishonest Group				
Following OpW vs. Following No-OpW	1	15	0.931	.350
Following OpW vs. Following OpL	1	15	1.111	.310
Honest Group				
Following OpW vs. Following No-OpW	1	19	0.014	.908
Following OpW vs. Following OpL	1	19	0.036	.853

Note. Op = Opportunity, No-Op = No-Opportunity, W = Win, L = Loss. Significant effects were only found for the FRN response of Dishonest subjects during Opportunity Win trials that were preceded by another Opportunity Win trial.

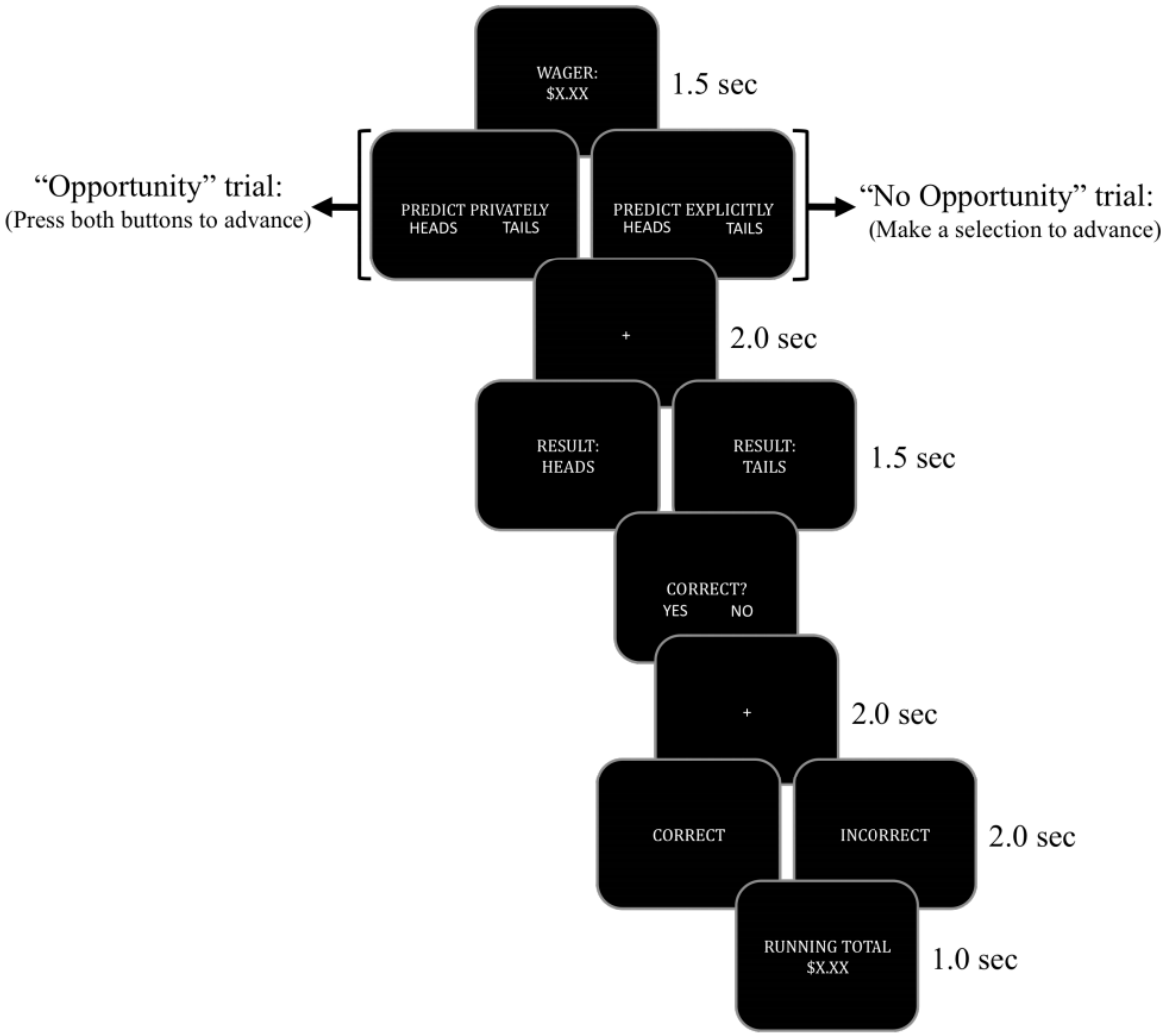

Figure 1. Appearance and order of stimuli for Opportunity and No-Opportunity trials.

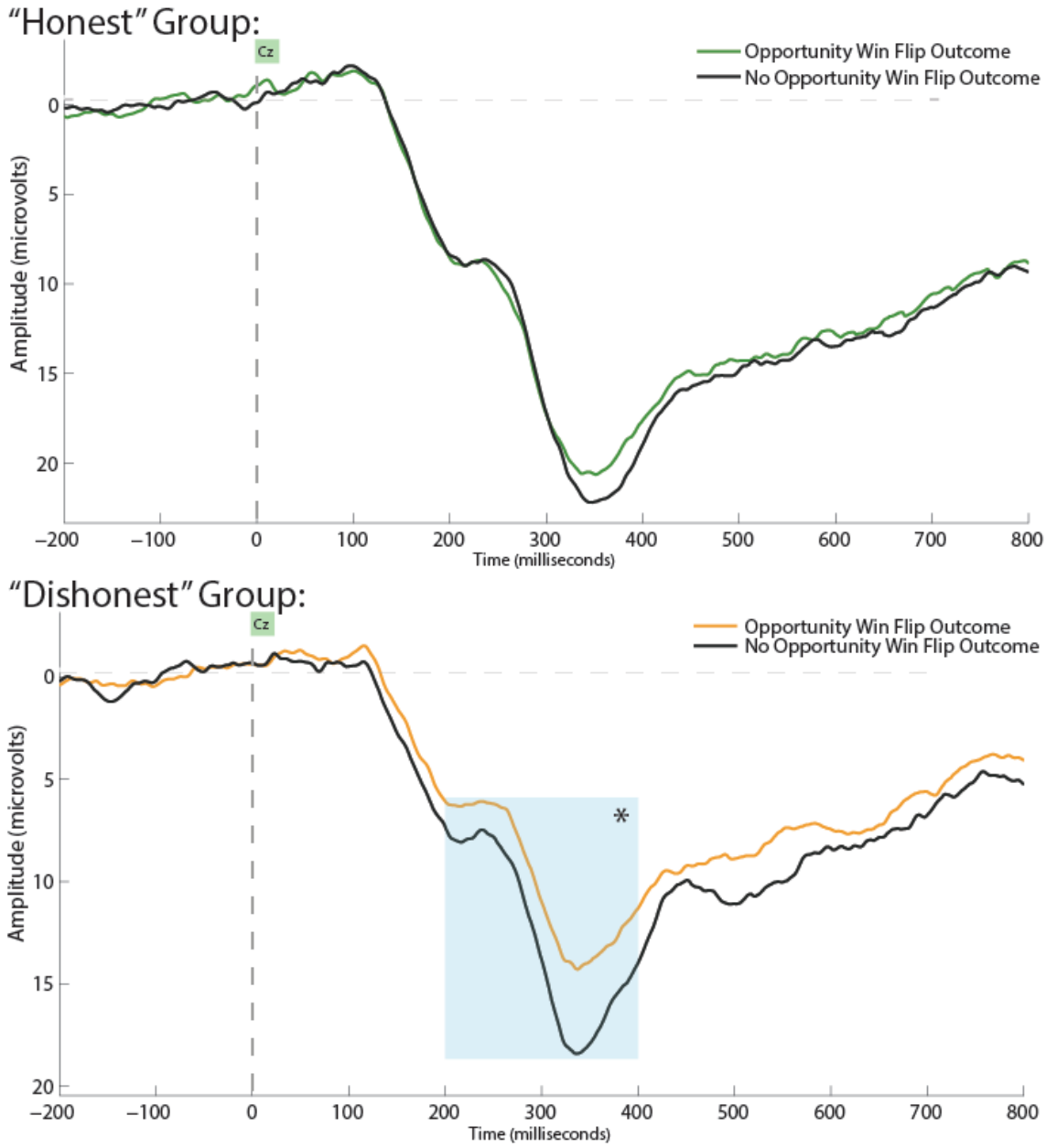

Figure 2. Group average waveforms from electrode Cz. FRN amplitudes are shown by data in the 200-300ms range, and P3 amplitudes are shown by data in the 300ms-400m range.

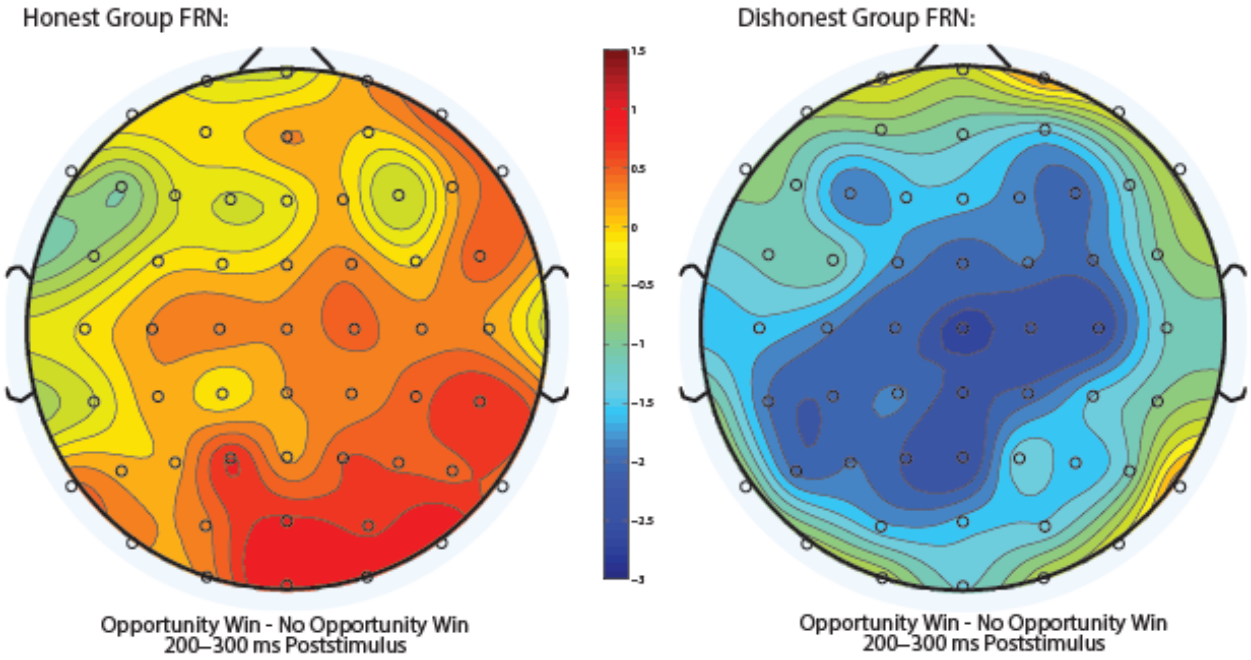

Figure 3.Topographic distribution of charge for Honest and Dishonest FRN responses.

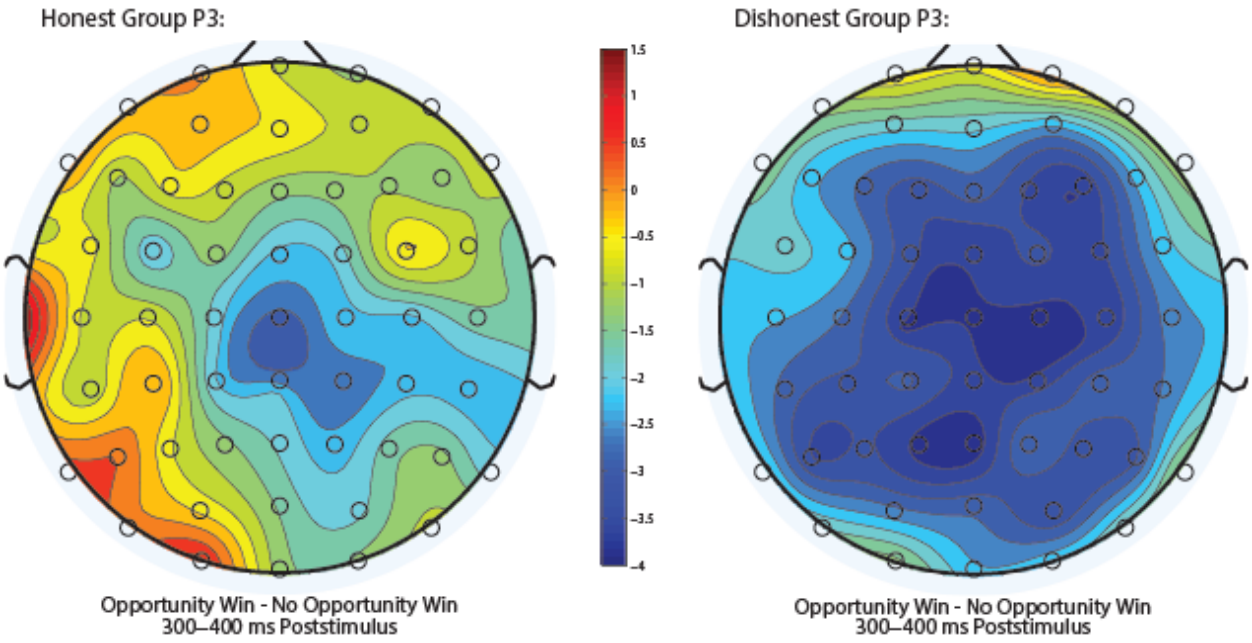

Figure 4. Topographic distribution of charge for Honest and Dishonest P3 responses.

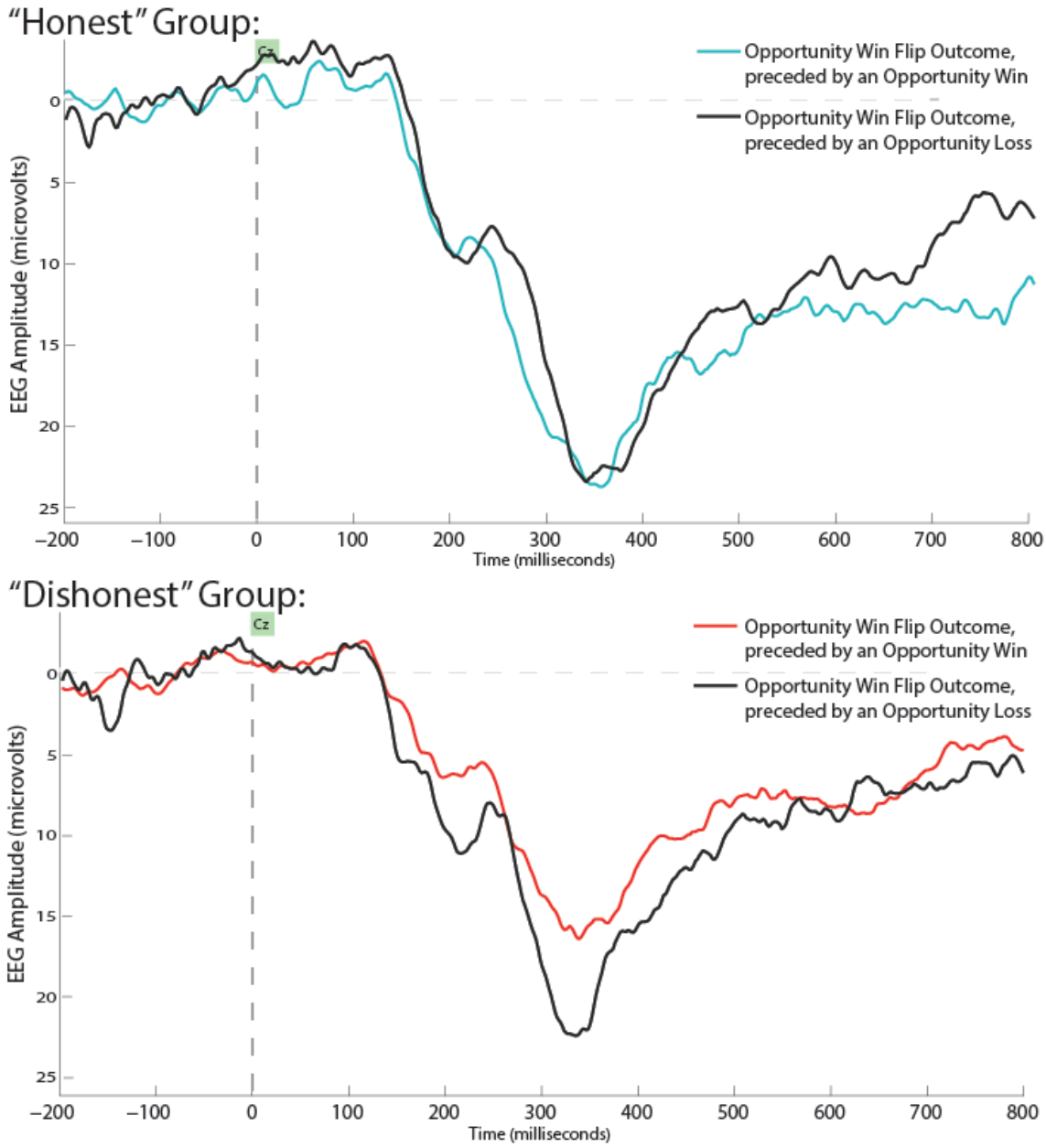

Figure 5.Waveforms measured at electrode Cz that show a dependence on preceding trial type.

Appendix A

Post-task Questionnaire

Subject Number _____ *Committing to private and public predictions*

Thank you for participating in our study!

Now that you've completed the task, we'd appreciate it if you answered a few questions. Again, feel free to skip any question you're not comfortable answering. Your answers will in no way affect your pay or winnings, and your answers will not be associated with your name - only a subject number. Circle or check applicable answer(s):

1. Did it occur to you that you could cheat on this task (i.e. report a correct guess when your private guess was incorrect)?

Yes **No**

2. On a scale of 1 - 10, were you *tempted* to cheat on this task? (1 = never, 10 = always)

1 2 3 4 5 6 7 8 9 10

On a scale of 1 - 10, how often did you cheat when possible? (1 = never, 10 = always)

1 2 3 4 5 6 7 8 9 10

If and when you did cheat, did you feel anything resembling guilt when answering?

Yes **No** **Not Applicable (did not cheat)**

Did you respond haphazardly or carelessly at any point due to lack of interest or boredom?

Yes **No** **If yes, how often and when, generally?**

_____

_____

Do you think behavior in this sort of task reflects honesty tendencies for people in more personally relevant choices?

Yes **Somewhat** **Hardly** **No**

Do you think that any of the following had an impact on your honesty?

Concern for, or relationship towards, the researchers and their work
Unfamiliarity with what EEG electrodes can detect (lie detection?)
Considering lies in this context as small or insignificant
Other (please share:)