Behavioral/Cognitive

Response to Anticipated Reward in the Nucleus Accumbens Predicts Behavior in an Independent Test of Honesty

Nobuhito Abe[1] and Joshua D. Greene[2]

[1]Kokoro Research Center, Kyoto University, Kyoto 606-8501, Japan, and [2]Department of Psychology, Harvard University, Cambridge, Massachusetts 02138

This study examines the cognitive and neural determinants of honesty and dishonesty. Human subjects undergoing fMRI completed a monetary incentive delay task eliciting responses to anticipated reward in the nucleus accumbens. Subjects next performed an incentivized prediction task, giving them real and repeated opportunities for dishonest gain. Subjects attempted to predict the outcomes of random computerized coin-flips and were financially rewarded for accuracy. In some trials, subjects were rewarded based on self-reported accuracy, allowing them to gain money dishonestly by lying. Dishonest behavior was indexed by improbably high levels of self-reported accuracy. Nucleus accumbens response in the first task, involving only honest rewards, accounted for ~25% of the variance in dishonest behavior in the prediction task. Individuals showing relatively strong nucleus accumbens responses to anticipated reward also exhibited increased dorsolateral prefrontal activity (bilateral) in response to opportunities for dishonest gain. These results address two hypotheses concerning (dis)honesty. According to the "Will" hypothesis, honesty results from the active deployment of self-control. According to the "Grace" hypothesis, honesty flows more automatically. The present results suggest a reconciliation between these two hypotheses while explaining (dis)honesty in terms of more basic neural mechanisms: relatively weak responses to anticipated rewards make people morally "Graceful," but individuals who respond more strongly may resist temptation by force of Will.

Key words: **dishonesty; fMRI; moral; morality; nucleus accumbens; reward**

Introduction

What makes people behave honestly or dishonestly? And can variability in honesty be explained in terms of familiar neurobiological mechanisms? The present investigation begins with two hypotheses concerning the cognitive nature of (dis)honesty. According to the "Will" hypothesis, honest behavior results from the active resistance of temptation, comparable to the controlled cognitive processes that enable the delay of reward (Metcalfe and Mischel, 1999; McClure et al., 2004). According to the "Grace" hypothesis, honest behavior happens more automatically, without the need for active self-control at the time of choice (Bargh and Chartrand, 1999; Haidt, 2001). Both hypotheses have received empirical support (Greene and Paxton, 2009; Mead et al., 2009; Gino et al., 2011; Shalvi et al., 2012). The Grace hypothesis is supported by fMRI and reaction time data indicating that consistently honest behavior involves no additional cognitive work (Greene and Paxton, 2009). This naturally raises the question: What makes consistently honest individuals morally "Graceful?" This question is particularly intriguing given that previous research has identified no distinctive neural signature of honest behavior, no pattern of neural activity corresponding to the proverbial "voice of conscience." In light of this, we hypothesized that consistent honest behavior arises, not from the presence of a neural voice of conscience, but from the absence of its opposite, a neural "voice of greed." In more concrete terms, we hypothesized that moral Grace results, at least in part, from relatively weak responses to anticipated rewards, not only when the rewards would be gained dishonestly, but more generally. We used fMRI to test the prediction that nucleus accumbens response to anticipated rewards predicts dishonest behavior, even when such responses occur in an independent task involving no opportunity for dishonest behavior.

Subjects undergoing fMRI completed the monetary incentive delay (MID) task, during which they experienced brief delays before claiming monetary rewards of variable value (Knutson et al., 2001a,b; Fig. 1*A*). Specifically, the mean percentage signal change in blood oxygenation level-dependent (BOLD) signal in anatomically defined nucleus accumbens was calculated for each subject during reward anticipation trials (reward > neutral; Buckholtz et al., 2010). The MID task was originally developed to maximize affective and motivational aspects of reward processing by using rapid presentation of stimuli and rewards contingent on behavior (Knutson et al., 2000). During a subsequent prediction task, subjects attempted to predict the outcomes of random computerized coin-flips and were financially rewarded for accuracy and punished for inaccuracy (Greene and Paxton, 2009; Fig. 1*B*). In the No-Opportunity condition, subjects recorded their predictions in advance, denying them the opportunity to cheat by lying about their accuracy. In the Opportunity condition, subjects made their predictions privately and were rewarded based

Received Jan. 16, 2014; revised June 3, 2014; accepted June 27, 2014.

Author contributions: N.A. and J.D.G. designed research; N.A. performed research; N.A. analyzed data; N.A. and J.D.G. wrote the paper.

We are grateful to Joe Paxton, Ryan Halprin, Ming Cheung, John Kwon, Fiery Cushman, and Joshua Buckholtz for their comments/assistance. This research was supported by the Richard Hodgson Memorial Fund at Harvard University. N.A. was supported by JSPS Postdoctoral Fellowships for Research Abroad.

Correspondence should be addressed to Nobuhito Abe, Kokoro Research Center, Kyoto University, 46 Shimoadachi-cho, Yoshida Sakyo-ku, Kyoto 606-8501, Japan. E-mail: abe.nobuhito.7s@kyoto-u.ac.jp.

DOI:10.1523/JNEUROSCI.0217-14.2014

 0270-6474/14/3410564-09$15.00/0

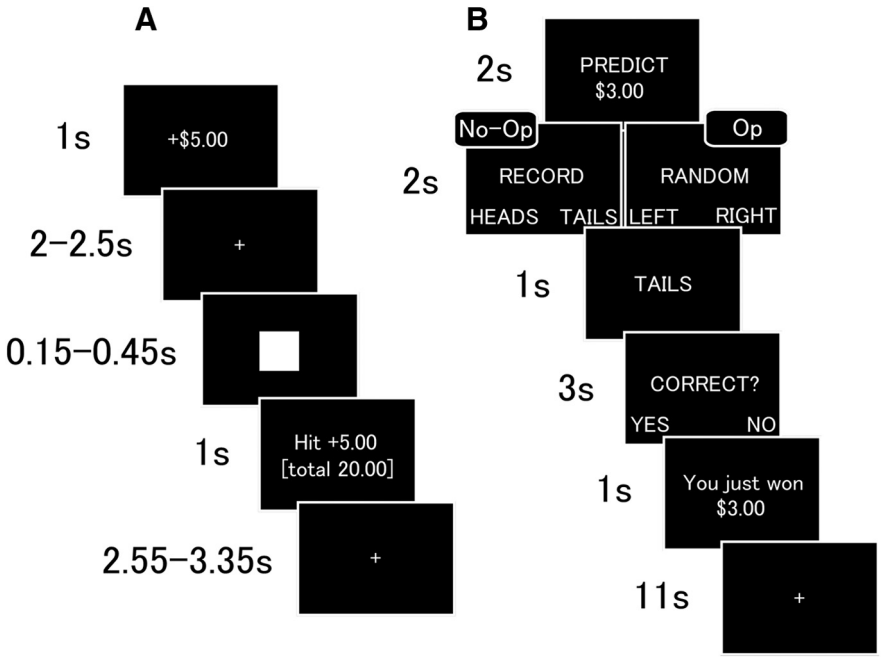

Figure 1. ***A***, ***B***, Task sequence of MID task (***A***) and coin-flip task (***B***). In the MID task (***A***), the subject observes the trial's monetary value, followed by a variable-duration fixation cross. After the fixation cross, a target square is briefly presented. The subject presses a button while the square is on the screen to get a financial reward or to avoid a financial loss. A feedback message with current and cumulative winnings/losses is presented. This is followed by a fixation interval. In the coin-flip task (***B***), the subject observes the trial's monetary value and privately predicts the outcome of the upcoming coin-flip. The subject records this prediction by pressing one of two buttons (No-Opportunity condition) or presses one of these buttons randomly (Opportunity condition). The subject then observes the outcome of the coin-flip. The subject then indicates whether the prediction was accurate and observes the amount of money won/lost based on the recorded prediction (No-Opportunity) or the self-reported accuracy (Opportunity). This is followed by a fixation interval.

on their self-reported accuracy, affording them the opportunity to gain money dishonestly by lying. In contrast with nearly all fMRI studies of deception (Abe, 2009, 2011), the lying observed here is genuinely dishonest lying because the present subjects were not explicitly instructed to lie. Dishonest behavior was indexed by improbably high levels of self-reported accuracy.

Materials and Methods

Subjects. The present results are based on data from 28 subjects (18 females and 10 males, mean age 21.3 years, age range 18–34 years). All subjects were right-handed, native English speakers who had no history of neurological or psychiatric disease. Our analyses required the classification of subjects as honest, dishonest, or ambiguous based on self-reported accuracy in the Opportunity condition of the coin-flip task (the data were normally distributed; Kolmogorov–Smirnov normality test, $p > 0.05$). Consistent with procedures used previously (Greene and Paxton, 2009), eight subjects reporting improbably high levels of accuracy at the individual level (binomial test, $p < 0.001$) were classified as dishonest (mean "accuracy" = 83.6%). This conservative threshold was used to ensure a sufficient number of cheat trials per dishonest subject. The 13 lowest-accuracy subjects (binomial test, $p > 0.05$ for the entire group of 13) were classified as honest (mean accuracy = 50.1%). This is the largest group of subjects that, at the group level, exhibit no significant evidence of cheating. The remaining 7 subjects were classified as ambiguous (mean accuracy = 67.1%). Although it is clear that at least some individuals within this group behaved dishonestly (323 of 481 trials, group binomial test, $p < 0.000001$), we classified these individuals as "ambiguous" because none of them met our conservative threshold for confirmed dishonest behavior at the individual level. The classification of subjects was used in the analysis of the data to test the Grace hypothesis and to identify subjects for exclusion. Subjects were paid $50 for participating, in addition to the bonus pay based on performance during the experimental tasks. Subjects gave written informed consent in accordance with a protocol approved by Harvard University's Committee on the Use of Human Subjects.

In addition to the data drawn from the 28 subjects analyzed, the data from a total of 11 subjects were discarded for reasons described below. The exclusion criteria used in the present study were identical to those used previously and yielded similar results (Greene and Paxton, 2009). We emphasize that our behavioral paradigm, which involves deception concerning the interests of the experimenters (though not of the payoff structures), inevitably requires higher rates of exclusion than those of fMRI experiments involving more typical behavioral tasks.

First, in debriefing, subjects were asked what they thought the experiment was about in an open-ended way. At this point in debriefing, seven subjects classified as dishonest and two subjects classified as honest voiced suspicions that the experiment was about cheating, lying, or dishonesty. We discarded the data from the seven dishonest subjects, but not the others. This was done to exclude data from subjects who may have seen themselves as morally justified in deceiving the experimenters because they believed that the experimenters were attempting to deceive them. We adopted this policy as a conservative measure, anticipating that some may hesitate to call such deception dishonest. We included the remaining two honest subjects because it is not essential to our design that honest behavior be motivated by purely moral considerations. Second, subjects were eventually informed of the purpose of the experiment and were asked whether they were aware that they could cheat. All but two subjects indicated that they were aware of the possibility of cheating. Data from these two subjects were excluded because our aim was to investigate honest behavior in the face of opportunity for dishonest gain, and these subjects were not aware of the opportunity. Third, data from two subjects were discarded due to excessive response failure (>30%).

Finally, we conducted tests to identify and exclude subjects who strategically underreported their accuracy. In the present paradigm, it is possible to gain money dishonestly while maintaining a chance level of accuracy by cheating on the Opportunity trials that are worth the most (i.e., $6.00 and $7.00) and deliberately underreporting accuracy for the Opportunity trials that are worth the least (i.e., $3.00 and $4.00). Subjects using this strategy can exhibit improbably high levels of cumulative monetary reward given their win/loss percentages. To identify such subjects, we compared the winnings of each honest subject to those of simulated honest subjects (10,000 permutations) with win/loss percentages individually matched to the subject being tested. The winnings of all honest subjects were consistent with their respective win/loss percentages ($p > 0.05$). Therefore, in the present study, no subjects were excluded for the strategic underreporting of accuracy.

General procedures. To measure neural response to anticipated reward, we used the MID task in which subjects anticipated a monetary reward, no reward, or the avoidance of monetary loss (Knutson et al., 2001a,b). To measure dishonesty, we used a coin-flip prediction task in which subjects had opportunities to gain money dishonestly by lying about the accuracy of their predictions (Greene and Paxton, 2009). We used a cover story to justify our giving subjects obvious opportunities for dishonest gain. This study was presented as a study of paranormal abilities to predict the future, aimed at testing the hypothesis that people are better able to predict the future when their predictions are (1) private and (2) financially incentivized. Thus, subjects were implicitly led to believe that the

opportunity for dishonest gain was a known but unintended by-product of the experiment's design and that they were expected to behave honestly. We note that in using this cover story, subjects were deceived about the experimenters' interests, but not about the economic structure of the task. Subjects were not presented with the cover story until after they had been recruited, thus avoiding self-selection for subjects with interests in parapsychology. An earlier study (Greene and Paxton, 2009) used a variety of personality scales in hopes of identifying familiar psychological traits that predict dishonest behavior. None of these yielded significant results. Thus, the present study did not include personality scales.

Before starting the experiment, we had subjects complete the Paranormal Belief scale (Tobacyk and Milford, 1983) to support our cover story. Subjects were given a thorough explanation of the task procedure and were familiarized with the MID task and coin-flip task by completing practice trials. At this point, some subjects mentioned to the experimenter that it was possible to cheat in the coin-flip task. The experimenter responded by acknowledging his awareness of that possibility. The experimenter explained that the possibility of cheating was a necessary by-product of the experimental design and encouraged the subject to follow the directions, which preclude cheating if followed.

MID task. In the MID task, participants had the opportunity to win money or avoid losing money by pressing a button during the brief presentation of target stimulus. The MID task session consisted of a total of 100 trials. During each trial, participants were shown one of five cues for 1000 ms, indicating the reward value of the trial. There were 20 high-reward trials ($5), 20 low-reward trials ($0.25), 20 high-loss trials ($5), 20 low-loss trials ($0.25), and 20 neutral trials ($0.00). Participants were then presented with a fixation cross during a variable interval (anticipatory delay phase, 2000–2500 ms). Subjects responded with a button press to a white target square that appeared for a variable length of time (target phase, 150–450 ms). For reward trials, subjects gained money by responding while the target was onscreen (a "hit"). On reward trials, there was no penalty for failing to press the button during this time (a "miss"). For loss trials, hits resulted in neither gain nor loss, but misses caused the subject to lose the amount indicated by the cue for that trial. Although no money was at stake in neutral trials, participants were instructed to rapidly press the button in response to the target square. Next, a feedback screen (outcome phase, 1000 ms) notified participants of the amount won/lost on that trial, as well as their cumulative winnings at that point. A variable intertrial interval (2550–3350 ms) followed each trial. The MID task session lasted ~12.5 min. Consistent with prior procedures (Buckholtz et al., 2010), we contrasted the neural activity for reward versus neutral trials in the nucleus accumbens during the anticipatory delay phase. We emphasize that this analysis focuses on responses to possible future rewards, perhaps dependent on motivation (Knutson et al., 2000), rather than responses to the receipt of reward.

To approximately equate MID task performance across subjects, we used an adaptive algorithm that dynamically adjusted the duration of the target presentation as a function of subject performance (Kuhl et al., 2010; Hahn et al., 2011). Five independent "trains" were used, representing the five different reward or loss values. For each train, the target accuracy was 66.0% and the duration of the target square, which was initialized to 300 ms, was adjusted on a trial-by-trial basis, depending on whether the running accuracy for that train was greater than or less than 66.0%. For instance, if mean accuracy in the high-reward condition after trial n was equal to 80%, then the square duration for trial $n + 1$ in the high-reward condition was shortened (making the trial more difficult). In contrast, if mean accuracy in the high-reward condition after trial n was equal to 50%, then the square duration for trial $n + 1$ in the high-reward condition was lengthened (making the trial easier). In this manner, the square duration was shortened or lengthened by 30 ms increments. In addition, target duration was set as to never fall below 150 ms and to never exceed 450 ms. Since this adaptive algorithm was used to alter target durations, reaction times cannot be meaningfully interpreted and are therefore not analyzed. This algorithm ensured that net earnings were positive for all of the subjects.

Coin-flip task. In the coin-flip task, subjects attempted to predict the outcomes of random computerized coin-flips and were financially rewarded for accuracy and punished for inaccuracy. The subject (1) observes the trial's monetary value and privately predicts the outcome of the upcoming coin-flip (2 s), (2) records this prediction by pressing one of two buttons (No-Opportunity condition) or presses one of these buttons randomly (Opportunity condition; 2 s), (3) observes the outcome of the coin-flip (1 s), (4) indicates whether the prediction was accurate (3 s), (5) observes the amount of money won/lost based on the recorded prediction (No-Opportunity condition) or the reported accuracy (Opportunity condition; 1 s), and (6) waits for the next trial (11 s). Thus, in the No-Opportunity condition, subjects recorded their predictions in advance, denying them the opportunity to cheat by lying about their accuracy. In the Opportunity condition, subjects made their predictions privately and were rewarded based on their self-reported accuracy, affording them the opportunity to cheat. Subjects completed a total of 210 trials. Within the 70 Opportunity trials, the values $3, $4, $5, $6, or $7 USD each appeared 14 times, as was the case for the 70 No-Opportunity trials. We included an additional set of 70 Low-Value-Opportunity trials that were worth $0.02, $0.10, $0.25, $0.35, and $0.50 USD. Each of these values also appeared 14 times. Neuroimaging data from these trials were not analyzed because the contrasts involving this condition cannot be controlled for monetary value. They were included to provide dishonest subjects with additional opportunities to behave honestly at little cost, thus giving them cover for cheating in the regular (higher-value) Opportunity trials. Subjects were paid the cumulative value of their winnings/losses. Net losses were capped at $0, and net winnings were capped at $75 (not including participation payment and MID bonus money). Trials appeared in random order in a series of 7 blocks of 30 trials each. Each block of the coin-flip task lasted ~10 min. Subjects' understanding of the experiment was assessed in debriefing (see above). They were asked about their thoughts and experiences during the experiment in an open-ended way. Subsequently, subjects were informed of the true nature of the experiment and were asked whether they were aware of the possibility of cheating.

In the present version of the coin-flip task, the buttons for random responding in the Opportunity condition are labeled "left" and "right" rather than "heads" and "tails." This change from past procedures (Greene and Paxton, 2009) was implemented to further reduce the (already small) proportion of subjects who are unaware of the possibility of cheating.

The following instructions were presented to subjects on a computer.

> *Thank you for participating. In this study your job is to predict the outcomes of computerized random coin flips. You may not think that you have the ability to do this, and that's okay. Just do your best. You may be surprised at what you can do! Press any key to continue.*
>
> *It has been suggested that people make more accurate predictions when they are motivated to predict accurately. To test this idea, we will be providing you with varying levels of financial incentive. Before each coin flip happens, an amount of money will appear on the screen (e.g., $0.25 or $5.00). This is the amount of money that you will win or lose depending on whether you accurately predict the outcome of the coin flip. If your prediction is correct, then you win the amount of money shown. If your prediction is incorrect, you lose the amount of money shown. The computer will keep track of all of your wins and losses. If, at the end of the experiment, your money total is positive, you will be paid that amount. If your total is negative or zero, you will not win any additional money. This is not pretend money. This is real money that you will be paid based on your performance in the experiment. However, your winnings cannot exceed $75. Press any key to continue.*
>
> *It has been suggested that people's ability to predict the future is disrupted if they have to record their predictions externally (i.e., outside of their minds). To test this idea, we will sometimes ask you to report your prediction in advance. In other cases, you will simply tell us after the fact whether or not your prediction was correct. Press any key to continue.*
>
> *Before each coin flip you will see the dollar amount that the trial is worth and, below it, the word "PREDICT" on the screen. At that*

Table 1. The mean proportions and the reaction times of subjects' responses

		Proportion (%)		Reaction time (ms)	
Group	Condition	Mean	SD	Mean	SD
Honest	Op Win	50.1	6.6	496	103
	Op Loss	49.9	6.6	569	98
	No-Op Win	49.9	6.0	503	98
	No-Op Loss	50.1	6.0	571	134
	LV-Op Win	49.6	7.2	498	98
	LV-Op Loss	50.4	7.2	604	108
Ambiguous	Op Win	67.1	2.0	602	144
	Op Loss	32.9	2.0	719	171
	No-Op Win	50.5	5.0	594	117
	No-Op Loss	49.5	5.0	639	143
	LV-Op Win	52.4	6.6	624	178
	LV-Op Loss	47.6	6.6	732	152
Dishonest	Op Win	83.6	8.8	539	148
	Op Loss	16.4	8.8	775	365
	No-Op Win	50.8	5.9	477	111
	No-Op Loss	49.2	5.9	559	179
	LV-Op Win	55.9	18.8	584	248
	LV-Op Loss	44.1	18.8	608	191

Op, Opportunity; LV, low-value.

point you should make your prediction in your mind. Next you will either see the word "RECORD" or the word "RANDOM." If you see the word "RECORD" you should press the button on the LEFT to indicate that you are predicting HEADS or the button on the RIGHT to indicate that you are predicting TAILS. If you see the word "RANDOM" then you should randomly press either the LEFT button or the RIGHT button. When you make random responses, you should not follow any fixed pattern. Press any key to continue.

Next you will see the word "HEADS" or "TAILS" appear on the screen. This is the outcome of the computerized coin flip. After that you will see a screen that says "CORRECT?" At that point you must indicate whether or not your prediction was correct. If you were told to hit a random button, it does not matter which button you hit. Whether or not your prediction was correct depends only on the prediction you made in your mind and the outcome of the coin flip. Press the LEFT key to indicate (YES) that your prediction was correct. Press the RIGHT key to indicate (NO) that your prediction was incorrect. If you failed to form a prediction before the outcome of the coin flip was revealed, then you should indicate that your prediction was incorrect. After you have said whether your prediction was correct, the computer will tell you how much money you won or lost on that coin flip. Press any key to continue.

You will make a total of 210 predictions. You will do these in 7 groups of 30 trials. There will be about 10 s between the end of one trial and the beginning of the next one. After each group of trials you will have a chance to rest. The whole task will take a little less than 90 min. Press any key to continue.

You are now ready to practice. Remember, first comes the dollar amount telling you what the coin flip is worth and the word "PREDICT." At that point you will make your prediction privately to yourself. (Note that the dollar amounts presented here will not count toward your final total.) Then you will see either "RECORD" or "RANDOM." If you see "RECORD" enter your prediction (LEFT key for HEADS, RIGHT key for TAILS). If you see "RANDOM" press either the LEFT key or the RIGHT key randomly. Then you will see the outcome of the coin flip (HEADS or TAILS). Then you will see the word "CORRECT?" on the screen. At that point you indicate whether the prediction you made in your mind was correct. Press the LEFT key (YES) if your prediction was correct or the RIGHT key (NO) if your prediction was incorrect. Then the computer will tell you how much money you won or lost on that coin flip. Then you wait for the next coin flip, which will begin with a dollar amount, as before. Press any key to begin practicing.

Image acquisition and data preprocessing. Whole-brain imaging was performed with a 3.0 tesla Siemens Magnetom Tim Trio MRI scanner with a 12-channel head coil. A T2*-weighted echoplanar imaging (EPI) sequence sensitive to BOLD contrast was used for functional imaging with the following parameters: repetition time (TR) = 2500 ms, echo time (TE) = 30 ms, flip angle = 90°, 72 × 72 acquisition matrix, field of view (FOV) = 216 mm, and in-plane resolution = 3 × 3 mm. Thirty-nine axial slices, with a slice thickness of 3 mm, were obtained. A high-resolution (spatial resolution 1.2 × 1.2 × 1.2 mm) structural image was also acquired using a T1-weighted magnetization-prepared rapid-acquisition gradient echo (MP-RAGE) pulse sequence. Head motion was restricted using firm padding that surrounded the head. Visual stimuli were projected onto a screen and were viewed through a mirror attached to the head coil. The subjects' responses were collected using a magnet-compatible response box. The EPI images were acquired in eight consecutive runs (i.e., one for the MID task and seven for the coin-flip task). The first four scans in each run were discarded to allow for T1 equilibration effects.

Data preprocessing and statistical analyses were performed using SPM8 (Wellcome Department of Imaging Neuroscience, London, UK). All volumes acquired from each subject were corrected for different slice acquisition times. The resultant images were then realigned to correct for small movements occurring between scans. This process generated an aligned set of images and a mean image per subject. Each participant's T1-weighted structural MRI was coregistered to the mean of the realigned EPI images and segmented to separate out the gray matter, which was normalized to the gray matter in a template image based on the Montreal Neurological Institute (MNI) reference brain. Using the parameters from this normalization process, the EPI images were also normalized to the MNI template (resampled voxel size 2 mm × 2 mm × 2 mm) and smoothed with an 8 mm full-width at half-maximum Gaussian kernel. A high-pass filter of 1/128 Hz was used to remove low-frequency noise, and an AR(1) (autoregressive 1) model was used to correct for temporal autocorrelations.

Statistical analysis. The fMRI data were analyzed using an event-related model. Each task (MID task and coin-flip task) was analyzed separately. For the MID task, all reward trials (high-reward and low-reward), loss trials (high-loss and low-loss), and neutral trials were pooled. Onsets for the anticipatory delay period of each of the trial types were separately modeled using a canonical hemodynamic response function. The right and left anatomical nucleus accumbens regions of interest (ROIs) were derived from Individual Brain Atlases using Statistical Parametric Mapping Software (IBASPM; Alemán-Gómez et al., 2006) implemented in the WFU PickAtlas (Wake Forest University, Winston-Salem, NC; Maldjian et al., 2003). To quantify neural response to anticipated reward across subjects, we used MarsBaR software (Brett et al., 2002) to extract percentage change in BOLD signal of the nucleus accumbens for each condition (i.e., averaged across all trials of a given condition) for each subject. The percentage change values for neutral trials during the delay period were subtracted from those of the reward trials (collapsed across monetary value). We used this mean signal change value for each subject to predict each subject's level of dishonesty, i.e., each subject's self-reported % Wins in Opportunity condition of the coin-flip task.

For the coin-flip task's fMRI data, all events of interest were modeled through convolution with a canonical hemodynamic response function temporally indexed by participants' responses. The parameter estimates (betas) for each condition were calculated for all brain voxels, and the following two contrasts of parameter estimates were computed: Opportunity Win vs No-Opportunity Win and Opportunity Loss vs No-Opportunity Loss. The first contrast identifies signal differences associated with (but not exclusively associated with) dishonest behavior. The second contrast identifies signal associated with honest behavior in the presence of opportunity for dishonest gain. In the neuroimaging analysis of the coin-flip task, the data from two subjects were excluded because of their extremely low number of Opportunity-Loss trials (two for both subjects). This low number of Opportunity-Loss trials prevented us from obtaining a stable activation map for these subjects. The exclusion of these two subjects explains why analyses using fMRI data from the coin-flip task are based on 26 subjects, instead of 28, as in the

analysis correlating response to reward in the MID task with dishonest behavior. The contrast images for the remaining 26 subjects were then entered into a series of multiple-regression analyses, in which the results generated by the independent MID task are used as predictors of activity in the prefrontal control network. Specifically, we examined the relationship between the response to reward in the MID task (the signal change averaged across right and left nucleus accumbens) and the activity across brain regions for Opportunity Win vs No-Opportunity Win and Opportunity Loss vs No-Opportunity Loss in the coin-flip task. The significant activations were identified at the statistical threshold of $p < 0.001$ (uncorrected for multiple comparisons) with the cluster size of 10 or more voxels. The peak voxels of clusters exhibiting reliable effects are reported in MNI coordinates. We also generated graphs showing time courses of percentage change in BOLD signal after participants' responses. Data for these graphs were generated by modeling decision-related BOLD data using a finite impulse response function. The finite impulse response model makes no assumptions about the shape of activations, thereby providing unbiased estimates of the average signal intensity at each time point for each event type. In each subject, the mean percentage change in BOLD signal was estimated for each of six scan acquisitions after each decision (0–15 s after decision). Time courses were subsequently averaged across participants and event types.

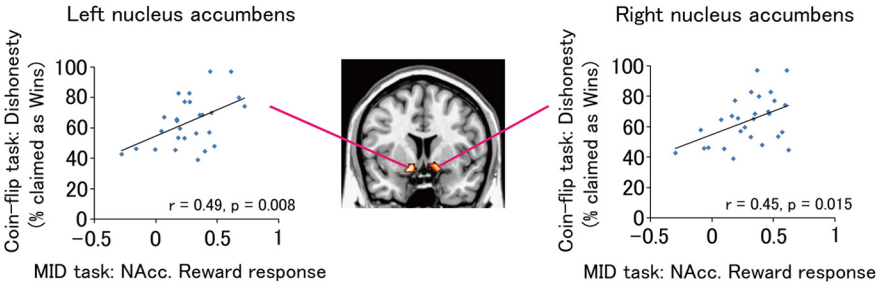

Figure 2. Response to anticipated reward in the nucleus accumbens predicts the frequency of dishonest behavior in an independent task (n = 28). The x-axis shows for each subject the mean difference in the nucleus accumbens' BOLD response to reward versus neutral trials during the MID task. The y-axis shows each subject's self-reported % Wins in the Opportunity condition of the coin-flip task, an index of dishonesty. Coloration shows anatomically defined ROIs superimposed on a standard brain. NAcc, Nucleus accumbens.

Table 2. Regions exhibiting positive correlations between response to anticipated reward in the nucleus accumbens in the MID task and difference in mean signal change for chosen (Opportunity) Loss trials versus forced (No-Opportunity) Loss trials

Region (Brodmann's area)	Coordinates			Z value	Cluster size
	x	y	z		
Right superior parietal lobule (7)	20	−74	52	3.75	49
Right middle frontal gyrus (9)	34	14	54	3.52	13
Left middle frontal gyrus (46)	−38	30	42	3.33	17
Left inferior occipital gyrus (18)	−28	−94	−8	3.17	15

p < 0.001 uncorrected; minimum cluster size = 10.

Results

Behavioral data

During the MID task, participants succeeded on an average of 63.5% (SD = 4.8) of the trials. Thus, the proportion of hits is highly consistent with the target value selected based on previous reports (Knutson et al., 2001a,b; Kuhl et al., 2010; Hahn et al., 2011). There was no correlation between the winnings in the MID task and the self-reported % Wins in the Opportunity condition across subjects ($r = -0.23, p = 0.244$). Thus, we succeeded in minimizing differences in reward history before the coin-flip task and prevented such differences from exerting a detectable influence on subsequent behavior.

The results of the coin-flip task are summarized in Table 1. All three groups of subjects (Honest, Dishonest, Ambiguous) were at chance performance in the No-Opportunity condition. Thus, we found no evidence for subjects having paranormal abilities to predict the future (Bem, 2011). To determine whether the reaction time data support the Grace hypothesis, we conducted planned contrasts following a 3 (group: Honest, Ambiguous, Dishonest) × 3 (condition: Opportunity, Low-Value-Opportunity, No-Opportunity) × 2 (outcome: Win, Loss) ANOVA. A Greenhouse–Geisser correction for sphericity was used when necessary. As expected, the ANOVA revealed a significant three-way interaction ($F_{(3.17,39.64)} = 2.95$, partial $\eta^2 = 0.19$, $p = 0.042$).

Following up on this three-way ANOVA, we first consider Win trials. In the first of our planned contrasts, we compared Opportunity Win trials (which include both honest and dishonest Wins) with No-Opportunity Win trials (which include only forced honest Wins). Within the dishonest group we found a significant difference in reaction time between these two conditions ($t_{(7)} = 2.60$, $p = 0.035$). This finding raises the possibility that dishonest Wins involve additional controlled processing, leading to longer reaction times. Within the ambiguous group, we found no significant difference in reaction time between Opportunity Win trials and No-Opportunity Win trials ($t_{(6)} = 0.32$, $p = 0.762$). Similarly, within the honest group, we found no significant difference in reaction time between these two conditions ($t_{(12)} = -0.34$, $p = 0.743$). Here the critical test is to determine whether a group × condition interaction was significant within Win trials from honest and dishonest groups. As the results of these contrasts suggest, there was a significant group × condition interaction ($F_{(1,19)} = 4.80$, partial $\eta^2 = 0.20$, $p = 0.041$). We also compared Opportunity Win trials with Low-Value-Opportunity Win trials. Here we found no significant differences in reaction time between these two conditions for honest group ($t_{(12)} = -0.12, p = 0.906$), ambiguous group ($t_{(6)} = -0.69, p = 0.516$), and dishonest group ($t_{(7)} = -0.96, p = 0.370$).

Next we consider Loss trials. Within the dishonest group, Opportunity Loss trials involve decisions to refrain from dishonest behavior, whereas No-Opportunity Loss trials involve only forced Losses. We found a significant difference in reaction time between these two conditions ($t_{(7)} = 2.56$, $p = 0.037$). This finding indicates that additional controlled processing is required when dishonest subjects forgo opportunities for dishonest gain. Notably, this effect was also observed in the ambiguous group. We found a significant difference in reaction time between Opportunity Loss trials and No-Opportunity Loss trials ($t_{(6)} = 2.53$, $p = 0.045$). Critically, within the honest group we found no significant difference in reaction time between Opportunity Loss trials and No-Opportunity Loss trials ($t_{(12)} = -0.07$, $p = 0.948$). Here the critical test is to determine whether a group × condition interaction was significant within Loss trials from honest and dishonest groups. As the results of these contrasts suggest, there was a significant group × condition interaction ($F_{(1,19)} = 9.32$, partial $\eta^2 = 0.33$, $p = 0.007$). These findings replicate the results of previous work (Greene and Paxton, 2009) and clearly support the Grace hypothesis, suggesting that consistently honest subjects engage no additional processing when they forgo the opportunities for dishonest gain. We note that the Grace hypothesis and the

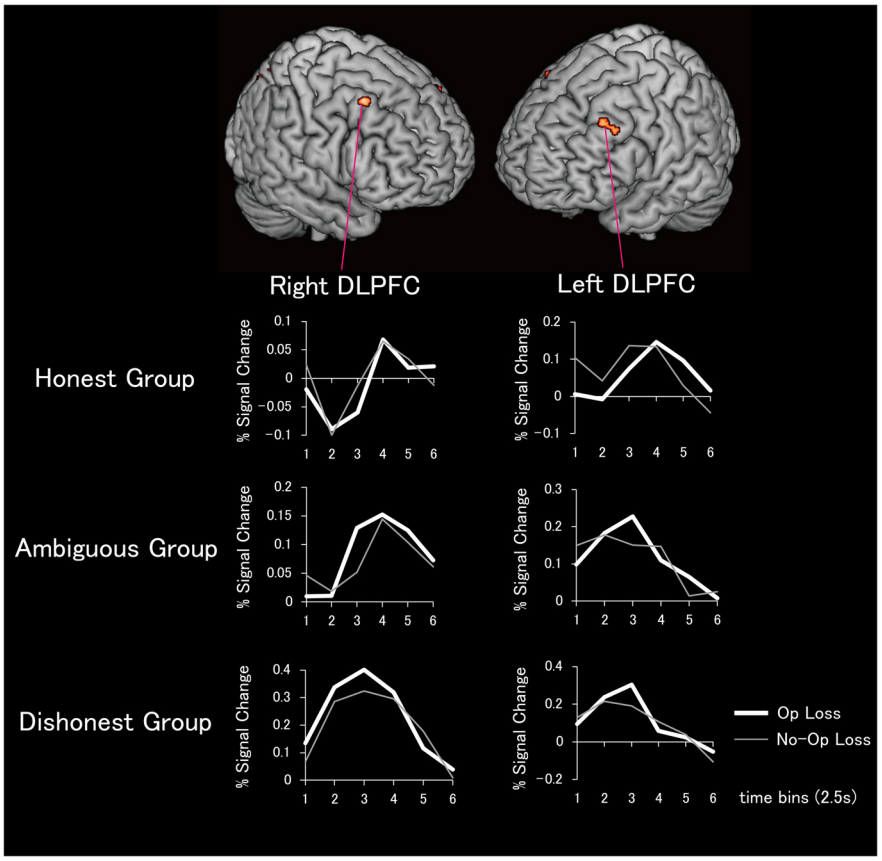

Figure 3. Response to anticipated reward in the nucleus accumbens predicts DLPFC activity when refraining from gaining money dishonestly ($n = 26$). Bilateral DLPFC regions exhibited positive correlations ($p < 0.001$, uncorrected) between mean response to anticipated reward in the nucleus accumbens (averaged across right and left regions) during the MID task and the difference in mean signal change for chosen (Opportunity) Loss trials versus forced (No-Opportunity) Loss trials. Graphs show time courses of mean decision-related percentage change in BOLD signal during the coin-flip task for the honest, ambiguous, and dishonest groups. Signal time courses are displayed across 6 time-bins of 2.5 s each.

data supporting it concern only the cognitive processes engaged at the time of the behavioral response. This leaves open the possibility that subjects in the honest group made "willful" decisions to behave honestly at the outset of the task or at some earlier point in their lives.

We also compared Opportunity Loss trials with Low-Value-Opportunity Loss trials. Although we found no significant differences in reaction time between these two conditions for the ambiguous group ($t_{(6)} = -0.29$, $p = 0.785$), the analyses from the honest and dishonest groups yielded notable results. Within the dishonest group, the reaction time for Opportunity Loss trials was marginally longer than that for Low-Value-Opportunity Loss trials ($t_{(7)} = 2.31$, $p = 0.054$). Within the honest group, the reaction time for Opportunity Loss trials was shorter than that for Low-Value-Opportunity Loss trials ($t_{(12)} = -2.39$, $p = 0.034$). As the results of these two contrasts suggest, there was a significant group × condition interaction ($F_{(1,19)} = 11.76$, partial $\eta^2 = 0.38$, $p = 0.003$). This interaction suggests that the reaction time effects observed in the present study depend critically on monetary value. Moreover, it provides additional support for the claim that additional controlled processing is required when forgoing dishonest gain for the dishonest group, but not for the honest group (Greene and Paxton, 2009).

In the present study, we also tested for correlations between the frequency of dishonest behavior and reaction times for the various trial types (Opportunity Win, Opportunity Loss, No-Opportunity Win, and No-Opportunity Loss). Here we examine all subjects together. As expected, we found no significant correlations between the frequency of dishonest behavior and reaction times for No-Opportunity Win trials ($r = -0.06$, $p = 0.773$) and No-Opportunity Loss trials ($r = 0.07$, $p = 0.739$), respectively. Likewise, we found no significant correlation between the frequency of dishonest behavior and reaction times for Opportunity Win trials ($r = 0.22$, $p = 0.264$). However, we did find a positive correlation between the frequency of dishonest behavior and reaction times for Opportunity Loss trials ($r = 0.53$, $p = 0.004$). These results again support the Grace hypothesis: honest subjects do not engage additional cognitive control in any case (i.e., are honest "Gracefully"), but dishonest subjects engage more control, particularly when refraining from behaving dishonestly.

fMRI data

Following the method of Buckholtz et al. (2010), we first calculated for each subject the mean difference in nucleus accumbens BOLD signal for the reward versus neutral trials in the MID task. Here, the nucleus accumbens was delimited using bilateral a priori anatomical ROIs. We confirmed that the nucleus accumbens activity was significantly higher for high-reward trials ($5) than for low-reward trials ($0.25; left nucleus accumbens, $t_{(27)} = 8.14$, $p < 0.000001$; right nucleus accumbens, $t_{(27)} = 8.96$, $p < 0.000001$; normality of the data was confirmed for all parametric tests; Kolmogorov–Smirnov normality tests, all $p > 0.05$), indicating that the present MID task is a valid measure for neural responses associated with reward anticipation.

We then tested our main hypothesis by calculating the correlation between this measure of neural response to anticipated reward and our measure of dishonest behavior, subjects' self-reported % Wins in the Opportunity condition of the prediction task. (Once again, not all self-reported Wins are dishonest. Rather, self-reported % Wins is correlated with the level of dishonesty.) As predicted, nucleus accumbens response correlated positively with the frequency of dishonest behavior (left nucleus accumbens, $r = 0.49$, $p = 0.008$; right nucleus accumbens, $r = 0.45$, $p = 0.015$; bilateral average, $r = 0.50$, $p = 0.007$; normality of the data was confirmed for all parametric tests; Kolmogorov–Smirnov normality tests, all $p > 0.05$). Thus, the nucleus accumbens signal during the MID task accounted for ~25% of the variance in dishonest behavior (bilateral average $R^2 = 0.25$; Fig. 2). We emphasize that the MID provides a measure of reward-related response that is independent of subjects' responses to the rewards available in the coin-flip task.

Second, we asked whether response to anticipated reward in the nucleus accumbens predicts activity within the prefrontal control network during the coin-flip task. Here, our hypothesis is that individuals with relatively large nucleus accumbens re-

Table 3. Results of planned fMRI contrasts

Group/contrast/region (Brodmann's Area)	Coordinates x	y	z	Zvalue	Cluster size
Honest					
Op Win > No-Op Win					
No suprathreshold activation					
Op Loss > No-Op Loss					
Left lingual gyrus (18)	−24	−66	−10	4.35	42
Left fusiform gyrus (37)	−30	−32	−22	3.70	23
Left cerebellum	−4	−52	−24	3.32	14
Ambiguous					
Op Win > No-Op Win					
No suprathreshold activation					
Op Loss > No-Op Loss					
Left intraparietal sulcus (7)	−26	−50	44	3.88	19
Left precentral gyrus (6)	−36	−10	52	3.77	23
Right precentral gyrus (6)	62	12	24	3.62	25
Left inferior parietal lobule (2/40)	−52	−28	38	3.51	11
Dishonest					
Op Win > No-Op Win					
Right inferior frontal gyrus (45)	46	22	4	3.57	13
Op Loss > No-Op Loss					
Left anterior cingulate cortex (25)	−4	32	0	4.38	16
Left hippocampus	−14	−16	−14	4.15	18
Right middle frontal gyrus (45/46)	42	34	32	3.97	16
Ambiguous and dishonest					
Op Win > No-Op Win					
Left anterior cingulate cortex (32)	−8	20	38	3.56	17
Op Loss > No-Op Loss					
Right anterior cingulate cortex (32)	16	14	38	3.79	11
Left orbitofrontal cortex (47)	−32	26	−12	3.72	73
Left middle frontal gyrus (46)	−40	52	10	3.65	31
Left anterior cingulate cortex (24)	−6	34	14	3.63	88
Left insula	−38	6	0	3.55	48
Left medial superior frontal gyrus (32)	−6	30	36	3.53	86
Left inferior frontal gyrus (45)	−38	36	12	3.30	11

$p < 0.001$ uncorrected; minimum cluster size = 10. Op, Opportunity.

sponses to anticipated reward will require additional cognitive control to forgo available rewards. We tested this hypothesis using a whole-brain analysis. More specifically, the nucleus accumbens signal in the MID task (the signal change averaged across right and left nucleus accumbens) was entered as a covariate of interest in a regression analysis contrasting Opportunity Loss vs No-Opportunity Loss trials. As predicted, we observed effects bilaterally in the middle frontal gyrus (dorsolateral prefrontal cortex; DLPFC; Table 2, Fig. 3). The effects observed in the DLPFC do not survive correction for multiple comparisons ($p < 0.001$ uncorrected) and should therefore be interpreted with caution. Nevertheless, the fact that these effects are bilateral and consistent with a strong a priori hypothesis reduces the likelihood that they are due to chance. These effects were not observed in a regression analysis contrasting Opportunity Win vs No-Opportunity Win trials. Thus, it appears that individuals with greater nucleus accumbens responses to anticipated reward in the MID task also exhibit greater engagement of DLPFC when forgoing opportunities for dishonest gain during the coin-flip task.

We also conducted subtraction analyses of Opportunity Win vs No-Opportunity Win trials (to identify neural activity associated with choosing to behave dishonestly) and Opportunity Loss vs No-Opportunity Loss trials (to identify neural activity associated with choosing to refrain from dishonest behavior) for honest, ambiguous, and dishonest groups (Table 3). The critical test for the Will and Grace hypotheses is the comparison between Opportunity Loss trials and No-Opportunity Loss trials. Consistent with previous work (Greene and Paxton, 2009), we predicted increased engagement of DLPFC during honest decisions in the dishonest group, but not in the honest group. Consistent with this prediction, we found significant activation in the right middle frontal gyrus (DLPFC) in the contrast of Opportunity Loss vs No-Opportunity Loss in the dishonest group. Combining the data from dishonest and ambiguous groups, we found significant activation in left middle frontal gyrus (DLPFC). Parallel DLPFC effects were not observed in the honest group.

Discussion

We used fMRI and two independent behavioral tasks to test the prediction that response to anticipated reward in the nucleus accumbens predicts behavior in a laboratory test of honesty. Individual differences in reward-related response were indexed by the level of fMRI BOLD signal in the nucleus accumbens during the anticipation of reward in the MID task. Dishonest behavior was indexed by improbably high levels of self-reported accuracy in our incentivized coin-flip prediction task. As predicted, individuals exhibiting relatively strong nucleus accumbens responses to anticipated reward exhibited higher rates of dishonest behavior. Such individuals also exhibited (at an uncorrected threshold) increased bilateral engagement of a key region within the prefrontal control network (DLPFC) when refraining from dishonesty.

These findings illuminate the cognitive and neural determinants of honesty and dishonesty in three key ways. First, they link honesty and dishonesty to individual variation in a core mammalian neural system, the mesolimbic reward pathway, which uses mechanisms that have been conserved across evolutionary time (Schultz et al., 1997; O'Doherty, 2004; Rangel et al., 2008; Haber and Knutson, 2010; Shohat-Ophir et al., 2012). These findings also link everyday dishonesty to clinically relevant conditions. Previous studies using the MID task have linked reward-related responses in the nucleus accumbens to psychopathic traits (Buckholtz et al., 2010) and Gray's impulsivity (Hahn et al., 2009). These results, along with more recent evidence concerning trait-positive arousal in healthy individuals (Wu et al., 2014), indicate that responses to anticipated reward as measured by the MID reflect stable traits. However, further research will be needed to determine whether the neural signals of the kind observed here can predict dishonest behavior at significant delays.

Second, these findings support the Grace hypothesis, while refining it in an interesting way. Consistent with the more general Grace hypothesis (Greene and Paxton, 2009), our results show that variation in automatic processing is associated with the tendency to be honest and to engage a key part of the prefrontal control network (MacDonald et al., 2000; Miller and Cohen, 2001; Seeley et al., 2007; Badre, 2008) when behaving honestly. The present results take this hypothesis a step further, indicating that consistent honesty is associated with automatic dispositions that are domain-general, i.e., not specific to the moral domain (Shenhav and Greene, 2010). As an alternative, one might hypothesize that consistently honest individuals are Gracefully honest because they have a specific lack of attraction to dishonest rewards, much as people have a specific lack of sexual attraction to close kin (Lieberman et al., 2007). For example, an honest financier might respond no less than others to the prospect of honestly earned profits, but automatically discount the value of money to be gained by insider trading. Our results, however, are consistent with the hypothesis that Graceful honesty arises, at least in part, from a more general tendency to place less value on, or to be less motivated by, monetary rewards, independent of the reward's moral status.

We emphasize that none of our results, from either the MID task or the coin-flip task, can be explained by systematic differences in reward history produced by performance on a prior task. We used two key design features to preclude effects of reward history. First, all subjects performed the MID task first, followed by the coin-flip task. Second, by using a staircase design, MID task performance and, thus, reward history across subjects were approximately equated (see above; Materials and Methods).

Third, our results suggest a reconciliation between the evidence supporting the Will and Grace hypotheses. Although the present results indicate that relatively weak responses to anticipated reward are associated with Graceful honesty, other evidence suggests that Will (active self-control) also plays a role in honest behavior, with "moral identity" (Gino et al., 2011) and the availability of justifications (Shalvi et al., 2012) as moderating factors. The present correlational evidence is consistent with an alternative (though not mutually exclusive) hypothesis: relatively weak responses to anticipated reward make people morally Graceful, but individuals with stronger responses may resist temptation by force of Will. This is consistent with our finding that nucleus accumbens response predicts dishonest behavior. It is also consistent with our (more tentative) finding that nucleus accumbens response predicts engagement of the DLPFC when people forego opportunities for dishonest gain. Although we believe that this interpretation provides the most coherent account of the present results in light of the literature, the present results do not rule out an earlier interpretation (Greene and Paxton, 2009) according to which the engagement of the DLPFC reflects additional controlled processing that is not preferentially associated with the Willful resistance of temptation.

To infer from the observed DLPFC effect the engagement of cognitive control in this specific task requires a "reverse inference" (Poldrack, 2006), but reverse inferences are by no means categorically fallacious (Hutzler, 2014; Machery, 2014). Tasks requiring high levels of control reliably engage the DLPFC (MacDonald et al., 2000; Miller and Cohen, 2001; Seeley et al., 2007; Badre, 2008). However, the extent to which the aforementioned inference is justified depends on the extent to which the engagement of DLPFC selectively indicates the engagement of cognitive control. At the very least, the observed DLPFC effect, along with concomitant reaction time effects, is highly consistent with the engagement of cognitive control, and Willful self-control more specifically.

Three further limitations of the present study warrant attention. First, although our task design allows us to identify dishonest behavior at the level of individual subjects (by identifying improbably high levels of self-reported Wins), it does not allow us to identify individual lies. This is because most Opportunity Win trials are won honestly, with only a minority of Opportunity Win trials involving decisions to lie. Second, we do not know whether the reward-related responses measured here generalize to non-monetary rewards or to monetary rewards available in other contexts. Third, our primary results are correlational, preventing us from drawing firm conclusions concerning causal relationships between neural responses and (dis)honest behavior.

Despite these limitation, the present findings do suggest that the neural responses to reward are important cognitive and neurobiological determinants of (dis)honesty. More specifically, it appears that honesty gets a boost if one's response to available rewards—both honest and dishonest—is somewhat tepid.

References

Abe N (2009) The neurobiology of deception: evidence from neuroimaging and loss-of-function studies. Curr Opin Neurol 22:594–600. CrossRef Medline

Abe N (2011) How the brain shapes deception: an integrated review of the literature. Neuroscientist 17:560–574. CrossRef Medline

Alemán-Gómez Y, Melie-García L, Valdés-Hernandez P (2006) IBASPM: toolbox for automatic parcellation of brain structures. Paper presented at 12th Annual Meeting of the Organization for Human Brain Mapping, Florence, Italy, June. Available on CD-Rom in Neuroimage 27.

Badre D (2008) Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. Trends Cogn Sci 12:193–200. CrossRef Medline

Bargh JA, Chartrand TL (1999) The unbearable automaticity of being. Am Psychol 54:462–479. CrossRef

Bem DJ (2011) Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. J Pers Soc Psychol 100:407–425. CrossRef Medline

Brett M, Anton JL, Valabregue R, Poline JB (2002) Region of interest analysis using an SPM toolbox [abstract]. Neuroimage 16 [Suppl 1]:497.

Buckholtz JW, Treadway MT, Cowan RL, Woodward ND, Benning SD, Li R, Ansari MS, Baldwin RM, Schwartzman AN, Shelby ES, Smith CE, Cole D, Kessler RM, Zald DH (2010) Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. Nat Neurosci 13:419–421. CrossRef Medline

Gino F, Schweitzer ME, Mead NL, Ariely D (2011) Unable to resist temptation: how self-control depletion promotes unethical behavior. Org Behav Hum Decis Process 115:191–203. CrossRef

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

Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35:4–26. CrossRef Medline

Hahn T, Dresler T, Ehlis AC, Plichta MM, Heinzel S, Polak T, Lesch KP, Breuer F, Jakob PM, Fallgatter AJ (2009) Neural response to reward anticipation is modulated by Gray's impulsivity. Neuroimage 46:1148–1153. CrossRef Medline

Hahn T, Heinzel S, Dresler T, Plichta MM, Renner TJ, Markulin F, Jakob PM, Lesch KP, Fallgatter AJ (2011) Association between reward-related activation in the ventral striatum and trait reward sensitivity is moderated by dopamine transporter genotype. Hum Brain Mapp 32:1557–1565. CrossRef Medline

Haidt J (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment. Psychol Rev 108:814–834. CrossRef Medline

Hutzler F (2014) Reverse inference is not a fallacy per se: cognitive processes can be inferred from functional imaging data. Neuroimage 84:1061–1069. CrossRef Medline

Knutson B, Westdorp A, Kaiser E, Hommer D (2000) FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 12:20–27. CrossRef Medline

Knutson B, Adams CM, Fong GW, Hommer D (2001a) Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J Neurosci 21:RC159(1–6). Medline

Knutson B, Fong GW, Adams CM, Varner JL, Hommer D (2001b) Dissociation of reward anticipation and outcome with event-related fMRI. Neuroreport 12:3683–3687. CrossRef Medline

Kuhl BA, Shah AT, DuBrow S, Wagner AD (2010) Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. Nat Neurosci 13:501–506. CrossRef Medline

Lieberman D, Tooby J, Cosmides L (2007) The architecture of human kin detection. Nature 445:727–731. CrossRef Medline

MacDonald AW 3rd, Cohen JD, Stenger VA, Carter CS (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 288:1835–1838. CrossRef Medline

Machery E (2014) In defense of reverse inference. Br J Philos Sci 65:251–267. CrossRef

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19:1233–1239. CrossRef Medline

McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural

systems value immediate and delayed monetary rewards. Science 306: 503–507. CrossRef Medline

Mead NL, Baumeister RF, Gino F, Schweitzer ME, Ariely D (2009) Too tired to tell the truth: self-control resource depletion and dishonesty. J Exp Soc Psychol 45:594–597. CrossRef Medline

Metcalfe J, Mischel W (1999) A hot/cool-system analysis of delay of gratification: dynamics of willpower. Psychol Rev 106:3–19. CrossRef Medline

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24:167–202. CrossRef Medline

O'Doherty JP (2004) Reward representations and reward-related learning in the human brain: insights from neuroimaging. Curr Opin Neurobiol 14:769–776. CrossRef Medline

Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci 10:59–63. CrossRef Medline

Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9:545–556. CrossRef Medline

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599. CrossRef Medline

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD (2007) Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci 27:2349–2356. CrossRef Medline

Shalvi S, Eldar O, Bereby-Meyer Y (2012) Honesty requires time (and lack of justifications). Psychol Sci 23:1264–1270. CrossRef Medline

Shenhav A, Greene JD (2010) Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. Neuron 67:667–677. CrossRef Medline

Shohat-Ophir G, Kaun KR, Azanchi R, Mohammed H, Heberlein U (2012) Sexual deprivation increases ethanol intake in *Drosophila*. Science 335: 1351–1355. CrossRef Medline

Tobacyk J, Milford G (1983) Belief in paranormal phenomena: Assessment instrument development and implications for personality functioning. J Pers Soc Psychol 44:1029–1037. CrossRef

Wu CC, Samanez-Larkin GR, Katovich K, Knutson B (2014) Affective traits link to reliable neural markers of incentive anticipation. Neuroimage 84: 279–289. CrossRef Medline