Neuroscience 200 (2012) 50–58

TO BELIEVE OR NOT TO BELIEVE: TRUST CHOICE MODULATES BRAIN RESPONSES IN OUTCOME EVALUATION

Y. LONG,[a] X. JIANG[a] AND X. ZHOU[a,b]*

[a]*Center for Brain and Cognitive Sciences and Department of Psychology, Peking University, Beijing 100871, China*

[b]*Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China*

Abstract—Making a trust decision in interpersonal relationship involves forming positive expectation toward the decision outcome. Previous studies have suggested that trust and distrust are qualitatively distinct and have differential neurocognitive substrates. In this study, we investigated how trust choice would modulate brain responses to decision outcome in a modified coin-toss game. Participants received statements from partners concerning the results of coin-toss and decided whether to believe the truthfulness of the statements. In two experiments, event-related potentials (ERPs) to the real results revealed after the trust choice demonstrated differential patterns following trust and distrust choices. Both the feedback-related negativity (FRN) and the P300 showed effects of outcome valence following trust choices, but the FRN effect was reduced following distrust choices. Thus, trust choice creates different contexts in which aspects of decision outcome can be encoded simultaneously by the FRN. The FRN may reflect the subjective evaluation of decision outcome in a specific context rather than a general expectancy towards the outcome.

Key words: outcome evaluation, trust, distrust, ERP, FRN, P300.

Building a trust relationship between individuals or between parties is important to interpersonal exchange and to the stability of social and economic systems. Trust can be described as a rational decision making process involving a certain amount of risk (Morrison and Firmstone, 2000). The degree to which one party trusts another is a measure of belief in the honesty, fairness, or benevolence of the other party. A truster may form reasonable expectations toward or have confidence in the trustee that the trustee will behave in a way beneficial to the truster. In making a trust decision, the truster is in the risk of being harmed if the trustee does not behave accordingly.

Previous research has found that the level of trust in a country correlates positively with the national economic performance (Knack and Keefer, 1997) and that trust as a personality trait is associated with subjective well-being (DeNeve and Cooper, 1998). Trust can enable cooperative behaviors (Gambetta, 1988), promote network relations (Miles and Snow, 1992), and facilitate rapid formulation of ad hoc work groups (Meyerson et al., 1996). Situational factors (Boudreau et al., 2009; Lewicki et al., 1998), characteristics of the truster (Rotter, 1967), and information concerning the trustee (King-Casas et al., 2005; Phan et al., 2010) can affect whether and how a trust behavior takes place.

Comparatively, little attention has been paid to the processes of making a distrust decision and to its potential functions in social exchange, although accumulating evidence suggests that distrust is not a simple absence of trust but is qualitatively distinct from trust (Cho, 2006; Dimoka, 2010; Kramer and Cook, 2004; McKnight and Choudhury, 2006). Moreover, few studies have been conducted to investigate how a trust or distrust decision would affect the evaluation of decision outcomes. It is conceivable that the same outcome following a trust or a distrust decision may have different subjective significance to the truster and may guide future behavior in different ways. Moreover, clarifying how trust choice modulates the brain activity in evaluating decision outcomes would help us understand the nature of neural encoding processes in outcome evaluation.

The present study was to investigate how trust choice affects the brain activity in outcome evaluation, an issue that has not been addressed before. To this end, we measured electrophysiological responses on participants who took part in a coin-toss game (Fig. 1) modeled after Lupia and McCubbins (1998). In this game, a participant first receives a statement from a partner (dubbed "reporter"), indicating whether a coin tossed has landed on head or tail, and decides whether to believe the truthfulness of the statement. The real result of the coin toss is then revealed, serving as an (implicit) feedback to the correctness of the trust choice. Brain responses to the real result (i.e. outcome) are recorded through the event-related potentials (ERPs).

We focused on two ERP components that have been shown to be particularly sensitive to neurocognitive processes involved in outcome evaluation and performance monitoring (Gehring and Willoughby, 2002; Holroyd and Coles, 2002; Miltner et al., 1997; Nieuwenhuis et al., 2004). The first component, the feedback-related negativity (FRN), is a negative deflection between 200 and 350 ms following the onset of feedback stimulus. The FRN is more pronounced for negative feedback associated with unfavorable outcomes, such as monetary losses (Gehring and

*Correspondence to: X. Zhou, Department of Psychology, Peking University, Beijing 100871, China. Tel: +86-10-6275-6599; fax: +86-10-6276-1081.
E-mail address: xz104@pku.edu.cn (X. Zhou).

Abbreviations: ANOVA, analysis of variance; EOGs, electro-oculograms; ERPs, event-related potentials; FRN, feedback-related negativity.

0306-4522/12 $36.00 © 2011 IBRO. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuroscience.2011.10.035

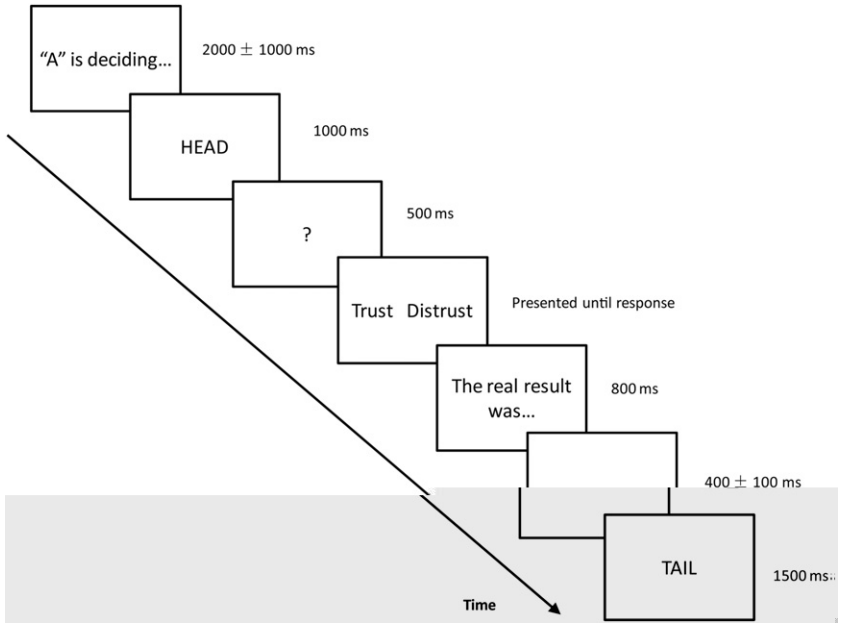

Fig. 1. Sequence of events in a single trial.

Willoughby, 2002), unexpected outcomes (Heldmann et al., 2008; Wu and Zhou, 2009), and incorrect responses (Miltner et al., 1997). Importantly, the FRN effect in outcome evaluation has been found to be affected by social factors that influence the decision process, including interpersonal relationship in reward processing (Leng and Zhou, 2010; Ma et al., 2011; Marco-Pallarés et al., 2010; Wu et al., 2011), the extent of others included in the "self" concept (Kang et al., 2010), and the extent of personal responsibility for the outcome (Li et al., 2010; Zhou et al., 2010). For example, when the ERP participants observe others performing a gambling task, the FRN effect elicited by the observed loss and gain feedback is larger for friends than for strangers performing the task (Ma et al., 2011). Previous studies also found that there is a correlation between the FRN amplitude and the participants' rating on how much they feel to be involved in the task, with larger FRN amplitudes corresponding to higher involvement ratings (Yeung et al., 2005). Since compared with a distrust decision, a trust decision involves stronger expectation toward the partner's intention (Mayer et al., 1995; McKnight et al., 2003; Morrison and Firmstone, 2000; Pavlou and Gefen, 2004) and a greater sense of self-involvement, we expected to observe greater ERP differentiation (i.e. the FRN effect) between negative and positive outcomes following trust choices than following distrust choices.

The second ERP component is the P300, which is usually defined as the most positive peak or mean amplitude in the 200–600 ms time window post-onset of feedback. The P300 has been shown to encode various aspects of feedback stimuli, including the magnitude of reward (Sato et al., 2005; Yeung and Sanfey, 2004), expectancy towards outcome (Hajcak et al., 2005, 2007; Wu and Zhou, 2009), and arguably the valence of feedback (Hajcak et al., 2005, 2007; Leng and Zhou, 2010; Wu and Zhou, 2009). The magnitude of the P300 has also been shown to be sensitive to social factors, with larger P300 being associated with closer interpersonal relationship (Leng and Zhou, 2010; Ma et al., 2011) and higher level of personal responsibility (Li et al., 2010; Zhou et al., 2010) in decision making. As trust behaviors are related to shorter social distance between individuals (Buchan et al., 2002) and stronger sense of personal involvement and responsibility, we expected to observe more positive P300 responses to outcomes following trust choices than to outcomes following distrust choices.

We conducted two experiments. Experiment 1 manipulated trust choice and the valence of outcome, whereas Experiment 2 further manipulated the intention of the reporter in addition to trust choice and outcome valence. The two experiments produced convergent evidence for the impact of trust choice upon brain responses to decision outcomes.

EXPERIMENT 1

Experimental procedures

Participants. Twenty-four undergraduate students (10 males) from Beijing Forestry University, aged 19–25 years, were recruited. All the participants were healthy and right-handed. Eight undergraduate students (four males), who were strangers to the participants, were recruited as confederates. Informed consent was obtained from each participant. This study was approved by the Academic Committee of the Department of Psychology, Peking University.

Stimuli and procedures. As is shown in Fig. 2, the experiment had a two (trust choice: trust vs. distrust) by two (outcome valence: gain vs. no gain) factorial design. In addition, the proportion of trials in which the confederates lied about the result of coin toss was manipulated, such that two confederates (A and B) lied 50% of the times and two other confederates (C and D) gave false

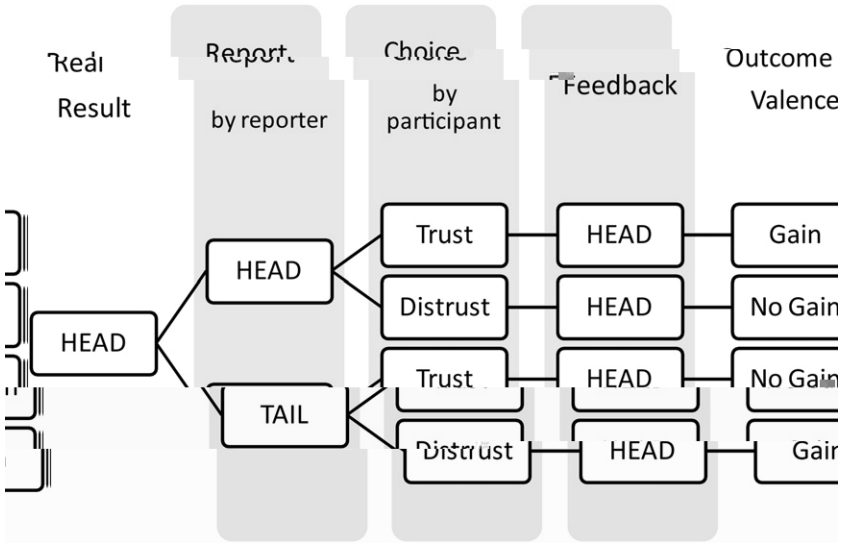

Fig. 2. Illustration of game rules. Gay areas indicate contents that were shown to participants in each trial.

reports 70% and 30% of the times, respectively. This manipulation was to ensure the believability of the scenario and to avoid weariness of the participant interacting with a single reporter. Although participants knew with whom they were interacting in each trial (because of the presentation of cue), they did not know the distribution of lies beforehand and could only learn this distribution through the game.

Each participant played 100 trials with one of the four reporters in turn, creating four test blocks. The order of the four blocks was counter-balanced over participants, using a Latin Square design. Moreover, the spatial positions of the "trust" and "distrust" cues in each trial, presented on the left and the right side of the screen and requiring the left and right hand responses respectively, were also counter-balanced over participants. Within each block, trials from different conditions were randomized for each participant, with the restriction that no more than eight consecutive trials had the same feedback. Feedback that a participant observed was predetermined by a pseudo-randomized sequence.

When a participant and the four confederates (out of eight, the same sex as the participant) came to the laboratory, they were told to take part in a game in which four of them would be "reporters" while the other one would be a "receiver." The role of each person was ostensively decided by a lottery. They were told that they would sit in different rooms and interact through intranet. In each trial, the reporter, upon observing the outcome of a coin flip, could tell the receiver the result of toss being either head or tail, and this message would appear on the computer screen in front of the receiver, who shall decide whether to trust the report. The receiver was made to believe that each person would gain an extra point for his/her success in getting it right (for the receiver) or in deceiving the receiver (for the reporter). He/she was also made to believe that each point was associated with extra monetary reward.

After the briefing of the general rules of the game and after the lottery, the participant was led to the EEG room and was assigned the role of "receiver." The participant did not know which partner he/she met would be A, B, C, or D. In each trial (Fig. 1), the participant would see first a message "A is deciding . . ." for about 3 s (varied between 2 and 4 s) and then a report, being "head" or "tail," for another 1 s. After the sign "?" for 0.5 s, the "trust" and "distrust" cues were presented on the screen until the participant made choice by pressing a corresponding button on a joy stick. After he/she hit the button, a message "the real result was . . ." appeared for 0.8 s and then screen went blank for 0.3–0.5 s. Finally, the feedback with the word "head" or "tail" appeared at the center of the screen for 1.5 s.

EEG recording and analysis. EEGs were recorded from 64 scalp sites using tin electrodes mounted in an elastic cap (Brain Products, Munich, Germany) according to the international 10–20 system. Vertical electro-oculograms (EOGs) were recorded supra-orbitally from the right eye. The horizontal EOG (HEOG) was recorded from electrodes placed at the outer canthus of left eye. All EEGs and EOGs were referenced online to an external electrode placed on the tip of nose and were re-referenced offline to the mean of the left and right mastoids. Electrode impedance was kept below 10 kΩ for EOG channels and below 5 kΩ for all other electrodes. The biosignals were amplified with a band pass from 0.016 to 100 Hz and digitized online with a sampling frequency of 500 Hz.

The EEG data were preprocessed with Brain Vision Analyzer software. Ocular artifacts were corrected with an eye-movement correction algorithm (Gratton et al., 1983). Continuous EEGs were segmented with an epoch of 1000 ms time-locked to the onset of feedback stimulus (from 200 ms before to 800 ms after the feedback). Trials exceeding ±90 μV in amplitude, containing a transient of over 100 μV in a period of 100 ms, or containing activity lower than 0.5 μV in a period of 100 ms were rejected. Data were then filtered offline with a 30 Hz low-pass filter. Roll-off of the band-pass filter was 24 dB/oct. The segmented EEGs were baseline corrected according to the mean amplitude of the activity pre-onset of the feedback.

For statistical analyses, we focused on two representative electrodes, Fz and CPz (Fig. 4A), although we also conducted analyses for amplitudes on a group of electrodes. Time windows were selected for analysis based on visual inspection of the waveforms. For the FRN, we analyzed the mean amplitudes in the time window of 230–310 ms on Fz; for the P300, we took the peak amplitudes in the time window of 250–450 ms on CPz. We selected these two electrodes because the FRN effect and the P300 responses were the largest on these electrodes, respectively. Effects over the whole scalp are depicted in Fig. 4B. Analyses of variance (ANOVAs) were conducted with two within-participant factors: trust choice and outcome valence. The Greenhouse-Geisser correction was applied when the assumption of sphericity was violated.

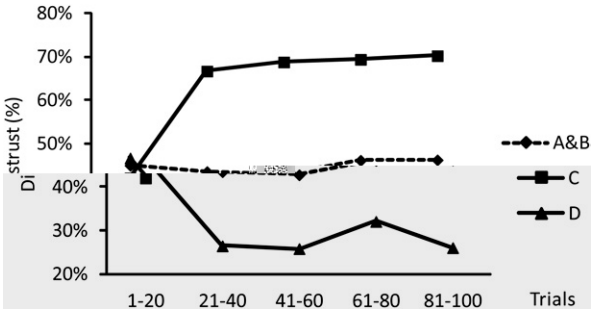

Fig. 3. Trends of responses (distrust choices) for the four types of reporters. A and B were the two reporters lied 50% of the time (percentages illustrated collapsed over the two persons); C lied 70% of the times, and D lied 30% of the times.

Results

Behavioral results. For the 24 participants, we compared the extent to which they trusted reports from the confederate A, B, C, or D by dividing the 100 trials into five bins according to the order of trials. As can be seen from Fig. 3, participants rapidly learnt whom they should trust most. After 20 trials, the percentage of trials on which the participants made the distrust choice was very similar to the objective manipulations. For the last 80 trials, one-sample *t*-test revealed that the probability of participants choosing distrust was equivalent to the objective distribution while playing against Reporter C (M=68.75%, SD=14.28%, $t(23)<1$) or Reporter D (M=27.60%, SD=14.41%, $t(23)<1$). However, participants' percentage of distrust choices was significantly lower than the objective probability (i.e. 50%) when they played with Reporter A (and B) (M=44.69%, SD=9.48%, $t(47)=-3.88$, $P<0.001$). In the analysis of EEG data, we included only the last 80 trials from each reporter, although the same pattern of effects was obtained when all the 100 trials were included.

ERP results. The reporter type was not treated as a variable in this experiment because there were not enough trials for this analysis. ERP responses to feedback were then sorted into four groups and entered into statistical analysis: trust–gain, trust–no-gain, distrust–gain, and distrust–no-gain. Waveforms for the four conditions on Fz and CPz are depicted in Fig. 4A. The FRN and P300 effects over the whole scalp are depicted in Fig. 4B.

ANOVA on the mean amplitudes of the FRN on Fz revealed a main effect of trust choice, $F(1,23)=8.80$, $P<0.01$, and a main effect of outcome valence, $F(1,23)=13.01$, $P<0.001$. "Distrust" decisions elicited a more negative-going FRN (8.46 μV) than "trust" decisions (9.92 μV); incorrect guesses elicited more negative-going FRN (7.92 μV) than correct guesses (10.46 μV). Importantly, the interaction between trust choice and outcome valence was significant, $F(1,23)=5.53$, $P<0.05$. Further tests showed that the FRN effect (i.e. the difference between "no gain" and "gain" trials) following trust choices was highly significant (3.97 μV), $F(1,23)=17.21$, $P<0.001$; the FRN following distrust choices, however, did not reach significance (1.11 μV), $F(1,23)=1.50$, $P>0.1$. The same pattern of effects was obtained when we included an array of nine electrodes surrounding Fz (FP1, FPz, FP2, F1, Fz, F2, FC1, FCz, FC2) into the analysis.

Similar analyses were conducted for the peak values of the P300. Here we found a significant main effect of trust

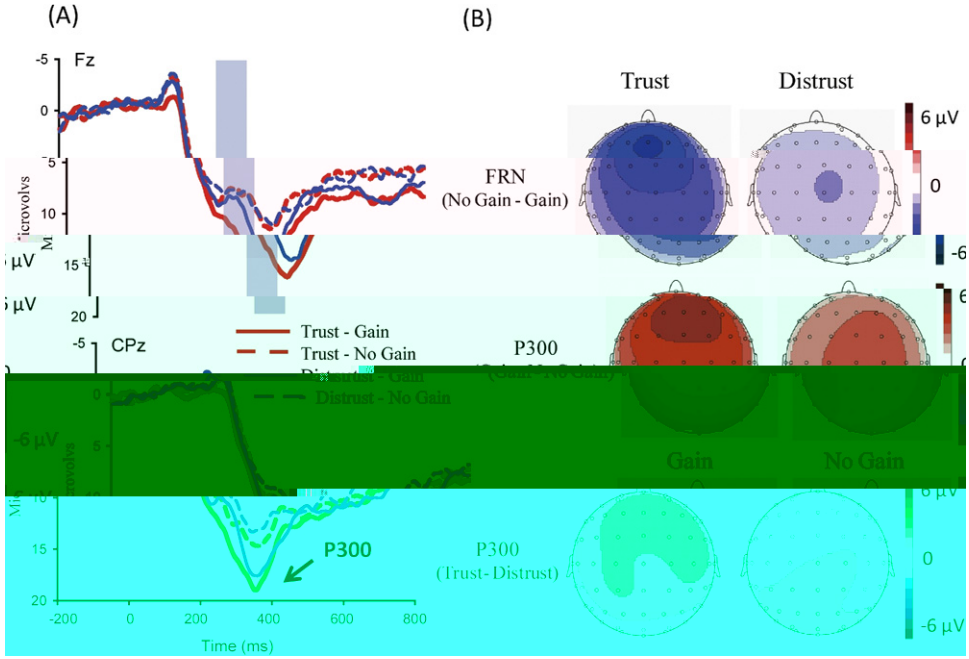

Fig. 4. (A) ERP waveforms time-locked to the onset of feedback stimuli in Experiment 1, sorted according to the participants' choices and outcomes. (B) Scalp topographies of the difference waves between ERP responses to the no-gain vs. gain outcomes averaged for the 230–310 ms time window (the upper panels) and between the gain vs. no-gain outcomes for peak values in the 250–450 ms time window (the lower panels).

choice, $F(1,23)=6.72$, $P<0.05$, and a main effect of outcome valence, $F(1,23)=16.22$, $P<0.001$. The P300 was more positive following "trust" decisions (19.06 μV) than following "distrust" decisions (17.39 μV) and was more positive for the gain trials (19.72 μV) than for the no-gain trials (16.72 μV). The interaction between the two factors was not significant, $F(1,23)<1$. Again, the same pattern of effects was obtained when we analyzed data from an array of nine electrodes centering on CPz (C1, Cz, C2, CP1, CPz, CP2, P1, Pz, P2). Note, although on Fig. 4B the P300 effect appeared to be the largest on some frontal electrodes, this increased effect was very likely due to the contamination by the FRN effect in a slightly earlier time window.

Discussion

The main effect of outcome valence on the FRN replicated many previous studies (Gehring and Willoughby, 2002; Holroyd and Coles, 2002; Leng and Zhou, 2010; Yeung et al., 2005; Zhou et al., 2010). Importantly, this FRN effect was modulated by trust choice, with a significant FRN effect following trust choices but a non-significant effect following distrust choices. Given the novelty of this finding, we decided to replicate it in the next experiment. Accounts for this finding shall be given in General Discussion. On the other hand, the valence effect on the P300 replicated some previous studies (Hajcak et al., 2005, 2007; Wu and Zhou, 2009), whereas the trust choice effect on the P300 was consistent with our prediction that making a trust decision leads to greater self-involvement and more devoted attentional processing of the outcome.

One methodological limitation of this experiment was that in order to have enough trials for ERP averaging we did not distinguish reporter types that were associated with different probabilities of participants receiving positive or negative outcome feedback. Previous studies on outcome evaluation showed that the probability of the gain or loss outcome may affect the manifestation of the FRN (Holroyd et al., 2003) and the P300 (Linden, 2005). It is plausible that the variation of the outcome probability between different reporter types had somehow affected the valence and the trust choice effects on the FRN and/or P300 even although the overall probability of the gain or no-gain outcome in this experiment was 50%, collapsing over reporter types. In Experiment 2, we used only reporters that lied in 50% of the trials.

Another methodological limitation of Experiment 1 was that we blocked the reporter types when we presented coin-toss trials to the participants. It is plausible that the participants may have got accustomed to the behavioral pattern of the partner in each block and developed some response strategies that could affect the manifestation of the FRN and P300 effects. In Experiment 2, we randomly mixed trials for different types of reporters. The empiric question was whether the patterns of the FRN and P300 effects observed in Experiment 1 would be replicated in Experiment 2.

EXPERIMENT 2

Besides randomly mixing trials for different types of reporters and changing the percentage of reporters lying over the result of coin-toss, Experiment 2 manipulated the intentionality of the reporter. Participants would play against either a human or a computer partner. While reports from a human partner can be regarded as resulting from intentional decisions, reports from a computer partner can be seen as resulting from random selection of options and as having no particular motivation to overpower the participants. It is not clear yet whether the perceived intentionality of the partner would affect the participants' behavioral responses to the truthfulness of statements and their brain responses to the outcomes.

Experimental procedures

Participants. Eighteen undergraduate students (eight males) were recruited from Beijing Normal University, Beijing Jiaotong University, and Beijing Science and Technology College, aged 20–24 years. All the participants were healthy and right-handed. Informed consent was obtained from each participant.

Stimuli and procedures. This experiment had a two (reporter type: intentional vs. unintentional) by two (trust choice: trust vs. distrust) by two (outcome valence: gain vs. no-gain) factorial design. Both types of reporters gave truthful statements for 50% of the times. There were 200 trials for each reporter type. Trials from different conditions were randomly mixed, with the restriction that the first 40 trials were composed of five trials from each condition. These trials were considered as practice trials and were excluded from data analysis. Within each type, half of the trials following the "trust" or "distrust" choice had feedback consistent with the initial coin-toss results (i.e. head or tail).

When two participants of the same sex came to the laboratory, they were told to take part in a game in which one of them would be "reporter" while the other one would be a "receiver." The role of each person was ostensively decided by a lottery. They were told that they would sit in different rooms and interact through intranet with each other or with a computer program. In each trial, upon tossing a coin the reporter (the human or computer partner) would tell the receiver the result of toss as being either head or tail, and this message would appear on the computer screen in front of the receiver, who shall decide whether to trust the report. Each person would gain an extra point for his/her success in getting it right (for the receiver) or in deceiving the receiver (for the human reporter).

After the briefing of the general rules of the game and after the lottery, the participants were led to different EEG rooms and were assigned the role of "receiver" (i.e. both acting as EEG participants). In each trial, a participant would see first a message of either "The partner is deciding . . ." or "The computer is deciding . . ." for about 2.3 s. This time varied between 1 and 5 s for the human partner and was kept a constant 1.5 s for the computer partner. Then a report, being "head" or "tail," was presented for 1 s. The

"trust" and "distrust" cues then were presented on the screen until the participant made choice by pressing a corresponding button on a joy stick. Note, unlike Experiment 1, we did not present a sign of "?" for 0.5 s before the "trust" and "distrust" cues. After he/she hit the button, a line of words "the real result was . . ." appeared for 0.8 s and the screen went blank for 0.3–0.5 s. Finally, the feedback with the word "head" or "tail" appeared at the center of the screen for 1.5 s.

After the EEG session, each participant was given a post-experiment questionnaire asking them to write down his/her perceived percentage of truthful reports from each reporter. The participant was also required to indicate on a 7-point scale the level of trust for the two reporters and the level of self-involvement when playing with the two reporters.

EEG recording and analysis. EEG recording and statistical analyses were conducted in the same way as Experiment 1. We again focused on two representative electrodes, Fz and CPz (Fig. 5). Time windows for analysis was determined according to visual inspection of the waveforms. For the FRN, we analyzed the mean amplitudes in the time window of 270–350 ms on Fz; for the P300, we took the peak amplitudes in the time window of 280–500 ms on CPz. We selected these two electrodes because the FRN effects and the P300 responses were the largest on these electrodes, respectively. Note also that for the FRN and P300 here we used time windows that differed from those in Experiment 1 in order to maximize the possibility of observing differential responses for conditions. Experiment 2 differed from Experiment 1 not only on the how many types of reporters were used but also on whether a "?" frame was presented for a trial. It is plausible that because Experiment 2 omitted the "?" frame, participants were less prepared for the processes of outcome evaluation and hence the time window for appearance of the FRN effect was delayed, compared with Experiment 1. ANOVAs were conducted with three within-participant factors: reporter type, trust choice, and outcome valence. The Greenhouse-Geisser correction was applied when the assumption of sphericity was violated.

Results

Behavioral results. The post-experiment questionnaire showed that the perceived percentages of truthful reports from intentional reporters (M=47.78%, SD=11.4%) and unintentional reporters (M=48.61%, SD=13.48%) did not differ from each other, $t(17)<1$, neither the level of trust for intentional reporters (M=3.82, SD=0.99) and unintentional reporters (M=4.39, SD=0.78), $t(17)<1$. These results may indicate that the perceived intentionality of partners play no obvious role in determining participants' behavioral responses to the truthfulness of reports. Moreover, there was a strong correlation between participants' perceived percentage of the reporter's truthful statements and participants' rating of the reporter's trustworthiness, r=.646, $P<0.005$ for the human partner; r=.560, $P<0.05$ for the computer partner. The variation of the perceived percentage of true statements over participants and the correlations between the two measurements strongly suggest that participants cared for the truthfulness of the re-

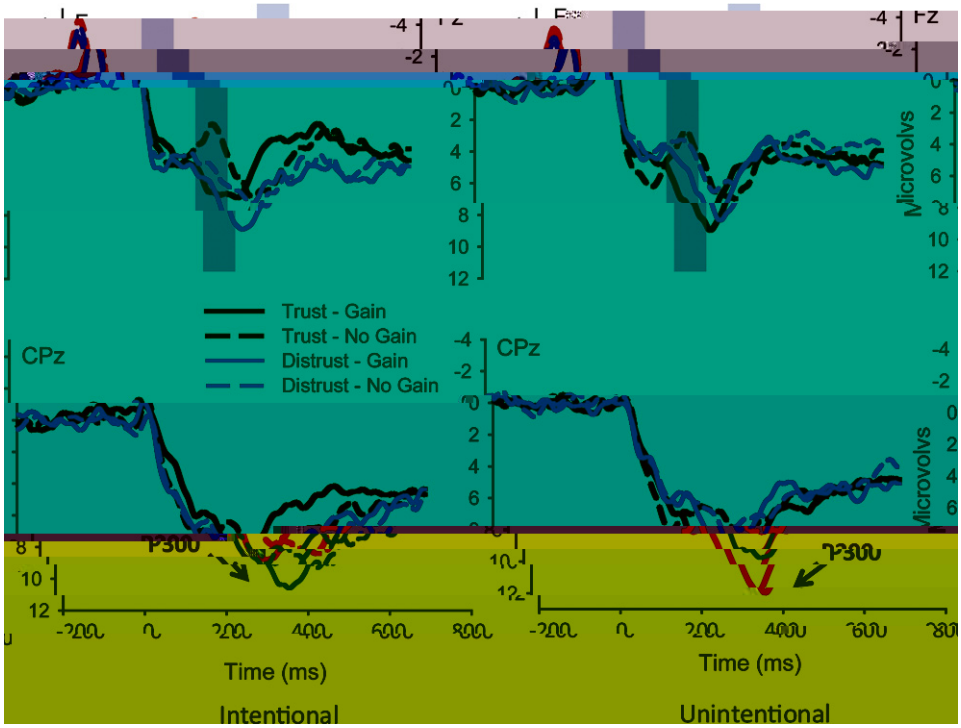

Fig. 5. ERP waveforms time-locked to the onset of feedback stimuli in Experiment 2, sorted as a function of reporter type, the participants' choice, and the valence of outcome.

porters' statements, rather than treating the feedback as simple match versus mismatch information.

On the other hand, there was a significant difference between participants' level of self-involvement when playing with the human or computer partner, $t(17)=4.12$, $P=0.001$. Ratings for playing with intentional reporters ($M=5.94$, SD=0.94) were higher than those for playing with unintentional reporters ($M=5.11$, SD=1.18).

ERP results. ERP responses to feedback were sorted into eight groups and entered into statistical analysis, crossing the three variables. Waveforms for the eight conditions on Fz and CPz are depicted in Fig. 5.

ANOVA on the mean amplitudes of the FRN on Fz revealed a significant main effect of outcome valence, $F(1,17)=26.44$, $P<0.001$. Incorrect guesses elicited more negative-going FRN (3.81 μV) than correct guesses (6.06 μV). Importantly, the interaction between trust choice and outcome valence was significant, $F(1,17)=5.22$, $P<0.05$. Further tests showed that the FRN effect following trust choices was highly significant (3.21 μV), $F(1,17)=36.93$, $P<0.001$; the FRN effect following distrust choices (1.31 μV) was only marginally significance, $F(1,17)=4.29$, $P=0.054$. The main effect of trust choice did not reach significance, neither the main effect of reporter type, both $F(1,17)<1$.

Similar analyses were conducted for the peak amplitudes of the P300. Here we found a significant main effect of outcome valence, $F(1,17)=9.60$, $P<0.01$, with the P300 being more positive for the gain (12.11 μV) than for the no-gain trials (10.69 μV). Moreover, we found a significant interaction between reporter type and outcome valence, $F(1,17)=19.58$, $P<0.01$. For intentional reporters, gain trials (11.40 μV) did not differ from no-gain trials (11.11 μV), $F(1,17)<1$; for unintentional reporters, however, gain trials (12.82 μV) were larger than no-gain trials (10.27 μV), $F(1,17)=17.39$, $P=0.001$.

On the other hand, although the interaction between reporter type and trust choice was significant, $F(1,17)=5.34$, $P<0.05$, further tests did not show any significant simple effects, all $P>0.1$. The main effects of reporter type and trust choice were not significant, both $F(1,17)<1$, neither the three-way interaction between reporter type, trust choice, and outcome valence, $F(1,17)=1.08$, $P>0.1$.

Discussion

The main effect of outcome valence on both the FRN and P300 replicated Experiment 1. Importantly, the interaction between trust choice and outcome valence on the FRN was also replicated. As in Experiment 1, the FRN responses to gain and no-gain trials differed significantly following trust choices but not much so following distrust choices. Moreover, this pattern of the FRN effect was held whether the participants were playing with human or computer partners. However, the reporter type did affect the P300 responses, with a significant valence effect for the computer partner but not for the human partner. We will explore the significance of these findings in General Discussion.

We did not observe a significant main effect of reporter type on either the FRN or the P300. This null effect was inconsistent with the post-experiment subjective rating with higher self-involvement when playing with the human partner than with the computer partner. A possible explanation for this discrepancy is that the ERPs measure online brain responses, whereas the post-experiment questionnaire measures more global, reflective feeling. Nevertheless, we believe that the potential impact of intentionality upon outcome evaluation and performance monitoring is worth further investigation.

Unlike Experiment 1, Experiment 2 did not find a trust choice effect on the P300. This null effect might be due to the change in experimental setup. In Experiment 1, although four types of reporters were included, two of them had comparatively predictable behavioral pattern (lying 70% and 30% of the times, respectively). This higher predictability was coupled with the block presentation, making it relatively easy for the participants to predict the reporters' behavior. In Experiment 2, however, the reporters lied 50% of the times and trials for different types of reporters were randomly mixed. Further experiments are needed to investigate how this change of experimental context would affect the encoding of trust choice on the P300.

GENERAL DISCUSSION

Two experiments obtained consistent findings concerning how trust choice could modulate brain responses to decision outcomes: the FRN was more negative going for no-gain outcomes than for gain outcomes following trust choices but this effect was reduced following distrust choices. Moreover, the intentionality of the partner had no apparent impact upon this pattern of FRN effects. On the other hand, the P300 also encoded outcome valence, but this valence effect was modulated by the partner's intentionality. In the following paragraphs, we explore the significance of these findings.

The differential FRN responses to gain and no-gain outcomes following trust choices may reflect the detection of social expectancy violation. The FRN effect is commonly accounted for by the reinforcement learning theory (Holroyd and Coles, 2002; Nieuwenhuis et al., 2004; Yeung et al., 2004), which suggests that the FRN encodes a reward prediction error that occurs when the ongoing event is worse than expected. Studies also showed that the prediction error can be defined not only in terms of the valence of outcome but also in terms of whether the outcome fits pre-established, non-valence expectancy (Jia et al., 2007; Wu and Zhou, 2009). A trust decision correlates with strong positive expectation towards the other party and the outcome (Rotter, 1967; Rousseau et al., 1998), even when the individual's trust choice has no effect upon the other party's behavior. Violation of this kind of social expectation could contribute to (i.e. enlarge) the encoding of outcome valence by the FRN, although this suggestion should be tested directly (see Yeung et al., 2005).

Importantly, the two experiments consistently observed a reduced FRN effect for the no-gain vs. gain

outcome following distrust choices. A simple account for this diminished effect is that a distrust choice signals a sense of aloofness, creating a context in which participants generate little, if at all, expectancy towards the upcoming events. Then whether the actual results (the feedback) are consistent or inconsistent with what the partners reported (i.e. with the reversed predictions towards the outcomes based on the "distrust" decision) makes no or not much difference to the participants and hence does not elicit differential FRN responses to the outcomes. This account is consistent with the motivational/affective hypothesis of the FRN (Gehring and Willoughby, 2002): since the FRN correlates with the level of self-involvement (Yeung et al., 2005) and reflects the motivational significance of outcomes, detached concern towards the outcomes would elicit no FRN effect.

Alternatively, the absence of a significant FRN effect following distrust choices may reflect the conflict between outcome valence and the evaluation of truthfulness of statements made by the partners. After a "distrust" decision, the outcome consistent with what the reporters stated (i.e. a truthful statement) was actually a negative outcome for the participants, because it resulted in "no gain" for participants, and should elicit negative-going FRN responses. However, finding out from the feedback that the partners have told the truth (i.e. the truth itself) should elicit a positive-going FRN response, as encoding of the trustworthy of others is an automatic process that are carried out within a few hundred milliseconds (Rudoy and Paller, 2009). That is, the FRN encodes two dimensions of the outcome, one dimension in terms of participants' self-interests (no gain vs. gain) and another dimension in terms of the truthfulness of the original statement (telling lie vs. telling truth). The two effects, for the two dimensions respectively, might (partially) cancel each other, resulting in an overall less negative-going response for the "distrust–no gain" condition. The same reasoning can be applied to the situation in which outcomes were inconsistent with what the reporters stated but would lead to gains for the participants after "distrust" decisions. Thus, when brain responses to the two situations are compared, the gain and no-gain outcomes following distrust choices produce little differential FRN responses. Noted that although following "distrust" choices, the FRN encoding of outcome valence and of the truthfulness of the original statement was in opposite directions, canceling each other, the encoding of the two dimensions were congruent following "trust" choices and a significant FRN effect could then be revealed for the contrast of "no gain vs. gain."

Detailed examination of the FRN effects suggests that the second account is more likely to stand than the first. Experiment 2 did observe a marginally significant FRN effect for outcomes following distrust choices. Moreover, when we pool together the FRN effects following distrust decisions from the two experiments (collapsing over the reporter type in Experiment 2), treating experiment as a between-participant factor, we found that the aggregated FRN effect (1.26 μV) was significant, $F(1,40)=4.61$, $P<0.05$. This finding is inconsistent with the aloofness account but consistent with the second account in which different aspects of the outcome are encoded simultaneously by the FRN. Moreover, the outcome valence effect on the P300 also indicate that the evaluation system is not aloof to aspects of outcomes following distrust choices.

The persistent outcome valence effect on the P300 in both experiments, replicating previous studies (Hajcak et al., 2005, 2007; Leng and Zhou, 2010; Wu and Zhou, 2009), suggests that in a later stage of outcome evaluation (Leng and Zhou, 2010), personal interests are taken into account by the system even when distrust choices have been made previously. The more positive responses to gain outcomes than to no-gain outcomes indicate that more attentional resources (Donchin and Coles, 1998; Gray et al., 2004; Linden, 2005) are devoted to outcomes that benefit oneself.

CONCLUSION

By presenting statements concerning results of coin-toss to participants and by asking them to make trust or distrust choices to the statements, this study demonstrated in two experiments that trust choice modulates the brain activity in evaluating gain and no-gain outcomes. The outcome valence effect was observed on both the FRN and the P300 following trust choices but the effect on the FRN was reduced following distrust choices. Making a trust choice generates strong expectation towards a positive outcome and enhances the level of self-involvement, whereas making a distrust choice would force the system to encode aspects of decision outcomes simultaneously. The FRN may reflect the subjective evaluation of decision outcome in a specific context rather than a general expectancy towards the outcome.

Acknowledgments—This study was supported by National Basic Research Program (973 Program: 2010CB833904) and by grants from Natural Science Foundation of China (30110972, J1103602). We thank Dr. Stephen Crites and an anonymous reviewer for their constructive comments on early versions of the manuscript.

REFERENCES

Boudreau C, McCubbins MD, Coulson S (2009) Knowing when to trust others: an ERP study of decision making after receiving information from unknown people. Soc Cong Affect Neurosci 4:23–34.

Buchan NR, Croson RTA, Dawes RM (2002) Swift neighbors and persistent strangers: a cross-cultural investigation of trust and reciprocity in social exchange. Am J Sociol 108(1):168–206.

Cho J (2006) The mechanism of trust and distrust formation and their relational outcomes. J Retailing 82(1):25–35.

DeNeve KM, Cooper H (1998) The happy personality: a meta-analysis of 137 personality traits and subjective well-being. Psychol Bull 124(2):197–229.

Dimoka A (2010) What does the brain tell us about trust and distrust? Evidence from a functional neuroimaging study. MIS Q 34(2): 373–396.

Donchin E, Coles MGH (1998) Is the P300 component a manifestation of context updating? Behav Brain Sci 11:357–374.

Gambetta D (1988) Trust: making and breaking cooperative relations. New York: Basil Blackwell.

Gehring WJ, Willoughby AR (2002) The medial frontal cortex and the rapid processing of monetary gains and losses. Science 295:2279–2282.
Gratton G, Coles MGH, Donchin E (1983) A new method for off-line removal of ocular artifacts. Electroenceph Clin Neurophysiol 55:468–484.
Gray HM, Ambady N, Lowenthal WT, Deldin P (2004) P300 as an index of attention to self-relevant stimuli. J Exp Soc Psychol 40:216–224.
Hajcak G, Holroyd CB, Moser JS, Simons RF (2005) Brain potentials associated with expected and unexpected good and bad outcomes. Psychophysiology 42(2):161–170.
Hajcak G, Moser JS, Holroyd CB, Simons RF (2007) It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. Psychophysiology 44(6):905–912.
Heldmann M, Rüsseler J, Münte TF (2008) Internal and external information in error processing. BMC Neurosci 9:33–40.
Holroyd CB, Coles MGH (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Review 109:679–709.
Holroyd CB, Nieuwenhuis S, Yeung N, Cohen JD (2003) Errors in reward prediction are reflected in the event-related brain potential. Neuroreport 14(18):2481–2484.
Jia S, Li H, Luo Y, Chen A, Wang B, Zhou X (2007) Detecting perceptual conflict by the feedback-related negativity in brain potentials. Neuroreport 18(791):1385–1388.
Kang SK, Hirsh JB, Chasteen AL (2010) Your mistakes are mine: self-other overlap predicts neural response to observed errors. J Exp Soc Psychol 46(1):229–232.
King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR (2005) Getting to know you: reputation and trust in a two-person economic exchange. Science 308:78–82.
Knack S, Keefer P (1997) Does social capital have an economic payoff? A cross-country investigation. Q J Econ 112(4):1251–1288.
Kramer RM, Cook KS (2004) Trust and distrust in organizations: dilemmas and approaches. New York: Russell Sage Foundation.
Leng Y, Zhou X (2010) Modulation of the brain activity in outcome evaluation by interpersonal relationship: an ERP study. Neuropsychologia 48:448–455.
Lewicki R, McAllister D, Bies R (1998) Trust and distrust: new relationships and realities. Acad Manag Rev 23(12):438–458.
Li P, Jia S, Feng T, Liu Q, Suo T, Li H (2010) The influence of the diffusion of responsibility effect on outcome evaluations: electrophysiological evidence from an ERP study. Neuroimage 52(4):1727–1733.
Linden DE (2005) The P300: where in the brain is it produced and what does it tell us? Neuroscientist 11:563–576.
Lupia A, McCubbins MD (1998) The democratic dilemma: can citizens learn what they need to know? New York: Cambridge University Press.
Ma Q, Shen Q, Xu Q, Li D, Shu L, Weber B (2011) Empathic responses to others' gains and losses: an electrophysiological investigation. Neuroimage 54(3):2472–2480.
Marco-Pallarés J, Krämer UM, Strehl S, Schröder A, Münte TF (2010) When decisions of others matter to me: an electrophysiological analysis. BMC Neurosci 11:86–93.
Mayer RC, Davis JH, Schoorman D (1995) An integrative model of organizational trust. Acad Manag Rev 20:709–734.
McKnight DH, Choudhury V (2006) Distrust and trust in B2C e-commerce: do they differ? In proceedings of the 2006 International Conference on Electronic Commerce, Fredericton, New Brunswick, Canada.
McKnight DH, Kacmar CJ, Choudhury V (2003) Whoops . . . did I use the wrong concept to predict e-commerce trust? Modeling the risk-related effects of trust versus distrust concepts. In proceedings of the 36th Hawaii International Conference on System Sciences, Los Alamitos, CA: IEEE Computer Society Press.
Meyerson D, Weick KE, Framer RM (1996) Swift trust and temporary groups. In: Trust in organizations: frontiers of theory and research (Framer RM, Tyler TR, eds), pp 166–195. Thousand Oaks, CA: Sage.
Miles RE, Snow CC (1992) Causes of failure in network organizations. Calif Manag Rev 34:53-72.
Miltner WHR, Braun CH, Coles MGH (1997) Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a "generic" neural system for error detection. J Cogn Neurosci 9:788–798.
Morrison DE, Firmstone J (2000) The social function of trust and implications for e-commerce. Int J Advertising 15(5):1–17.
Nieuwenhuis S, Yeung N, Holroyd CB, Schuiger A, Cohen JD (2004) Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. Cereb Cortex 14:741–747.
Pavlou PA, Gefen D (2004) Building effective online marketplaces with institution-based trust. Inf Syst Res 15(1):37–60.
Phan KL, Sripada CS, Angstadt M, McCabe K (2010) Reputation for reciprocity engages the brain reward center. Proc Natl Acad Sci U S A 107(29):13099–13104.
Rotter JB (1967) A new scale for the measurement of interpersonal trust. J Pers 35:615-554.
Rousseau DM, Sitkin SB, Burt RS, Camerer C (1998) Not so different after all: a cross-discipline view of trust. Acad Manag Rev 23(3):393–404.
Rudoy JD, Paller KA (2009) Who can you trust? Behavioral and neural differences between perceptual and memory-based influences. Front Hum Neurosci 3:16.
Sato A, Yasuda A, Ohira H, Miyawaki K, Nishikawa M, Kumano H, et al. (2005) Effects of value and reward magnitude on feedback negativity and P300. Neuroreport 16(4):407–411.
Wu Y, Leliveld MC, Zhou X (2011) Social distance modulates recipient's fairness consideration in the dictator game: an ERP study. Biol Psychol 88:253–262.
Wu Y, Zhou X (2009) The P300 and reward valence, magnitude, and expectancy in outcome evaluation. Brain Res 1286:114–122.
Yeung N, Cohen JD, Botvinick MM (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol Review 111(4):931–959.
Yeung N, Holroyd CB, Cohen JD (2005) ERP correlates of feedback and reward processing in the presence and absence of response choice. Cereb Cortex 15(5):535–544.
Yeung N, Sanfey AG (2004) Independent coding of reward magnitude and valence in the human brain. J Neurosci 24(28):6258–6264.
Zhou Z, Yu R, Zhou X (2010) To do or not to do? Action enlarges the FRN and P300 effects in outcome evaluation. Neuropsychologia 48(12):3606–3613.

(Accepted 19 October 2011)
(Available online 30 October 2011)