*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

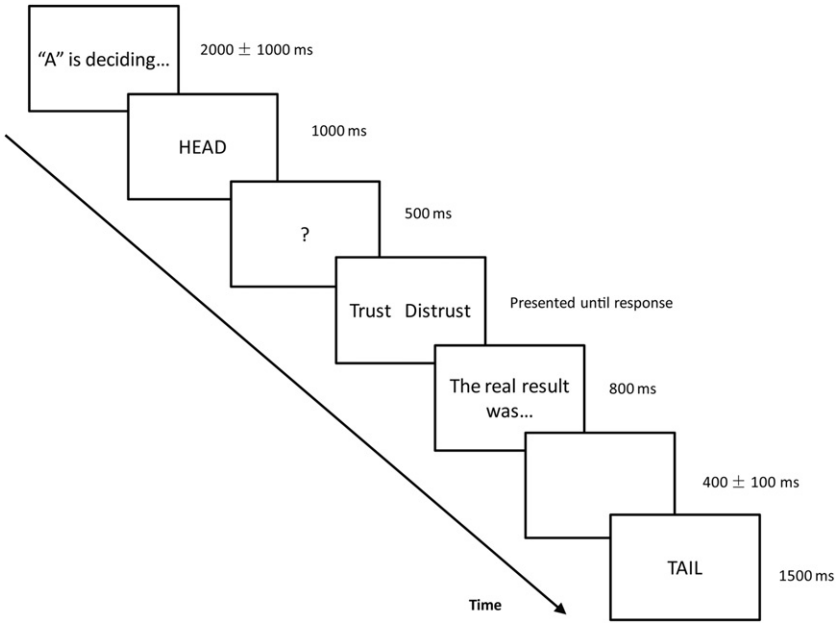


**Fig. 1.** Sequence of events in a single trial.

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