Biological Psychology 114 (2016) 13–22

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Novel, ERP-based, concealed information detection: Combining recognition-based and feedback-evoked ERPs

Liyang Sai[a,f,1], Xiaohong Lin[c,1], J.Peter Rosenfeld[d], Biao Sang[a], Xiaoqing Hu[e,*], Genyue Fu[b,f,**]

[a] School of Psychology & Cognitive Science, East China Normal University, Shanghai, China
[b] Department of Psychology, Hangzhou Normal University, Hangzhou, China
[c] Bioimaging Core, Faculty of Health Sciences, University of Macau, Macau SAR, China
[d] Department of Psychology, Northwestern University, Evanston, USA
[e] Department of Psychology, University of Texas at Austin, Austin, USA
[f] Department of Psychology, Zhejiang Normal University, Jinhua, China

ARTICLE INFO

Article history:
Received 26 December 2014
Received in revised form
18 November 2015
Accepted 24 November 2015
Available online 27 November 2015

Keywords:
Memory detection
Feedback-related negativity
P300
Concealed information detection
Deception detection

ABSTRACT

The present study introduced a novel variant of the concealed information test (CIT), called the feedback-CIT. By providing participants with feedbacks regarding their memory concealment performance during the CIT, we investigated the feedback-related neural activity underlying memory concealment. Participants acquired crime-relevant memories via enacting a lab crime, and were tested with the feedback-CIT while EEGs were recorded. We found that probes (e.g., crime-relevant memories) elicited larger recognition-P300s than irrelevants among guilty participants. Moreover, feedback-related negativity (FRN) and feedback-P300 could also discriminate probes from irrelevants among guilty participants. Both recognition- and feedback-ERPs were highly effective in distinguishing between guilty and innocent participants (recognition-P300: AUC = .73; FRN: AUC = .95; feedback-P300: AUC = .97). This study sheds new light on brain-based memory detection, such that feedback-related neural signals can be employed to detect concealed memories.

1. Introduction

An accurate technique in detecting individual memories stored in the brain can have significant implications for fields such as neuroscience and law, criminal justice and forensic assessment. Research in neuroscience-based memory detection can help develop valid paradigms and sensitive neural markers to reveal memory status of certain information (i.e., old vs. new). Memory detection can be particularly helpful when an examinee is motivated to conceal past experiences (e.g., deception, malingering, memory concealment; Farah, Hutchinson, Phelps, & Wagner, 2014; Rosenfeld, Hu, Labkovsky, Meixner, & Winograd, 2013), or even when the examinee does not have conscious access to past experiences (as in prosopagnosia or dissociative identity disorders; Allen, 2011; Tranel & Damasio, 1985). Here, we present novel evidence that event-related brain potentials (ERPs) that are associated with feedback processing can accurately identify concealed, crime-relevant memories. Moreover, these feedback-ERPs add incremental validity in memory detection above and beyond traditionally used, recognition-based ERPs.

The concealed information test (CIT, also known as the guilty knowledge test, Lykken, 1959) has been studied extensively as a scientifically valid tool to identify concealed memories (for an overview, see Verschuere, Ben-Shakhar, & Meijer, 2011). During a CIT, participants are presented with at least two types of stimuli: (1) a rarely presented, crime-relevant item (i.e., a probe) that is meaningful to a guilty examinee who possesses crime-related knowledge, (2) a series of frequently presented crime-irrelevant items (i.e., irrelevants) that should be meaningless to any examinee, guilty or innocent. The assumption of the CIT is that for criminals who commit the crime, the crime-relevant probes will elicit a distinctive profile of behavioral and physiological responses compared with those elicited by meaningless irrelevant stimuli. On the other hand, for examinees who do not possess the crime-relevant information (e.g., innocent people who are not aware of the crime), the probe is simply another meaningless irrelevant. Thus, for

* Corresponding author.
** Corresponding author at: Department of Psychology, Hangzhou Normal University, Hangzhou, China.
E-mail addresses: xqhu@utexas.edu (X. Hu), fugenyue@hznu.edu.cn (G. Fu).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.biopsycho.2015.11.011
0301-0511/

innocent examinees, responses associated with the probe should not differ from those associated with irrelevant stimuli. The CIT has been studied with autonomic nervous system activities (e.g., electrodermal activity or heart rate, Ben-Shakhar & Elaad, 2003; Gamer, Verschuere, Crombez, & Vossel, 2008); central nervous system activities (e.g., ERPs and especially the P300, Allen, Iacono, & Danielson, 1992; Farwell & Donchin, 1991; Rosenfeld et al., 1988, Rosenfeld, Angell, Johnson, & Qian, 1991; for reviews, see Rosenfeld, 2011, Rosenfeld et al., 2013 blood-oxygen level dependent (BOLD) responses, Gamer, Klimecki, Bauermann, Stoeter, & Vossel, 2012; Langleben et al., 2002; Sai, Zhou, Ding, Fu, & Sang, 2014b for a review, see Gamer, 2011).

Orienting response theory (OR, Siddle, 1991; Sokolov, 1963) has been proposed as one theoretical foundation for the CIT (Verschuere & Ben-Shakhar, 2011). That is, because the probe triggers automatic recognition or memory retrieval (e.g., either a participant's own name, or a crime-relevant detail for criminals), such recognition will increase the probe's signal value and motivational significance and will thus elicit a distinctive profile of behavioral and physiological responses (Verschuere & Ben-Shakhar 2011). Despite the OR's putatively dominant role in explaining the CIT, memory concealment may involve processes other than memory retrieval or recognition, such as response inhibition or performance monitoring (Verschuere, Crombez, Koster, Van Bockstaele, & De Clercq, 2007; Hu, Wu, & Fu, 2011; Hu, Pornpattananangkul, & Rosenfeld, 2013). Recently, a series of studies found that feedback or reward processing plays an important role during deception or memory concealment (Ding, Sai, Fu, Liu, & Lee, 2014; Hu, Pornpattananangkul, & Nusslock, 2015a; Sai, Lin, Hu, & Fu, 2014a; Sip et al., 2012). In these studies, participants were typically provided with "win" or "lose" feedback following either deceptive or truthful responses, or following probe or irrelevant stimuli. Across different neuroimaging techniques (e.g., ERPs, functional near infrared spectroscopy, functional magnetic resonance imaging), these studies consistently found that feedback elicited non-overlapping neural activities as a function of deceptive or honest response, or as function of memory status (Ding et al., 2014; Hu et al., 2015a; Sai et al., 2014a; Sip et al., 2012).

Based on this evidence, we hypothesized that during a CIT, guilty participants will process feedback differently than their innocent counterparts because probes will be meaningful and motivationally significant only among guilty participants. Moreover, given that feedback processing engages performance evaluation and/or reward-related processing whereas memory processing engages recognition and/or retrieval, we hypothesized that feedback-related ERPs should effectively complement existing, recognition-based ERPs such as the P300. To test these hypotheses, we incorporated performance feedback into an ERP-based CIT, called the feedback-CIT (fCIT). Therefore, this novel fCIT task allows us to investigate ERPs that are time-locked to positive/negative feedback stimuli following the probe and irrelevant stimuli. This will be the first time that we incorporate the feedback-ERPs with recognition-ERPs in identifying incidentally acquired, crime-relevant memories.

Regarding feedback-related ERPs, a considerable amount of research reports that an ERP component, known as the feedback-related negativity (FRN), is sensitive to feedback valence such as correct vs. incorrect, win vs. loss, better or worse (Bismark, Hajcak, Whitworth, & Allen, 2013; Gehring & Willoughby, 2002; Holroyd & Coles, 2002; Miltner, Braun, & Coles, 1997). FRN is typically observed during a 200-350 ms time window after the onset of feedback stimuli. In particular, negative feedback such as "lose" elicits a negative deflection of on-going ERP waveforms than positive feedback such as "win", and this FRN is characterized with a fronto-central scalp distribution (Gehring & Willoughby, 2002; Miltner et al., 1997; Walsh & Anderson, 2012). Recent studies, however, employed more sophisticated analyses to decompose ERPs and found that this FRN was driven by a positive potential elicited by a positive feedback, rather than a negative potential elicited by a negative feedback. For example, when using temporal principal component analyses to decompose overlapping ERP activities (e.g., FRN, P300, late slow waves), studies showed that positive feedback such as reward cues elicited positive ERPs while negative feedback elicited dampened ERPs (Foti, Weinberg, Dien, & Hajcak, 2011; Proudfit, 2015). Moreover, a recent simultaneous EEG-fMRI recording study revealed that the ERPs elicited by positive feedback was positively correlated with BOLD responses in reward-related brain regions such as the ventral striatum and the mid-frontal brain regions (Becker, Nitsch, Miltner, & Straube, 2014; see also Carlson, Foti, Mujica-Parodi, Harmon-Jones, & Hajcak, 2011).

In the context of memory detection, we hypothesized that the guilty participants would be intrinsically motivated to pay more attention to positive/negative feedback during the CIT especially after probe, when compared to their innocent counterparts who are naïve about probe. This is because for guilty participants, a probe has significant signal value due to its meaningfulness among other meaningless irrelevant stimuli. For innocent participants, since the probe is simply another meaningless irrelevant, they should process feedback equally following either the probe or irrelevant stimulus. We thus expected that the feedback-related ERPs could discriminate probe from irrelevants only among guilty but not innocent participants.

2. Methods

2.1. Participants

Thirty-two participants (16 males, Mean age = 19.5 years, S.D. = 1.3 years) were recruited for monetary compensation (30 RMB, ∼5 US$). All participants had normal or corrected-to-normal vision, and were right-handed. All Participants reported that they had no neurological or psychiatric disorders history. The study was approved by the Ethics Committee of Zhejiang Normal University.

2.2. Procedures

After completing consent forms, participants were randomly assigned to one of two groups: a guilty group or an innocent group. Participants from the guilty group were instructed to enact a lab crime: to enter a lab office and steal an object hidden in a drawer. This specific object (a ring) was never mentioned to participants in the instruction. Thus, participants acquired the crime-relevant knowledge solely during the lab crime (Hu et al., 2013). Innocent participants, on the other hand, were simply asked to walk around the same room without committing any crime. In the room, a wallet was put on the desk so that both guilty and innocent participants would notice it. This "wallet" item was used as a target for all participants in the subsequent ERP-based feedback-concealed information test (fCIT). To ensure that all participants saw the wallet, we asked participants about whether they saw a wallet on the desk while they were in the room. One participant who reported not seeing the wallet was sent them back to the room to locate the wallet.

The fCIT contained six different stimuli: wallet, ring, watch, necklace, bracelet and earring. All participants were exposed to wallet; only guilty participants were exposed to ring; all participants were not exposed to the remaining four irrelevant stimuli until they saw them during the CIT. The "wallet" was therefore designated as a *target,* which all participants were instructed to respond to by pressing an "F" key with their left index finger. Participants were then instructed to press a "J" key to remaining stimuli (non-target) with their right index finger. Participants were further instructed that the "F" key would indicate "Yes, I recognize this item" whereas the "J" key would indicate "No, I do not recognize this item" (for similar instructions, see Rosenfeld, Hu, & Pederson, 2012; Verschuere, Rosenfeld, Winograd, Labkovsky, & Wiersema, 2009). Response buttons were counterbalanced across participants in both guilty and innocent groups. Because only guilty participants stole the ring, they would be lying when they pressed the "J" key to the word "ring" as they denied their knowledge of the crime.

Following this response assignment instruction, participants from both groups were instructed that the subsequent brainwave test aimed to detect whether they were lying or not based on their brainwaves during the test. Moreover, participants were told that the brainwave test would provide them with test outcomes following each single trial. Specifically, there were two possible feedbacks: a feedback "+4" would indicate that the participants are telling a truth and thus pass the trial, whereas a feedback "−2" would indicate that the participants are lying and thus fail the trial. Unbeknownst to participants, the feedbacks were bogus and were given randomly. This fCIT allows us to focus on two ERP epochs: (1) one epoch that followed either probe or target or irrelevant (i.e., recognition process), and (2) one epoch that followed the feedback stimulus (i.e., outcome evaluation process).

During the fCIT, participants were seated about 1m in front of a computer monitor. Each stimulus was presented in black font on a white background. Each trial began with a 500-ms long fixation point. One of the six stimuli (one probe, one target, and four irrelevants) was then randomly presented on the center of the monitor for 300 ms. Participants were instructed to press one of two response buttons (either "J" or "F" keys on a standard keyboard) to this stimulus as quickly and accurately as possible. Following a 1000 ms blank screen, a word "detecting" was presented for 500 ± 100 ms, suggesting that the test was analyzing participants' brainwaves. Following another 500 ms blank, a feedback ("+4" or "−2") was presented for 1000 ms. This experimental design resulted in four conditions in the feedback stage: probe-success (probe followed by " +4"); probe-failure (probe followed by "−2"); irrelevant-success (irrelevant followed by " +4"); irrelevant-failure (irrelevant followed by "−2"). Each stimulus (probe or irrelevant) were repeated 60 times, with half of them followed by a success feedback " +4", and the remaining half of the stimuli followed by a failure feedback "−2". The target was also repeated 60 times, but feedbacks following the target were based on participants' actual performance (i.e., " +4" for correct responses while "−2" for incorrect responses). In total, there were $6 \times 60 = 360$ trials. Participants were allowed to take a break every 60 trials. The fCIT lasted about 30 min.

2.3. EEG acquisition and ERP quantification

Continuous EEGs were recorded from 32 sites using Ag/AgCl electrodes embedded in an elastic cap (Neuroscan Inc., USA) according to the International 10–20 system. On-line recordings were referenced to the left mastoid and then re-referenced offline to the average of the left and right mastoids. Electrode impedances were kept below 5 kΩ. The vertical electro-oculograms (EOGs) were recorded above and below the right eye; the horizontal EOGs were recorded from electrodes placed at the outer canthus of the left eye and right eye. The EEG and EOG were bandpassed from 0.1 to 70 Hz, with an on-line sampling rate of 1000 Hz.

For off-line analyses, continuous EEGs were first filtered with a 30 Hz low-pass filter (24 dB/ct) to minimize the influences of high-frequency noises (bandpass was thus from 0.1 to 30 Hz). Continuous EEGs were then segmented into 1200 ms epochs, and were time-locked to either the probe/irrelevant stimuli or the positive/negative feedback stimuli. This 1200 ms epoch contained a 200 ms pre-stimulus baseline and a 1000 ms time window after stimulus onset. The irrelevant ERPs were averaged across all four irrelevant items. Trials were baseline corrected, and those with signals exceeding ± 100 μV were defined as artifact trials and were excluded from averaging. Number of averaged ERPs trials for both probe/irrelevant stimuli and feedback stimuli are presented in Table 1.

Table 1
The number of averaged trials in each condition.

Guilty	Number of trials (S.D.)	Innocent	Number of trials (S.D.)
Probe	53(4)	Probe	52(4)
Irrelevants	208(20)	Irrelevants	208(17)
Probe-success	28(2)	Probe-success	28(3)
Probe-failure	28(3)	Probe-failure	26(3)
Irrelevants-success	105(15)	Irrelevants-success	106(10)
Irrelevants-failure	105(10)	Irrelevants-failure	105(10)

For ERPs time-locked to CIT stimuli (i.e., probe vs. irrelevant), we focused on the baseline–peak (b–p) measured P300 amplitude at Pz, the amplitude of which was calculated as the mean amplitude between 400–600 ms after stimulus onset. Moreover, because previous P300-based CIT studies also measured the late posterior negativity that followed the P300 (see Hu, Bergström, Bodenhausen, & Rosenfeld, 2015b; Soskins, Rosenfeld, & Niendam, 2001), we additionally measured the amplitude of this late negativity at Pz and combined it with the b–p P300 as a peak–peak measured P300. Specifically, an algorithm searched from 300 to 800 ms for the maximal positive 100-ms segment average. The midpoint of the segment was defined as the P300 latency; the algorithm continued to search from this P300 latency to 1000 ms for the maximal negative 100-ms segment average. The difference between the maximal positive segment and the maximal negative segment was defined as the P300 p–p amplitude. Both b–p and p–p P300 measures were employed for receiver operating characteristic (ROC) analyses to examine individual classification efficiency (i.e., how well do ERPs distinguish between guilty and innocent participants at an individual level), but only b–p P300 was used in group-level analyses.

For feedback-locked ERPs, we employed a temporal principal component analysis (PCA) to separate the FRN and the feedback-P300. Given their different scalp distributions (e.g., Yeung & Sanfey, 2004), this temporal PCA was conducted at Fz to extract the FRN and at Pz to extract the feedback-P300. This temporal PCA was conducted using the ERP PCA Toolbox (Version 2.4.6, Dien, 2010). Since there may exist large ERP latency variations between guilty and innocent group that would confound temporal PCA (Holroyd, Pakzad-Vaezi, & Krigolson, 2008), we conducted PCA in guilty and innocent group separately. The temporal PCA used all time points from each participant's averaged ERP as variables (1200 time points), with participants and conditions as observations. Promax rotation was then applied to extract virtual temporal components (e.g. Dien, Khoe, & Mangun, 2007). Based on scree plots, 18 factors were extracted in the guilty group, and 16 factors were extracted in the innocent group. Of these factors, 12 factors in the guilty group and 14 factors in the innocent group accounted for at least 1% of the total variance in the data. Based on visual inspection of the waveforms associated with these factors, we selected two factors that most readily corresponded to the earlier FRN and the later feedback-P300: In guilty group, a positivity peaking at 308 ms for FRN and a positivity peaking at 383 ms for feedback-P300. In innocent group, a positivity peaking at 281 ms for FRN and a positivity peaking at 374 ms for feedback-P300 (see Table 2). The waveforms for each factor were reconstructed (i.e., converted to microvolts) by multiplying the factor pattern matrix with the standard deviations. These factors were scored using the peak values of each virtual com-

Table 2
PCA factors selected for statistical analysis.

Corresponding ERP components	Temporal factors	Temporal loading peaks (ms)	Variances explained (%)
Guilty			
FRN	TF7	308	2.36
P300	TF4	383	10.18
Innocent			
FRN	TF11	281	1.26
P300	TF2	374	28.38

ponent (see also Foti et al., 2011). These peak values were used in all subsequent analyses, including analyses of variance (ANOVA), correlational analyses, as well as ROC analyses. For ANOVA, we applied Greenhouse-Geisser correction when assumption of sphericity was violated. Post-hoc comparisons were computed with Fisher's Protected Least Significant Difference.

3. Results

3.1. Behavioral results

A 2 by 2 mixed ANOVA was conducted on RTs to probe and irrelevant stimuli, with the stimulus type (probe vs. irrelevant) as a within-subject variable and group (guilty vs. innocent) as a between-subject variable. Results revealed a marginally significant main effect of stimulus type: $F(1, 30) = 3.80$, $p = .06$, $\eta_p{}^2 = .11$: participants took longer to respond to the probe than to the irrelevants (Mean ± S.E.M.: 548.68 ± 4.60 ms vs. 531.31 ± 4.60 ms). There was also a marginally significant interaction between stimulus type and group, $F(1, 30) = 3.07$, $p = .09$, $\eta_p{}^2 = .09$. Follow-up tests found that there was a significant RT difference between the probe and the irrelevants in the guilty group (560.94 ± 8.42 vs. 527.96 ± 8.42 ms, $F(1, 30) = 6.85$, $p = .014$, $\eta_p{}^2 = .19$); while no RT difference was found between the probe and the irrelevants in the innocent group (536.42 ± 2.91 vs. 534.66 ± 2.91 ms, $F(1, 30) = 0.02$, $p = 89$).

3.2. ERP results

3.2.1. Recognition-P300s to probes and irrelevants

A 2 (within-subject: probe vs. irrelevant) by 2 (between-subject: guilty vs. innocent) mixed ANOVA revealed a significant main effect of stimulus type, $F(1, 30) = 6.24$, $p = .018$, $\eta_p{}^2 = .17$, with larger P300 elicited by the probe than P300 elicited by the irrelevants (5.63 ± 0.21 μV vs. 4.68 ± 0.21 μV). This main effect was qualified by a significant stimulus type by group interaction $F(1, 30) = 6.53$, $p = .016$, $\eta_p{}^2 = 0.18$. Follow-up tests showed that among guilty participants, probe elicited significantly larger P300 than irrelevants (6.71 ± 0.30 vs. 4.78 ± 0.30 μV, $F(1, 30) = 12.77$, $p < .001$, $\eta_p{}^2 = .30$). In contrast, no P300 amplitude differences were found between probe and irrelevants for innocent participants (4.56 ± 0.24 vs. 4.58 ± 0.24 μV, $F(1, 30) < 0.001$, $p = .969$). For grand averaged ERPs and their scalp distributions that are locked to CIT stimuli, see Figs. 1 and 2.

FRN[1] *at Fz (a positivity peaking at 308 ms in the guilty group and a positivity peaking at 281 ms in the innocent group corresponded to FRN)* (Fig. 3)

A mixed 2 (stimulus type: probe vs. irrelevant) by 2 (feedback type: success "+4" vs. failure "−2") by 2 (group: guilty vs. innocent) ANOVA was performed on the FRN amplitude with the first two variables as within-subject variables and the third variable as a between-subject variable. Results revealed a significant stimulus type effect, $F(1, 30) = 48.4$, $p < .001$, $\eta_p{}^2 = .62$: feedback following the probe elicited a more positive FRN than feedback following the irrelevant (2.95 ± 0.19 vs. 1.21 ± 0.19 μV). A significant main effect of feedback was also found, $F(1, 30) = 9.72$, $p = .004$, $\eta_p{}^2 = .25$, with a more positive FRN following success than following failure feedbacks (2.55 ± 0.15 vs. 1.61 ± 0.15 μV). A significant main effect of group was also revealed: FRN in the guilty group was more positive than that in the innocent group: $F(1, 30) = 8.68$, $p = .006$, $\eta_p{}^2 = .22$, 2.87 ± 0.38 vs. 1.29 ± 0.38 μV. Regarding interactions, there was a significant two-way interaction between stimulus and group, $F(1, 30) = 43.03$, $p < .001$, $\eta_p{}^2 = .59$. Follow-up tests found that among guilty participants, the probe elicited a more positive FRN than irrelevant $F(1, 30) = 91.35$, $p < .001$, $\eta_p{}^2 = .75$, 4.56 ± 0.24 vs. 1.18 ± 0.24 μV. This FRN difference was absent in the innocent group $F(1, 30) = 0.08$, $p = .78$, 1.34 ± 0.59 μV vs. 1.24 ± 0.59 μV. No other interactions were significant (stimulus × feedback, $F(1, 30) = 3.001$, $p = .094$; feedback × group $F(1, 30) = 0.924$, $p = .344$; stimulus × feedback × group $F(1, 30) = 0.279$, $p = .601$).

Feedback-P300 (a positivity peaking at 383 ms in the guilty group and a positivity peaking at 374 ms in the innocent group corresponded to feedback-locked P300 at Pz, See Fig. 3)

The same mixed 2 by 2 by 2 ANOVA was performed on the feedback-P300 amplitude, with stimulus type (probe vs. irrelevant) and feedback (success vs. failure) as two within-subject variables, and group (guilty vs. innocent) as one between-subject variable. Results revealed a significant main effect of stimulus type, $F(1, 30) = 48.78$, $p < .001$, $\eta_p{}^2 = .62$, with a more positive feedback-P300 following feedbacks regarding the probe than irrelevant (10.65 ± 0.41 vs. 6.77 ± 0.41 μV). There was a significant main effect of group, $F(1, 30) = 5.06$, $p = .032$, $\eta_p{}^2 = .14$, with a more positive feedback-P300 in the guilty group than in the innocent group (10.36 ± 1.04 vs. 7.09 ± 1.04 μV). No effect of feedback was found, $F(1, 30) = 1.08$, $p = .307$. Importantly, these two main effects were qualified by a significant stimulus by group interaction $F(1, 30) = 38.13$, $p < .001$, $\eta_p{}^2 = .56$. Follow-up tests indicated that the probe elicited a more positive feedback-P300 than did the irrelevant in the guilty group: 14.02 ± 0.50 vs. 6.71 ± 0.50 μV, $F(1, 30) = 86.58$, $p < .001$, $\eta_p{}^2 = .74$; however, among innocent participants, there was no feedback-P300 difference between the probe and the irrelevant (7.29 ± 0.24 vs. 6.84 ±0.24 μV), $F(1, 30) = 0.33$, $p = .571$. There was no other significant two-way interaction between the stimulus and feedback ($F(1, 30) = 1.581$, $p = .218$); or feedback and group ($F(1, 30) = 1.136$, $p = .295$) or three-way interaction ($F(1, 30) = 2.654$, $p = .114$). (For grand averaged ERPs and their scalp distributions elicited by feedback stimuli, see Fig. 2 and Fig. 4)

3.3. Correlations between recognition-P300 and feedback-ERPs

We explored whether neural activity locked to recognition phase (probe vs. irrelevant) and feedback phase (win vs. loss following probe and irrelevant) correlated with each other in guilty and innocent participants. Specifically, we computed difference scores (probe-minus-irrelevant) for recognition-P300, FRN and feedback-P300. Note that the FRN and feedback-P300 were collapsed across loss and wins as the primary interest of the study is to discriminate between crime-relevant and crime-irrelevant memories. Results showed that among guilty participants, there were no correlations between recognition-P300 and FRN: $r(16) = .12$, $p = .653$. There was a moderate correlation between recognition-P300 and feedback-P300, but this result is not quite significant at the .05 level, $r(16) = .45$, $p = .079$. Moreover, there was a significant correlation between FRN and the feedback-P300, $r(16) = .70$, $p = .002$. For the innocent group, there were no significant correlations between

[1] Recent studies suggest that FRN is driven by a positive ERP elicited by positive feedback (Becker et al., 2014; Carlson et al., 2011; Proudfit, 2015), and our temporal PCA analysis is consistent with this finding. Here, we chose to use the FRN term to be consistent with most previous ERP research focusing on feedback processing.

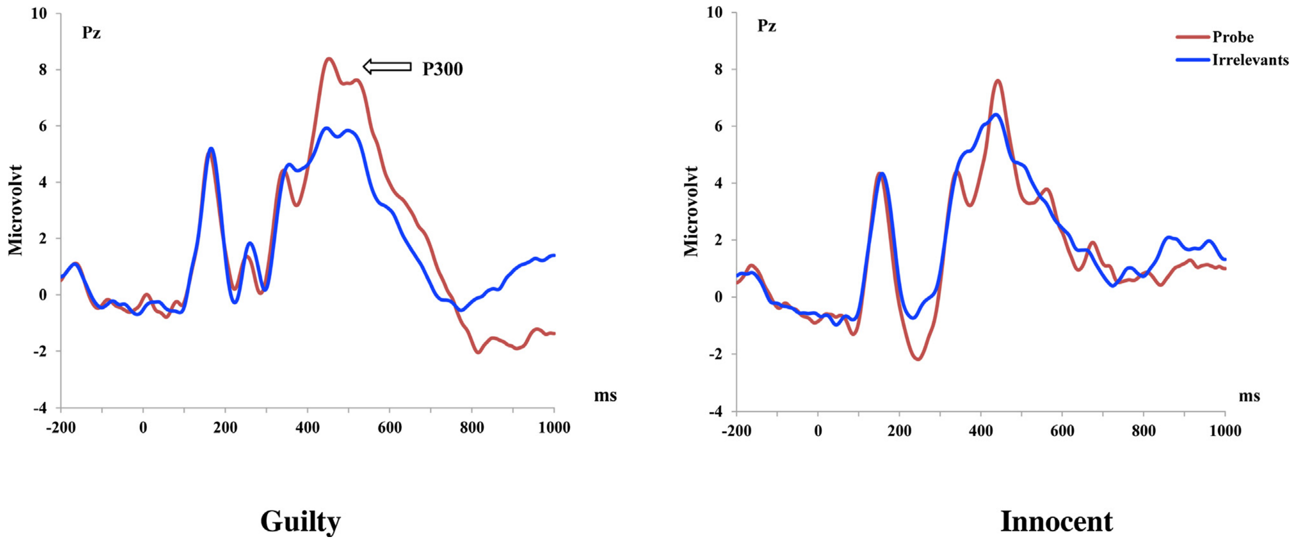

Fig. 1. Grand-average ERP waveforms from Pz during CIT stimuli.

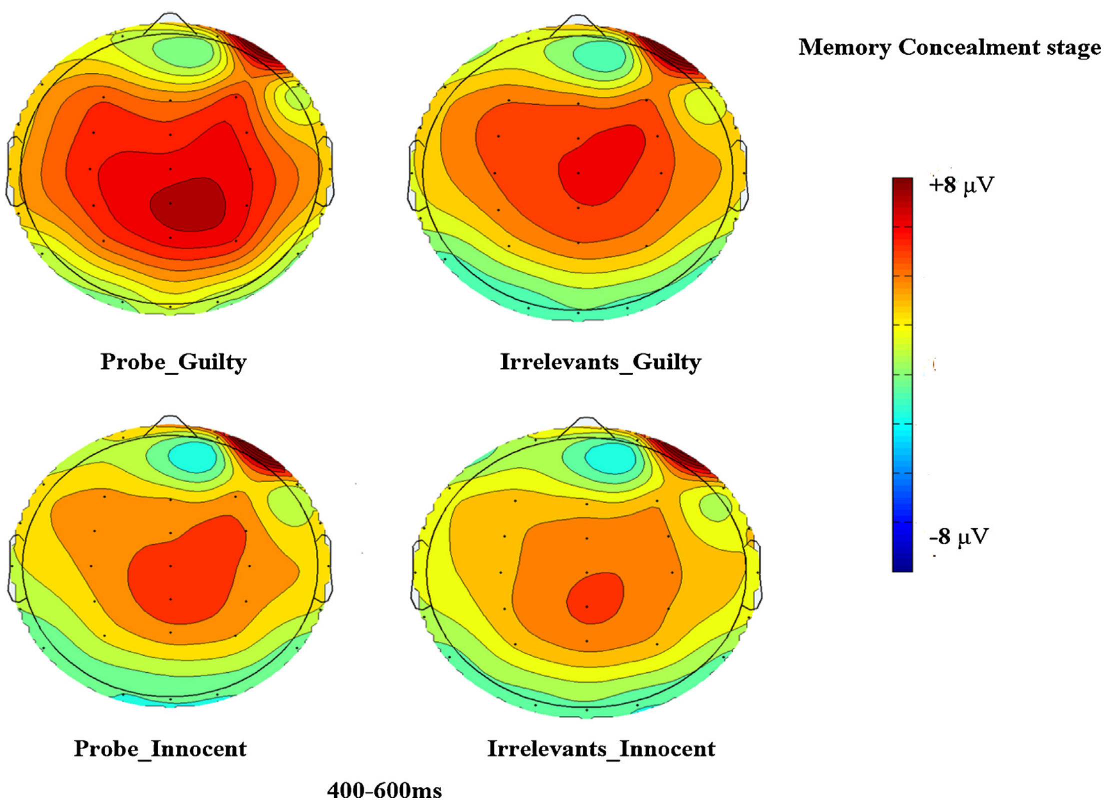

Fig. 2. Scalp distribution of the P300 during CIT stimuli.

ERPs during feedback epoch.

ERP activities (r s < .43, p s > .09, for detailed statistical results, see Table 3).

3.4. Individual classification efficiency

To determine individual classification efficiency of each ERP component, we conducted a receiver operating characteristic (ROC) analysis. This ROC analysis was based on the probe-minus-irrelevant amplitude for each ERP component (recognition-P300 at Pz, FRN at Fz, and feedback-P300 at Pz). Guilty or innocent group status was used as the dependent variable. Results (see Table 4) showed that all three ERP components can effectively

Table 3
Correlations between P300 during CIT, FRN and P300 during feedback stage.

	FRN	Feedback-P300
Guilty group		
Recognition-P300	.12	.45
FRN		.70**
Innocent group		
Recognition-P300	−.03	−.43
FRN		−.34

** $p < .01$, $n = 16$ in each group.

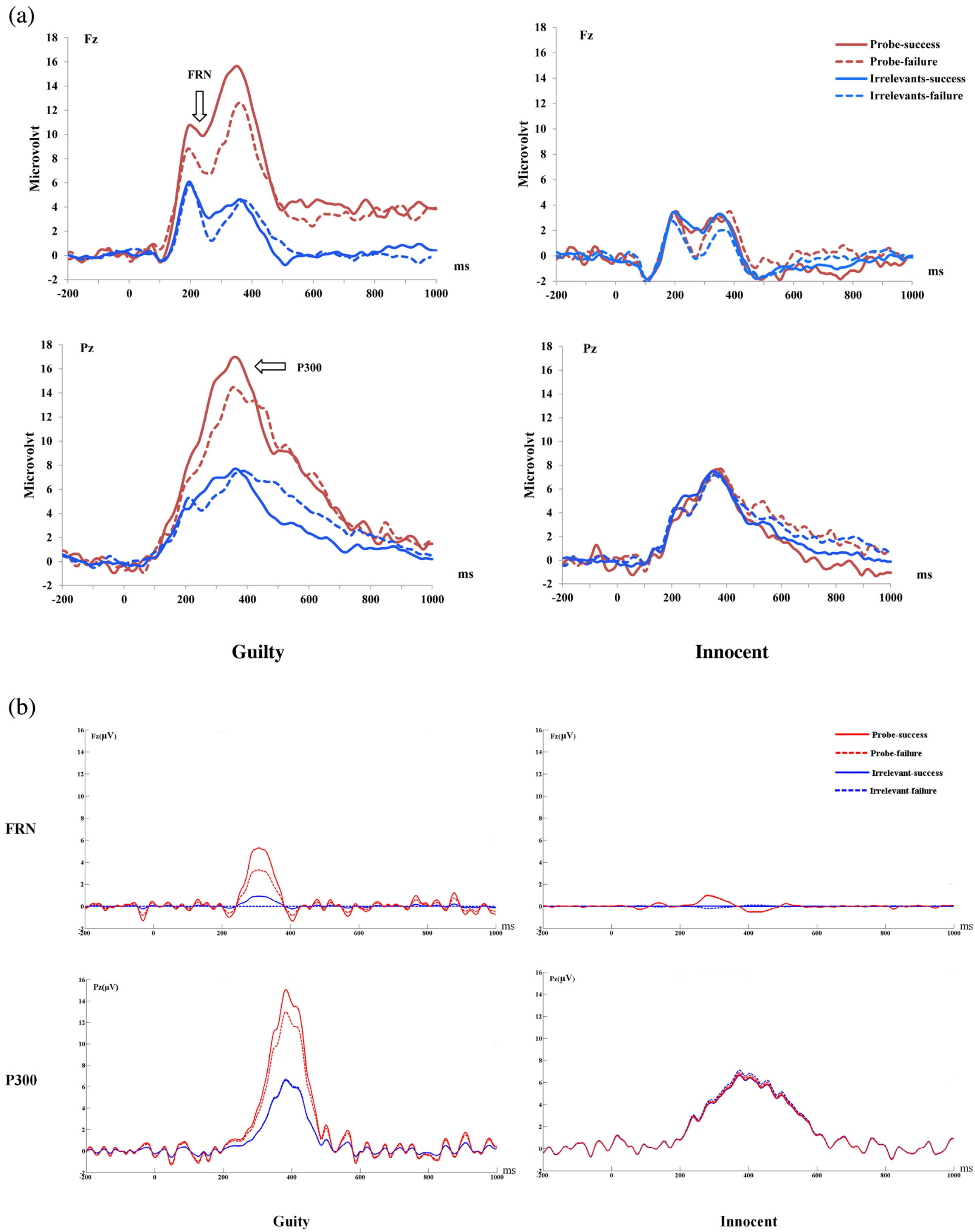

Fig. 3. (a) Grand-average ERP waveforms (before PCA transformation) during feedback stage. (b) PCA-extracted ERP waveforms during feedback stage.

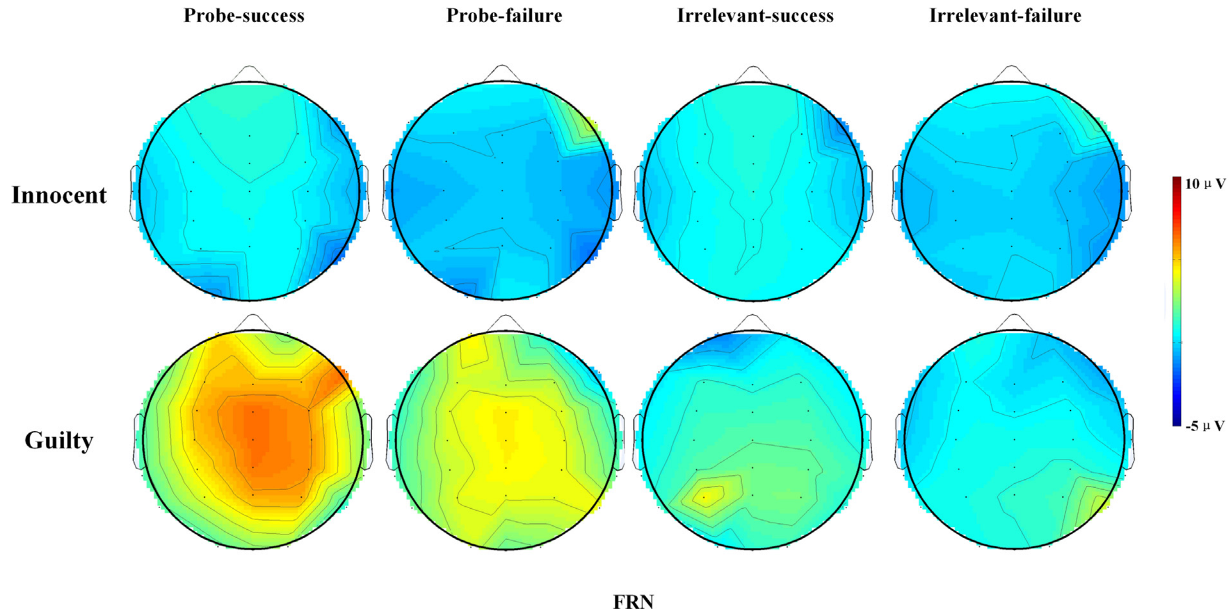

Fig. 4. PCA-based Scalp distribution of the FRN during feedback stage.

Table 4
Receiver operation characteristic (ROC) analyses.

	AUC	95% confidence intervals
(b-p) Recognition-P300	.73*	.53–.89
(p–p) Recognition-P300	.68[1]	.49–.87
FRN	.95***	.86–1.00
Feedback-P300	.97***	.92–1.00
(b–p) Recognition-P300 and FRN	.92***	.83–1.00
(b–p) Recognition-P300 andFeedback-P300	.89***	.77–1.00
FRN and Feedback-P300	.99***	.98–1.00
All three indices	.98***	.95–1.00

1, $p = .09$, *$p < .05$, ***$p < .001$.

distinguish guilty from innocent participants above chance (.5) level, AUCs = .73–.97, p s < .05, except for the p–p recognition-P300, AUC = .68, $p = .09$. Lastly, we examined whether combining all three ERP components would further improve individual classification efficiency.[2] Specifically, probe-irrelevant differences for each ERP activity were transformed into standardized z-scores across all guilty and innocent participants. Then we averaged these three z-scores into a single measure for each participant (see Hu et al., 2013). This ROC analysis yielded an AUC = .98. And the highest AUC (.99) was obtained when combining FRN and feedback-P300.

4. Discussion

The present study investigated whether feedback-related ERPs can be used to detect concealed information in a novel feedback-based concealed information test (fCIT, modified from Hu et al., 2013; Rosenfeld, 1992; Rosenfeld et al., 2012). This fCIT allows researchers to utilize (1) recognition-P300 as a neural marker of concealed memories and (2) feedback-related FRN/P300 that underlie the evaluation of memory concealment during the CIT, i.e., whether concealment was successful or not. We replicated the classic ERP-CIT findings that the P300 was responsive to concealed crime-relevant memories despite participants' overt denial of recognition. This recognition-P300 could accurately distinguish between guilty and innocent individuals (but not as accurate as usual; see Rosenfeld, 2011 as explained below). Most importantly, we reported a novel finding that feedback-related ERPs such as FRN/P300 varied as a function of the stimulus type (probe vs. irrelevant) that the feedback was about. Together with Sai et al. (2014a), these findings provide strong evidence supporting the use of feedback-related brain activities in memory detection.

4.1. P300s in memory concealment

Consistent with previous P300-based CIT studies, we found that the probe elicited a larger P300 than the irrelevants in guilty, but not innocent group. The P300 is typically regarded as a neural signal with multiple neural generators (Johnson, 1993), and may reflect more than one process such as enhanced attention, working memory updating and explicit recollection (for a review, see Polich, 2007). In this memory detection context, this enlarged P300 to the probe could be unambiguously due to meaningfulness and/or automatic recognition of the probe because of guilty participants' prior crime experience (Allen et al., 1992; Farwell & Donchin, 1991; Rosenfeld et al., 1991, 1998; for a recent review, see Rosenfeld et al., 2013).

4.2. Feedback-related ERPs during feedback processing

One of the challenges in quantifying FRN is that its amplitude can be influenced by later, larger, positive-going P300s. A temporal PCA is thus employed to separate the FRN and the feedback-P300. Consistent with previous studies that used temporal PCAs to extract FRNs, we found that the positive feedback elicited a more positive potential than the negative feedback (Foti et al., 2011; Proudfit, 2015). Holroyd et al. (2008) similarly found that positive feedback elicited a positive-going ERP in the same time window of FRN; whereas the FRN was similar to oddball-related N200. Our temporal PCA results further corroborate the argument that the FRN is driven by a positive ERP deflection elicited by positive feedback.

Most relevant for the current study, we found that memory status such as recognizable or not (probe and irrelevant) modulated subsequent FRNs. Specifically, feedback stimuli following the probe elicited a more positive FRN than feedback following the

[2] When combining these indices, we did not include (p–p) recognition-P300, because (p–p) recognition-P300 cannot discriminate guilty from innocent participants at .05 significance level ($p = .09$).

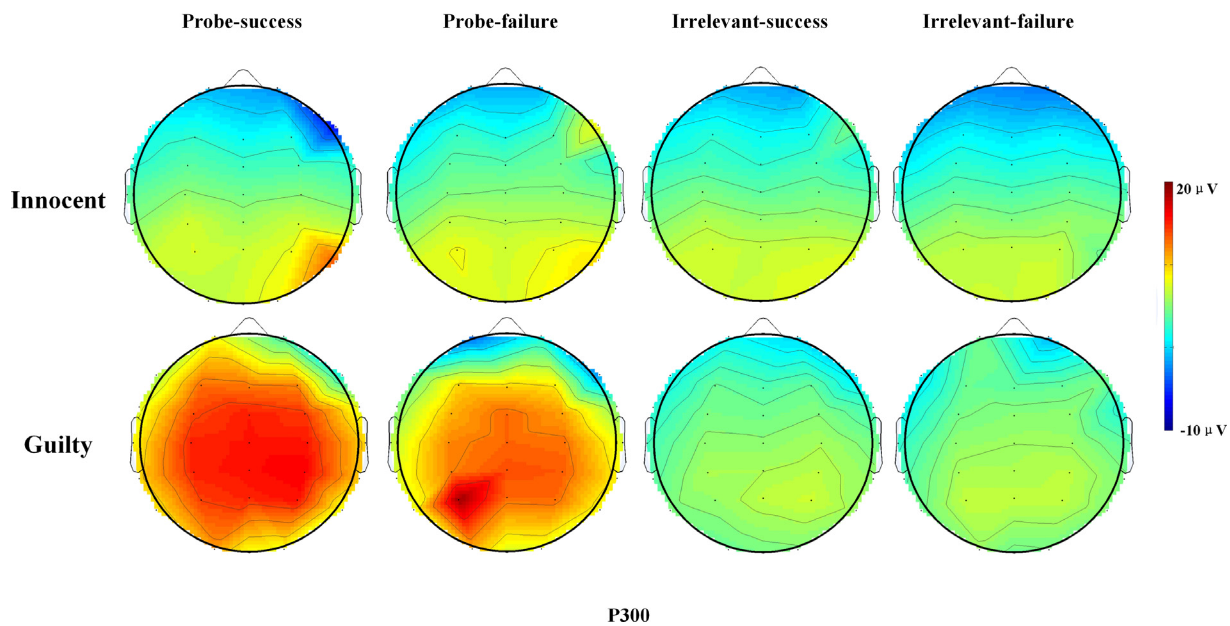

Fig. 5. Scalp distribution of the feedback-P300 during feedback stage.

irrelevants, a pattern emerged only in the guilty group but not in the innocent group. Recent empirical findings have linked reward positivity (i.e., the positive potential elicited by positive feedback) with hemodynamic responses in reward-related neural network including ventral striatum and mid-frontal cortex (Becker et al., 2014; Carlson et al., 2011). In the present study, guilty participants were motivated to conceal their criminal memories by denying their knowledge of the crime-relevant probes. Thus, it is reasonable to infer that guilty participants would find positive feedback (i.e., they lied successfully) more rewarding and satisfying than negative feedback, and this difference should be most significant following probes but not irrelevants. Finally, the indistinguishable feedback-ERPs for innocent participants were expected, because the probe was simply one of the meaningless irrelevant stimuli for them.

Unlike FRN, feedback-P300 did not respond to feedback valence. However, we found that this feedback-P300 were larger following probes than irrelevants among guilty participants, which also replicated Sai et al. (2014a). To date, researchers still lack a clear understanding of the psychological processes implicated by feedback-P300. Previous studies have yielded inconsistent results. For example, some previous studies found a valence effect on feedback-P300, with positive feedback elicited larger feedback-P300s than negative feedback (e.g., Bismark, Hajcak, Whitworth, & Allen, 2005). In contrast, some other studies found that the feedback-P300 seems to encode the magnitude of reward instead of reward valence, such that winning $10 elicited larger feedback-P300 than winning $1 (e.g., Sato et al., 2005; Yeung & Sanfey, 2004). The present study is in line with this latter literature, suggesting that this feedback-P300 may not be sensitive to feedback valence. Moreover, our finding suggests that the feedback-P300 amplitude could be modulated by participants' motivation and task engagement, as the feedback-P300 was significantly larger following the recognizable probe than meaningless irrelevant. Similarly, Luo, Sun, Mai, Gu, and Zhang (2011) found that feedback stimulus elicit a more positive feedback-P300 when participants received feedback about their deceptive responses. Yeung, Holroyd, and Cohen (2005) showed that the feedback-P300 was larger in a monetary gambling task when participants had to make active choices compared to when participants passively viewed gambling outcomes. These findings consistently suggested that the feedback-P300 could be modulated by stimulus or response type (probe vs. irrelevant, deception vs. truth telling), which may be mediated by enhanced attention engagement to feedback following deception.

It should be noted that an alternative PCA strategy is to conduct the analysis across both guilty and innocent participants. However, in the present dataset, this PCA strategy did not yield meaningful ERP components in the FRN time window. One possible explanation is that PCA results can be influenced by latency variances across different groups (Donchin & Heffley, 1978; Holroyd et al., 2008). Therefore combining both guilty and innocent participants' ERP data may introduce large variances to the temporal PCA analyses. Although applying PCAs to separate groups is not uncommon (e.g., Liao et al., 2010; Sorg et al., 2007), future replications are warranted to further confirm our findings. Despite this potential limitation in PCA, our central aim is to test whether feedback-related neural activity can effectively discriminate innocent from guilty participants, and our data showed promising results even before PCA transformation (see Fig. 3a).

For neuroscience-based forensic assessment, it is important to document how well these feedback-related ERPs can discriminate guilty from innocent participants at an individual level. Our ROC analyses showed that the FRN and the feedback-P300 could discriminate guilty from innocent participants highly effectively (AUCs: .95–.97). These findings, together with Sai et al. (2014a), are promising because we documented a novel neural signal in detecting concealed memory and deception. However, it should be noted that providing random or bogus feedback to examinees in the field may raise ethical concerns. Future research should tackle the question that how to utilize feedback and thus elicit feedback-related ERPs in field testing (Fig. 5).

It also should be noted that the detection rate of the peak–peak P300 during the first, memory concealment epoch was lower than that in most previous P300-based CIT studies (Rosenfeld, 2011; Rosenfeld et al., 2013). One likely explanation is that the trial-by-trial feedback could distract participants from processing the probe/irrelevants during the memory concealment epoch, which reduced the P300 difference between probe and irrelevants in the guilty group. In contrast, Rosenfeld et al. (2012) gave feedback after every 10–15 trials and achieved 100% accuracy. Further studies could directly compare CITs with and without feedback to examine this hypothesis. Another possible explanation for the low detection rate is that we set the high pass filter at 0.1 Hz, while previous studies usually used 0.3 Hz (e.g., Rosenfeld et al., 2012). Whether or not continuous feedback may impact recognition-P300, the current study provided evidence that feedback-related ERPs can effectively discriminate guilty from innocent participants.

Acknowledgment

The present study was supported by the National Science Foundation of China to G. Fu (31070894 and 31371041)

References

Allen, J. J. (2011). Clinical applications of the concealed information test. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: theory and application of the concealed information test* (pp. 231–252). New York, NY: Cambridge University Press.

Allen, J. J., Iacono, W. G., & Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and omplicit behavioral measures: a methodology for prediction in the face of individual differences. *Psychophysiology*, *29*(5), 504–522.

Becker, M. P., Nitsch, A. M., Miltner, W. H., & Straube, T. (2014). A single-trial estimation of the feedback-related negativity and its relation to BOLD responses in a time-estimation task. *The Journal of Neuroscience*, *34*(8), 3005–3012.

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: a meta-analytic review. *Journal of Applied Psychology*, *88*(1), 131.

Bismark, A. W., Hajcak, G., Whitworth, N. M., & Allen, J. J. (2013). The role of outcome expectations in the generation of the feedback-related negativity. *Psychophysiology*, *50*(2), 125–133.

Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: a combined ERP and fMRI study. *Neuroimage*, *57*(4), 1608–1616.

Dien, J. (2010). The ERP PCA toolkit: an open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, *187*(1), 138–145.

Dien, J., Khoe, W., & Mangun, G. R. (2007). Evaluation of PCA and ICA of simulated ERPs: promax vs. infomax rotations. *Human Brain Mapping*, *28*(8), 742–763.

Ding, X. P., Sai, L., Fu, G., Liu, J., & Lee, K. (2014). Neural correlates of second-order verbal deception: a functional near-infrared spectroscopy (fNIRS) study. *Neuroimage*, *87*, 505–514.

Donchin, E., & Heffley, E. F., III. (1978). Multivariate analysis of event-related potential data: a tutorial review. In D. Otto (Ed.), *Multidisciplinary perspectives in event-related potential research* (pp. 555–572). Washington, DC: U.S. Government Printing Office.

Farah, M. J., Hutchinson, J. B., Phelps, E. A., & Wagner, A. D. (2014). Functional MRI-based lie detection: scientific and societal challenges. *Nature Reviews Neuroscience*, *15*(2), 123–131.

Farwell, L. A., & Donchin, E. (1991). The truth will out: interrogative polygraphy (lie detection) with event-related brain potentials. *Psychophysiology*, *28*(5), 531–547.

Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: temporospatial principal components analysis and source localization of the feedback negativity. *Human Brain Mapping*, *32*, 2207–2216.

Gamer, M., Klimecki, O., Bauermann, T., Stoeter, P., & Vossel, G. (2012). fMRI-activation patterns in the detection of concealed information rely on memory-related effects. *Social Cognitive and Affective Neuroscience*, *7*(5), 506–515.

Gamer, M. (2011). Detecting of deception and concealed information using neuroimaging techniques. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: theory and application of the concealed information test* (pp. 90–113). New York, NY: Cambridge University Press.

Gamer, M., Verschuere, B., Crombez, G., & Vossel, G. (2008). Combining physiological measures in the detection of concealed information. *Physiology & Behavior*, *95*(3), 333–340.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, *295*(5563), 2279–2282.

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology*, *42*(2), 161–170.

Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology*, *45*(5), 688–697.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709.

Hu, X., Pornpattananangkul, N., & Nusslock, R. (2015). Executive control-and reward-related neural processes associated with the opportunity to engage in voluntary dishonest moral decision making. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 475–491.

Hu, X., Bergström, Z. M., Bodenhausen, G. V., & Rosenfeld, J. P. (2015). Suppressing unwanted autobiographical memories reduces their automatic influences: evidence from electrophysiology and an implicit autobiographical memory test. *Psychological Science*, *26*, 1098–1106.

Hu, X., Pornpattananangkul, N., & Rosenfeld, J. P. (2013). N200 and P300 as orthogonal and integrable indicators of distinct awareness and recognition processes in memory detection. *Psychophysiology*, *50*(5), 454–464.

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

Johnson, R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology*, *30*(1), 90–97.

Langleben, D., Schroeder, L., Maldjian, J., Gur, R., McDonald, S., Ragland, J., & Childress, A. (2002). Brain activity during simulated deception: an event-related functional magnetic resonance study. *Neuroimage*, *15*(3), 727–732.

Liao, W., Chen, H., Feng, Y., Mantini, D., Gentili, C., Pan, Z., & Lui, S. (2010). Selective aberrant functional connectivity of resting state networks in social anxiety disorder. *Neuroimage*, *52*(4), 1549–1558.

Luo, Y.-J., Sun, S.-Y., Mai, X.-Q., Gu, R.-L., & Zhang, H.-J. (2011). Outcome evaluation in decision making: ERP studies. In Han, Shihui, Pöppel, & Ernst (Eds.), *Culture and neural frames of cognition and communication* (pp. 249–285). Springer.

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*(6), 385–388.

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a generic neural system for error detection. *Journal of Cognitive Neuroscience*, *9*(6), 788–798.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148.

Proudfit, G. H. (2015). The reward positivity: from basic research on reward to a biomarker for depression. *Psychophysiology*, *52*(4), 449–459.

Rosenfeld, J. P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: theory and application of the concealed information test* (pp. 63–89). New York, NY: Cambridge University Press.

Rosenfeld, J. P., Angell, A., Johnson, M., & Qian, J. H. (1991). An ERP-based, control-question lie detector analog: algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology*, *28*(3), 319–335.

Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience*, *42*(1–2), 157–161.

Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J., & Winograd, M. R. (2013). Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology*, *90*(2), 118–134.

Rosenfeld, J. P., Hu, X., & Pederson, K. (2012). Deception awareness improves P300-based deception detection in concealed information tests. *International Journal of Psychophysiology*, *86*, 114–121, 2.

Rosenfeld, J.P., (1992). Method of credibility assessment based on feedback elicited P3 responses of brain. U.S. Patent #5,170,780.

Sai, L., Lin, X., Hu, X., & Fu, G. (2014). Detecting concealed information using feedback related event-related brain potentials. *Brain and Cognition*, *90*, 142–150.

Sai, L., Zhou, X., Ding, X. P., Fu, G., & Sang, B. (2014). Detecting concealed information using functional near-infrared spectroscopy. *Brain Topography*, *27*(5), 652–662.

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., & Kuboki, T. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport*, *16*(4), 407–411.

Siddle, D. A. (1991). Orienting, habituation, and resource allocation: an associative analysis. *Psychophysiology*, *28*(3), 245–259.

Sip, K. E., Skewes, J. C., Marchant, J. L., McGregor, W. B., Roepstorff, A., & Frith, C. D. (2012). What if I get busted? Deception, choice, and decision-making in social interaction. *Frontiers in Neuroscience*, *6*(58).

Sorg, C., Riedl, V., Mühlau, M., Calhoun, V. D., Eichele, T., Läer, L., & Zimmer, C. (2007). Selective changes of resting-state networks in individuals at risk for Alzheimer's disease. *Proceedings of the National Academy of Sciences*, *104*(47), 18760–18765.

Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). The case for peak-topeak measurement of P300 recorded at. 3 Hz high pass filter settings in detection of deception. *International Journal of Psychophysiology*, *40*, 173–180. http://dx.doi.org/10.1016/S0167-8760(00)00154-9

Tranel, D., & Damasio, A. R. (1985). Knowledge without awareness: an autonomic index of facial recognition by prosopagnosics. *Science*, *228*(4706), 1453–1454.

Verschuere, B., Ben-Shakhar, G., & Meijer, E. (2011). *Memory detection: theory and application of the concealed information test*. New York, NY: Cambridge University Press.

Verschuere, B., & Ben-Shakhar, G. (2011). Theory of the concealed information test. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: theory and application of the concealed information test* (pp. 128–150). New York, NY: Cambridge University Press.

Verschuere, B., Crombez, G., Koster, E. H. W., Van Bockstaele, B., & De Clercq, A. (2007). Startling secrets: startle eye blink modulation by concealed crime information. *Biological Psychology*, *76*, 52–60. http://dx.doi.org/10.1016/j.biopsycho.2007.06.001

Verschuere, B., Rosenfeld, J. P., Winograd, M. R., Labkovsky, E., & Wiersema, R. (2009). The role of deception in P300 memory detection. *Legal and Criminological Psychology*, *14*(2), 253–262.

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews*, *36*(8), 1870–1884.

Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex*, *15*(5), 535–544.

Yeung, N., & Sanfey, A. G. (2004). Independent coding of reward magnitude and valence in the human brain. *The Journal of Neuroscience*, *24*(28), 6258–6264.