available at www.sciencedirect.com
ScienceDirect
www.elsevier.com/locate/brainres
BRAIN RESEARCH

Research Report

Electrophysiological correlates of hypothesis evaluation: Revealed with a modified Wason's selection task

Xueli Cai, Fuhong Li, Yifeng Wang, Todd Jackson, Jie Chen, Lili Zhang, Hong Li*

School of Psychology, Southwest University, Chongqing, 400715, China
Key Laboratory of Cognition and Personality (SWU), Ministry of Education, China

ARTICLE INFO

Article history:
Accepted 23 June 2011
Available online 30 June 2011

Keywords:
Hypothesis evaluation
Wason's selection task
N2
P3
LPC

ABSTRACT

A modified Wason's selection task was used to explore brain correlates of hypothesis evaluation, a core process of hypothesis testing. Twenty-two undergraduate participants (11 males, 11 females) were provided with a proposition (hypothesis) and a card. They were asked to evaluate whether the card verified or falsified the given proposition while event related potentials (ERP) were measured. Behavioral results showed that participants required less time to make correct responses in verification conditions than in falsification conditions. The ERPs time-locked to the second side of each card showed that (1) smaller amplitudes of P2 were elicited in *backward falsification* than in *backward verification*, which reflected a lower intensity of perception; (2) a profound negative deflection was found in falsification conditions compared to verification conditions during the N2 time window, which implied the processing of conflicting information; (3) in comparison to verification conditions, falsification conditions evoked a decreased P3 component, which was linked to the process of hypothesis evaluation; and (4) a late positive component (400–600 ms) was only triggered in the forward falsification condition, reflecting the manipulation of cognitive context.

1. Introduction

Hypothesis testing (HT) plays an important role in high-level cognitive processes, such as concept formation, decision making and problem solving, during which people generate and select new rules or apply previously tested ones (Klayman and Ha, 1987; Wason, 1960, 1968). Recently, research has focused on elucidating the cognitive and neural underlying of hypothesis testing.

In an early study (Elliott and Dolan, 1998), participants were required to discover a rule governing which one from a choice of two black and white checkerboard stimuli was the correct response. Positron emission tomography (PET) results indicated that hypothesis testing activated the cerebellum, left anterior cingulate, and right precuneus. Similarly, Papo et al. (2003) presented participants with a hypothesis testing task and recorded scalp potentials. A hidden rule about numbers was presented and participants were required to judge whether a triplet of numbers was an instance of that rule or not, followed by feedback. Electrophysiological results revealed significant differences between responses to positive and to negative feedback. Papo et al. (2003) focused only on the process of feedback, without detailed discussion of hypothesis testing. Li et al. (2009a, 2009b, in press) used inductive generalization tasks to examine brain activity related to hypothesis testing, and found an increased late positive

* *Corresponding author at:* School of Psychology, Southwest University, Chongqing, 400715, China. Fax: +86 23 68367960.
E-mail addresses: fuhonglee@vip.qq.com (F. Li), lihong@swu.edu.cn (H. Li).

0006-8993/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2011.06.056

complex (LPC) when hypotheses were rejected, reflecting a process of hypothesis evaluation and updating of memory-context. However, these ERP studies did not dissociate the process of hypothesis evaluation from the process of working memory.

The purpose of the present study was to investigate the neural correlates of hypothesis evaluation by utilizing a modified Wason's selection task, a classic paradigm to explore hypothesis testing (Wason, 1960, 1968). In contrast to the classic Wason's selection task, test cards were presented individually in our modified task, with the sides of each card displayed sequentially. Participants were required to evaluate whether a card verified or falsified the preceding proposition. This modification allowed us to segment the cognitive processes of HT and explore the electrophysiological correlates of HT using ERP method with high-time resolution (Li et al., 2009a, 2009b; Papo et al., 2003). Specifically, the modified hypothesis-testing task was divided into three following stages. In the first stage, proposition apprehension, participants processed and comprehended the proposition. To confirm they had understood the proposition, they pressed any key to enter into the second stage, during which one side of a test card was processed within a set time limit. Finally, participants entered into the third stage, in which they were presented with the second side of the card and were required to determine whether the contents of the test card were congruent with the preceding proposition. The present study focused specifically on the third stage, corresponding to the time course of integrating the contents of the two sides of a test card and evaluating the preceding hypothesis. Our pilot experiment revealed that participants could complete the third stage within approximately 1000 ms from the onset of the second side of the test card, confirming that the design met the requirements of ERP method.

It has been suggested that the process of hypothesis testing correlates closely to the process of feedback (Papo et al., 2003). Positive feedback helps to keep the hypothesis active, whereas negative feedback leads to the rejection of an invalid hypothesis. In comparison to positive feedback, negative feedback produces a more negative deflection during the time window of 230–330 ms from the onset of feedback (Carter et al., 1998; Holroyd and Coles, 2002; Papo et al., 2003). Accordingly, it was possible that feedback-related negativity (FRN) would be detected in the present study, because the process of hypothesis evaluation also correlates with the process of feedback (Li et al., 2009b). However, several other researchers have proposed that FRN is more related to conflict detection (Jia et al., 2007; Li et al., 2009b; Veen and Carter, 2002). Therefore, it was possible that conflict-related negativity such as N2 would also be found in the present study (Chen et al., 2007; Shi, et al., 2005; Wang et al., 2001).

In addition, previous studies (Hajcak et al., 2005, 2007; Kutas et al., 1977; McCarthy and Donchin, 1981; Nasman and Rosenfeld, 1990) have demonstrated that the P3 component is related to outcome evaluation. Hence, we predicted that an effect for P3 would also be observed. Moreover, the results of previous relevant studies (Li et al., 2009a, 2009b, Li et al., 2011), suggest that increased LPC is closely related to the updating of working memory (WM) and hypothesis evaluation. If LPC is only correlated to the updating of WM, we predicted that an LPC effect would be absent in the present study, because there was no requirement for WM updating. On the contrary, if the LPC could also be triggered by the process of hypothesis evaluation alone, then we expected that there would be a significant difference between falsification and verification conditions during the LPC time window.

2. Results

2.1. Behavioral data

Reaction time (RT) and accuracy data were recorded for each trial in the experimental condition. RT was defined as the time between the onset of the second side (S2) of the card and key pressing. Accuracy was defined as the percentage of correct responses out of the total number of trials in each condition. Mean RT and accuracy rate were calculated for each of the four experimental conditions and shown in Table 1. The high accuracy levels across experimental conditions suggested that participants easily completed the hypothesis testing task.

The repeated measures ANOVA with 2 function (verification vs. falsification)×2 directionality (forward vs. backward) as within subject factors showed a significant effect of directionality for accuracy, $F(1, 21)=5.10$, $p=0.035$. There was also a significant interaction, $F(1, 21)=5.05$, $p=0.036$. The further analysis showed that accuracy in the forward falsification condition was significantly higher than in the backward falsification condition, $F(1, 21)=6.22$, $p=0.021$. The accuracy in the backward verification condition was also significantly higher than in the backward falsification condition, $F(1, 21)=5.77$, $p=0.026$.

For RT of correct trials, the main effect of function was significant, $F(1, 21)=114.625$, $p<0.001$. RT in the verification condition was substantially shorter than in the falsification condition. There was no main effect of directionality, $F(1, 21)=1.14$, $p=0.297$.

2.2. ERP results

2.2.1. Early components

As shown in Fig. 1, the N1, P2, N2 and P3 were elicited in all conditions. The repeated measures ANOVA showed that no

Table 1 – Performance accuracy and mean reaction time (RT) for each experimental condition.

	Forward verification	Backward verification	Forward falsification	Backward falsification
Accuracy	0.98 (0.15)	0.97 (0.16)	0.98 (0.15)	0.95 (0.21)
RT (ms)	853 (491)	865 (489)	1097 (562)	1146 (599)

Note: Standard deviations are in parentheses.

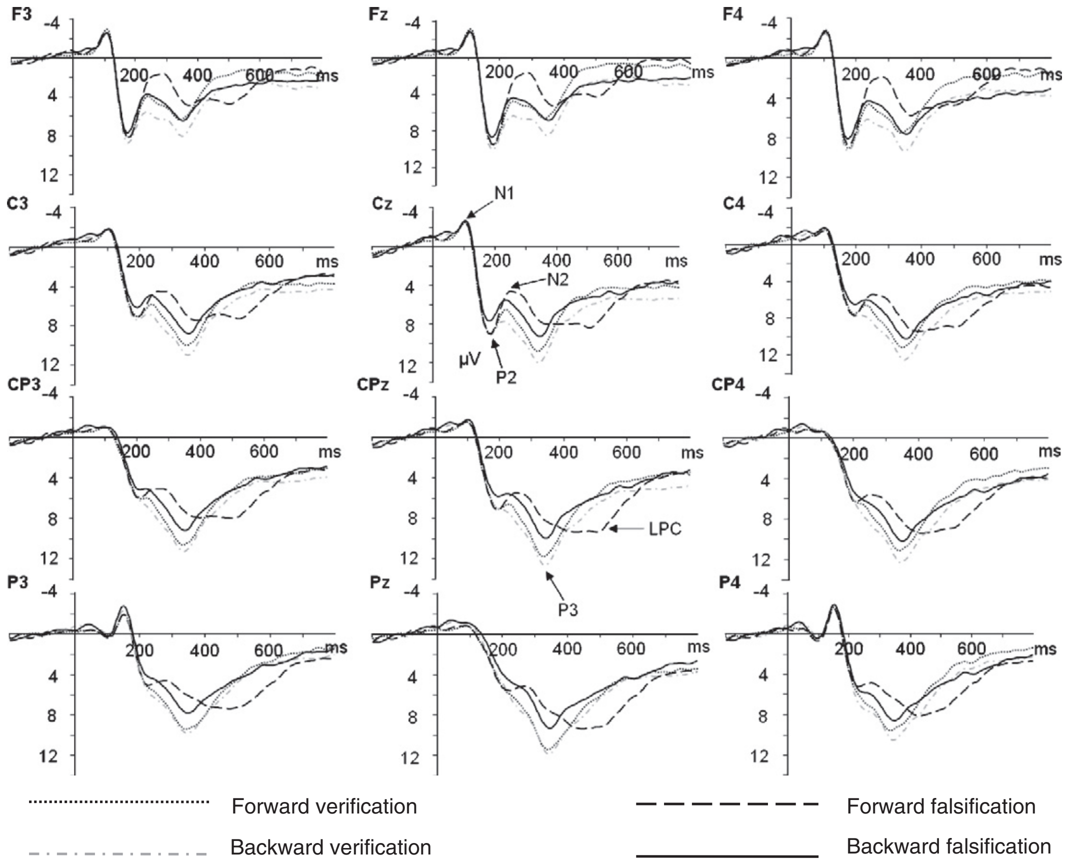

Fig. 1 – Grand average (n = 22) ERP waveforms for the four experimental conditions. The selected electrodes were F3, Fz, F4, C3, Cz, C4, CP3, CPz, CP4, P3, Pz and P4.

significant main effects of condition were found for the N1 latency or amplitude.

The main effects of function on P2 latency and amplitude were statistically significant. The P2 latency was significantly longer in the falsification condition than in the verification condition, $F(1, 21) = 5.53$, $p = 0.029$. P2 amplitude was larger in the verification condition than in the falsification condition, $F(1, 21) = 7.87$, $p = 0.011$. There was an interaction between function and directionality for P2 amplitude, $F(1, 21) = 5.67$, $p = 0.027$. Further analysis revealed a significant difference between backward verification and backward falsification, with a frontal distribution on the scalp, $F(1, 21) = 11.49$, $p = 0.003$. However, no significant difference was found between forward verification and forward falsification, $F(1, 21) = 0.01$, $p = 0.930$.

There were significant main effects of function and directionality on the latency and amplitude of N2 [latency: $F(1, 21) = 18.33$, $p < 0.001$ and $F(1, 21) = 8.20$, $p = 0.009$ for function and directionality respectively; amplitude: $F(1, 21) = 21.92$, $p = 0.001$ and $F(1, 21) = 6.22$, $p = 0.021$ for function and directionality respectively]. Interactions between function and directionality were not significant for the latency or amplitude of N2 [latency: $F(1, 21) = 2.39$, $p = 0.137$; amplitude: $F(1, 21) = 0.23$, $p = 0.636$]. As shown in the first row of Fig. 2, the obvious difference in N2 amplitude between forward verification and forward falsification appeared with a frontal distribution on the scalp, $F(1, 21) = 11.75$, $p = 0.003$. There was also significant difference for N2 amplitude between backward verification and backward falsification, which was distributed at the central sites, $F(1, 21) = 17.08$, $p < 0.001$.

Peak-to-peak test for N2 minus P2 revealed main effects of function [$F(1, 21) = 20.26$, $p < 0.001$], directionality [$F(1, 21) = 29.01$, $p < 0.001$] and an interaction [$F(1, 21) = 4.90$, $p = 0.038$]. The further analysis revealed significant differences between verification and falsification conditions, regardless of directionality [Forward: $F(1, 21) = 83.26$, $p < 0.001$; Backward: $F(1, 21) = 83.00$, $p < 0.001$].

2.2.2. *P3 and late components*

The main effects of function and directionality were significant with respect to P3 latency and amplitude. P3 latencies were shorter in the verification conditions than in the falsification conditions [$F(1, 21) = 15.99$, $p = 0.001$] and were longer in the forward conditions than in the backward conditions [$F(1, 21) = 5.25$, $p = 0.032$]. The P3 amplitudes were larger in the verification conditions than in the falsification conditions [$F(1, 21) = 26.85$, $p < 0.001$] and were smaller in the forward conditions than in the backward conditions [$F(1, 21) = 8.61$, $p = 0.008$]. The P3 latencies of the four conditions were 348.156 ± 3.457 ms (M ± SE) in forward verification, 349.065 ± 3.075 ms in backward verification, 361.847 ± 3.364 ms in forward falsification, and 353.655 ± 3.139 ms in backward falsification. The P3 amplitudes of the four conditions were 10.390 ± 1.131 μV in forward verification, 11.321 ± 1.109 μV in backward verification, 8.009 ± 1.115 μV in forward falsification,

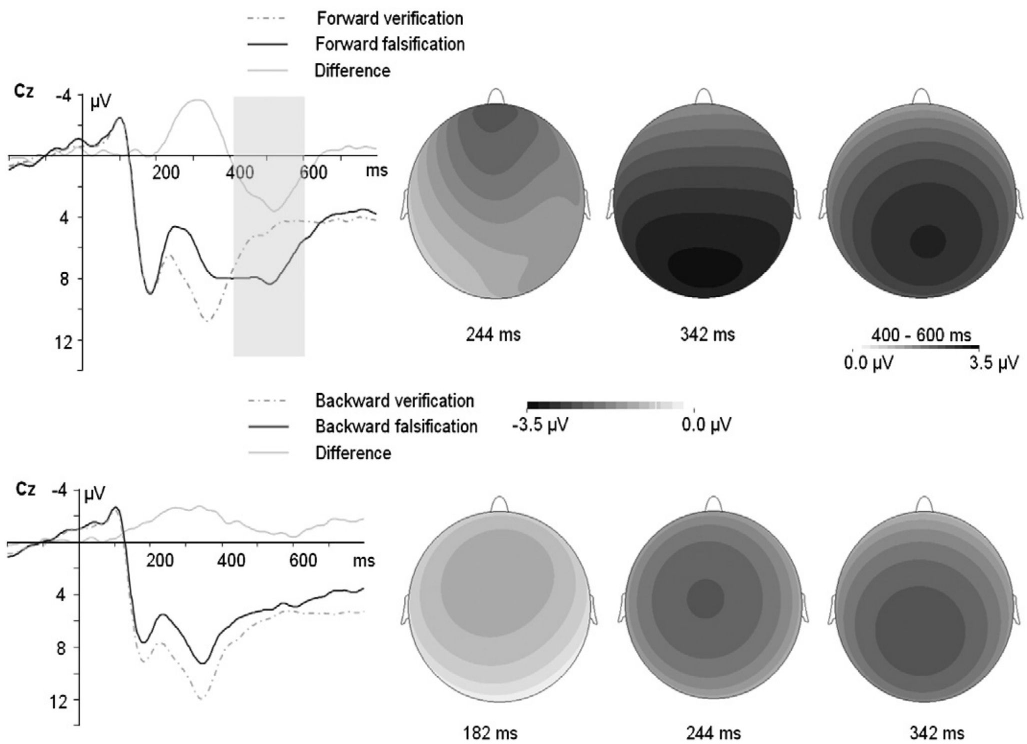

Fig. 2 – The difference waves of falsification minus verification and corresponding topographies. Up. Difference wave for forward condition and topographies in N2 (244 ms), P3 (342 ms), and LPC (400–600 ms) time windows; Bottom. Difference wave for backward condition and topographies in P2 (182 ms), N2 (244 ms), and P3 (342 ms) time windows.

and 9.479±1.126 μV in backward falsification. The interaction between function and directionality was significant for P3 latency, $F(1, 21)=6.59$, $p=0.018$. Further analysis showed a significant difference between forward verification and forward falsification, $F(1, 21)=21.70$, $p<0.001$. No significant difference in latency was found between backward verification and backward falsification, $F(1, 21)=3.32$, $p=0.083$. The interaction between function and directionality was not significant for the P3 amplitude, $F(1, 21)=0.72$, $p=0.406$.

As shown in Figs. 1 and 2, there was a more positive deflection in the forward falsification condition than in the other three conditions during 400–600 ms time windows. Mean amplitudes for LPC in the forward falsification condition were larger than those of other three conditions. With regard to mean amplitudes of LPC for every 50-ms time window from 400 to 600 ms, the main effect of function was significant during 500–550 ms, $F(2, 21)=7.42$, $p=0.014$. There were interactions between function and directionality in each time window [400–450 ms, $F(1, 21)=11.59$, $p=0.003$; 450–500 ms, $F(1, 21)=26.87$, $p<0.001$; 500–550 ms, $F(1, 21)=29.02$, $p<0.001$; 550–600 ms, $F(2, 21)=8.42$, $p=0.009$]. Further analysis indicated that there were significant differences between forward verification and forward falsification within most of the LPC time windows except 400–450 ms [400–450 ms, $F(1, 21)=2.83$, $p=0.107$; 450–500 ms, $F(1, 21)=18.84$, $p<0.001$; 500–550 ms, $F(1, 21)=21.52$, $p<0.001$; 550–600 ms, $F(2, 21)=7.69$, $p=0.011$] (see Table 2 and Fig. 2). Compared with backward verification, backward falsification was more negative but no significant difference was found between them within most LPC time windows except for 400–450 ms [400–450 ms, $F(1, 21)=6.67$, $p=0.017$; 450–500 ms, $F(1, 21)=2.03$, $p=0.169$; 500–550 ms, $F(1, 21)=0.37$, $p=0.548$; 550–600 ms, $F(1, 21)=0.09$, $p=0.766$].

2.2.3. *PCA analysis in forward verification and forward falsification condition*

Results of PCA in the forward verification condition indicated that three eigenvectors, explaining 89.2% of the total variance of difference waveforms, were identified (Fig. 3). Fig. 3a presents the eigenvectors (rotated factor loadings) and their eigenvalues (percentage of total variance explained). The three eigenvectors had distinct scalp distributions, showing different patterns of activity at frontal, parietal, and central locations. Eigenvector 1, which accounted for 62.0% of the total variance and peaked at approximately 350 ms, contributed differentially to the waveforms at centro-parietal locations than at frontal sites and was prominent at Pz, $F(2, 42)=29.80$,

Table 2 – Pair-wise comparisons of mean amplitudes of LPC between verification and falsification.

Interval (ms)	Forward verification vs. forward falsification		Backward verification vs. backward falsification	
	$F(1, 21)$	p	$F(1, 21)$	p
400–450	2.83	0.107	6.67	0.017
450–500	18.84	0.000	2.03	0.169
500–550	21.52	0.000	0.37	0.548
550–600	7.69	0.011	0.09	0.766

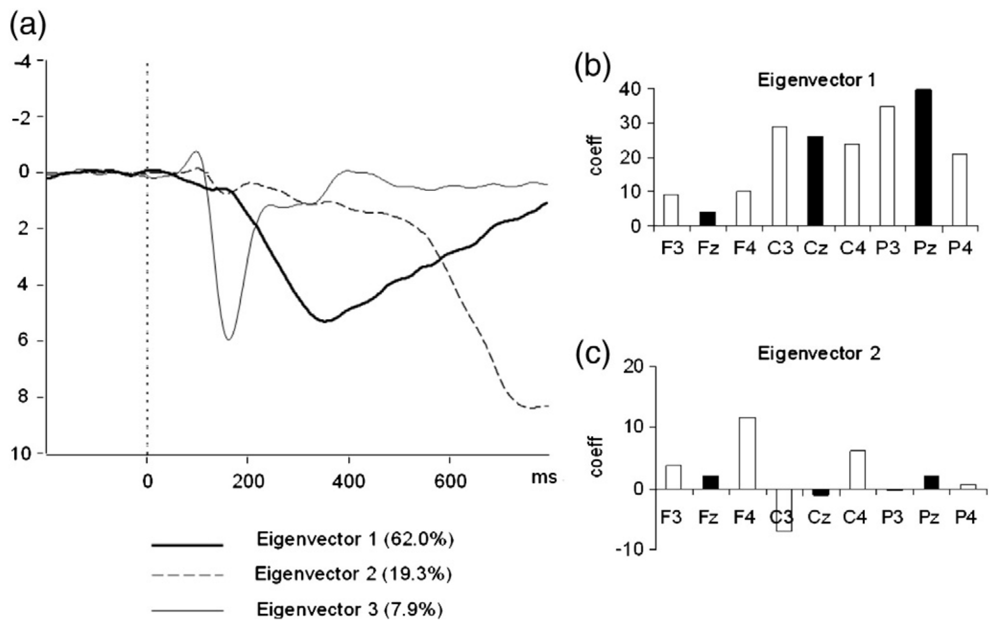

Fig. 3 – The results of PCA in forward verification condition. (a) Eigenvectors obtained from PCA. Included were ERPs elicited by 22 participants in response to forward verification at nine electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4). Shown here are the three eigenvectors with the largest eigenvalues (listed in parentheses). (b) Coefficients of eigenvector 1. (c) Coefficients of eigenvector 2.

$p<0.001$ (Fig. 3b). Eigenvector 2, explaining 19.3% of the total variance, showed no early activity with energy increase beginning at 600 ms and showed distinct effect at right scalp recordings, which were prominent at F4, $F(2, 42)=7.87$, $p=0.011$ (Fig. 3c). Eigenvector 3, explaining 7.9% of the total variance, contributed prominently to the waveforms frontally and was prominent at Fz, $F(2, 42)=20.03$, $p<0.001$, with a peak contribution at about 160 ms.

Three eigenvectors were identified in the forward falsification condition (Fig. 4), explaining 86.6% of the total variance of the difference waveforms. Fig. 4a presents the eigenvectors (rotated factor loadings) and their eigenvalues. Eigenvector 1, accounting for 67.7% of the total variance, showed early activity with an energy increase beginning at 100 ms and a peak contribution at about 160 ms. The relative contribution to the waveform (reconstruction coefficients) was larger at Fz

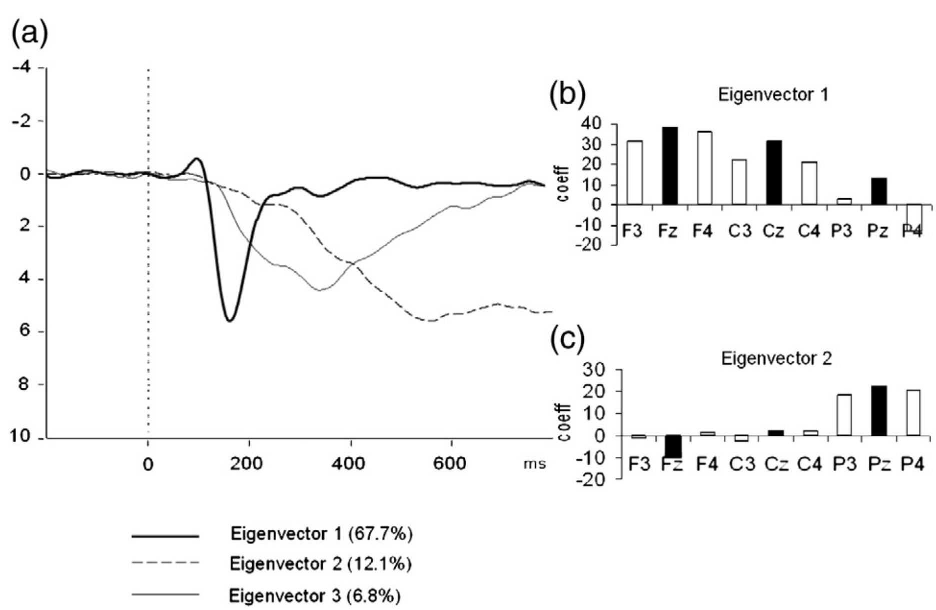

Fig. 4 – The results of PCA in the forward falsification condition. (a) Eigenvector obtained from PCA. Included are ERPs elicited by 22 participants in response to forward falsification at nine electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4). Shown here are the three eigenvectors with the largest eigenvalues (listed in parentheses). (b) Coefficients of eigenvector 1. (c) Coefficients of eigenvector 2.

than at all other recording locations, $F(2, 42)=8.34$, $p=0.004$ (Fig. 4b). Eigenvector 2, explaining 12.1% of the total variance, showed an energy increase beginning at 300 ms and a peak contribution at about 550 ms. This eigenvector contributed differentially to the waveforms at parietal locations rather than at frontal and central sites, and was prominent at Pz, $F(2, 42)=5.73$, $p=0.011$ (Fig. 4c). Eigenvector 3, explaining 6.8% of the total variance, peaked at 340 ms and showed distinct effect at centro-parietal rather than frontal sites, $F(2, 42)=6.97$, $p=0.003$.

In summary, PCA analysis revealed that there was only one source of activity (i.e. eigenvector 1) during the P3-LPC (320–600 ms) time window for the forward verification condition. In contrast, there were two sources of activity (i.e. eigenvector 2 and eigenvector 3) during the same time window for the forward falsification condition.

3. Discussion

In the present study, the temporal course and neural correlates of hypothesis evaluation were examined by comparing effects of four experimental conditions in a modified Wason's selection task. The high accuracy in each of the four experimental conditions suggested that participants successfully solved the HT problems. Consistent with previous studies (Klayman and Ha, 1987; Nickerson, 1996, 1998; Wason, 1960, 1968), the RT was shorter in the verification conditions than in the falsification conditions, reflecting the confirmation bias in hypothesis testing.

3.1. *Early components*

There were no significant effects of function or directionality with respect to the N1 amplitude and latency. This suggests that early in the process, attention allocation was similar across the four experimental conditions (Mangun, 1995). Conversely, significant effects were observed for the P2 amplitude, which is related to perceptual processing (Bigman and Pratt, 2004; Chen et al., 2007; Hillyard and Anllo-Vento, 1998; Li et al., 2009a). The P2 amplitude was higher in the backward verification condition than in the backward falsification condition. This effect may be related to differences in perceptual processing as a result of the preceding stimulus, i.e., the first side (S1) of a test card. Specifically, S1 in the backward verification condition shared the same attributes with the consequent of the given proposition. It is possible that participants may have supposed that the card could verify the stated proposition and thus paid more mental resources toward S2, such that perceptual processing was more intensive when S2 was presented. However, S1 did not share the same attributes with the antecedent or consequent of the given proposition in backward falsification. Hence, participants may have presumed that the card was inconsistent or irrelevant with the stated proposition and thus used fewer cognitive resources when S2 was presented due to the universal existence of confirmation bias (Klayman and Ha, 1987; Nickerson, 1996, 1998; Wason, 1960, 1968).

In comparison with verification conditions, the falsification conditions elicited a FRN-like negativity during the time window of 210–310 ms after stimulus onset, consistent with the idea that hypothesis testing is related to feedback (Papo et al., 2003). However, it was difficult to interpret the difference between verification and falsification conditions, because no corrective feedback was provided in the current study. FRN is elicited when the signal of positive or negative feedback is given explicitly (Papo et al., 2003), but some recent studies (Holroyd, et al., 2006; Li et al., 2009b; Muller et al., 2005) have observed that neutral feedback, as well as negative feedback, may elicit a negative deflection. In addition, other researchers have argued that the FRN is related to conflict detection (Chen et al., 2007, 2008; Folstein and Van Petten, 2008; Nieuwenhuis et al., 2003; Veen and Carter, 2002; Yeung and Cohen, 2006). Controversial views about the nature of FRN suggest that the negative deflection for negative feedback may reflect different cognitive processes. The negative deflection found for the falsification condition in the present study may not be the function of feedback, at least not directly.

In contrast, the effect of condition on the mean amplitudes during 210–310 ms was most likely correlated to conflict detection. In the verification condition, test cards were congruent with the representation of given hypotheses, suggesting that the test cards were matched well to the context (i.e., the given hypothesis). Hence, there was no conflict between test cards and the context. However, in the falsification condition, test cards were not congruent with the context (the given hypothesis). As a result, conflicting information was processed intensively.

3.2. *P3 and late components, index of hypothesis evaluation*

In the current study, larger P3 amplitude was found in the verification than in the falsification condition, irrespective of directionality. There is compelling evidence that the P3 component is related to the process of outcome evaluation or decision making (Hajcak et al., 2005, 2007; Luu et al., 2009; Sato et al., 2005; Toyomaki and Murohashi, 2005; Wu and Zhou, 2009; Yeung et al., 2005, Yeung and Sanfey, 2004; Yu et al., 2007). Some ERP studies suggest that P3 amplitude in outcome evaluation is stronger for positive outcomes (or positive feedback) than for negative outcomes (Hajcak et al., 2005, 2007; Johnson and Donchin, 1985; Wu and Zhou, 2009). Consistent with these studies, P3 in the present study may be correlated with the evaluation process in hypothesis testing tasks. Specifically, in the falsification condition, test cards disconfirmed the evidence (or outcome) for given hypotheses, while those in the verification condition confirmed the evidence. These differences between outcome evaluations were reflected by different P3 amplitudes. Moreover, intrinsic difficulties of hypothesis evaluation are reflected by RT and P3 latencies (Kutas et al., 1977; McCarthy and Donchin, 1981; Nasman and Rosenfeld, 1990). In this study, P3 latencies were significantly longer in falsification than verification conditions, suggesting more resources were required in the former conditions.

Interestingly, an exaggerated positive potential peaking at around 500 ms was only found in the forward falsification condition. PCA results showed that the late positive component (LPC) was not the same component as P3, suggesting that the LPC may be correlated to further independent processing of hypothesis evaluation, central to the manipulation of cognitive context. Diverging from previous studies (Donchin, 1981, Donchin and Coles, 1988; Kiss et al., 1998; Kusak et al., 2000; Li et al., 2009a, 2009b; Polich, 2003, 2007), participants in the present study were required only to judge whether a stimulus was congruent with the cognitive context (i.e. the preceding proposition). As a result, the process of updating of working memory was not salient, and therefore the LPC may be not correlated to this process. Presumably, LPC is related to the "central executive" component of WM (Baddeley, 1992; Morris and Jones, 1990). A number of studies have found that late positive wave correlates with executive activity (Garcia-Larrea and Cezanne-Bert, 1998; Gevins et al., 1996; Kiss et al., 1998, 2001). Indeed, online information monitoring is needed in the hypothesis evaluation stage. When evaluating the relationship between a card and a given hypothesis, the central executive system is called upon to monitor the process of evaluation. In the forward falsification condition, wherein the new stimuli violated the cognitive context, the central executive system monitored this violation and elicited the increased LPC.

To summarize, by utilizing a modified Wason's selection task, this experiment explored brain activation underlying hypothesis evaluation under different conditions. The P2 amplitude was smaller in the backward falsification than in the backward verification condition, reflecting less intense perceptual processing. During the N2 component, there were larger negative deflections in the falsification conditions than in the verification conditions, irrespective of directionality, suggesting the occurrence of more intensive conflict. In addition, the P3 amplitudes were smaller in the falsification condition than in the verification conditions, implying differential processing of hypothesis evaluation. A late positive component was found only in the forward falsification condition during 400–600 ms post-stimulus, which may be a reflection of the central manipulation of cognitive context.

4. Experimental procedures

4.1. *Participants*

Participants were 22 native Chinese college students (11 males) between the ages of 19 and 23 years (mean = 21 years; SD = 1.1). All participants were right-handed, had normal or corrected-to-normal vision and were free from any psychiatric diagnoses or medication. Each participant signed an informed consent form prior to the study and was paid for participation.

4.2. *Materials and design*

Participants were presented with a proposition and two sides of a card sequentially, displayed in the center of a 17-inch screen (CRT monitor). Within each trial, the proposition was a statement about the content on the two sides of a card, describing a relationship between a geometric figure and a digital number. All stimuli were presented on a computer screen with a resolution of 800 × 600 pixels. Shapes of geometric figures included rectangles, squares, triangles, parallelograms, trapezoids, circles, crescents, hearts, arrows, crosses, diamonds, pentagons, and hexagons. The Arabic digits were integer numbers (1 to 9, −1 to −9). The stimuli were all drawn in CorelDraw 11 (Corel Corporation, Ottawa, Canada), individually exported, and saved as bitmap files. Sizes of the figures and numbers were approximately 2.21 cm in height and 2.21 cm in width. During the experiment, the distance between participant eye level and the screen was approximately 80 cm. The horizontal and vertical angles were both less than 3°.

Participants were informed that an "if" proposition and two sides of a card would be presented sequentially in each trial. Their task was to judge whether the card supported the preceding proposition by pressing one of three keys when the second side of the card was presented.

As detailed in Table 3, independent variables were (1) the function of the test cards (two levels: verification, falsification) and (2) directionality of presentation of the two sides of a card (two levels: forward, backward). Therefore, four experimental conditions [forward verification (PQ), backward verification (QP), forward falsification (P–Q), backward falsification (–QP)] were subjected to statistical analysis. The four experimental conditions and other irrelevant conditions were arranged randomly for each participant.

Prior to the experimental session, participants were required to practice until accuracy reached 90% across the four experimental conditions (PQ, P–Q, QP, –QP). The formal experimental session included 80 trials for each experimental condition and 20 trials for each irrelevant condition (resulting in a total of 400 trials for each participant).

Table 3 – The example of experimental design.

Proposition	Condition		First side of a card	Second side of a card	Number of trials
	Directionality	Function			
If one side of a card is a rectangular, the other side is an even number.	Forward	Verification	Rectangular(P)	Even number(Q)	80
		Falsification	Rectangular(P)	Odd number(−Q)	80
		Irrelevance	Other shape(−P)	Even number(Q)	20
		Irrelevance	Other shape(−P)	Odd number(−Q)	20
	Backward	Verification	Even number(Q)	Rectangular(P)	80
		Irrelevance	Even number(Q)	Other shape(−P)	20
		Falsification	Odd number(−Q)	Rectangular (P)	80
		Irrelevance	Odd number(−Q)	Other shape(−P)	20

Note: P means the antecedent of a proposition; Q means the consequent of a proposition.

At the beginning of each trial, a proposition appeared on the computer monitor for a maximum of 5000 ms and disappeared when the participant confirmed comprehension by pressing any key. This was followed by a fixation point (a white cross) for 300 ms which was replaced by a blank screen for 800–1000 ms. Then the first side of a card was presented for 1000 ms, followed by a blank screen for 800–1000 ms. Finally, the second side of the card was presented until a response was made (Fig. 5). Participants were informed to use their right forefinger to press key "J" when the card supported the proposition (verification), and use their left forefinger to press key "F" when the card rejected the proposition (falsification), and use their thumb to press key "Space" when the card was irrelevant to (i.e. neither supported nor rejected) the proposition. Response hand (left and right key pressing) was counterbalanced across participants.

Participants were instructed to remain still and to avoid head movements during the task. They were instructed to blink only while reading a proposition and after responding to S2. Breaks were given every 60 trials and the length of the break for each participant was based on individual preference.

4.3. *Electrophysiological recording and analysis*

Electroencephalography (EEG) was conducted with a 64-channel EEG recording system (Brain Products, Germany), with linked references on the left and right mastoids (average mastoid reference, Luck, 2005). A ground electrode was placed on the medial aspect of the frontal region. The vertical electrooculogram (VEOG) was recorded with electrodes placed above and below the left eye and the horizontal electrooculogram (HEOG) with electrodes placed by the right side of the right eye and the left side of left eye.

The EEG data was analyzed off-line using Vision Analyzer software (Brain Vision analyzer). First, the EEG data was subjected to the eye movement correction algorithm of Gratton et al. (1983), implemented using the Vision Analyzer software (Brain Vision analyzer). Consistent with previous studies (Li et al., 2009a; Sim and Kiefer, 2005), the corrected data was then segmented and filtered (band-pass 1–16 Hz, 24-bit analog-to-digital converter). Then the filtered data was corrected to a 200 ms baseline prior to the onset of stimulus (i.e., the second side of the test card), followed with a rejection of artifactual trials (exceeding 120 μV). As a result, approximately 8.8% of the total trials were excluded (8.9% in the forward verification condition, 8.6% in the backward verification condition, 7.6% in the forward falsification condition, and 10.4% in the backward falsification condition). Finally, the artifact-free ERPs time-locked to the onset of S2 were averaged separately in each condition for each participant.

As seen in the grand averaged waveforms and topographical maps (Figs. 1 and 2), the four peaks were clearly identified among all the four conditions in most of the fronto-central and parietal sites. Accordingly, the following 25 electrode sites were chosen for statistical analysis: F1, F3, Fz, F2, F4, FC1, FC3, FCz, and FC2, FC4 (10 frontal to central sites) and C1, C3, Cz, C2, C4, CP1, CP3, CPz, CP2, CP4, P1, P3, Pz, P2, P4 (15 central to parietal sites). The peak amplitudes were measured with respect to the mean voltages during the 200 ms pre-stimulus interval. Peak latencies were measured relative to stimulus onset. The N1 component was measured in the 80–150 time window, P2 was assessed in the 150–200 ms time window, N2 was detected in the 210–310 ms time window and P3 was detected in the 320–370 ms time window. For the N1, P2, N2 and P3, we selected the average amplitude of the respective ERP components in a time window from 20 ms before the peak until 20 after the peak for statistical analysis. Additionally, since there were P2 differences between conditions, in order to exclude a possible influence of the preceding P2 on the N2, we conducted a peak-to-peak measurement on P2–N2 complex. As no clear peak was found after 400 ms, the mean amplitudes of the LPC (400–600 ms) were measured at successive 50 ms intervals.

Latencies and amplitudes (baseline to peak) of the components (N1, P2, N2, P3), as well as the mean amplitudes of LPC were analyzed using a two function (verification vs. falsification) × two directionality (forward vs. backward) × 25 (electrodes) repeated measures analysis of variance (ANOVA). When significant effects were found, post-hoc examination of the nature of the effects was performed using pair-wise comparison tests. For all analyses, P-values were corrected for deviation from sphericity according to the Greenhouse–Geisser method.

Defining P3 and the LPC in the forward verification and forward falsification condition could potentially be problematic due to the relatively low signal to noise ratio of the component. Standard peak picking techniques are often unable to pick the signal out from the surrounding noise. To overcome this difficulty and provide a quantification of the differences of the topographical distribution between P3 and LPC, we used principal components analysis (PCA) to separate the overlapping components.

PCA has been utilized repeatedly in ERP research as an additional tool to aid in the analysis of the scalp-recorded ERP and differentiation of overlapping components (Bigman and Pratt, 2004; Spencer et al., 1999, 2001). In PCA, a set of averaged

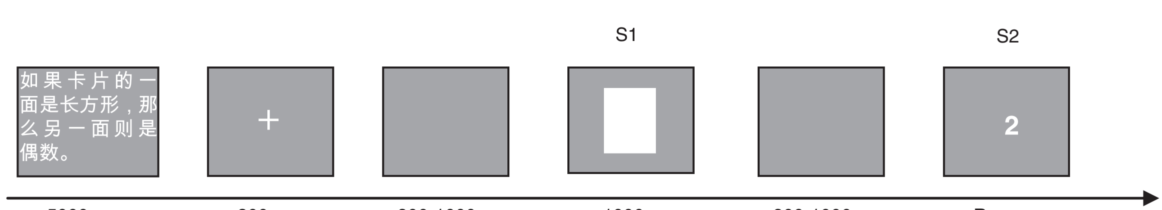

Fig. 5 – The experimental procedure for one sample trial in the present study. The proposition was written in Chinese and meant that "If one side of a card is a rectangular, then the other side is an even number".

ERP waveforms is statistically examined for covariation in amplitude across time points.

In this study, PCA was performed on the whole amplitude of early peaks and late components (500 sampling points, 200 ms pre-stimulus until 800 ms post-stimulus), including P3 and LPC. PCA was performed for the forward verification condition and the forward falsification condition independently, with an ensemble of waveforms from all 22 participants. For each participant, the PCA included waveforms from 64 scalp electrodes in both conditions. Following PCA, ANOVA was performed on the coefficients of the eigenvectors that accounted for at least 5% of the total variance.

Acknowledgments

This work was supported by an NNSF grant 31070985 and a Key Discipline Fund of National 211 Project, China (NSKD08014).

REFERENCES

Baddeley, A., 1992. Working memory. Science 255, 556–559.
Bigman, Z., Pratt, H., 2004. Time course and nature of stimulus evaluation in category induction as revealed by visual event-related potentials. Biol. Psychol. 66, 99–127.
Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., Cohen, J.D., 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280, 747–749.
Chen, A.T., Luo, Y.J., Wang, Q.H., Yuan, J.J., Yao, D.Z., Li, H., 2007. Electrophysiological correlates of category induction: PSW amplitude as an index of identifying shared attributes. Biol. Psychol. 76, 230–238.
Chen, A.T., Xu, P., Wang, Q.H., Luo, Y.J., Yuan, J.J., Yao, D.Z., et al., 2008. The timing of cognitive control in partially incongruent categorization. Hum. Brain Mapp. 29, 1028–1039.
Donchin, E., 1981. Surprise surprise? Psychophysiology 18, 493–513.
Donchin, E., Coles, M.G., 1988. Is the P300 component a manifestation of context updating? Behav. and Brain Sci. 11, 357–427.
Elliott, R., Dolan, R.J., 1998. Activation of different anterior cingulate foci in association with hypothesis testing and response selection. NeuroImage 8, 17–29.
Folstein, J.R., Van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology 45, 152–170.
Garcia-Larrea, L., Cezanne-Bert, G., 1998. P3, positive slow wave and working memory load: a study on the functional correlates of slow wave activity. Electroencephalogr. Clin. Neurophysiol. 108, 260–273.
Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artifact. Electroencephalogr. Clin. Neurophysiol. 55, 468–484.
Gevins, A., Smith, M.E., Le, J., et al., 1996. High resolution evoked potential imaging of the cortical dynamics of human working memory. Electroencephalogr. Clin. Neurophysiol. 98, 327–348.
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, 161–170.
Hajcak, G., Moser, J.S., Holroyd, C.B., Simons, R.F., 2007. It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. Psychophysiology 44, 905–912.
Hillyard, S.A., Anllo-Vento, L., 1998. Event-related brain potentials in the study of visual selective attention. Proc. Natl. Acad. Sci. 95, 781–787.
Holroyd, C.B., Coles, M.G.H., 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol. Rev. 109, 679–709.
Holroyd, C.B., Hajcak, G., Larsen, J.T., 2006. The good, the bad and the neutral: electrophysiological responses to feedback stimuli. Brain Res. 1105, 93–101.
Jia, S.W., Li, H., Luo, Y.J., Chen, A.T., Wang, B.X., Zhou, X.L., 2007. Detecting perceptual conflict by the feedback-related negativity in brain potentials. Neuroreport 18, 1385–1388.
Johnson, R., Donchin, E., 1985. Second thoughts: multiple P300s elicited by a single stimulus. Psychophysiology 22, 182–194.
Kiss, I., Pazderka-Robinson, H., Floden, D., 2001. Event-related brain potentials and central executive function: further evidence for Baddeley's model. J. Psychophysiol. 15, 1–13.
Kiss, I., Pisio, C., Francois, A., Schopflocher, D., 1998. Central executive function in working memory: event-related brain potential studies. Cogn. Brain Res. 6, 235–247.
Klayman, J., Ha, Y.W., 1987. Confirmation, disconfirmation, and information in hypothesis testing. Psychol. Rev. 94, 211–228.
Kusak, G., Grune, K., Hagendorf, H., Metz, A.M., 2000. Updating of working memory in a running memory task: an event-related potential study. Int. J. Psychophysiol. 39, 51–65.
Kutas, M., McCarthy, G., Donchin, E., 1977. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197, 792–795.
Li, F.H., Cao, B.H., Lei, Y., Li, H., 2009a. Electrophysiological correlates of hypothesis testing. NeuroReport 20, 197–201.
Li, F.H., Cao, B.H., Gao, X.M., Xiao, F., Qing, J.L., Li, H., 2009b. Temporal course and the electrophysiological correlates of hypothesis testing as revealed in a modified category induction task. Brain Res. 1301, 61–67.
Li, F.H., Cao, B.H., Cai, X.L., Li, H., in press. Similar brain mechanism of hypothesis testing between children and adults. Developmental Neuropsychology.
Luck, S.J., 2005. An introduction to event-related potentials and their neural origins. In: Luch, S.J. (Ed.), An Introduction to the Event-related Potential Technique. MIT Press, Cambridge, MA, p. 107.
Luu, P., Shane, M., Pratt, N., Tucker, D.M., 2009. Corticolimbic mechanisms in the control of trial and error learning. Brain Res. 1247, 100–113.
Mangun, G.R., 1995. Neural mechanisms of visual selective attention. Psychophysiology 32, 4–18.
McCarthy, G., Donchin, E., 1981. A metric for thought: a comparison of P300 latency and reaction time. Science 211, 77–80.
Morris, N., Jones, D.M., 1990. Memory updating in working memory: the role of the central executive. Br. J. Psychol. 81, 111–121.
Muller, S.V., Moller, J., Rodriguez-Fornells, A., Munte, T.F., 2005. Brain potentials related to self-generated and external information used for performance monitoring Clin. Neurophysiol. 116, 63–74.
Nasman, V.T., Rosenfeld, J.P., 1990. Parietal P3 response as an indicator of stimulus categorization: increased P3 amplitude to categorically deviant target and non-target stimuli Psychophysiology 27, 338–350.
Nickerson, R.S., 1996. Hempel's paradox and Wason's selection task: logical and psychological puzzles of confirmation. Thinking and Reasoning 2, 1–31.
Nickerson, R.S., 1998. Confirmation bias: a ubiquitous phenomenon in many guises. Rev. Gen. Psychol. 2, 175–220.
Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., Ridderinkhof, K.R., 2003. Electrophysiological correlates of

anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. Cogn. Affect. & Behav. Neurosci. 3, 17–26.
Papo, D., Baudonnie' re, P.M., Hugueville, L., Caverni, J.P., 2003. Feedback in hypothesis testing: an ERP study. J. Cogn. Neurosci. 15, 508–522.
Polich, J., 2003. Overview of P3a and P3b. In: Polich, J. (Ed.), Detection of Change: Event-related Potential and fMRI Findings. Kluwer Academic Press, Boston, pp. 83–98.
Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148.
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 4, 407–411.
Shi, L.X., Wang, Y.P., Yang, J., 2005. The effects of different hands' responses on main ERP components. J. Clin. Neurophysiol. 1, 74–78.
Sim, E.-J., Kiefer, M., 2005. Category-related brain activity to natural categories is associated with the retrieval of visual features: evidence from repetition effects during visual and functional judgments. Cogn. Brain Res. 24, 260–273.
Spencer, K.M., Dien, J., Donchin, E., 1999. A componential analysis of the ERP elicited by novel events using a dense electrode array. Psychophysiology 36, 409–414.
Spencer, K.M., Dien, J., Donchin, E., 2001. Spatiotemporal analysis of the late ERP responses to deviant stimuli. Psychophysiology 38, 343–358.
Toyomaki, A., Murohashi, H., 2005. The ERPs to feedback indicating monetary loss and gain on the game of modified "rock–paper–scissors". Int. Congr. Ser. 1278, 381–384.
Veen, V., Carter, C.S., 2002. The timing of action-monitoring processes in the anterior cingulate cortex. J. Cogn. Neurosci. 14, 593–602.
Wason, P.C., 1960. On the failure to eliminate hypotheses in a conceptual task. Q. J. Exp. Psychol. 12, 129–140.
Wason, P.C., 1968. On the failure to eliminate hypotheses—a second look. In: Wason, P.C., Johnson-Laird, P.N. (Eds.), Thinking and Reasoning. Penguin, Harmondsworth, Middlesex, England, pp. 165–174.
Wang, H.J., Wang, Y.P., Kong, J., Cui, L.L., Tian, S.J., 2001. Enhancement of conflict processing activity in human brain under task relevant condition. Neurosci. Lett. 298, 155–158.
Wu, Y., Zhou, X.L., 2009. The P300 and reward valence, magnitude, and expectancy in outcome evaluation. Brain Res. 1286, 114–122.
Yeung, N., Cohen, J.D., 2006. The impact of cognitive deficits on conflict monitoring: predictable dissociations between the error-related negativity and N2. Psychol. Sci. 17, 164–171.
Yeung, N., Holroyd, C.B., Cohen, J.D., 2005. ERP correlates of feedback and reward processing in the presence and absence of response choice. Cereb. Cortex 15, 535–544.
Yeung, N., Sanfey, A.G., 2004. Independent coding of reward magnitude and valence in the human brain. J. Neurosci. 24, 6258–6264.
Yu, R.J., Luo, Y.J., Ye, Z., Zhou, X.L., 2007. Does the FRN in brain potentials reflect motivational/affective consequence of outcome evaluation? Prog. Nat. Sci. 17, 136–143.