Effects of value and reward magnitude on feedback negativity and P300

Atsushi Sato,[CA*] Asako Yasuda,[*] Hideki Ohira, Kaori Miyawaki, Masami Nishikawa, Hiroaki Kumano and Tomifusa Kuboki

Department of Psychosomatic Medicine, Graduate School of Medicine, The University of Tokyo, Japan

[CA]Corresponding Author: atsuchan-tky@umin.ac.jp

[*]A. Sato and A. Yasuda contributed equally to this work.

Received 15 November 2004; accepted 10 January 2005

Feedback negativity is a negative component of the event-related brain potential observed 250–300 ms after feedback stimuli. The present study investigated the effects of value (correct or incorrect) and reward magnitude (no, small or large) on feedback negativity and P300. Feedback negativity was larger after incorrect feedback than after correct feedback, irrespective of reward magnitude. In contrast, P300 amplitude increased with reward magnitude, irrespective of value. The amplitude of feedback negativity was correlated with a trait score of negative affect and not positive affect, whereas P300 amplitude was correlated with positive affect and not negative affect. These results suggest that value and reward magnitude are processed separately in the brain. *NeuroReport* 16:407–411 © 2005 Lippincott Williams & Wilkins.

Key words: Feedback negativity; Medial frontal negativity; Positive and negative affect scales; P300; Reward magnitude

INTRODUCTION

For advantageous decision-making under uncertain circumstances, it is necessary to evaluate the affective or motivational significance of ongoing events rapidly and to use this evaluation to guide future decision-making. Recent studies of event-related brain potentials (ERPs) have identified neural correlates of this rapid evaluative function. Studies of feedback negativity or medial frontal negativity have contributed to this evidence [1–6]. Feedback negativity is a negative component of ERP that occurs when participants receive feedback stimuli indicating outcomes of their previous choices in simple gambling tasks. It is maximal over medial frontal scalp locations, peaking 250–300 ms after feedback presentation. The results of previous brain electrical source analyses have suggested that feedback negativity is generated in the anterior cingulate cortex (ACC) [1].

Feedback negativity was originally observed following feedback stimuli indicating incorrect performance irrespective of the modality of the feedback stimuli [2], leading to the proposal that this component reflects the detection of errors. However, a recent study, in which correctness was defined in terms of whether the participant's chosen outcome was better or worse than the alternative outcome, showed that feedback negativity was most pronounced following monetary losses as opposed to monetary gains, whereas the correct/error status did not influence the amplitude [1]. More recently, it has been reported that feedback negativity is sensitive to both the gain/loss aspect and the correct/incorrect value of the feedback depending on which aspect is most salient [3], suggesting that this component reflects the evaluation of events along a good–bad dimension, rather than in terms of correctness or gain/loss. An alternative theory holds that feedback negativity is elicited by the impact of reward prediction error signal to the ACC *via* the mesencephalic dopamine system [4]. The mesencephalic dopamine neurons are known to code errors in prediction of reward, that is the discrepancy between the predicted and the actual reward. According to this reinforcement learning theory, these reward signals are used to guide action selection mediated by the ACC, through the reinforcement of actions associated with reward and punishment of actions associated with penalty or no reward. However, feedback negativity was observed following negative outcomes even in a task context in which participants made no overt actions [5]. This result suggests that feedback negativity reflects the evaluative function alone, and not its use in reinforcing or punishing a recently produced response.

Taken together, these data are consistent with the proposal that feedback negativity reflects rapid evaluation of the affective or motivational impact of outcome events. However, the nature of the evaluative process reflected in feedback negativity remains unclear. Recently, it has been reported that feedback negativity was larger after monetary losses than after monetary gains, irrespective of the magnitude of reward or penalty [6]. In contrast, the amplitude of a later component of the ERP, P300, was shown to be sensitive to the absolute magnitude of the reward, independently of the value of the outcome [6]. These results suggest that the evaluative function reflected in feedback negativity may be limited to the classification of events as being good or bad. The purpose of the present study was to replicate and enlarge this finding. In

a two-choice decision-making task in which participants were asked to guess whether the next card would be higher or lower than the cue card, the value of outcome and the magnitude of reward or penalty were manipulated. Under the large reward or penalty conditions, participants could receive a large monetary reward for a correct decision but lose it for an incorrect decision; under the small reward or penalty conditions, they could receive or lose a small monetary reward for correct or wrong decisions, respectively; under the no reward or penalty conditions, they neither received nor lost any money related to their choice. If feedback negativity selectively reflects the classification of events as being good or bad, it should be sensitive only to the correct/incorrect aspect of the feedback. That is, even under no penalty conditions, error feedback should elicit feedback negativity as large in amplitude as under small and large penalty conditions. In this study we also investigated whether individual differences in negative affectivity, the dispositional tendency to experience negative affect, affected the amplitude of feedback negativity. Previous studies showed that individuals with high trait negative affect displayed an increased tendency to classify an ambiguous stimulus as negative. This tendency is supposed to be a vulnerable factor to anxiety disorder and depression [7]. Thus, we hypothesized that there were significant positive correlations between trait negative affect scores and the amplitude of feedback negativity.

METHOD

Participants: Eighteen healthy right-handed undergraduates (eight female), ranging in age from 20 to 31 years (mean=25.7 years, SD=3.4), participated in this study. All participants had normal or corrected-to-normal vision. None had a history of neurological or psychiatric disease. This research was approved by the local ethics committee and was conducted in accordance with the Declaration of Helsinki. All the participants gave their written informed consent to participate in the experiment.

Procedure: Prior to the experiment, participants were asked to answer the trait version of the positive and negative affect scales (PANAS) [8]. Participants were seated approximately 50 cm from a screen in a shielded room and electrodes were attached to the head. In the experiment, each participant performed a two-choice decision-making task in which they had to guess whether the next card number would be higher or lower than the cue card number. Participants were informed that the card numbers ranged from 0 to 9. Each trial began with a fixation point (a plus sign; 2.29° high, 2.29° wide, black against a white background) presented for 500 ms, which was then replaced by the instructive stimulus (9.15° high, 18.18° wide, black against a white background). The instructive stimulus told the participants how much money (0, 10 or 50 yen; US$1 is equivalent to approximately 105 yen) they would receive for correct choices and lose for wrong choices in the trial. Presentation of the instructive stimulus for 500 ms was followed by a cue card (15.71° high, 13.46° wide, black against a white background). The face values of the cue cards were 4 and 5, where participants could not predict whether the response was correct or not. The participants were required to press the right button with the right index finger if they thought the next card would be higher than the cue card and to press the left button with the left index finger if they guessed the next card would be lower. The assignment of choices to buttons was counterbalanced across participants. The cue card remained on the screen until the participants responded. Three seconds after a button press, a feedback stimulus that indicated whether the preceding decision had been correct or incorrect was presented for 500 ms (i.e. '○' for correct or '×' for error; 5.72° high, 5.72° wide, black against a white background). As feedback negativity is insensitive to the physical nature of the eliciting stimulus [2], correct and error feedback stimuli were not counterbalanced across participants. The next trial began 1000 ms after the offset of feedback stimulus. Single-trial epochs were extracted offline for a period from 100 ms before until 600 ms after feedback stimuli.

There were two within-participant factors: (1) value of outcome (correct and error) and (2) reward or penalty magnitude of feedback stimuli (large, small and no reward/penalty). Under the high reward or penalty conditions, participants could receive 50 yen for a correct decision but lose it for an incorrect decision; under the small reward or penalty conditions, they could receive or lose 10 yen for correct or wrong decisions, respectively; under the no reward or penalty conditions, they neither received nor lost any money related to their choice. The task had a fixed outcome. Hence, participants received feedback indicating exactly 50 correct and 50 incorrect responses for each condition, although they were not informed of this rule. The entire experiment consisted of six blocks of 60 trials including 10 dummy trials (the face values of the cue cards were 2, 3, 6 and 7) for each block. The trials were presented in a randomized order within each block. All the participants began with a credit of 1000 yen and were instructed to earn more money through the experiment. Participants were given opportunities to rest between the blocks.

Recordings and analysis: An electroencephalogram (EEG) was recorded from 13 Ag/AgCl cup electrodes (10 mm in diameter) placed at Fp1, Fp2, F3, F4, C3, C4, P3, P4, T3, T4, Fz, Cz and Pz according to the International 10/20 system. However, only data from Fz, Cz and Pz were analyzed in the present study. All scalp electrodes were referenced to an average of the two ear lobes with a nasion ground. EEG impedances were kept below 5 kΩ. EEG signals were amplified using a Nihon-Kohden system with a band pass of 0.1–70 Hz. The signals were digitized with a sample rate of 200 Hz. After eye movement and eye blink artifacts were corrected using a spatial filtering method [9], ERPs were extracted by averaging EEG separately for participants according to each condition. Each participant had a minimum of 40 valid epochs per condition. The component of feedback negativity was defined by the difference in the ERP value between the most positive value within a 150–220 ms window following presentation of the feedback stimulus and the most negative value of the ERP within a window extending from the onset of the negativity to 325 ms following presentation of the feedback stimulus. In the present study, the P300 component was also measured as the average within 360–410 ms after the feedback stimulus, relative to the 100 ms prestimulus baseline. Repeated-measures analysis of variance (ANOVA) for

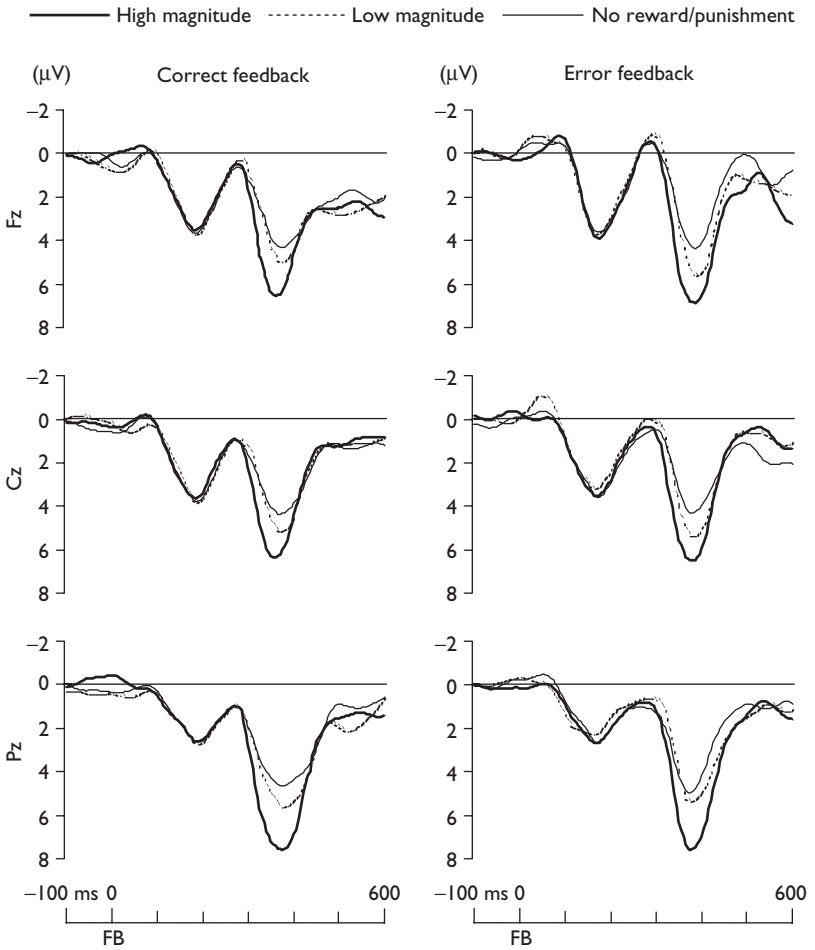

Fig. 1. Grand averages of electroencephalogram of all electrodes under the correct feedback condition (left side) and under the error feedback condition (right side), separately for magnitude. The time of feedback stimulus presentation is indicated by FB.

feedback negativity and P300 was performed with value of feedback (correct and error), reward/penalty magnitude of feedback stimuli (large, small and no reward/penalty) and location (Fz, Cz and Pz).

To control the increase in type I error, the degree of freedom was adjusted using the Greenhouse and Geisser [10] coefficient when appropriate. *Post hoc* tests of simple effects were performed using the Bonferroni correction with a significance level of $p<0.05$. In correlation analysis, the degree of association was measured by Pearson's correlation coefficient.

RESULTS

Feedback negativity: The grand averages for each electrode are shown in Fig. 1. Feedback negativity appeared as a peak at approximately 275 ms after the feedback stimulus under both correct and error feedback conditions. ANOVA conducted on feedback negativity amplitude revealed significant main effects of value [$F(1,17)=32.67$, $p<0.001$, $\varepsilon=1.00$, partial $\eta^2=0.79$] and location [$F(2,34)=52.19$, $p<0.001$, $\varepsilon=0.60$, partial $\eta^2=0.75$]. The main effect of magnitude was statistically insignificant [$F(2,34)=1.47$, ns, $\varepsilon=0.77$, partial $\eta^2=0.08$]. In addition, the interaction of value and location was significant [$F(2,34)=7.69$, $p<0.05$, $\varepsilon=0.62$, partial $\eta^2=0.31$]. None of the other interactions reached statistical significance. Post-hoc test further revealed that feedback negativity was consistently higher under the error feedback condition than under the correct feedback condition, which was prominent at Fz and Cz (see Table 1).

Feedback negativity and the dispositional positive and negative affect: Feedback negativity was consistently higher under the error feedback condition than under the correct feedback condition. To assess whether the differences in dispositional positive affect and negative affect might explain the disparity in the amplitudes of feedback negativity across values of feedback, correlation analysis was performed between the feedback negativity amplitude and the dispositional positive affect and negative affect. For this purpose, between-value differences in the amplitude of feedback negativity (error feedback minus correct feedback) were calculated separately for each participant. As the preceding analysis demonstrated that the main effect of magnitude on the feedback negativity was statistically insignificant, the difference measures of feedback negativity were averaged over magnitude. The scores of the trait positive affect and negative affect were calculated from

Table 1. Mean (±SD) amplitude of feedback negativity under each condition.

Location	Magnitude	Correct feedback	Error feedback
Fz	High	3.02 (1.63)	4.20 (1.17)
	Low	3.12 (1.96)	4.27 (1.34)
	No reward/penalty	2.89 (1.49)	4.09 (0.92)
Cz	High	2.65 (1.43)	3.37 (1.25)
	Low	2.64 (1.17)	3.38 (1.67)
	No reward/penalty	2.61 (0.79)	3.29 (1.02)
Pz	High	1.44 (1.34)	1.80 (0.94)
	Low	1.47 (1.05)	1.73 (1.62)
	No reward/penalty	1.48 (0.70)	1.75 (0.63)

Amplitudes are absolute values.

Table 2. Mean (±SD) amplitude of P300 under each condition.

Location	Magnitude	Correct feedback	Error feedback
Fz	High	6.63 (0.83)	6.83 (1.38)
	Low	4.97 (1.02)	5.43 (1.39)
	No reward/penalty	4.32 (0.84)	4.33 (1.70)
Cz	High	6.87 (2.02)	7.22 (1.43)
	Low	5.18 (0.95)	5.42 (1.10)
	No reward/penalty	4.35 (1.10)	4.35 (1.31)
Pz	High	7.59 (2.08)	7.60 (2.28)
	Low	5.55 (1.10)	5.38 (2.10)
	No reward/penalty	4.65 (1.35)	4.83 (1.79)

PANAS for each participant. In correlation analysis, a significant correlation was found in negative affect and between-value differences in the amplitude of feedback negativity at Fz (r=0.49, p<0.05). None of the other correlations reached significance.

P300 elicited by feedback stimuli: P300 appeared as a peak at approximately 380 ms after feedback stimulus (see Fig. 1). ANOVA conducted on P300 amplitude revealed significant main effects of magnitude [F(2,34)=138.88, p<0.001, ε=0.65, partial η^2=0.89] and location [F(2,34)=3.74, p<0.05, ε=0.80, partial η^2=0.18]. The main effect of value was statistically insignificant [F(1,17)=0.86, ns, ε=1.00, partial η^2=0.05]. None of the interactions reached significance. Post-hoc test further revealed that P300 was highest at Pz and lowest at Fz, and that P300 was enhanced when the magnitude was larger under both correct and error conditions (see Table 2).

P300 and the dispositional positive and negative affect: P300 was higher when the magnitude of feedback was larger. To assess whether the differences of positive affect and negative affect might explain the disparity in the amplitudes of P300 across magnitude of feedback, correlation analysis was performed between the P300 amplitude and positive affect or negative affect. For this purpose, between-magnitude differences in the amplitude of P300 (large minus small reward/penalty, small minus no reward/penalty and large minus no reward/penalty) were calculated separately for each participant. As the main effect of value on P300 was statistically insignificant, the difference measures of P300 were averaged over value. In correlation analysis, significant correlations were found between positive affect and large-minus-small differences in the amplitude of P300 at Pz (r=0.48, p<0.05) and between positive affect and large-minus-no differences in the amplitude of P300 at Pz (r=0.47, p<0.05). None of the other correlations reached significance.

DISCUSSION

In this study feedback negativity was consistently larger after incorrect feedback than after correct feedback, irrespective of reward magnitude. Feedback negativity was equally large for no, small and large monetary losses. These results suggest that feedback negativity reflects the classification of events as being good or bad, irrespective of the presence or absence of monetary reward. In contrast, the amplitude of P300 increased with the magnitude of reward or penalty but was unaffected by the value of outcome. As no, small and large outcomes were encountered with equal frequency, the observed P300 amplitude difference cannot be explained by a difference in the frequencies with which outcomes were encountered, and instead appears to be related specifically to the reward magnitude of the feedback stimuli.

In this study, a significant positive correlation was found between trait negative affect score and the amplitude of feedback negativity. In contrast, no significant correlation between trait positive affect score and the amplitude of feedback negativity was found. Given that feedback negativity reflects a discrete evaluation of events as good or bad, it follows that individuals with higher trait negative affect are more likely to assign a negative value to experienced events. In support of this suggestion, previous studies demonstrated that negative affect was related to the tendency to make negative judgments [7,11]. Taken together, these results suggest that the rostral part of the ACC, the most likely neural generator of feedback negativity [1], may be the neural basis of a negative interpretive bias observed in patients with anxiety disorder. Further studies are required to prove the validity of this suggestion.

Unpredictably, significant positive correlations were found between trait positive affect scores and the amplitudes of P300. That is, in individuals with higher trait positive affect score P300 was enhanced with increases in motivational significance. Given that higher trait positive affect is associated with heightened appetitive or incentive motivation [7], these results suggest that P300 amplitude varies with reward magnitude because of the increased motivational significance of greater rewards and penalties. Further studies are required to reveal the mechanism by which P300 varies with reward magnitude.

In conclusion, these results suggest that value and reward magnitude are processed separately in the brain, and that feedback negativity and P300 reflect evaluation of different aspects of motivational significance of outcome events; the former reflects evaluation of value and the latter reflects evaluation of magnitude.

REFERENCES

1. Gehring WJ, Willoughby AR. The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 2002; **295**:2279–2282.
2. Miltner WHR, Braun CH, Coles MGH. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for

a 'generic' neural system for error detection. *J Cogn Neurosci* 1997; **9**:788–798.
3. Nieuwenhuis S, Yeung N, Holroyd CB, Schurger A, Cohen JD. Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cereb Cortex* 2004; **14**:741–747.
4. Holroyd CB, Cole MGH. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev* 2002; **109**:679–709.
5. Yeung N, Holroyd CB, Cohen JD. ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb Cortex* (in press).
6. Yeung N, Sanfey AG. Independent coding of reward magnitude and valence in the human brain. *J Neurosci* 2004; **24**:6258–6264.
7. Mineka S, Watson D, Clark LA. Comorbidity of anxiety and unipolar mood disorders. *Annu Rev Psychol* 1998; **49**:377–412.
8. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol* 1988; **54**:1063–1070.
9. Ille N, Berg P, Scherg M. Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *J Clin Neurophysiol* 2002; **19**:113–124.
10. Greenhouse SW, Geisser S. On the methods in the analysis of profile data. *Psychometrika* 1959; **24**:95–112.
11. MacLeod AK, Byrne A. Anxiety, depression, and the anticipation of future positive and negative experiences. *J Abnorm Psychol* 1996; **105**:286–289.