Psychophysiology, 42 (2005), 161–170. Blackwell Publishing Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2005.00278.x

Brain potentials associated with expected and unexpected good and bad outcomes

GREG HAJCAK,[a] CLAY B. HOLROYD,[b] JASON S. MOSER,[a] AND ROBERT F. SIMONS[a]

[a]Department of Psychology, University of Delaware, Newark, Delaware, USA
[b]Department of Psychology, University of Victoria, Victoria, British Columbia, Canada

Abstract

The error-related negativity (ERN) is an event-related brain potential observed when subjects receive feedback indicating errors or monetary losses. Evidence suggests that the ERN is larger for unexpected negative feedback. The P300 has also been shown to be enhanced for unexpected feedback, but does not appear to be sensitive to feedback valence. The present study evaluated the role of expectations on the ERN and P300 in two experiments that manipulated the probability of negative feedback (25%, 50%, or 75%) on a trial-by-trial basis in experiment 1, and by varying the frequency of positive and negative feedback across blocks of trials in experiment 2. In both experiments, P300 amplitude was larger for unexpected feedback; however, the ERN was equally large for expected and unexpected negative feedback. These results are discussed in terms of the potential role of expectations in processing errors and negative feedback.

Descriptors: Expectations, Feedback, Event-related brain potential, Error-related negativity, Ne, P300, Reinforcement learning, Response monitoring

A number of recent event-related brain potential (ERP) studies have focused on neural activity related to errors and negative feedback. In particular, the response-locked ERP at fronto-central recording sites is characterized by a negative deflection that begins around the time of an erroneous response and peaks approximately 50 ms later (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Gehring, Coles, Meyer, & Donchin, 1990; Gehring, Goss, Coles, Meyer, & Donchin, 1993). This error-related negativity (ERN or Ne) has been observed across various stimulus and response modalities, and consequently appears to reflect the activity of a generic response monitoring system (Bernstein, Scheffers, & Coles, 1995; Falkenstein et al., 1991; Holroyd, Dien, & Coles, 1998; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). Studies utilizing whole-head ERP recording systems and source-localization software have consistently indicated that the ERN is generated by a single source in the anterior cingulate cortex (ACC; Dehaene, Posner, & Tucker, 1994; Holroyd et al., 1998).

Following early response-locked ERN studies, Miltner, Braun, and Coles (1997) reported an ERN-like component following the presentation of negative feedback (cf. Ruchsow, Grothe, Spitzer, & Kiefer, 2002; see Nieuwenhuis, Holroyd, Mol, & Coles, 2004, for a review). Miltner et al. found that when subjects received negative feedback regarding the accuracy of their performance, the ERP following negative feedback was characterized by a negative deflection at fronto-central recording sites with a peak latency of approximately 250 ms. Like the response ERN, this feedback ERN was source localized to near the ACC (Miltner et al., 1997). In considering the topographical and morphological similarity of this feedback ERN and the response ERN, Miltner et al. proposed that both ERNs might represent the activity of a single error detection system.

More recently, Holroyd and Coles (2002) argued that both the response ERN and feedback ERN reflect the activity of a reinforcement learning system that continually evaluates ongoing events against expected outcomes. This reinforcement learning theory of the ERN is predicated on previous research implicating the basal ganglia and the midbrain dopamine system in reward prediction and reinforcement learning. According to this previous research (Barto, 1995; Montague, Dayan, & Sejnowski, 1996; for review, see Schultz, 2002), the basal ganglia evaluate ongoing events and predict whether future events will be favorable or unfavorable. When the basal ganglia revise their predictions for the better or for the worse, they induce a phasic increase or decrease, respectively, in the activity of midbrain dopamine neurons. These phasic increases and decreases in dopamine activity indicate that ongoing events are "better than expected" or "worse than expected," respectively.

The reinforcement learning theory of the ERN extends this theoretical framework by proposing that the impact of the

This research was supported in part by National Institutes of Mental Health (NIMH) predoctoral fellowship MH069047 (G.H.), NIMH postdoctoral fellowship MH63550 (C.B.H.), and NIMH grant MH62196. Portions of this article were presented at the 44th annual meeting of the Society for Psychophysiological Research, Santa Fe, New Mexico, October, 2004.

Address reprint requests to: Greg Hajcak, Department of Psychology, University of Delaware, Newark, DE 19716, USA. E-mail: hajcak@psych.udel.edu.

dopamine signals on motor-related areas of the ACC modulates the amplitude of the ERN, such that phasic decreases in dopamine activity (indicating that ongoing events are worse than expected) are associated with large ERNs, and phasic increases in dopamine activity (indicating that ongoing events are better than expected) are associated with small ERNs (Holroyd & Coles, 2002; see also Holroyd, 2004). According to the theory, then, the response ERN and the feedback ERN are both elicited by the first (unpredicted) indication that an unfavorable event is occurring or is about to occur, the former by error responses and the latter by undesired outcomes. The theory further holds that the dopamine signals are used by the motor-related areas in the ACC to improve performance on the task at hand according to principles of reinforcement learning.

Although the reinforcement learning theory of the ERN proposes that the ERN is associated with events that are worse than expected, the theory is nonspecific about what actually constitutes an "unfavorable" outcome. Thus, for example, the theory does not distinguish between financial rewards and punishments on the one hand (as indicated by "utilitarian" feedback) and correct trials and error trials on the other (as indicated by "performance" feedback). Instead, the theory focuses on the reinforcing properties that these outcomes share in common (Holroyd, Coles, & Nieuwenhuis, 2002). For example, in a recent experiment Nieuwenhuis, Yeung, Holroyd, Schurger, and Cohen (2004) found that when feedback stimuli conveyed both utilitarian and performance information, the feedback ERN could reflect *either* dimension of the information, depending on which aspect of the feedback was emphasized. The authors argued that their findings indicated that "In the context of this [reinforcement learning] theory, utilitarian and performance aspects of feedback are functionally equivalent: both evaluate outcomes along a good–bad dimension, and hence both can elicit an ERN" (p. 745, cf. Gehring & Willoughby, 2002).

Recently, we have begun to explore the conditions by which external outcomes are categorized by the system as favorable or unfavorable. In a recent experiment, for example, Holroyd, Larsen, and Cohen (2004) found that the ERN was sensitive to the *relative* value of feedback. In this experiment, feedback indicating that participants received nothing generated a large ERN in a task context in which participants could win money, but that same feedback did not generate an ERN in a task context in which participants could lose money. Thus, the ERN was elicited by unfavorable outcomes, where the favorableness of each outcome was determined by the context in which the outcome was delivered.

In the present study we tested a core prediction of the theory: that the amplitude of the ERN is positively related to the size of the outcome prediction error, and thus depends on the difference between expected and actual outcomes. Evidence consistent with this prediction was previously obtained in the original reinforcement learning task adopted by Holroyd and Coles (2002; also see Nieuwenhuis et al., 2002) and in a subsequent guessing task (Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003). These studies found that unexpected negative feedback produced the largest ERNs. In the present study, we further tested this prediction by manipulating participants' expectancies on a trial-by-trial basis.

Specifically, participants performed a gambling experiment in which they selected between four response options and were presented with a feedback stimulus indicating that they either received a reward (positive feedback) or received nothing (negative feedback); however, before making the response the participants were presented with a cue indicating that they had either a 25%, 50%, or 75% chance of receiving a financial reward on that trial. We predicted that the cues would induce expectancies of future rewards, and that negative outcomes would elicit the largest feedback ERNs on those trials in which rewards were most expected.

To verify the success of this manipulation, we also measured the amplitude of the P300, a component of the ERP that is sensitive to expectancies (Courchesne, Hillyard, & Courchesne, 1977; Duncan-Johnson & Donchin, 1977; Johnson & Donchin, 1980). We reasoned that if strong expectancies were induced in the participants by the stimulus cues, then these expectancies would be reflected in the amplitude of the P300. Furthermore, based on a recent study by Yeung and Sanfey (2004), who found equally large P300s following equiprobable negative and positive feedback in a gambling experiment, we predicted that the P300 would not differ for positive and negative feedback.

EXPERIMENT 1

In this experiment, participants performed a guessing task similar to that used by Holroyd et al. (2003): Participants were instructed to guess which of four doors hid a prize; however, unlike the Holroyd et al. study, at the start of each trial participants received a cue indicating how many of the doors (1, 2, or 3) hid a prize. Following each response, participants were presented with feedback indicating whether or not their guess was accurate. Consistent with the cue, the probability of negative feedback was 75%, 50%, and 25%, respectively, and thus the overall probability of negative feedback was 50%. We reasoned that participants would tend to expect negative feedback on one-cue trials and expect positive feedback on three-cue trials. If the ERN is sensitive to expectations, as predicted by the reinforcement learning theory of the ERN, then negative feedback on three-cue trials should be associated with a larger ERN relative to negative feedback on one-cue trials and the ERN to negative feedback on two-cue trials should be intermediate. Additionally, we expected the P300 to be larger for unexpected outcomes; based on the recent results of Yeung and Sanfey (2004), we hypothesized that P300 amplitude would not differ for positive versus negative feedback.

Method

Participants

Eighteen undergraduate students (5 men) in an upper level psychology class at the University of Delaware participated in the current experiment for extra credit. Additionally, participants were told that they could earn between $0.00 and $24.00 in bonus money based on their performance. All participants were paid $12.00. Data from 1 participant were not included due to a technical malfunction.

Task

The task was administered on a Pentium I class computer, using Presentation software (Neurobehavioral Systems, Inc.) to control the presentation and timing of all stimuli. Throughout the task, participants were shown a graphic representing four doors in a horizontal line, and were instructed to guess which door hid a prize. Participants were instructed to press the left and right

"ctrl" and "alt" keys to select a door. Following each choice, a feedback stimulus appeared on the screen that informed subjects about the accuracy of their guess. A green "+" feedback indicated a correct guess, and a green "o" feedback indicated an incorrect guess. Prior to each trial, a white "1," "2," or "3" cue appeared on the screen to inform the subjects how many doors contained prizes. All stimuli were presented against a black background, and were positioned in the center of the screen. All cue and feedback stimuli occupied approximately 2° of visual angle horizontally, and 2° vertically. A fixation mark (+) was presented just prior to the onset of each stimulus.

In terms of stimulus timing, the cue remained on the screen for 1000 ms; the doors appeared immediately following cue offset, and remained on the screen until participants responded. Finally, the feedback appeared 500 ms following response, and remained on the screen for 1000 ms. The interval between offset of the feedback stimulus and the following cue was 1000 ms.

Unbeknownst to the participants, the veracity of each trial was predetermined and pseudorandom such that overall the participants received exactly 50% correct feedback; negative feedback was delivered on 25% of three-cue trials, 50% of two-cue trials, and 75% of one-cue trials.

Procedure

After a brief description of the experiment, EEG sensors were attached and the participant was given detailed task instructions. To become familiar with the task, participants were given a practice block consisting of 40 trials, and were instructed to guess which door hid a prize. Following the practice, participants were told that they would earn $0.10 for each correct guess. The actual experiment consisted of six blocks of 40 trials (240 total trials) with each block initiated by the participant. The experimenter entered the room every 80 trials to inform the participant how much money he or she had earned. Upon completion of the task, participants were asked to fill out a brief questionnaire.

Psychophysiological Recording, Data Reduction, and Analysis

The electroencephalogram (EEG) was recorded using a Neurosoft Quik-Cap. Recordings were taken from three locations along the midline: frontal (Fz), central (Cz), and parietal (Pz). In addition, Med-Associates tin electrodes were placed on the left and right mastoids (A1 and A2, respectively). During the recording, all activity was referenced to Cz. The electrooculogram (EOG) generated from blinks and vertical eye movements was also recorded using Med-Associates miniature electrodes placed approximately 1 cm above and below the participant's right eye. The right earlobe served as a ground site. All EEG/EOG electrode impedances were below 10 KΩ and the data from all channels were recorded by a Grass Model 7D polygraph with Grass Model 7P1F preamplifiers (bandpass = 0.05–35 Hz).

All bioelectric signals were digitized on a laboratory microcomputer using VPM software (Cook, 1999). The EEG was sampled at 200 Hz. Data collection began with the participants' response (500 ms prior to feedback) and continued for 1500 ms. Off-line, the EEG for each trial was corrected for vertical EOG artifacts using the method developed by Gratton, Coles, and Donchin (1983; Miller, Gratton, & Yee, 1988) and then re-referenced to the average activity of the mastoid electrodes. Trials were rejected and not counted in subsequent analysis if there was excessive physiological artifact (i.e., 25 ms of invariant analog data on any channel or A/D values on any channel that equaled that converters minimum or maximum values). Single-trial EEG data were lowpass filtered at 20 Hz with a FIR digital filter as per Cook and Miller (1992).

Finally, stimulus-locked ERPs were averaged based on expectancy and feedback valence. Specifically, averages were computed for each of six feedback types: expected, neutral, and unexpected negative feedback stimuli and expected, neutral, and unexpected positive feedback stimuli.

The ERN was quantified at Fz, Cz, and Pz as follows. First, each data point after feedback onset was subtracted from a baseline equal to the average activity in a 200-ms window prior to the feedback. Next, a difference wave was created by subtracting the ERP observed for positive feedback from the ERP observed for negative feedback; this was done for expected outcomes (negative feedback on a one-cue trial minus positive feedback on a three-cue trial), neutral outcomes (negative feedback on a two-cue trial minus positive feedback on a two-cue trial), and unexpected outcomes (negative feedback on a three-cue trial minus positive feedback on a one-cue trial). ERNs for each level of expectancy were then defined as the maximum negative amplitude of these difference waves within a window between 0.2 and 0.5 s following feedback. This procedure controlled for the main effect of stimulus frequency on the ERP, ensuring that the ERP measure was sensitive to the interaction of feedback frequency and valence (Holroyd, 2004).

The P300 was also evaluated for each cue and outcome type at Pz, where it was maximal. The P300 was defined as the most positive point in the ERP 200 to 600 ms following feedback onset. The ERN and P300 were statistically evaluated using SPSS (version 10.1) General Linear Model software with the Greenhouse–Geisser correction applied to p values associated with multiple df repeated-measures comparisons.

Results

Behavioral Results

The feedback was presented in a pseudorandom order such that performance had no relationship to feedback. However, participants were asked to complete a post-task questionnaire in which they rated how much attention they paid to both the cue and the feedback stimuli on a scale from 1 (*the stimuli were ignored*) to 7 (*paid close attention*). On average, participants rated their attention to cue and feedback stimuli as 5.69 ($SD = 1.14$) and 5.50 ($SD = 1.32$), respectively. These self-report data suggest that participants were engaged in the task and paid attention to both the cue and feedback stimuli (for similar self-report results, see Holroyd, Larsen, et al., 2004). All participants reported that they believed they performed better on the three-cue trials than on the one-cue trials, suggesting that all participants were aware of the reward contingencies and paid attention to both the cue and feedback.

The Feedback ERN

Figure 1 presents stimulus-locked average ERPs for positive and negative feedback for expected (left), neutral (middle), and unexpected (right) outcomes at Fz (top), Cz (middle), and Pz (bottom). Consistent with previous studies, negative feedback was associated with a frontally maximal negative deflection that peaked approximately 300 ms following feedback. Figure 2 presents the difference wave obtained by subtracting positive from negative feedback for all levels of expectancy (expected negative minus expected positive, neutral negative minus neutral positive, and unexpected negative minus unexpected positive) at

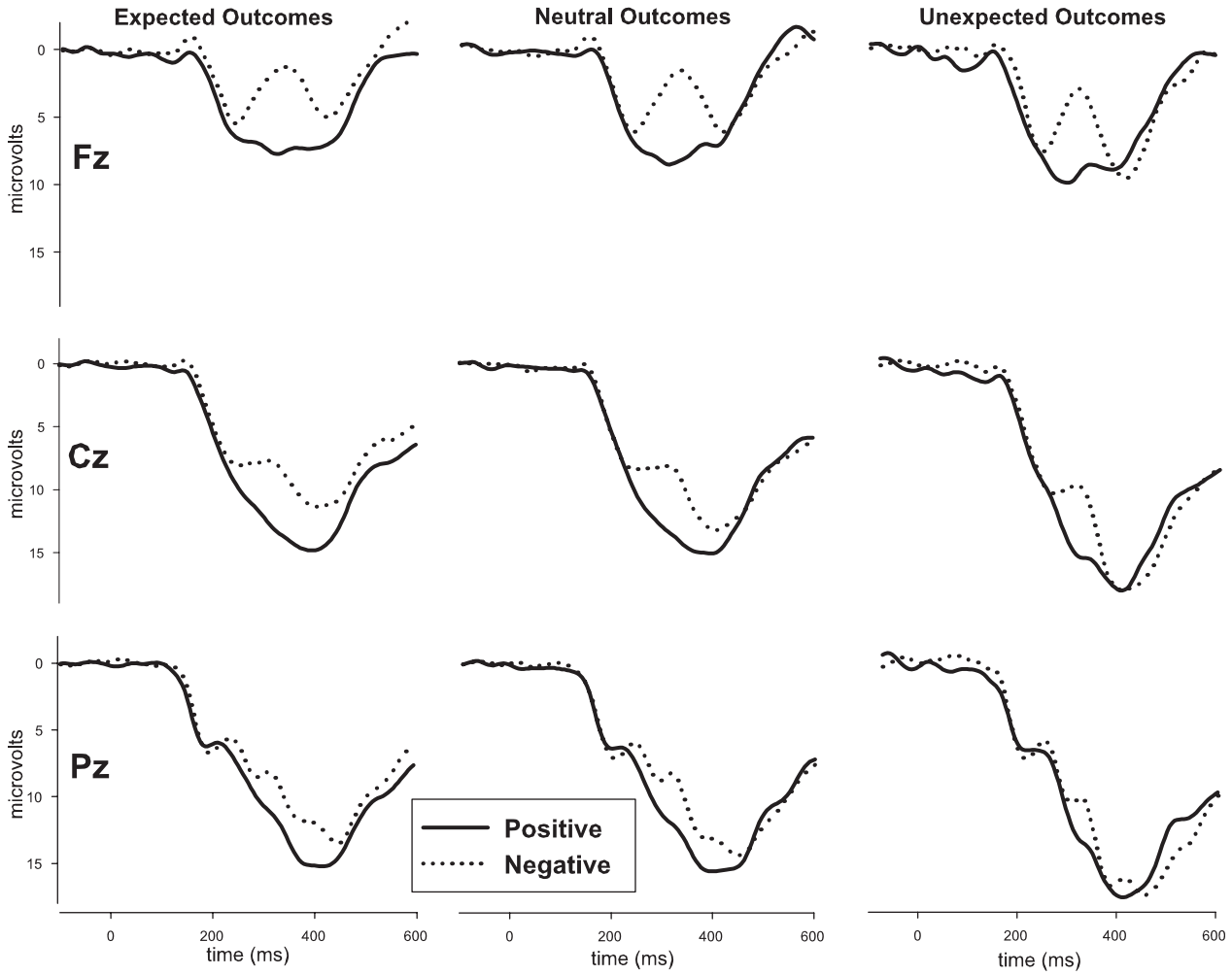

Figure 1. ERPs for expected (left), neutral (middle), and unexpected (right) positive and negative feedback at Fz (top), Cz (middle), and Pz (bottom) in Experiment 1. Feedback onset occurred at 0 ms.

Fz (top), Cz (middle), and Pz (bottom), and Table 1 presents the average ERN amplitudes. From both Figures 1 and 2, it is apparent that the feedback ERN was evident for negative feedback, regardless of expectancy. A 3 (location) × 3 (expectancy) repeated-measures ANOVA of the feedback ERN confirmed the impression from Figure 2 that the feedback ERN was largest at frontal recording sites, $F(2,32) = 17.41$, $p < .001$, $\varepsilon = .83$; however, the feedback ERN did not differ as a function of expectancy, $F(2,32) < 1$, and there was no interaction between location and expectancy, $F(4,64) < 1$. Consistent with the fronto-central maximum reported in previous studies, post hoc tests indicated that the ERN was larger at Fz and Cz than at Pz, $t(17) = 6.19$, $p < .001$ and $t(17) = 5.02$, $p < .001$, respectively, but the ERN was equally large at Fz and Cz, $t(17) = 1.21$, $p > .20$.[1]

P300

The mean P300 amplitudes for expected, neutral, and unexpected feedback at Pz are presented in Table 1. A 2 (feedback valence) × 3 (expectancy) repeated-measures ANOVA confirmed that P300 amplitude varied with respect to expectancy, $F(2,32) = 45.48$, $p < .001$, $\varepsilon = .82$, but did not differ overall with respect to the feedback valence, $F(1,16) < 1$. Thus, the P300 was, in fact, larger for more unexpected outcomes. Finally, the interaction between feedback valence and expectancy, $F(2,32) = 2.88$, $p > .09$, did not reach significance. Post hoc analyses confirmed that unexpected outcomes had larger P300s than both neutral, $t(17) = 5.94$, $p < .001$, and expected outcomes, $t(17) = 8.71$, $p < .001$; additionally, neutral outcomes had larger P300s than expected outcomes, $t(17) = 3.09$, $p < .01$.

Discussion

In the present study, the P300 was larger for unexpected outcomes, and self-report data confirmed that participants were aware of the reward contingencies in the present study. Specifically, all participants reported performing best on three-cue (high reward probability) trials and worst on one-cue (low reward probability) trials. Taken together, these data indicate that the expectancy manipulation in the present study was successful—and influenced both self-report and ERP data. In addition, negative feedback in the present study was uniformly associated with a negative deflection at the frontal recording site that peaked approximately 300 ms after feedback onset. Because this negativity was virtually absent on trials with positive feedback, its

[1]Even when the ERN was evaluated where it was maximal (Fz and Cz), as in Holroyd et al. (2003), the effect of expectancy did not reach significance (at both Fz and Cz: $F[2,32] < 1$).

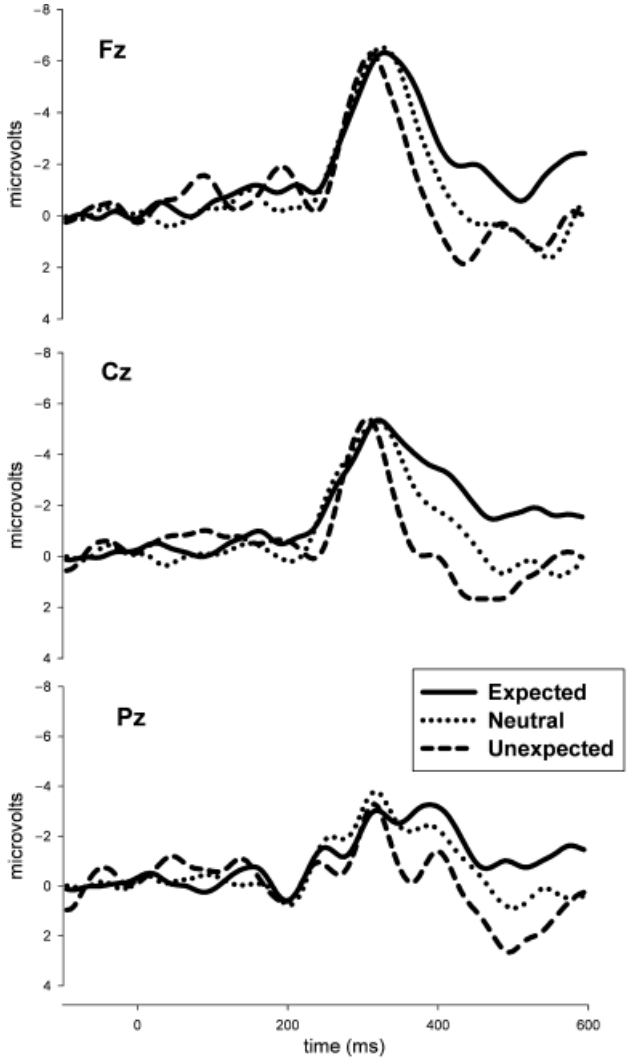

Figure 2. Negative minus positive difference waves for expected, neutral, and unexpected outcomes at Fz (top), Cz (middle), and Pz (bottom) from Experiment 1. Feedback onset occurred at 0 ms.

morphology, topography, and functional role are consistent with previous reports on the feedback ERN. However, unexpected feedback did not elicit larger magnitude ERNs, indicating that the amplitude of the feedback ERN was insensitive to expectations regarding negative feedback.

The present study failed to find an effect of expectations on the amplitude of the feedback ERN when expectations were manipulated on a trial-by-trial basis with a cue that indicated the conditional probability of positive and negative feedback. These results contrast with those reported by Holroyd et al. (2003), who found a larger ERN for unexpected negative feedback, when expectations were manipulated by the frequency of positive and negative feedback. One possible explanation for the difference between the present study and the Holroyd et al. study is the method used to manipulate expectations of positive and negative outcomes. Because Holroyd et al. (2003) found that infrequent negative feedback was related to a larger ERN than frequent negative feedback, we reasoned that a frequency manipulation may engender stronger expectations regarding feedback. Accordingly, we further investigated the role of expectancy on the ERN by manipulating expectations in a way similar to that used by Holroyd et al. in Experiment 2 by varying the frequency of negative feedback (25%, 50%, and 75%) between experimental blocks.

EXPERIMENT 2

In this experiment, we induced expectancies of reward by manipulating the frequency of positive and negative feedback. On each trial of the task, participants selected between four "balloons" that appeared on a computer screen and were presented with a feedback stimulus indicating that they either received a reward or received nothing. Unbeknownst to the participants, this feedback was delivered according to a random schedule that varied by condition: Positive feedback was presented at random on 25%, 50%, and 75% of the trials in each of three different blocks. This design was similar to the Holroyd et al. (2003) experiment but included a neutral (50%) condition in addition to the high (75%) and low (25%) reward conditions to mirror the levels of expectation from Experiment 1. As in Experiment 1, we predicted that the amplitude of the ERN would increase in proportion to the probability of receiving positive feedback. We predicted that the participants would develop expectancies of future reward based on the frequency of reward delivered in each condition, and that negative outcomes would elicit the largest feedback ERNs in those conditions in which the rewards were most frequent. Additionally, we expected that P300 amplitude would be larger for more infrequent feedback.

Note that the critical difference between Experiments 1 and 2 is that, in the first experiment, the expectancies were induced on a trial-by-trial basis by the predictive cues, whereas in the second

Table 1. *Mean (M) and Standard Deviations (SD) for ERN Magnitudes at Fz, Cz, and Pz and P300 Magnitudes at Pz in Studies 1 and 2*

	Study 1						Study 2					
	Expected		Neutral		Unexpected		Frequent		Neutral		Infrequent	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ERN												
Fz	− 7.29	2.68	− 7.16	2.62	− 7.37	5.03	− 8.63	4.74	− 10.77	5.62	− 8.82	4.54
Cz	− 6.65	3.50	− 6.66	3.84	− 6.99	4.83	− 10.13	5.22	− 12.65	5.95	− 10.66	7.02
Pz	− 5.07	2.94	− 5.57	3.09	− 5.30	4.10	− 8.95	5.16	− 10.99	7.29	− 9.98	5.90
P300												
Reward	16.53	5.40	17.50	5.81	19.11	5.97	18.13	6.42	21.83	7.86	21.88	7.07
Nonreward	15.45	4.64	16.25	5.16	19.91	5.79	12.38	3.97	16.27	5.80	17.17	5.34

experiment the expectancies were induced by exposure to sequences of positive and negative feedback throughout each block of trials.

Method

Participants

Twelve undergraduate students (3 men) at Princeton University participated for pay ($20.00) or class credit. In addition, all participants earned a small monetary bonus (about $7.00), as described below.

Task

Participants sat comfortably about 1.5 m in front of a computer screen in an electromagnetically shielded room. On each trial of the task, participants saw an imperative stimulus (0.6° high, 5.0° wide, blue color against a black background) consisting of four circles in a row (''O O O O''). Participants were asked to imagine that these circles were balloons, and were told that one of the balloons contained 3 cents and the others were empty. The imperative stimulus remained on the screen until the participant selected a balloon by pressing one of four buttons on a response pad. At the time of the response, the imperative stimulus was replaced by a second stimulus (0.6° high, 5.0° wide, blue color against a black background, 1 s duration) in which the selected balloon was replaced by an asterisk (e.g., ''* O O O'' if the subject selected the leftmost balloon). The purpose of the asterisk was to indicate to the participant which balloon they had selected. Following the offset of the second stimulus, a feedback stimulus appeared (0.6° high, red color, 1 s duration) directly above the center of the previous stimulus. The interstimulus interval (ISI) between the offset of the feedback stimulus and the onset of the imperative stimulus was 0.5 s. The feedback stimuli were the symbols ''+'' and ''o.'' Participants were told that presentation of ''+'' (positive feedback) and ''o'' (negative feedback) stimuli indicated that the balloon they chose on that trial contained 3 cents or 0 cents, respectively. Participants were told that they should try to maximize the total amount of money earned, and that they would receive that money at the end of the experiment. Unbeknownst to the participants, the feedback was random. Participants engaged in three blocks of 160 trials each; in each block, positive feedback was delivered (at random, with replacement) on either 25%, 50%, or 75% of the trials on that block. The order of the blocks was systematically varied across participants for counterbalancing.[2] At the end of each block, participants were informed the total amount of money they earned during that block; across the experiment, participants earned a total of about $7.00 in bonus money.

Data Acquisition

An electrode cap with Ag/AgCl electrodes was applied to each participant. The EEG was recorded along the midline according to the 10–20 system (Jasper, 1958) from channels FPz, AFz, Fz, FCz, Cz, CPz, Pz, POz, Oz, and Iz. Other electrodes were placed on the right mastoid, above and below the right eye, and on the outer canthi of both eyes. The ground electrode was placed on the chin or on the cheek. All electrode recordings were referenced to an electrode placed on the left mastoid. EEG data were recorded with Sensorium Inc. (Charlotte, VT) EPA-6 128-Channel Electrophysiology Amplifiers at a sample rate of 250 Hz. Impedances were less than 40 KΩ. Experimental control and data acquisition were controlled by E-Prime (Psychology Software Tools, Inc., Pittsburgh, PA) and Cogniscan (Newfoundland, NJ), respectively. Participants answered a short questionnaire upon completion of the experiment.

Data Analysis

For each feedback stimulus, a 1-s epoch of data (200 ms baseline) was extracted from the continuous data file for analysis. Ocular artifact was corrected with an eye-movement correction algorithm (Gratton et al., 1983). The EEG data were re-referenced off-line to average mastoid electrodes by subtracting, from each sample of data recorded at each channel, one-half of the activity recorded from the right mastoid. The re-referenced data were then baseline corrected by subtracting, from each sample of data recorded at each channel, the average activity of that channel during the baseline period. Single-trial EEG data were lowpass filtered below 20 Hz with the Interactive Data Language (Research Systems, Inc., Boulder, CO) digital filter algorithm. ERPs were created for each participant by averaging the single-trial EEG according to feedback type and condition.

The ERN was evaluated in the same manner as described in Experiment 1 at Fz, Cz, and Pz as the peak of the difference waves for all levels of feedback frequency (frequent negative minus frequent positive, neutral negative minus neutral positive, and infrequent negative minus infrequent positive). The P300 amplitude was quantified at channel Pz, where it was maximal, and was evaluated in the same manner described in Experiment 1.

The data were submitted to ANOVA with repeated measures. The Greenhouse–Geisser correction for repeated measure was applied where appropriate.

Results

Behavior

Because the feedback in this experiment was delivered randomly, the experiment did not provide a meaningful behavioral measure. However, upon completion of the experiment, participants were asked to evaluate their interest in the task. On a scale from 1 (*the feedback stimuli were generally ignored*) to 5 (*the feedback stimuli were evaluated closely; participants were interested whether or not they won or lost money on each trial*), participants rated their interest in the feedback as 4.3 ± 0.7, suggesting that they attended to the task. In addition, debriefing revealed that many subjects believed that they exercised some degree of control over the feedback, suggesting that they tried to use the feedback to guide their behavior (cf. Holroyd et al., 2004).

The feedback ERN

Figure 3 presents the ERPs for positive and negative feedback for the frequent (left), neutral (middle), and infrequent (right) outcomes at Fz (top), Cz (middle), and Pz (bottom). The feedback ERN is characterized by the negative deflection that peaked about 300 ms following negative feedback. Figure 4 presents the negative feedback minus positive feedback difference waves for all levels of frequency at Fz (top), Cz (middle), and Pz (bottom) and Table 1 presents the average ERN amplitudes. Inspection of

[2]As a control (see Holroyd, 2004), participants also engaged in an ''oddball'' task in which they counted the occurrence of a target stimulus that appeared on 10% of the trials (stimuli: ''+'' and ''o,'' 1 s duration, target type counterbalanced within subject; intertrial stimulus interval = 0.5 s). The oddball task consisted of two blocks of 200 trials each. The order of tasks was counterbalanced across participants, such that half of the participants performed the oddball task before the reward expectancy task.

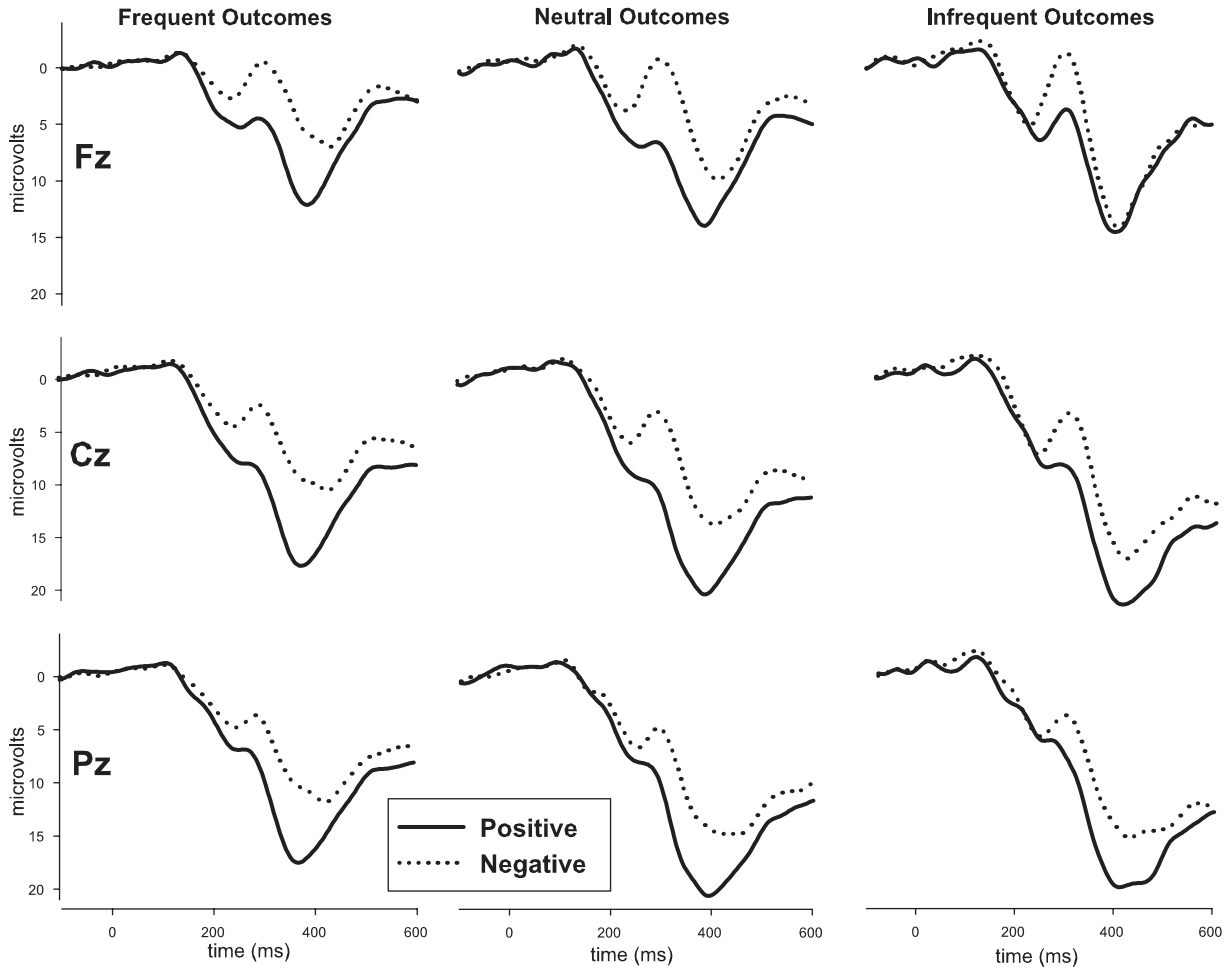

Figure 3. ERPs for frequent (left), neutral (middle), and infrequent (right) positive and negative feedback at Fz (top), Cz (middle), and Pz (bottom) in Experiment 2. Feedback onset occurred at 0 ms.

Figures 3 and 4 suggest that the ERPs associated with positive and negative feedback were about equally different from one another irrespective of frequency. A 3 (electrode site) × 3 (frequency) repeated-measures ANOVA revealed a main effect of electrode site, $F(2,22) = 4.17$, $p < 0.05$, $\varepsilon = .79$, but no main effect of frequency, $F(2,22) < 1$, nor an Frequency × Electrode interaction, $F(4,44) < 1$.[3] Consistent with a fronto-central maximum, post hoc tests indicated that the ERN was larger at Cz than both Fz and Pz, $t(11) = 2.98$, $p < .05$ and $t(11) = 2.43$, $p < .05$, respectively, and that the ERN did not differ at the Fz and Pz recording sites, $t(11) < 1$.[4]

P300

The mean P300 amplitudes for frequent, neutral, and infrequent feedback are presented in Table 1. Inspection of these data and ERP waveforms in Figure 3 suggests that, consistent with previous results, P300 amplitude was larger for infrequent outcomes than for frequent outcomes (Donchin & Coles, 1988). This observation was confirmed by a two-way repeated-measures ANOVA on feedback valence (positive, negative) and frequency (frequent, neutral, infrequent), which indicated that P300 amplitude varied with respect to frequency, $F(2,22) = 9.46$, $p < .005$, $\varepsilon = .84$. Post hoc paired-sample t tests indicated that both neutral and infrequent outcomes generated larger P300s than frequent outcomes, $t(11) = 3.61$, $p < .01$ and $t(11) = 4.91$, $p < .001$, respectively; P300s did not differ for neutral compared to infrequent outcomes, $t(11) < 1$. Furthermore, a main effect of valence indicated that P300 was larger to positive feedback than to negative feedback, $F(1,11) = 22.11$, $p < .001$, $\varepsilon = 1.00$. In contrast, there was no interaction of valence and frequency, $F(2,22) < 1.0$.

Discussion

In contrast to a previous report that found an enhanced ERN for infrequent negative outcomes (Holroyd et al., 2003), these results failed to show evidence that the ERN amplitude is sensitive to expectancy. That is, processing associated with feedback valence (as measured by the maximal difference between ERPs associated with positive and negative feedback) was not modulated by frequency of occurrence. In contrast, P300 amplitude revealed main effects of both valence and expectancy, but no interaction. These results are discussed further below.

[3]It is possible that these effects were not significant because the participants only developed frequency-based expectations toward the end of each block. To address this possibility, we analyzed ERNs from the second half of each block; this analysis indicated no main effect of frequency, $F(2,22) < 1$, nor a Frequency × Electrode interaction, $F(4,44) < 1$.

[4]We also evaluated the ERN where it was maximal (Cz), as in Holroyd et al. (2003); the effect of expectancy did not reach significance, $F(2,22) < 1$.

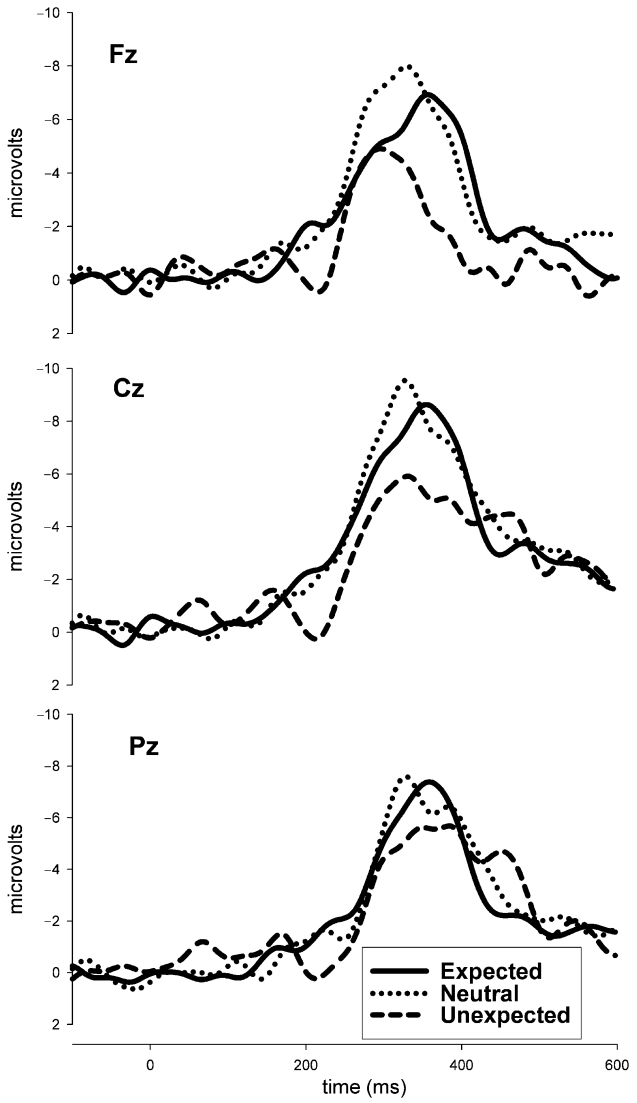

Figure 4. Negative minus positive difference waves for frequent, neutral, and infrequent outcomes at Fz (top), Cz (middle), and Pz (bottom) from Experiment 2. Feedback onset occurred at 0 ms.

General Discussion

The effect of expectancy on feedback processing was evaluated in two experiments. First, expectancies were induced trial by trial by cues that indicated the probability of positive feedback on each trial. Second, the expectancies were induced by varying the frequency of occurrence of positive feedback across blocks of trials. In both experiments, the P300 amplitude was largest for the unexpected outcomes, confirming that participants indeed formed expectations regarding feedback. Contrary to our prediction, however, the results of both experiments indicated that the amplitude of the feedback ERN was insensitive to expectancy.

It is important to note that there were several differences between the results of Experiments 1 and 2. Although the scalp topography of the ERN in both Experiments 1 and 2 are consistent with previous studies that report a fronto-central maximum, the difference wave was clearly fronto-centrally maximal in Experiment 1, but was centrally maximal in Experiment 2. These results may have been due to sequence and timing differences between the experiments. Specifically, in Experiment 1 a 500-ms delay occurred between response and feedback, whereas in Experiment 2 an intervening stimulus occurred for 1 s between the response and feedback. These differences may have contributed to the topographical differences in the scalp distribution of the ERN between the experiments. However, despite large differences in the implementation of the two experiments—including the location where the experiments were carried out, the equipment used to conduct the experiments, and the task itself—the two experiments converged on the same essential result: Unexpected negative feedback was not associated with an enhanced ERN. This convergence provides an indication of the robustness of these results.

These results contrast with findings from previous investigations from trial-and-error learning experiments that demonstrate an enhanced ERN for unexpected negative feedback (Holroyd & Coles, 2002; Nieuwenhuis et al., 2002). In these studies, subjects learned stimulus–response mappings based on feedback; however, subjects occasionally received negative feedback that was inconsistent with the learned stimulus–response mapping, and this type of unexpected negative feedback generated an enhanced ERN. Our findings can be reconciled with these observations if one assumes that the sensitivity of the system that generates the ERN to expectations is highly nonlinear, such that only extreme violations of expectancy (between 75% and 100%) exercise the amplitude of the ERN.

Although both trial-and-error learning tasks and gambling tasks involve feedback that indicates positive and negative outcomes, these tasks appear to differ insofar as subjects cannot learn systematically in gambling tasks because feedback is delivered randomly. Nevertheless, from an operant perspective, even a gambling task is a trial-and-error learning task as participants do not know a priori that feedback is not useful. In fact, many participants reported both using the feedback to inform their decisions and finding patterns in the feedback.

The present studies also contrast with the results of Holroyd et al. (2003), who found that infrequent negative feedback in a gambling task elicited a larger ERN than frequent negative feedback. Importantly, the Holroyd et al. study was analogous to the present Experiment 2, but without the neutral (50%) condition. Again, it is possible that the feedback ERN is most sensitive to probabilities greater than 75%. Although the feedback ERN would also be sensitive to intermediate probabilities, differences in feedback ERN amplitude would be more difficult to discriminate. If such were the case, then the results of the Holroyd et al. experiment might be due to a statistical anomaly.

Somewhat surprisingly, we note that in Experiment 2 P300 amplitude was larger to positive feedback than to negative feedback, irrespective of feedback expectancy (Figure 3, bottom). A similar result was found in the study by Holroyd et al. (2004; see their Figures 3a, 3b), although the finding was not discussed in that paper. Taken together, these results suggest that P300 amplitude is larger for positive outcomes than for negative outcomes (cf. Johnson & Donchin, 1985). These results contrast with recent research suggesting that the P300 is insensitive to feedback valence (Yeung & Sanfey, 2004) and with the results of our Experiment 1 in which a main effect of valence on P300 amplitude was not observed. Likewise, these results appear to contradict a proposal by Ito, Larsen, Smith, and Cacioppo (1998; Ito & Cacioppo, 2000) that unfavorable events elicit larger P300s than favorable events because of a "negativity bias" (however, see Johnson & Donchin, 1985, who report larger P300 amplitudes to positive feedback).

These differing results on the P300 may have to do with subjective expectations regarding the frequency of positive and negative feedback. For instance, subjects may believe that positive feedback is more likely overall than negative feedback—even when the objective frequency of feedback is equal (e.g., in Experiment 1). If this were the case, the P300 to negative feedback may be enhanced because it is perceived as more infrequent. That is, an enhanced P300 to positive feedback may be obscured by the effects of expectations on the P300, if negative feedback is perceived as relatively infrequent.

Given that the feedback ERN was defined in this experiment as the maximum amplitude of the difference wave associated with positive and negative feedback, it might be asked whether or not our measure was sensitive to component overlap with the P300 rather than to the feedback ERN itself. In this regard we can make several observations. First, the size of the difference was largest at frontal (Experiment 1) and central (Experiment 2) scalp locations, suggesting that the variation in the component's amplitude was due to something other than P300, which is largest over parietal cortex (Donchin & Coles, 1988). Second, the reinforcement learning theory of the ERN is nonspecific about whether or not unpredicted positive events induce a positive-going deflection of the ERP (Holroyd, 2004). For this reason it is important to evaluate ERPs associated with both positive and negative feedback when testing the theory. Third, the difference measure is an appropriate method to isolate the effects of valence on the ERP from other effects that are purely frequency related. For instance, it is apparent in both Figures 1 and 3 that positive feedback on unexpected trials was also characterized by a slightly enhanced negativity in the time range of the ERN (cf. Holroyd, 2004). Accordingly, the difference wave measure is sensitive to variance in the activation of the cognitive process of interest (valence processing), rather than to particular deflections in the ERP that may or may not be correlated with that process.

In conclusion, this study does not provide support for the reinforcement learning theory of the ERN, which predicts that variation in the amplitude of the feedback ERN should be larger for unpredicted outcomes than for predicted outcomes. It is unclear why these results contrast with prior findings, particularly with those of the study by Holroyd et al. (2003). It is possible that the probabilities in the present study were not extreme enough to exercise the amplitude of the ERN, despite the observed sensitivity of the P300 to those probabilities. On the other hand, it must be admitted that the present range of probabilities (25%–75%) seems rather large; if the system that produces the ERN is mainly sensitive to more extreme probabilities, then it must be highly nonlinear. A future study could examine this question by using probabilities that induce even stronger expectations.

REFERENCES

Barto, A. G. (1995). Adaptive critics and the basal ganglia. In J. Houk, J. Davis, & D. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 215–232). Cambridge, MA: MIT Press.

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. H. (1995). "Where did I go wrong?" A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1312–1322.

Cook, E. W. III (1999). *VPM reference manual*. Birmingham, AL: Author.

Cook, E. W. III, & Miller, G. A. (1992). Digital Filtering—Background and tutorial for psychophysiologists. *Psychophsyiology*, *29*, 350–367.

Courchesne, E., Hillyard, S. A., & Corchesne, R. Y. (1977). P3 waves to the discrimination of targets in homogenous and heterogeneous stimulus sequences. *Psychophysiology*, *14*, 590–597.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science*, *5*, 303–305.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 355–372.

Duncan-Johnson, C. C., & Donchin, E. (1980). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophsyiology*, *14*, 456–467.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of cross-modal divided attention on late ERP components: II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*, 447–455.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology*, *51*, 87–107.

Gehring, W. J., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1990). The error-related negativity: An event-related brain potential accompanying errors. *Psychophysiology*, *27*, S34.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–390.

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

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.

Holroyd, C. B. (2004). A note on the oddball N200 and feedback ERN. In M. Ullsperger & M. Falkenstein (Eds.), *Errors, conflicts, and the brain: Current opinions on response monitoring*. Leipzig: MPI of Cognitive Neuroscience.

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

Holroyd, C. B., Coles, M. G. H., & Nieuwenhuis, S. (2002). Medial prefrontal cortex and error potentials. *Science*, *296*, 1610–1611.

Holroyd, C. B., Dien, J., & Coles, M. G. H. (1998). Error-related scalp potentials elicited by hand and foot movements: Evidence for an output-independent error-processing system in humans. *Neuroscience Letters*, *242*, 65–68.

Holroyd, C. B., Larsen, J. T., & Cohen, J. D. (2004). Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology*, *41*, 245–253.

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *NeuroReport*, *14*, 2481–2484.

Ito, T. A., & Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology*, *36*, 660–676.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality & Social Psychology*, *75*, 887–900.

Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Johnson, R., & Donchin, E. (1980). P300 and stimulus categorization: Two plus one is not so different from one plus one. *Psychophysiology*, *17*, 167–178.

Johnson, R., & Donchin, E. (1985). Second thoughts: Multiple P300s elicited by a single stimulus. *Psychophysiology*, *22*, 182–194.

Miller, G. A., Gratton, G., & Yee, C. M. (1988). Generalized implementation of an eye movement correction procedure. *Psychophysiology*, *25*, 241–243.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (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*, 788–798.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*, 1936–1947.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reward-related brain potentials from medial frontal cortex: A review. *Neuroscience & Biobehavioral Reviews*, *28*, 441–448.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*, 752–760.

Nieuwenhuis, S., Ridderinkhof, K. R., Talsma, D., Coles, M. G. H., Holroyd, C. B., & Kok, A., et al. (2002). A computational account of altered error processing in older age: Dopamine and the error-related negativity. *Cognitive, Affective, and Behavioral Neuroscience*, *2*, 19–36.

Nieuwenhuis, S., Yeung, N., Holroyd, C. B., Schurger, A., & Cohen, J. D. (2004). Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cerebral Cortex*, *14*, 741–747.

Ruchsow, M., Grothe, J., Spitzer, M., & Kiefer, M. (2002). Human anterior cingulate cortex is activated by negative feedback: Evidence from event-related brain potentials in a guessing task. *Neuroscience Letters*, *325*, 203–206.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*, 241–263.

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

(Received May 24, 2004; Accepted December 20, 2004)