*Psychophysiology, 46* (2009), 313–320. Wiley Periodicals, Inc. Printed in the USA.
Copyright © 2009 Society for Psychophysiological Research
DOI: 10.1111/j.1469-8986.2008.00760.x

# The neural consequences of flip-flopping: The feedback-related negativity and salience of reward prediction

JASON S. MOSER AND ROBERT F. SIMONS
Department of Psychology, University of Delaware, Newark, Delaware, USA

**Abstract**

Research suggests that the feedback-related negativity (FRN) is only sensitive to expectancy when the prediction violation is salient. To further examine this issue, we asked participants to guess which of two virtual doors hid 5 cents. Prior to *and* after making their guess, participants were asked whether or not they expected to win. We extracted four conditions based on subjects' predictions before and after their choice of door: (1) win/win, (2) lose/lose, (3) win/lose, and (4) lose/win. Results revealed that the FRN was largest on trials in which subjects predicted they *were not* going to win and then changed their minds and predicted that they *were* going to win (i.e., the lose/win condition). Together with previous FRN results and findings in the area of regret and decision making, we suggest that the FRN reflects a context-sensitive signal that integrates information about current and past actions, thoughts, and emotions.

**Descriptors:** Feedback, Predictions, ERP, FRN, Emotion

The feedback related negativity (FRN) is a medial frontal negative deflection of the visually evoked event-related brain potential (ERP) peaking approximately 250 ms following feedback, indicating negative compared to positive performance feedback and monetary loss or nonreward compared to reward (Gehring & Willoughby, 2002; Hajcak, Holroyd, Moser, & Simons, 2005; Hajcak, Moser, Holroyd, & Simons, 2006; Holroyd & Coles, 2002; Holroyd, Hajcak, & Larsen, 2006; Luu, Tucker, Derryberry, Reed, & Poulsen, 2003; Nieuwenhuis, Holroyd, Mol, & Coles, 2004; Ruchsow, Grothe, Spitzer, & Kiefer, 2002; Yeung, Holroyd, & Cohen, 2005; Yeung & Sanfey, 2004). Localization studies suggest that the FRN is generated near the anterior cingulate cortex (ACC; Gehring & Willoughby, 2002; Miltner, Braun, & Coles, 1997; Nieuwenhuis, Yeung, Holroyd, Schurger, & Cohen, 2004), consistent with fMRI results implicating the ACC in negative feedback processing (e.g., Nieuwenhuis, Heslenfeld, et al., 2005).

The prevailing theory of the FRN, proposed by Holroyd and Coles (2002), suggests that the FRN is a reward prediction signal, reflecting the activity of a reinforcement learning system. The reinforcement learning theory (RL theory) further proposes that the amplitude of the feedback negativity is determined by the impact of a phasic midbrain dopamine response on the ACC, such that unexpected negative feedback is associated with a large negativity and unexpected positive feedback is associated with a small negativity. Thus, the FRN is seen to represent an early frontal brain response indexing the interaction between valence and expectancy of feedback. Although several studies have confirmed the general predictions of the RL theory, showing the largest FRNs to unexpected negative compared to positive feedback (Gibson, Krigolson, & Holroyd, 2006; Holroyd & Coles, 2002; Holroyd, Nieuwenhuis, Yeung, & Cohen, 2003; Nieuwenhuis, Nielen, Mol, Hajcak, & Veltman, 2005), recent experiments have revealed that the modulation of the FRN by the interaction of valence and expectancy is more nuanced (Hajcak et al., 2005; Hajcak, Moser, Holroyd, & Simons, 2007).

First, Hajcak et al. (2005) failed to find an interaction effect of valence and expectancy on the FRN in a guessing task that induced differential expectations via a cue indicating the probability of reward (25%, 50%, or 75%). In this study, subjects first received the predictive cue and then were asked to guess which of four doors hid a prize of 5 cents, following which subjects were given feedback indicating either reward (5 cents) or nonreward (0 cents). Results of this study revealed that the magnitude of the feedback negativity was larger for negative compared to positive feedback, but was not further modulated by expectancy. Following this null finding, we surmised that our paradigm left subjects' expectations relatively weak or not consistent with objective reward probability, whereas other tasks such as reinforcement learning tasks require subjects to pay significant attention to stimulus–response mappings and learn reward probabilities accordingly throughout the course of the task (Gibson et al., 2006; Holroyd & Coles, 2002; Nieuwenhuis, Heslenfeld, et al., 2005). For example, the predictive cue in the

The authors thank Jenna Dietz and Natalie Shroyer for their help with data collection. We also thank Natalie Rosenthal, Robert Rosenthal, and Beth Morling for their intellectual contributions to this article. Portions of this article were presented at the 47th annual meeting of the Society for Psychophysiological Research, Savannah, Georgia, October, 2007. This research was supported by National Institute of Mental Health (NIMH) predoctoral fellowship MH077388 (J.S.M.).

Address reprint requests to: Jason S. Moser, Department of Psychology, 108 Wolf Hall, University of Delaware, Newark, DE 19716, USA. E-mail: jmoser@udel.edu

Hajcak et al. (2005) study might not have induced the intended modulation in expectations because participants might have succumbed to the ''gambler's fallacy'' and expected reward in the face of low objective reward probability (e.g., 25%) after receiving nonreward on several previous trials. Another possibility is that subjects formed differential expectations that were generally consistent with objective reward probability, but the strength of their expectations might have been relatively weak or uniform (e.g., 50% vs. 55% vs. 60%) compared to those prescribed by the cue (25% vs. 50% vs. 75%).

In a set of follow-up experiments using the same task, we assessed participants' expectations on each trial to more accurately capture subjective expectations that may or may not be consistent with objective reward probability. The first experiment asked subjects to make their reward prediction prior to making their guess, and the second experiment asked subjects to make their reward prediction after making their guess. Results of these two experiments showed that the FRN was modulated by the interaction of valence and expectancy only in the situation where subjects made their reward prediction after making their guess. We reasoned that when subjects made their predictions prior to guessing what door hid a prize, there was significant time between their prediction and the feedback; thus their prediction could have changed after they made their guess or, by the time feedback occurred, their prediction was no longer strong. On the other hand, when subjects made their predictions after guessing what door hid a prize, we hypothesized that because their prediction occurred in close temporal proximity to the delivery of the feedback as well as after they had committed to a decision about what door hid a prize, their prediction was consequently more salient.

Thus, the sensitivity of the FRN to the interaction between valence and expectancy appears highly dependent on the strength of the prediction–outcome association—this has been shown in reinforcement learning tasks that induce strong predictions because they require significant attention to actions and outcomes throughout task performance and in our doors task experiment when predictions were made following commitment to a guess and just prior to the delivery of feedback. However, because our previous experiments asked subjects about their predictions qeither before *or* after their guess and not both times, we were unable to test the notion that predictions might actually change across the course of a trial and that such changes might be associated with weak predictions. Additionally, we did not collect information on how strong predictions were in the two scenarios; our conclusion that the strength of predictions was stronger in the second study was inferred by the successful modulation of the FRN by valence and expectancy that was not found in the first and by the fact that the prediction was made after committing to a guess and just before the delivery of feedback—that is, we inferred the strength of prediction from the proximity of attention to the prediction and feedback. The current study aimed to address these issues by asking subjects to make subjective reward predictions both before *and* after making their guess of door.

By asking subjects to make subjective reward predictions before and after making their guess of door, we wanted to find instances of prediction consistency—that is, when subjects make the same prediction both before and after making their guess—and prediction inconsistency—that is, when subjects make different predictions before and after making their guess. Therefore, the current design allowed us to extract four types of trials: (1) trials in which subjects predict they will win both before and after making their guess of door (hereafter referred to as ''win/win'' trials); (2) trials in which subjects predict they will not win both before and after making their guess of door (hereafter referred to as ''lose/lose'' trials); (3) trials in which subjects predict they will win before making their choice of door, but then change their mind and predict they will not win after making their choice of door (hereafter referred to as ''win/lose'' trials); and (4) trials in which subjects predict they will not win before making their choice of door, but then change their mind and predict that they will win after making their choice of door (hereafter referred to as ''lose/win'' trials). In this way, we were also able to control for when during the trial attention was drawn to subjective predictions—which was the primary manipulation of prediction strength in our previous two studies—in that all subjects received prompts for subjective predictions *both* before and after their guess of door.

To test our hypothesis concerning prediction strength and modulation of the FRN by the interaction of valence and expectancy—that is, that the stronger the prediction, the larger the FRN modulation by valence and expectancy—we conducted two planned comparisons using the four trial types described above: (1) we compared FRN magnitude on win/win trials to FRN magnitude on lose/lose trials, and (2) we compared FRN magnitude on win/lose trials to FRN magnitude on lose/win trials. Based on our previous findings and conceptualizations, we generated two alternative hypotheses for the current study. If, as we suggested in Hajcak et al. (2007), weak expectations are reflected in opportunities for, or instances of, changes in subjective reward prediction, we hypothesized that the difference between FRN magnitude on win/lose trials and lose/win trials should be small and the difference between FRN magnitude on win/win and lose/lose trials should be large. We reasoned that the comparison of win/lose and lose/win trials would result in a small FRN difference because the flip-flopping of subjective reward predictions would indicate relatively weak expectations in both cases, whereas the comparison of win/win and lose/lose trials would result in a large FRN difference because the consistency of subjective reward predictions would indicate relatively strong expectations in opposite directions, thus representing the largest discrepancy. On the other hand, we also suggested in Hajcak et al. (2007) that the temporal proximity of reward prediction to feedback delivery is important in modulating the FRN because of the attention the reward prediction prompt draws—in the Hajcak et al. (2007) study attention was drawn to the reward prediction *either* early or late in the trial, relative to feedback delivery. Given that all four conditions in the present study required two reward predictions—one before *and* one after door guess—we reasoned that attention to the reward prediction prompts might differ across the trial. Thus, an alternative view on the current study would predict that if, in fact, participants pay differential attention to the two reward prediction prompts, it might be that in the win/win and lose/lose trials, attention is paid to the first prompt, and consistency actually indicates a waning of attention, as attention was drawn to the first question and the second prompt was met with a consistent, but not well thought out, response—that is, participants were acting as if they were on ''cruise control'' for the second prediction question. In this case, we would predict that the difference in FRN magnitude between win/win and lose/lose trials would be small, as predictions were ''set'' early on in the trial and temporally distant from feedback delivery, whereas, on win/lose and lose/win trials, attention might be drawn to the second prompt, as participants would have allocated significant resources to change their initial response, thus suggestive of deeper analysis of their prediction on that trial and thus their attention to their prediction would be temporally closer to feedback delivery. In this case, the difference in FRN magnitude between win/lose and lose/win trials would be large, and specifically larger FRNs would be expected on lose/win than win/lose trials because of violation of the expectation in the former instance.

## Method

### *Participants*

Forty undergraduate students in introductory level psychology classes at the University of Delaware participated in the current experiment for course credit. Given that our primary aim was to examine differences between specific types of trials—win/win versus lose/lose trials, and win/lose versus lose/win—and the number of each type was determined by subject behavior, we utilized a subset of 14 (8 women) subjects from the original set of 40 who had 20 or more usable trials in each category for all analyses described below. In this way, our final sample comprised subjects with an adequate number of trials of each type to generate reliable ERP averages.

### *Task*

The task was administered on a Pentium III class computer, using Presentation software (Neurobehavioral Systems, Inc.) to control the presentation and timing of all stimuli. Subjects' primary objective on each trial was to guess which of two doors presented horizontally in a color graphic hid a prize by pressing the left or right "ctrl" key. Each trial began with a fixation cross (+) presented at the center of the screen for 1000 ms. Then, the question: "Do you think you will win on this trial?" appeared on the screen and remained there until participants indicated yes or no using the left and right ctrl keys, respectively. Thus, prior to making their choice of door on each trial, participants predicted whether or not they thought they would chose correctly. Five hundred milliseconds after their initial subjective prediction, the graphic of the doors appeared until the participant chose a door. Then, 1500 ms after the participant's choice of door, the question: "Do you think you will win on this trial?" appeared again on the screen and remained until participants indicated yes or no using the ctrl keys mentioned above. This second question was posed to determine if subjects had changed their minds about their subjective prediction after having made their choice of door and were instructed as such. One thousand milliseconds following their second subjective prediction, a feedback stimulus appeared on the screen for 1000 msec: a green "+" feedback indicated a correct guess, and a green "o" feedback indicated an incorrect guess. All other stimuli were presented in white font against a black background; all stimuli were positioned in the center of the screen. The feedback stimuli occupied approximately 2° of visual angle horizontally, and 2° vertically. Following the presentation of the feedback, the intertrial interval was marked by the presentation of a blank screen for 1500 ms.

Participants were informed that they would earn $.05 for each correct guess and therefore would earn between $0.00 and $12.00 in bonus money based on their performance. Unbeknownst to the participants, the outcome of each trial was predetermined and pseudorandom such that, overall, participants received exactly 50% correct feedback. All participants were paid $6.00.

### *Procedure*

After a brief description of the experiment, EEG sensors were attached and the participant was given detailed task instructions. To familiarize participants with the task, each participant was given a practice block consisting of 20 trials and was instructed to guess which door hid a prize while also answering yes or no to the two subjective prediction questions. The experiment consisted of 12 blocks of 20 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 they had earned. Participants filled out a brief questionnaire at the completion of the experiment to ascertain level of interest (1 = *very boring*; 7 = *very interesting*), attention paid to the subjective predictions (1 = *ignored predictions*; 7 = *paid close attention*), attention paid to the outcomes (1 = *ignored outcomes*; 7 = *paid close attention*), and the intensity of emotional reaction to the outcomes (1 = *very unhappy*; 7 = *very happy*).

### *Psychophysiological Recording, Data Reduction, and Analysis*

The electroencephalogram (EEG) was recorded from the frontal (Fz), fronto-central (FCz), central (Cz), and parietal (Pz) recording sites using an ECI electrocap. In addition, tin disc electrodes were placed on the left and right mastoids (M1 and M2, respectively). During the recording, all activity was referenced to Cz. The electro-oculogram (EOG) generated from blinks and vertical eye movements was also recorded using Med-Associates miniature electrodes placed approximately 1 cm above and below the subject's right eye. The right earlobe served as a ground site. All EEG/EOG electrode impedances were kept below 10 kΩ and the data from all channels were recorded by a Grass Model 78D polygraph with Grass Model 7P511J preamplifiers (bandpass = 0.1–100 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 participant's response to the second question (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 converter's minimum or maximum values). Single trial EEG data were lowpass filtered at 20 Hz with a 19 weight FIR digital filter as per Cook and Miller (1992).

As the primary aim of the current study was to evaluate how the consistency of subjective prediction on an individual trial, as measured by participants' responses to the two questions posed before and after door choice, influenced the magnitude of the FRN, grand-average waveforms were created for the four trial types: (1) win/win, (2) win/lose, (3) lose/lose, and (4) lose/win. The average activity in the 100-ms prestimulus window served as a baseline. One outstanding issue in FRN research is whether variance in FRN amplitude between rewarding and nonrewarding feedback results mainly from activity elicited by nonrewards, by rewards, or by both. As argued at length by Luck (2005), the absolute amplitudes of ERP components are meaningless in and of themselves because, for example, an apparent decrease in the amplitude of a component can result from the superposition of a component with opposite polarity (see his rule #4, p. 56). For these reasons, we have used a difference-wave approach to isolate the valence-related variance in the ERP in a manner that is independent of the source of the variance (nonrewards, rewards, or both; Holroyd, 2004; Luck, 2005). Specifically, difference waves were created by subtracting the ERPs observed following rewards from the ERPs observed following nonrewards. These difference waves were created separately for each of the four trial types mentioned above (e.g., the difference between the ERPs generated for rewards and nonrewards on win/win trials). FRNs were then defined peak-to-peak as the difference between the most negative going peak and the most positive preceding peak in the 150–350-ms time window following feedback at each electrode site. This procedure controlled for the main effect of stimulus probability and prediction on the ERP, ensuring that the ERP measure was sensitive to the interaction of feedback expectations and valence (Holroyd, 2004).

The FRN was statistically evaluated using SPSS (Version 15.0) General Linear Model software with the Greenhouse–Geisser correction applied to $p$ values associated with multiple $df$ repeated measures comparisons. Partial eta squared ($\eta_p^2$) values are reported to demonstrate the size of effects in ANOVA models, where .05 represents a small effect, .1 represents a medium effect, and .2 represents a large effect (Cohen, 1973). Cohen's $d$ is reported for $t$ tests, where .2 represents a small effect, .5 represents a medium effect, and .8 represents a large effect size (Cohen, 1988). On the basis of our previous findings regarding the relationship between the FRN and reward prediction, we conducted two planned 2 (Trial) $\times$ 4 (Site: Fz, FCz, Cz, Pz) ANOVAs to test our hypotheses: (1) We tested the difference between win/win trials and lose/lose trials and (2) we tested the difference between win/lose trials and lose/win trials.

## Results

### Behavioral Measures

On average, participants responded win/win to the two questions on 39.75% ($SD = 12.47$) of trials and lose/lose on 24.85% ($SD = 11.95$) of trials, the difference between which was significant, $t(13) = 2.48$, $p = .028$, $d = .68$—reminiscent of the positive bias reported in Hajcak et al. (2007). On the other hand, participants responded win/lose on 19.96% ($SD = 8.62$) of trials and lose/win on 15.43% ($SD = 4.87$) of trials, the difference between which was not significant, $t(13) = 1.65$, $p = .124$, $d = .50$. In terms of responses to the posttask questionnaire, participants found the task, overall, somewhat interesting ($M = 3.00$, $SD = 1.47$) significantly more than the lowest possible ranking of 1 = *very boring*; $t(13) = 5.10$, $p < .001$. Participants reported having paid equal attention to the first and second subjective prediction questions ($M = 3.00$, $SD = 1.36$ for the first question and $M = 3.07$, $SD = 1.14$ for the second question), $t(13) < 1$, $d < .10$. Participants also reported having paid significant attention to the outcomes ($M = 4.64$, $SD = 1.50$) significantly more than the lowest possible ranking of 1 = *ignored outcomes*; $t(13) = 9.09$, $p < .001$. Participants' happiness ratings to the feedback, calculated as the difference between reactions to nonreward minus reward (to mirror FRN calculation), showed that participants felt equally unhappy when receiving nonreward (vs. reward) on win/win trials ($M = 1.79$, $SD = 1.42$, where larger numbers reflect more unhappiness) and lose/lose trials ($M = 1.43$, $SD = 1.45$), $t(13) = 1.10$, $p = .292$, $d = .29$. Interestingly, however, participants reported being more unhappy when receiving nonreward (vs. reward) on lose/win trials ($M = 1.86$, $SD = 1.46$) than on win/lose trials ($M = 1.21$, $SD = 1.31$), $t(13) = 3.23$, $p = .007$, $d = .86$.

### FRN

Figure 1 presents feedback-locked difference wave ERPs for the four trial types. Consistent with previous studies, all of the difference waves show a large negative deflection, indicative of significantly more negativity to nonreward than reward, peaking approximately 250 ms following feedback (all difference wave FRNs were significantly larger than 0), $t$s$(13) > 4.90$, $p$s $< .001$. Table 1 shows the descriptive statistics for the FRN difference wave score for each trial and site. The 2 (Trial) $\times$ 4 (Site) repeated measures ANOVA testing the difference between win/win and lose/lose trials revealed no significant effect of Trial, $F(1,13) = 1.89$, $p = .192$, $\eta_p^2 = .127$. Additionally, neither the main effect of Site nor the interaction between Trial and Site were significant, $F(3,39) = 1.17$, $p = .326$, $\eta_p^2 = .083$; $F(3,39) < 1$, respectively). Interestingly, however, the ANOVA testing the difference between win/lose and lose/win trials revealed a significant main effect of Trial, $F(1,13) = 4.82$, $p = .047$, $\eta_p^2 = .27$, indicating larger FRNs on lose/win trials than on win/lose trials (see Table 1). Neither the main effect of Site nor its interaction with Trial reached significance, $F(3,39) < 1$; $F(3,39) = 2.09$, $p = .119$, $\eta_p^2 = .139$, respectively.

Thus, our FRN findings support the notion that consistent prediction trials (i.e., win/win and lose/lose) are associated with diminished attention to predictions whereas inconsistent or flip-flop prediction trials (i.e., win/lose and lose/win) are associated with deeper analysis of and therefore heightened attention to predictions, possibly due to greater allocation of attentional resources to the flip-flop process. To further evaluate this "attentional" explanation of the current findings, we conducted a follow-up analysis of the P300, a classic attentional ERP component (Donchin, 1981), elicited by the feedback. If subjects in the current study engaged more attentional resources on flip-flop prediction trials (win/lose and lose/win) than consistent prediction trials (win/win and lose/lose), as we suggest, then the P300 should be larger on flip-flop than consistent trials.

The P300 was measured as the most positive peak occurring in the 200–600-ms postfeedback window, relative to a 100-ms prestimulus baseline, and submitted to a 2 (Condition: flip-flop vs. consistent) $\times$ 4 (Site: Fz, FCz, Cz, Pz) ANOVA. Results were consistent with our hypothesis, as the nearly significant main effect of Condition suggested that flip-flop trials were associated with larger P300s than consistent prediction trials, $F(1,13) = 4.02$, $p = .066$, $\eta_p^2 = .236$; see Table 2.

## Discussion

With the current study, we aimed to build on previous findings suggesting that modulation of the FRN by the interaction of feedback valence and expectancy is highly dependent on the strength of predictions and/or prediction–outcome associations. To this end, subjects in the current study engaged in a guessing task wherein they were asked to make reward predictions *both* before and after making their guess of which of two doors hid a prize of 5 cents. Although the FRN was always larger to nonreward than reward (consistent with Hajcak et al., 2005, 2007), the FRN was not further modulated by expectations on trials in

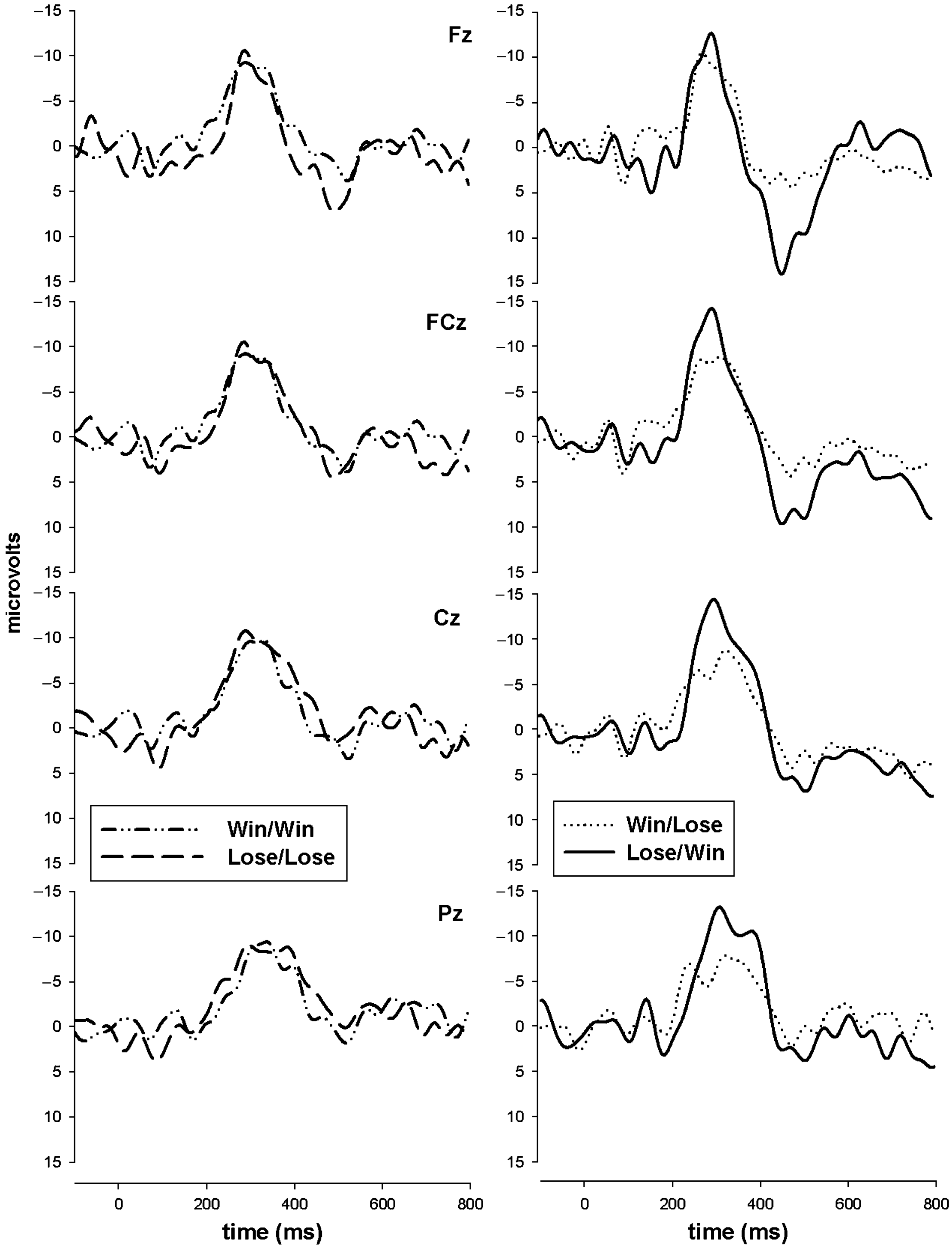


**Figure 1.** Stimulus-locked ERP difference waves for win/win and lose/lose trials (left) and win/lose and lose/win trials (right) at Fz, FCz, Cz, and Pz. Feedback onset occurred at 0 ms.

which subjects were consistent in their predictions—that is, FRN magnitude was comparable on win/win and lose/lose trials. Rather, the FRN *was* modulated by expectations on trials in which subjects were inconsistent in their predictions—that is, FRN magnitude differed between win/lose and lose/win trials such that the FRN was enhanced on lose/win trials. The posttask emotion ratings mirrored the FRN results in that participants reported equivalent unhappiness when receiving nonreward (vs. reward) on win/win and lose/lose trials, but reported significantly more unhappiness when receiving nonreward (vs. reward) on lose/win trials than on win/lose trials.

The current findings are consistent with our second set of predictions put forth in the introduction. Specifically, the significant difference between win/lose and lose/win trials suggests that subjects' attention was drawn to the second prediction because of their flip-flopping responses to the prompts. Because

**Table 1.** *Means (Standard Deviation) for FRN Magnitudes*

| Difference Score (nonreward–reward) | | Win/Win | Lose/Lose | Win/Lose | Lose/Win |
|---|---|---|---|---|---|
| FRN (μV) | Fz | 16.96 | 21.50 | 20.28 | 26.22 |
| | | (11.68) | (15.96) | (11.70) | (18.05) |
| | FCz | 17.80 | 22.28 | 19.45 | 26.56 |
| | | (12.05) | (15.00) | (8.72) | (18.68) |
| | Cz | 17.88 | 22.71 | 18.18 | 26.99 |
| | | (13.43) | (15.29) | (8.85) | (17.35) |
| | Pz | 16.33 | 20.47 | 18.06 | 27.04 |
| | | (11.17) | (10.23) | (8.97) | (16.43) |

participants changed their response, we hypothesized that they would engage in attention-demanding activity, rendering the salience of their reward prediction temporally close to the feedback, and thus engendering the modulation of the FRN in the expected direction—that is, larger on lose/win trials than win/lose trials because of the violation of expectancy in the former case. The fact that we found no difference in FRN magnitude between win/win and lose/lose trials is in line with our prediction that consistent reward prediction was indicative of the "setting" of attention to the first reward prediction and the waning of the salience of reward prediction across the course of the trial, thus engendering very little further modulation of the FRN by expectancy—that is, the subjects were on "cruise control" after their response to the first reward prediction prompt. That the P300 was larger on flip-flop than on consistent prediction trials lends further credence to this "attentional" interpretation of the current findings. Recent results suggesting that modulation of the FRN depends on the depth of processing of task stimuli, and associations are also consistent with this interpretation (cf. Baker, Krigolson, & Holroyd, 2006; Gibson et al., 2006; Lee, Krigolson, & Holroyd, 2006).

In addition to our attentional account of the present findings, research on negative emotions in decision making (cf. Zeelenberg, van Dijk, Manstead, & van der Pligt, 2000) provides further insights. Regret is a particularly salient emotion cited in this work and occurs when the outcome of a decision is worse than the outcome had an alternative decision been made. Perhaps most akin to our attentional account, then, a recent study examined self-regulatory orientation (Camacho, Higgins, & Luger, 2003; Kruglanski et al., 2000) and regret in decision making (Pierro et al., 2008). Two orientations or modes were examined: assessment and locomotion. When individuals are in the assessment mode they are characterized as more thoughtful, weighing options against each other, and critically evaluative. When individuals are in the locomotion mode, on the other hand, they are characterized as less thoughtful and more action oriented, making decisions and moving on, "committing the psychological resources that will initiate and maintain goal-directed progress in a straightforward manner, without undue distractions or delays" (Kruglanski et al., 2000, p. 794). In a series of studies, Pierro et al. found that individuals in the assessment mode were more regretful and engaged in more reflective or counterfactual thought after unfavorable outcomes than individuals in the locomotion mode. In the context of the Pierro et al. findings, we suggest that subjects were in a locomotion mode on win/win and lose/lose trials, making predictions and moving on, not considering their predictions too strongly—that is, they were on "cruise control" as we suggested. Thus, when negative feedback was delivered on both types of trials, subjects registered the negative feedback, as indexed by larger FRNs to nonreward compared to reward, but kept moving forward, allocating little attention to processing the feedback in the broader context of their predictions. In contrast, we suggest that the subjects were in an assessment mode on win/lose and lose/win trials, considering their predictions more thoroughly and weighing their commitment. Therefore when they received negative feedback after lose/win trials, not only was the feedback negative—registered by the larger FRN to nonreward compared to reward—but they foolishly changed their prediction for the worse, likely increasing regret—registered by the larger FRN compared to win/lose trials. Studies also show that when individuals switch purchasing decisions for the worse, they experience increased regret compared to instances when individuals repeat or maintain their initial decision even in the face of less than optimal outcomes (i.e., maintain the status quo; Inman & Zeelenberg, 2002; Tsiros & Mittal, 2000). Therefore, the fact that the FRN difference was enhanced between inconsistent prediction conditions might fit with this literature suggesting increased regret when switching decisions compared to maintaining the status quo.

Thus, the enhanced FRN found here for lose/win trials might be understood as reflecting the feeling of regret for flip-flopping and foolishly changing expectations in the positive direction when the first, more pessimistic one, was correct. The posttask emotion ratings further bolster this interpretation, as participants reported feeling increased unhappiness to nonreward (vs. reward) on lose/win trials. Interestingly, regret is associated with proactive tendencies including desires to correct one's mistake, undo the event, and get a second chance (for a review, see Zeelenberg et al., 2000). Given that the FRN is seen to be an ACC-generated warning signal that behavioral adjustments are needed for optimal performance (Holroyd & Coles, 2002; Holroyd & Krigolson, 2007) and regret is associated with tendencies to act, it stands to reason that the FRN should be largest on trials in which regret is largest. In this way, the current finding of enhanced FRN amplitude on lose/win trials can be interpreted as increased ACC activity when regret is elicited by a flip-flop for the worse and thus a more intense call to adjust performance, desire to correct one's response, or get a second chance at predicting correctly occurred. In line with this FRN–ACC–regret link, recent neuroimaging studies show enhanced ACC and orbitofrontal cortex activity under conditions of regret in decision making tasks (for a review, see Coricelli, Dolan, & Sirigu, 2007). Future studies with larger samples will be needed to further evaluate these hypotheses concerning the FRN, as the current study comprised a relatively small set of 14 participants who had

**Table 2.** *Means (Standard Deviation) for P300 magnitudes (in Microvolts)*

| | Consistent Trials | Flip-flop Trials |
|---|---|---|
| Fz | 23.85 | 29.05 |
| | (13.34) | (20.21) |
| FCz | 27.69 | 32.08 |
| | (16.62) | (21.78) |
| Cz | 28.29 | 31.89 |
| | (16.53) | (21.13) |
| Pz | 26.77 | 29.46 |
| | (15.97) | (20.57) |

sufficient numbers of trials for stable ERPs. Examining the possibility that individual differences in prediction strength, gambling behavior, and emotional reactivity to nonreward (vs. reward) might be related to FRN magnitude in such tasks will also be a fruitful avenue for future research.

The RL theory holds that FRN amplitude is determined by an interaction between feedback valence and expectancy, such that unexpected feedback induces greater variance in FRN amplitude relative to expected feedback. The results of the current study as well as those previously reported by our group (Hajcak et al., 2005, 2007), suggest that modulation of the FRN by the interaction of valence and expectancy deserves more attention. In general, the FRN seems to reliably differentiate negative from positive feedback; however, its further modulation by expectancy is more variable and depends on the depth or type of processing involved in the task (cf. Baker et al., 2006; Gibson et al., 2006; Lee et al., 2006). We therefore propose that the FRN likely reflects complex information related to current and past actions, thoughts, and emotions. Research on negative emotion, decision making, and consumer behavior (cf. Zeelenberg et al., 2000) provides additional insights regarding this interpretation of the FRN, which dovetails with current conceptions of the ACC's role in decision making and social behavior (cf. Rushworth, Behrens, Rudebeck, & Walton, 2007). Future studies testing more complex predictions regarding the FRN's significance in various behavioral contexts by unifying the vast literatures on emotion and decision making are needed.

## REFERENCES

Baker, R., Krigolson, O. E., & Holroyd, C. B. (2006). Examining the feedback error-related negativity using predictive stimuli. *Psychophysiology*, *43*, S22.

Camacho, C. J., Higgins, E. T., & Luger, L. (2003). Moral value transfer from regulatory fit: "What feels right is right" and "what feels wrong is wrong." *Journal of Personality and Social Psychology*, *84*, 498–510.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, *33*, 107–112.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cook, E. W. 3rd. (1999). *VPM reference manual*. Birmingham, AL: Author.

Cook, E. W., 3rd, & Miller, G. A. (1992). Digital filtering: Background and tutorial for psychophysiologists. *Psychophysiology*, *29*, 350–367.

Coricelli, G., Dolan, R. J., & Sirigu, A. (2007). Brain, emotion, and decision making: The paradigmatic example of regret. *Trends in Cognitive Sciences*, *11*, 258–265.

Donchin, E. (1981). Surprise! . . . surprise? *Psychophysiology*, *18*, 493–513.

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

Gibson, J., Krigolson, O. E., & Holroyd, C. B. (2006). Sensitivity of the feedback error-related negativity to reward probability. *Psychophysiology*, *43*, S41.

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

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. (2006). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology*, *71*, 148–154.

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.

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

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Reviews*, *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 Research*, *1105*, 93–101.

Holroyd, C. B., & Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology*, *44*, 913–917.

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.

Inman, J. J., & Zeelenberg, M. (2002). Regret in repeat purchase versus switching decisions: The attenuating role of decision justifiability. *Journal of Consumer Research*, *29*, 116–128.

Kruglanski, A. W., Thompson, E. P., Higgins, E. T., Atash, M. N., Pierro, A., Shah, J. Y., et al. (2000). To "do the right thing" or to "just do it": Locomotion and assessment as distinct self-regulatory imperatives. *Journal of Personality and Social Psychology*, *79*, 793–815.

Lee, K., Krigolson, O. E., & Holroyd, C. B. (2006). Modulation of the fERN amplitude by reward expectancy. *Psychophysiology*, *43*, S57.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science*, *14*, 47–53.

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

Miltner, W. H., 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 sysytem for error detection. *Journal of Cognitive Neuroscience*, *9*, 788–798.

Nieuwenhuis, S., Heslenfeld, D. J., von Geusau, N. J., Mars, R. B., Holroyd, C. B., & Yeung, N. (2005). Activity in human reward-sensitive brain areas is strongly context dependent. *Neuroimage*, *25*, 1302–1309.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neuroscience and Biobehavioral Reviews*, *28*, 441–448.

Nieuwenhuis, S., Nielen, M. M., Mol, N., Hajcak, G., & Veltman, D. J. (2005). Performance monitoring in obsessive-compulsive disorder. *Psychiatry Research*, *134*, 111–122.

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.

Pierro, A., Leder, S., Mannetti, L., Higgins, E. T., Kruglanski, A. W., & Aiello, A. (2008). Regulatory mode effects on counterfactual thinking and regret. *Journal of Experimental Social Psychology*, *44*, 321–329.

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

Rushworth, M. F. S., Behrens, T. E. J., Rudebeck, P. H., & Walton, M. E. (2007). Contrasting roles for cingulate and oribitofrontal cortex in decisions and social behaviour. *Trends in Cognitive Sciences*, *11*, 169–176.

Tsiros, M., & Mittal, V. (2000). Regret: A model of its antecedents and consequences in consumer decision. *Journal of Consumer Research*, *26*, 401–417.

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

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

Zeelenberg, M., van Dijk, W. W., Manstead, A. S. R., & van der Pligt, J. (2000). On bad decisions and disconfirmed expectancies: The psychology of regret and disappointment. *Cognition & Emotion*, *14*, 521–541.

(Received April 20, 2008; Accepted May 23, 2008)