# Loss feedback negativity elicited by single- versus conjoined-feature stimuli

Yanni Liu and William J. Gehring

**Event-related brain potential studies show that negative feedback in guessing tasks elicits a medial frontal negativity. Most theory and experimentation concerning this feedback-related negativity (FRN) has assumed that the FRN has little relationship to the perceptual characteristics of the feedback. This study challenges this assumption. We used a single visual feature or a conjunction of features to indicate reward feedback in a gambling task. In the single-feature condition, losses elicited a larger FRN than gains; in the conjoined-feature condition, that difference was not observed. The results are consistent with the proposal that the FRN is modulated by the deviation of feedback stimuli from a perceptual template. Future studies must not confound the perceptual properties and the valence of reward feedback. *NeuroReport* 20:632–636 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins.**

**NeuroReport** 2009, **20**:632–636

**Keywords: error-related negativity, feedback-related negativity, perceptual mismatch, reinforcement learning theory of the error-related negativity**

Department of Psychology, University of Michigan, Ann Arbor, Michigan, USA

Correspondence to Dr Yanni Liu, Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109-1043, USA
Tel: +1 734 883 9388; fax: +1 734 763 7480; e-mail: yanniliu@umich.edu

**Received** 23 December 2008 **accepted** 10 February 2009

## Introduction

People often rely on external feedback to improve performance. Event-related brain potential (ERP) studies have identified a negative ERP component elicited by error [1] or loss [2–5] feedback. This feedback-related negativity (FRN) is distributed over medial frontal regions of the scalp and reaches maximum amplitude around 250 ms following negative feedback. Holroyd and Coles [6] proposed the reinforcement-learning theory of the error-related negativity (RL-ERN) to explain both the FRN and the ERN, which accompanies error commission in speeded reaction time tasks [7,8]. According to the RL-ERN theory, a monitoring system located in the basal ganglia produces error signals upon the detection of events that are worse than expected. The error signal is carried by the mesencephalic dopamine system to the anterior cingulate cortex where the FRN is generated. The anterior cingulate cortex uses this signal to improve performance.

In this study, we sought to clarify the nature of the perceptual representation that is used to detect negative events when the FRN is generated. In many previous FRN studies, investigators assumed that the FRN was unaffected by the perceptual properties of the feedback stimuli [1,6,9,10]. Nevertheless, there is evidence against this assumption: Nieuwenhuis *et al.* [11] used a task originally developed by Gehring and Willoughby [2], in which the information delivered by a feedback stimulus regarding the accuracy of the response was dissociated from that regarding the reward associated with the response. (In these studies, a response can cause a loss but still be correct if the loss was the best outcome possible). They found that the perceptual salience of the feedback stimuli affected whether the FRN was sensitive to the gain/loss information or to the error/correct information [11]. Additional evidence that the perceptual characteristics can affect the size of the FRN comes from a study by Donkers *et al.* [12], who used a slot-machine task in which participants were asked to watch three digits presented successively on a computer screen; an FRN-like negativity was observed whenever a stimulus was different from the preceding stimulus, regardless of the relative gain or loss.

As it is currently formulated, the RL-ERN theory does not specify whether the perceptual attributes of the feedback should affect the FRN. For example, the theory does not specify whether categorizing a stimulus as rewarding or non-rewarding requires focused attention, or whether pre-attentive feature analysis alone is sufficient to provide the information required to elicit the FRN. Presumably, if the RL-ERN model were part of a larger computational framework that also described perceptual and motor processes (e.g., Executive Process/Interactive Control (EPIC) [13,14] or Adaptive Control of Thought–Rational (ACT-R) [15]) the theory would have to describe such details. Thus, one goal of the present study is to contribute to developing a broader theory by investigating the nature of the stimulus representation that elicits the FRN response.

An additional motivation for our study is an alternative hypothesis that is largely unexplored. This hypothesis states that the FRN is not sensitive to the reward properties of the feedback stimulus *per se*. Rather, the

0959-4965 © 2009 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/WNR.0b013e32832a3250

monitoring system eliciting the FRN detects the mismatch between the perceptual attributes of the expected stimulus and those that actually occur, not whether the stimulus has a particular association to gains or losses. Folstein and Van Petten [16] proposed a similar idea to explain the anterior N2, arguing that the N2 reflects a deviation from an actively maintained visual template (see also [17]). This perspective differs from the RL-ERN theory: if a rewarding stimulus is associated with a particular color (e.g., green), then the fast detection of non-reward could occur by maintaining a perceptual template such that any non-green stimulus signals a deviation. The FRN could thus represent a more general-purpose computation sensitive to task-related perceptual priming. Other kinds of stimulus distinctions could also cause an FRN to occur, and according to this view it just happens in the FRN paradigms that the distinction concerns the reward properties of the stimulus. Still, for this theory – like the RL-ERN theory – the particular stimulus attributes that constitute the perceptual template have not been specified in detail. Thus, both the RL-ERN theory and the perceptual-template hypothesis require further elaboration.

To examine this issue, the present experiment compares how two different levels of stimulus processing affect the FRN. In the experiment, either a single feature or a set of conjoined features was designated to indicate gain and loss reward information in a gambling task. When the reward information was indicated by a single feature, the perceptual mismatch between the actual outcome of a loss stimulus and the perceptual representation of the gain stimulus was easy to detect, because the detection of a single feature could be achieved by detecting distinctive stimulus features before they were conjoined to a visual location (as in the 'pop-out' in visual search tasks, [18]). When reward information was indicated by conjoined features, the detection of the reward properties of the feedback was dependent on focused attention [18]. We predicted that the FRN reward effect would be delayed or diminished when feedback was conveyed by conjoined features relative to feedback conveyed by a single distinctive feature.

## Methods

### Participants

There were 19 participants (8 males) aged between 18 and 21 years (mean age: 19.29 years). All were right-handed, had normal or corrected-to-normal vision, and normal color vision. Before the test, participants provided written informed consent in accordance with the Institutional Review Board of the University of Michigan. Each participant received course credit and a three- to five-dollar bonus for participation. The data from six participants (two males) were excluded from analyses: two because they were too sleepy to conduct the experiment, one because of an error by the experimenter, and two more because of excessive artifacts.

### Procedure

The participants were seated 60 cm in front of a 14-inch cathode ray tube computer monitor in a dimly lit, sound-attenuating and electromagnetically shielded room. They were instructed to remain as still as possible and to minimize eye blinks. On each trial, the participants were presented with two identical round chips displayed at the center of the screen following a 500 ms central fixation, and they were instructed that one chip indicated a gain and the other indicated a loss. Chips remained on the screen until the participants selected one by pressing a button with their left or right index finger, corresponding to the location of the chosen chip. 1800 ms after the response, the reward information indicating a gain or a loss on the trial appeared for 1000 ms. The intertrial interval was 1000 ms.

On each trial, a red circle, red square, blue circle or blue square served as the feedback stimulus. There were three feedback conditions, each characterized by the aspect of the stimulus that delivered reward information: 'single feature color feedback', in which the reward information was indicated by the color of the feedback stimulus (e.g., gain was indicated by a red circle or square); 'single feature shape feedback', in which the reward information was indicated by a certain shape (e.g., gain was indicated by a red or blue circle); 'conjoined color-shape feature feedback', in which the reward information was indicated by conjoined features (e.g., gain was indicated by a red circle or a blue square). The presentation sequence of the three feedback conditions was counterbalanced among the participants. Each feedback condition consisted of three blocks of trials, with the three blocks presented together. Each block started with 50 points as the initial allotment, and on each trial participants won or lost 5 points. Participants were informed that points would be converted into a small amount of money, although they did not know the exact conversion. On average the participants received a three- to five-dollar bonus. There were 110 trials per block, and participants were given summary information about the bonus they had earned after every 11 trials. The reward information each stimulus indicated was revealed to the participants at the beginning of the experiment and was counterbalanced among the participants. The feedback was randomly chosen from a set of equal numbers of gain and loss feedback. On average, the percentage of gain or loss stimuli was approximately 50%.

### Electrophysiological methods

The electroencephalogram (EEG) was recorded from 26 Ag/AgCl scalp electrodes embedded in a nylon mesh cap (Easy-Cap, Falk Minow Systems Inc., Herrsching-

Breitbrunn, Bavaria, Germany). The electrode locations consisted of FP1, AFz, FP2, F7, F3, Fz, F4, F8, FC3, FCz, FC4, T7, C3, Cz, C4, T8, CP3, CPz, CP4, P7, P3, Pz, P4, P8, O1 and O2. EEG data were recorded with a left mastoid reference and a forehead ground. An average mastoid reference was derived off-line using right mastoid data. The electro-oculogram (EOG) was recorded from Ag/AgCl electrodes above and below the left eye and external to the outer canthus of each eye. Impedances were kept below 10 KΩ. EEG and EOG were amplified by SYNAMPS DC amplifiers (Neuroscan Labs, Sterling, Virginia, USA) and filtered online from 0.01 to 100 Hz (half-amplitude cutoffs). The data were digitized at 500 Hz.

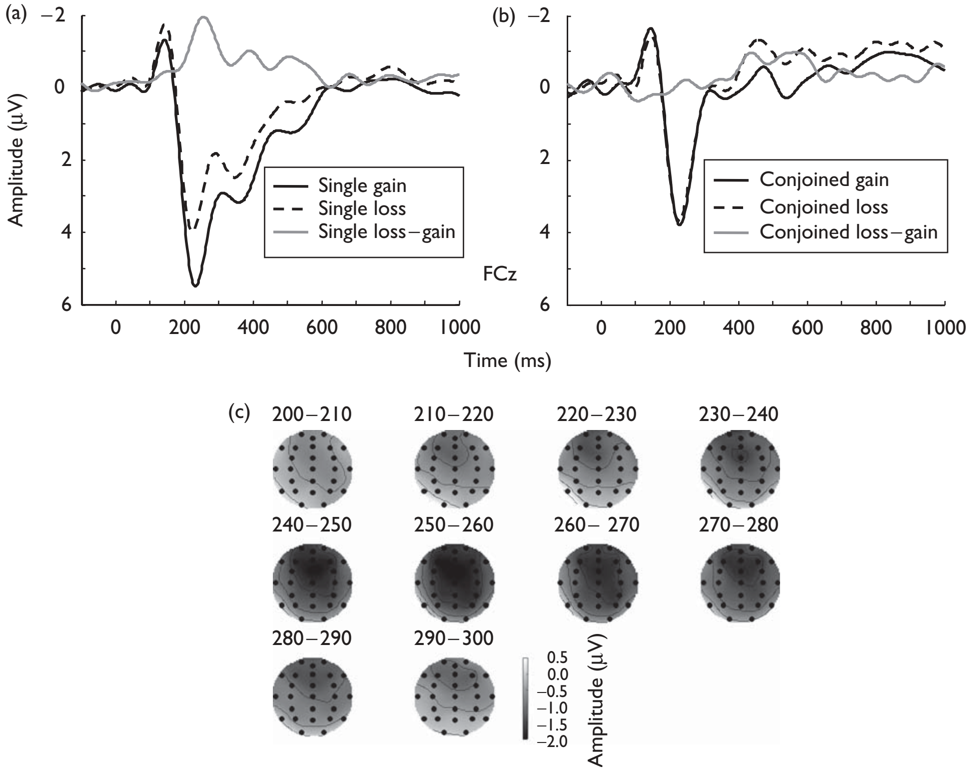
EEG epochs of 1100 ms (100 ms baseline) were extracted off-line from the continuous data file for analysis. Ocular artifacts were corrected using the algorithm described by Gratton *et al.* [19]. Statistical analyses were performed on the data without any additional filtering. The data presented in the figure were filtered with a nine-point Chebyshev II low-pass digital filter with a half-amplitude cutoff at 12 Hz.

## Results

In this task, there was no correct or incorrect response on each trial, and thus no objective way to evaluate performance accuracy. However, participants in gambling tasks often exhibit changes in response strategy in response to winning or losing [2]. We classified each trial according to whether the participant chose the same side as the previous trial and whether that previous trial had been a gain or loss. Trials where participants chose the side opposite to the one chosen on previous trials were labeled ‘switch’ trials; trials where participants chose the same side were labeled ‘no-switch’ trials. A 2 × 2 two-way repeated-measures analysis of variance with factors block type and reward valence revealed a main effect of reward valence [$F(1,12) = 8.86$, $P < 0.05$]. There was no main effect of block type or interaction (both $Fs < 1$). Thus, participants switched their responses more frequently after loss than after gain feedback in both the single-feature and conjoined-feature conditions, and the conditions did not differ in this switching effect.

We computed average waveforms for the single-feature gain and loss conditions, the conjoined-feature gain and loss conditions, and difference waveforms created by subtracting the gain waveform from the loss waveform for each condition. Inspection of data (Fig. 1a) suggests that the peak of the FRN occurred in the interval between the P2 (peak latency 224 ms) and N2 (peak latency 290 ms), because the loss–gain difference waveform showed a peak at 256 ms. Topographic scalp maps (Fig. 1c) supported

**Fig. 1**



Event-related brain potential (ERP) waveforms at electrode FCz for (a) single-feature blocks and (b) conjoined-feature blocks; (c) ERP topography of difference waveforms (loss–gain) between 200 and 300 ms for single-feature blocks.

this interpretation of the waveforms: during the period from 200 to 300 ms the maps show that the scalp topography of the FRN was consistently maximal at FCz throughout the epoch.

The FRN mean amplitude between 200 and 300 ms following the feedback was measured at FCz (see also [2]). A $2 \times 2$ two-way repeated-measures analysis of variance with factors block type and reward valence revealed main effects of block type [$F(1,12) = 18.98$, $P < 0.01$] and reward valence [$F(1,12) = 11.68$, $P < 0.01$], and an interaction between them [$F(1,12) = 16.68$, $P < 0.01$]. Subsequent analyses showed that in single-feature blocks, loss feedback elicited a larger FRN than gain feedback [$F(1,12) = 31.10$, $P < 0.001$]. In conjoined-feature blocks, there was no FRN difference between gain and loss feedback ($F < 1$). Inspection of the ERP waveforms (Fig. 1b) suggested a later negativity might exist in conjoined-feature blocks. Thus, the mean amplitude between 400 and 600 ms was measured as a possible late FRN, but there was no significant difference between gain and loss feedback [$F(1,12) = 2.00$, $P > 0.10$].

The preceding analyses collapsed across the color and shape single-feature conditions. Waveforms computed separately for each condition appeared virtually identical to those collapsed across the two conditions, with the exception that the peak of the color-condition difference waveform was earlier than that of the shape-condition difference [244 vs. 266 ms, $t(13) = 2.61$, $P < .05$]. Switching behavior also did not differ between the two conditions ($F < 1$).

## Discussion

In this study, participants received gain feedback and loss feedback conveyed by a single feature (color or shape) or by a conjunction of features. In the single-feature condition, loss feedback elicited a larger FRN than did gain feedback in the same latency window in which previous studies have identified the FRN (e.g., [2,6]). There was no difference in FRN amplitude between conjoined-feature loss and gain feedback. The behavioral results argue strongly against the possibility that participants in the conjoined-feature condition were unable to discriminate gain from loss feedback: participants switched their response choices more often after loss outcomes than after gain outcomes in both the single-feature and conjoined-feature conditions, suggesting that participants knew whether they had won or lost when presented with either type of feedback stimulus. Indeed, the single-feature and conjoined-feature conditions did not differ in switch rates.

These results are consistent with our proposal that the FRN is affected by the perceptual mismatch between gain and loss feedback stimuli [16]. In the gambling task, the perceptual system could be tuned for the attributes of gain stimuli (e.g., a certain gain-related feature); any feedback outcome that deviates from this primed/prepared gain-related perceptual representation would trigger the monitoring system to elicit an error signal.

The results would also be consistent with an amended RL-ERN theory in which the FRN is elicited only when the reward properties of the feedback are indicated by single, distinctive features (such as those that would produce pop-out in visual search, [18]). Such an amendment would pose a problem for RL-ERN theorists, however: the RL-ERN theory states that individuals learn the associations between stimuli and correct responses on the basis of the error detection that gives rise to the FRN. The stimulus-response associations subsumed by the model would seem to include those where the relevant stimulus properties are conjoined features. Our results suggest, however, that no FRN is elicited in those cases. Thus, a complete theory would need to include a different mechanism for detecting errors on the basis of conjoined features. Of course, it is possible that the FRN was elicited later in the conjoined-feature condition, but the two condition waveforms did not differ significantly in that later time window.

Recent theorizing has amended the RL-ERN theory in a different fashion. Investigators have suggested that a positivity associated with gains is either responsible for the FRN or overlaps it. Holroyd and Coles [20] have suggested that the difference between gain and loss trials occurs because of an overlap between a positivity that occurs on gain trials and an N2 that is present on all trials. An apparent negativity on loss trials occurs when positive reward information decreases the amplitude of the N2 on gain trials. Others have argued that a frontal P2 is elicited by gain stimuli and slightly precedes the FRN associated with losses [21]. The hypothesis that the gain–loss effect yielding the FRN is really a positivity that happens to coincide with the N2 is difficult to evaluate without additional experimentation that can establish the independence of these components.

## Conclusion

In this study, we found that when gain and loss feedback stimuli can be discriminated on the basis of a salient visual feature, the FRN is large; otherwise the FRN is not observed. In future studies, researchers must control the perceptual properties of the feedback stimuli when they use the FRN to investigate decision-making or reward-related processing.

## References

1 Miltner WHR, Braun CH, Coles MGH. Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a

'generic' neural system for error detection. *J Cogn Neurosci* 1997; **9**:788–798.

2 Gehring WJ, Willoughby AR. The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 2002; **295**: 2279–2282.

3 Hajcak G, Holroyd CB, Moser JS, Simons RF. Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 2005; **42**:161–170.

4 Yeung N, Holroyd CB, Cohen JD. ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cereb Cortex* 2005; **15**:535–544.

5 Yeung N, Sanfey AG. Independent coding of reward magnitude and valence in the human brain. *J Neurosci* 2004; **24**:6258–6264.

6 Holroyd CB, Coles MG. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev* 2002; **109**:679–709.

7 Falkenstein M, Hohnsbein J, Hoormann J. Late visual and auditory ERP components and choice reaction time. *Biol Psychol* 1993; **35**: 201–224.

8 Gehring WJ, Goss B, Coles MG, Meyer DE, Donchin E. A neural system for error detection and compensation. *Psychol Sci* 1993; **4**:385–390.

9 Hajcak G, Moser JS, Holroyd CB, Simons RF. The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biol Psychol* 2006; **71**:148–154.

10 Holroyd CB, Nieuwenhuis S, Yeung N, Cohen JD. Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport* 2003; **14**:2481–2484.

11 Nieuwenhuis S, Yeung N, Holroyd CB, Schurger A, Cohen JD. Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cereb Cortex* 2004; **14**:741–747.

12 Donkers FCL, Nieuwenhuis S, van Boxtel GJM. Mediofrontal negativities in the absence of responding. *Cogn Brain Res* 2005; **25**:777–787.

13 Meyer DE, Kieras DE. A computational theory of executive cognitive processes and multiple-task performance. 2. Accounts of psychological refractory-period phenomena. *Psychol Rev* 1997; **104**:749–791.

14 Meyer DE, Kieras DE. A computational theory of executive cognitive processes and multiple-task performance. 1. Basic mechanisms. *Psychol Rev* 1997; **104**:3–65.

15 Anderson JR. *Rules of the Mind*. Hillsdale, NJ: Erlbaum; 1993.

16 Folstein JR, Van Petten C. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 2008; **45**:152–170.

17 Gehring WJ, Gratton G, Coles MGH, Donchin E. Probability Effects on stimulus evaluation and response processes. *J Exp Psychol Hum Percept and Perform* 1992; **18**:198–216.

18 Treisman AM, Gelade G. Feature-Integration theory of attention. *Cogn Psychol* 1980; **12**:97–136.

19 Gratton G, Coles MGH, Donchin E. A new method for off-line removal of ocular artifact. *Electroencephalogr Clin Neurophysiol* 1983; **55**:468–484.

20 Holroyd CB, Coles MGH. Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behaviour. *Cortex* 2008; **44**:548–559.

21 Potts GF, Martin LE, Burton P, Montague PR. When things are better or worse than expected: the medial frontal cortex and the allocation of processing resources. *J Cogn Neurosci* 2006; **18**:1112–1119.