

# Sensitivity of Electrophysiological Activity from Medial Frontal Cortex to Utilitarian and Performance Feedback

Sander Nieuwenhuis<sup>1</sup>, Nick Yeung<sup>1</sup>, Clay B. Holroyd<sup>1</sup>, Aaron Schurger<sup>1</sup> and Jonathan D. Cohen<sup>1,2</sup>

<sup>1</sup>Department of Psychology, Princeton University, Princeton, NJ, USA and <sup>2</sup>Center for the Study of Brain, Mind, and Behavior, Princeton University, Princeton, NJ, USA

**A recent study has reported the observation in humans of an event-related brain potential component that is sensitive to the value of outcomes in a gambling task. This component, labeled medial frontal negativity (MFN), was most pronounced following monetary losses as opposed to monetary gains. In this study, we investigate the relationship between the MFN and the error-related negativity (ERN), a component elicited by feedback indicating incorrect choice performance. We argue that the two components can be understood in terms of a recently proposed theory that predicts the occurrence of such scalp negativities following stimuli that indicate that ongoing events are worse than expected. The results from two experiments using a gambling task demonstrate that the sensitivity of the MFN/ERN to the utilitarian and performance aspect of the feedback depends on which aspect is most salient. The results are consistent with the view that the two components are manifestations of the same underlying cognitive and neural process.**

**Keywords:** anterior cingulate cortex, ERP, error negativity, error-related negativity, gambling, utility

## Introduction

Gehring and Willoughby (2002a) have reported the observation of a negative event-related potential (ERP) component that is sensitive to feedback stimuli indicating monetary gains or losses. The participants in their study performed a gambling task. On each trial they were asked to choose one of two squares, each of which contained the numeral 5 or 25, indicating US cents. Each square then randomly turned either red or green. If the chosen stimulus turned one color (e.g. green), this meant that the participant gained the indicated number of cents. If the chosen stimulus turned the other color (e.g. red), this meant that the participant lost the indicated amount. Comparison of the ERP for all gain trials with the ERP for all loss trials revealed a negative component, peaking ~250 ms after the feedback, that was larger following losses than following gains. Dipole modeling of the midline–frontal scalp distribution of the component suggested a source in medial frontal cortex, in or near anterior cingulate cortex. Because of its purported medial-frontal generator, Gehring and Willoughby labelled the ERP component the *medial-frontal negativity* (MFN).

The negative brain potential reported by Gehring and Willoughby (2002a) bears a strong resemblance to an ERP component, the error-related negativity (ERN or Ne; Falkenstein *et al.*, 1991; Gehring *et al.*, 1993), that occurs following response errors in speeded response tasks (Falkenstein *et al.*, 2000; Yeung *et al.*, 2004), and following feedback stimuli indicating incorrect performance (Miltner *et al.*, 1997; Holroyd and Coles, 2002; Nieuwenhuis *et al.*, 2002; Ruchsov *et al.*, 2002).

Importantly, the ERN associated with negative feedback is similar in morphology, scalp topography and latency to the MFN reported by Gehring and Willoughby. In addition, although one study has produced inconsistent results (Luu *et al.*, 2003), most evidence suggests the anterior cingulate cortex as the likely generator of the feedback ERN (Miltner *et al.*, 1997; Holroyd and Coles, 2002; Ruchsov *et al.*, 2002; Ullsperger and von Cramon, 2003). These resemblances between the MFN and ERN, and their seemingly similar circumstances of occurrence, raise the question of whether the two components are perhaps manifestations of the same cognitive and neural process. This possibility is consistent with the theory of Holroyd and Coles (2002) that the ERN is generated whenever unexpected negative events occur. The theory predicts that one should observe an ERN following negative feedback regardless of whether the feedback indicates an incorrect response or a monetary loss, since both can indicate that events are worse than expected (Holroyd *et al.*, 2002).

Despite the clear similarities between the ERN and MFN, Gehring and Willoughby (2002b) suggested that they are not identical phenomena: They point out that the scalp topography of the MFN appears to have a more frontal distribution than the ERN observed in previous studies, suggesting that their underlying neural generators are different. Moreover, they showed that while the negative ERP component observed in their study was sensitive to the gain/loss (i.e. *utilitarian*) value of the chosen outcome, it was insensitive to whether subjects made a correct or incorrect choice (Gehring and Willoughby, 2002a). Correctness was defined in terms of whether the subject's chosen outcome was better or worse than the alternative outcome. For instance, a gain of +5 indicated an incorrect choice if the alternative outcome was +25. The finding that the MFN was insensitive to the correct/incorrect (i.e. *performance*) value of performance feedback suggests that the MFN does not reflect the kind of error-processing system that the feedback ERN is thought to index (Miltner *et al.*, 1997; Coles *et al.*, 2001). These results seem on the surface to be evidence that the ERN and MFN are distinct components.

However, a weakness of Gehring and Willoughby's design (Gehring and Willoughby, 2002a) was that the utilitarian information was easier to extract from the feedback display than the performance information: the most salient feature of the feedback, the color of the display, directly conveyed the utilitarian value of the chosen outcome. In contrast, determining whether the chosen outcome was the better or worse of the two required comparing two numbers and their associated colors, a cognitive operation that is more attention demanding and presumably also more time consuming. Thus, although Gehring and Willoughby's findings are consistent with their conclusion that the ERN and MFN are generated by separate

systems, the findings are also consistent with the idea that these components are produced by a single monitoring system that responds to basic, salient information in the environment about whether outcomes are good or bad (Holroyd and Coles, 2002): When the most salient aspect of feedback information indicates gain/loss, as in Gehring and Willoughby's study, the monitoring system is sensitive to this information. When the feedback indicates good or bad outcomes along a correct/incorrect dimension, as in typical studies of the ERN (Miltner *et al.*, 1997), the monitoring system behaves like an error-detecting system.

To test this account of Gehring and Willoughby's findings (Gehring and Willoughby, 2002a), we ran two experiments using a modified version of their gambling task. In both experiments, '+' and '-' symbols indicated a gain or loss of the chosen numeral (see Fig. 1A). Thus, gain/loss information was readily available in the feedback display for all participants. However, in one experiment, the colors green and red in the feedback display were used to emphasize the utilitarian (gain/loss) value (Experiment 1). This condition was essentially a replication of Gehring and Willoughby's experiment, and thus we expected to replicate their finding of a frontal midline negativity that is primarily sensitive to utilitarian information. The critical prediction concerned the results of Experiment 2, in which utilitarian information was available to the participants but performance (correct/error) information was made salient through the color of the feedback display. That is, participants saw one color (e.g. green) if the chosen outcome was better than the alternative outcome, and one color (e.g. red) if the chosen outcome was the worse of the two outcomes. We predicted that in this design an ERN-like negativity would be observed. That is, we expected that the monitoring system would be primarily sensitive to whether the chosen outcome was the better or worse of the two, regardless of whether this involved a win or a loss of money. In contrast, if there is a system that is truly sensitive to utilitarian information, the pattern of results should be similar to that found by Gehring and Willoughby, with an MFN-like negativity sensitive to the utilitarian value of the chosen outcome, and not to the correctness of the participant's choice.

To look ahead briefly, our results were consistent with the hypothesis that the ERN and MFN reflect the operation of a common system of performance monitoring: the scalp negativity observed in Experiment 1 showed sensitivity to utility information, while the scalp negativity observed in Experiment 2 was sensitive to whether participants' choices were correct or incorrect. This also allowed us to directly compare for the first time the scalp topography of the MFN and the ERN. This is important since the MFN and feedback ERN have been measured by different research groups using different equipment and different tasks. It seems that a more direct comparison is needed to evaluate potential differences between the scalp distributions. We therefore attempted such a comparison.

In sum, the present research aimed to test the hypothesis that the ERN and MFN reflect the operation of a common monitoring system that provides a rapid evaluation of ongoing events. According to this hypothesis, negative scalp potentials will be generated whenever the monitoring system signals that outcomes are worse than expected, e.g. following incorrect responses or monetary losses (Holroyd and Coles, 2002). Evidence consistent with this hypothesis would provide impor-

tant clues about the nature of the neural mechanisms contributing to human judgment and decision making.

## Materials and Methods

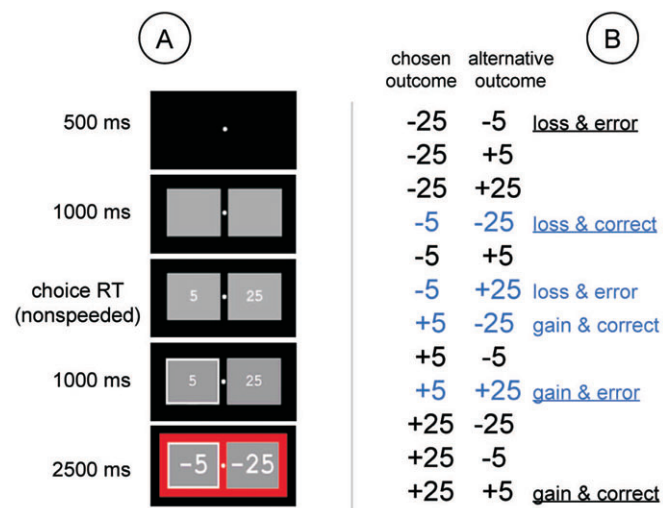
Except where indicated, the details of the gambling task were the same in Experiment 1 and 2.

### Participants

Participants in Experiment 1 were 14 young adults (eight women), ranging in age from 18 to 24 years (mean 22.0 years). Participants in Experiment 2 were 12 young adults (six women), ranging in age from 18 to 26 years (mean 22.4 years). All participants had normal or corrected-to-normal vision. They were paid \$25 as a basic salary, plus a feedback-related bonus, as described below. The experiment consisted of a single 2.5 h session.

### Task

Each trial (see Fig. 1A) started with the presentation of a fixation point, which remained on the screen during the whole trial. After 500 ms, two rectangles appeared on either side of the fixation point, marking the locations of the upcoming choice alternatives. After 1 s, the numeral '5' or '25' (indicating US cents) was presented in each of the rectangles. Participants then selected one of the two choice alternatives by pressing a corresponding response button with their left or right index finger. This choice was highlighted by a thickening of the white outline of the corresponding rectangle. One second after the choice response, the chosen and alternative outcomes were displayed by revealing the sign (+ or -) of each numeral. To emphasize the gain/loss (Experiment 1) or the correct/error (Experiment 2) value of the chosen outcome, a colored rectangle, red or green, was displayed around the two outcomes. [The nature of the feedback display differed somewhat from that used by Gehring and Willoughby (2002a). In their experiment, color was the only indication of whether each outcome was associated with a gain or a loss of money (i.e. there were no '+' and '-' symbols). We used '+' and '-' symbols to indicate the valence of the outcomes, so that color could be used independently to *emphasize* either the gain/loss dimension or the



**Figure 1.** (A) Example of stimulus events in the gambling task. The duration of each stimulus event is indicated. See text for details. (B) List of possible combinations of chosen outcome and alternative outcome. The blue conditions indicate the four conditions chosen here to analyze the effects of gain versus loss and correct versus error. The underlined conditions indicate the four conditions chosen by Gehring and Willoughby (2002a). 'Loss' and 'gain' indicate that the chosen outcome yielded a financial penalty or reward, respectively. 'Error' indicates that the alternative outcome would have yielded a larger reward or a smaller penalty, relative to the chosen outcome. 'Correct' indicates that the alternative outcome would have yielded a smaller reward or a larger penalty, relative to the chosen outcome.

correct/error dimension of the feedback.] The feedback display remained visible for 2.5 s, after which the next trial started.

Before the experiment, participants were informed that the value of each chosen outcome would be added to (or subtracted from) the total amount of bonus money awarded to them at the end of a block of trials. They were also instructed about the meaning of the color in the feedback display. Furthermore, they were told that their goal was to earn as much money as possible, and that they were free to choose any strategy to achieve that goal. Unbeknownst to the subjects, feedback was provided according to a prespecified pseudorandom sequence. Because the outcomes were determined randomly and therefore there was no strategy to learn, there was no meaningful performance measure in this task. Instead, the task simply provided a realistic context in which rewards and penalties, and correct and incorrect choices were experienced. Nevertheless, at debriefing, most participants reported that they had attempted to find a systematic pattern in the sequence of feedback, and that they had felt disappointed when testing of a specific hypothesis regarding this sequence led to negative outcomes.

### Stimuli

Stimuli (see Fig. 1A) were presented against a black background on a computer screen placed at a distance of ~150 cm from the participant. The fixation point was white and subtended  $0.4^\circ$ . The two rectangles on either side of the fixation point were gray with a thin white border. Each subtended  $2.3^\circ \times 3.2^\circ$ , and the visual angle between the centers of the rectangles was  $\sim 3.6^\circ$ . The numerals were presented in a white, 28 point, bold Courier font and subtended  $\sim 0.6^\circ$  vertically. The numerals in the feedback display were presented in font size 52. The colored rectangle that was displayed around the outcomes subtended  $5.3^\circ \times 9.6^\circ$ .

### Design and Procedure

The two outcomes on each trial, presented in the left and right rectangle, were never the same [e.g. (-25, -25)]. The remaining 12 possible combinations [(-25, -5), (-25, +5), etc.] were presented equally often across the experiment. The 12 experimental conditions that guided the data analysis (see Fig. 1B) were defined by crossing the four possible *chosen* outcomes (+25, +5, -5 and -25) with the three possible *alternative* outcomes given the chosen outcome (e.g. +25, +5 and -25 when the chosen outcome was -5). In Experiment 1, the color of the feedback display emphasized the utilitarian (gain/loss) value of the feedback. For half of the participants, the color green in the feedback display emphasized a positive chosen outcome (i.e. a gain of money), and the color red emphasized a negative chosen outcome (i.e. a loss). For the other half of the participants, this assignment was reversed. In Experiment 2, the color of the feedback display emphasized the performance (correct/error) value of the feedback. For half of the participants, the color green meant that they had chosen the better of the two outcomes (i.e. a correct choice), and the color red meant that they had chosen the worse of the two outcomes (i.e. an incorrect choice). For the other half of the participants, this assignment was reversed. Participants received 16 practice trials before entering the experimental phase, which consisted of 16 blocks of 36 trials each. At the end of each block, participants received visual feedback indicating the amount of money earned in that block and the accumulated total across blocks. When a block resulted in a net loss, this total was not subtracted from the accumulated total across blocks. There were 5 min breaks after every fourth block.

### Psychophysiological Recording and Data Analysis

EEG recordings were taken from 32 Ag/AgCl electrodes placed in an extended 10–20 system montage, referenced to linked mastoids. The electro-oculogram (EOG) was recorded from electrodes placed above and below the left eye, and from electrodes placed on the outer canthi of each eye. The ground electrode was placed on the chin. All electrode impedances were kept below 50 k $\Omega$ . The EEG signals were digitized at 250 Hz.

Single-trial epochs were extracted offline for a period from 100 ms before until 600 ms after the feedback stimuli. The EMCP method (Gratton *et al.*, 1983) was used to correct for EOG artifacts and to discard trials with recording artifacts. A prestimulus period of 100 ms

was subtracted as a baseline. For each participant and each condition, the EEG epochs were averaged with respect to feedback onset. Before subsequent analyses, the resulting ERP waveforms were lowpass filtered (<12 Hz) using a second-order digital Butterworth filter. ERN amplitude was defined as the average value of the signal at electrode FCz in the window 210–310 ms following the feedback, relative to a 100 ms prestimulus baseline. To facilitate our discussion of the results, we use the label 'ERN' rather than 'MFN' to describe both the gain/loss-related and the correct/error-related modulation of the ERP following the feedback. ERN data were analyzed using repeated measures analyses of variance (ANOVA) with factor condition (gain and correct, gain and error, loss and correct, loss and error; see below). Separate ANOVAs were conducted to statistically test whether the scalp distribution of the ERN differences of interest differed as a function of electrode location. Factors were anterior-posterior (F, FC, C, CP, P) and lateralization (3, z, 4). The Greenhouse-Geisser correction for violations of the ANOVA assumption of sphericity was applied where appropriate.

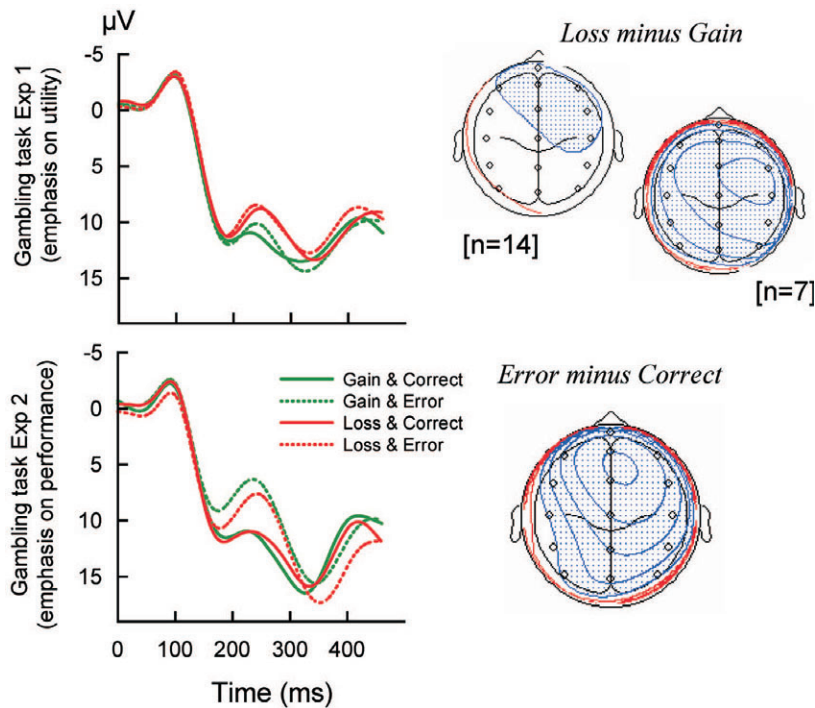
Following Gehring and Willoughby (2002a), we chose four conditions that allowed us to separately assess the effects of gain versus loss and correct versus error on ERN amplitude (see Fig. 1B). A first condition consisted of trials on which participants lost 5 $\epsilon$  and the alternative outcome was a loss of 25 $\epsilon$ . We call this the 'loss & correct' condition, because participants lost money but made a correct choice because they would have lost even more had they chosen the alternative outcome. A second condition comprised all trials on which participants lost 5 $\epsilon$  and the alternative outcome was a gain of 25 $\epsilon$ . We call this the 'loss and error' condition, because participants lost money and made an incorrect choice because the alternative outcome would have resulted in a gain of money. Following the same logic, the other two conditions were called 'gain and correct' (when participants gained 5 $\epsilon$  and the alternative was losing 25 $\epsilon$ ) and 'gain and error' (when participants gained 5 $\epsilon$ , but would have won 25 $\epsilon$  had they made the alternative choice). We evaluated the scalp topographies only for these four conditions: because gain versus loss and correct versus error were orthogonally varied, this allowed us to construct ERP difference waves that reflected the effect of one variable (e.g. gain versus loss) while controlling for the other variable (e.g. correct versus error). In contrast, since gains were generally associated with a correct choice, whereas losses were generally associated with an erroneous choice (see Fig. 1B), inclusion of all 12 conditions in the construction of the difference waves would have led to scalp topographies in which the effects of the two variables were confounded (cf. Gehring and Willoughby, 2002a).

Our 'gain and error' and 'loss and correct' conditions were the same as those used by Gehring and Willoughby (2002a), but we focused on different 'gain and correct' and 'loss and error' conditions. As can be seen in Figure 1B, Gehring and Willoughby's correct versus error comparison was confounded with differences in the amount of gain or loss. For instance, the 'gain and correct' condition was associated with a gain of 25 $\epsilon$ , whereas the 'gain and incorrect' condition was associated with a gain of 5 $\epsilon$ . Importantly, it has been shown that differences in absolute reward magnitude affect the amplitude of the P300 (Sutton *et al.*, 1978; N. Yeung and A. Sanfey, submitted). This effect is also apparent in Gehring and Willoughby's data, and in our data, as is evident in Figure 3. Because the feedback ERN is usually superimposed on the P300, differences between conditions in P300 amplitude complicate the measurement of ERN amplitudes. Therefore, although we will show the data from the four conditions chosen by Gehring and Willoughby (Fig. 3), the major part of our analyses involves a comparison of conditions that do not involve the confound discussed here.

## Results

### Experiment 1: Emphasis on Utility

In the gambling task of Experiment 1, color was used to emphasize the utilitarian value of the chosen outcome in the feedback display. The basic results, shown in Figure 2 (top

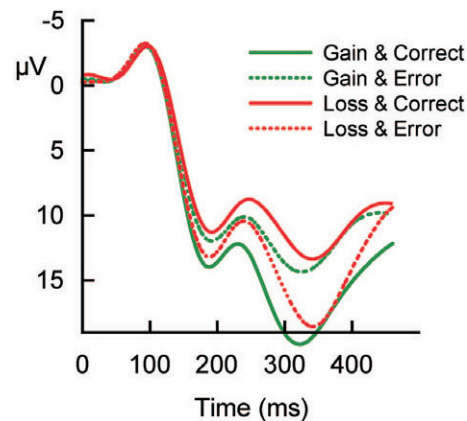


**Figure 2.** (Left panel) Feedback-locked ERPs at electrode FCz. Data from the gambling task in Experiment 1 (top) and 2 (bottom) are shown for four of the conditions. (Right panel) Voltage maps representing the scalp topography of the ERN difference waveform at its peak latency. The subtraction underlying each difference waveform is indicated next to each map. Blue regions indicate negative values, red regions indicate positive values. The decrease/increase in voltage represented by each isopotential line is  $0.04 \mu\text{V}$ .

panel), replicate those of Gehring and Willoughby. The ERN was more pronounced on loss trials than on gain trials [mean =  $9.8 \mu\text{V}$  versus  $11.8 \mu\text{V}$ ,  $F(1,13) = 16.6$ ,  $P < 0.001$ ], but did not substantially differ between correct and error trials [mean =  $11.0 \mu\text{V}$  versus  $10.7 \mu\text{V}$ ,  $F(1,13) = 1.5$ ,  $P = 0.38$ ]. Although significant, the average gain/loss effect was not very substantial. Indeed, the effect was absent in many individual participants. The small gain/loss effect observed is not simply a function of the conditions we chose for the analysis: as shown in Figure 3, the gain/loss effect was similarly small when examined in the conditions that were analyzed by Gehring and Willoughby. Inspection of the voltage map in Figure 2 reveals that the small ERN effect size is reflected in the scalp topography of the loss minus gain difference wave: The scalp distribution was almost flat. This was partly due to individual differences in the scalp distribution of the effect. We thought that a more interpretable scalp pattern should emerge if we focused our analysis on the participants ( $n = 7$ ) who showed the largest gain/loss effect (mean effect size =  $-3.5 \mu\text{V}$  versus  $-0.4 \mu\text{V}$  for the seven participants with the smallest effect). As can be seen in Figure 2 (top, rightmost panel), the resulting scalp map showed a somewhat right-lateralized frontocentral distribution.

### Experiment 2: Emphasis on Performance

In the gambling task of Experiment 2, color was used to emphasize the correct/error value of the chosen outcome in the feedback display. ERN amplitude in the four chosen conditions showed a qualitatively different pattern than in Experiment 1 (see Fig. 2, bottom panel). There was a clear effect on ERN



**Figure 3.** Feedback-locked ERPs at electrode FCz from the gambling task in Experiment 1. Data are shown for the conditions that were compared by Gehring and Willoughby (2002a). These data also illustrate the effect of differences in absolute reward magnitude on P300 amplitude, and the resulting difficulty of measuring ERN amplitude.

amplitude of correct versus error [mean =  $12.4 \mu\text{V}$  versus  $9.0 \mu\text{V}$ ,  $F(1,11) = 30.9$ ,  $P < 0.001$ ], but no effect of gain versus loss [mean =  $10.6 \mu\text{V}$  versus  $10.8 \mu\text{V}$ ,  $F(1,11) = 0.1$ ,  $P = 0.72$ ]. Unlike the gain/loss effect, the correct/error effect was present in most of the participants. The voltage map of the error minus correct difference wave indicated a frontocentral, slightly right-lateralized distribution. A pair-wise comparison confirmed that the ERN was larger over the right (electrodes '4') than over the left hemisphere (electrodes '3'),  $P = 0.003$ .

### Between-experiment Comparisons

Statistical between-experiment comparisons of the gambling task data confirmed that the gain/loss effect on ERN amplitude was larger when the utilitarian (gain/loss) value of the chosen outcome was emphasized (Experiment 1) than when the performance (correct/error) value of the chosen outcome was emphasized (Experiment 2),  $F(1,24) = 9.4$ ,  $P = 0.005$ . As predicted, the converse was true for the correct/error effect on ERN amplitude,  $F(1,24) = 14.7$ ,  $P = 0.001$ .

We also attempted to quantify the similarity between the scalp topographies of the gain/loss effect in Experiment 1 (based on the participants with the largest effect) and the correct/error effect in Experiment 2. To this end, we calculated the best-fitting regression line between the amplitudes of the two effects across electrodes (Yeung *et al.*, 2004). This analysis indicated that there was a high degree of similarity between the topographies,  $r = 0.67$ ,  $P = 0.006$ . We then performed a between-experiment ANOVA on the normalized voltage distributions to further compare the topographies. In line with previous analyses, this test revealed that the topographies did not reliably differ in terms of anterior-posterior orientation or lateralization, both  $F < 1$ .

### Discussion

The recent results of Gehring and Willoughby (2002a) have suggested a dissociation between two electrophysiological markers of medial frontal cortex involvement in processing external evaluative feedback: one ERP component, the MFN, that is sensitive to utility information in general and to losses in particular; and another ERP component, the ERN, that is associated with the evaluation of performance along a correct-error dimension. The ERN and MFN have a similar morphology and supposed neural generator in anterior cingulate cortex (Miltner *et al.*, 1997; Gehring and Willoughby, 2002a). Moreover, both scalp negativities occur very quickly following the feedback (~250 ms). The early latency of the two components raises the possibility that each is sensitive primarily to basic and salient evaluative information in the feedback stimulus. We conducted two experiments using a modified version of Gehring and Willoughby's gambling task, manipulating the relative salience of the utilitarian and performance (i.e. correct/error) aspect of the feedback. In both experiments, the feedback-evoked negativity was sensitive only to the most salient aspect of the feedback display: gain/loss in Experiment 1 and correct/error in Experiment 2. The scalp topography of these negativities did not reliably differ within the limited spatial resolution of our measurements.

In Gehring and Willoughby's study, the most salient information in the feedback display was the utilitarian (i.e. gain/loss) value of the chosen outcome, and the observed negativity was duly sensitive to this aspect of the feedback (Gehring and Willoughby, 2002a). In contrast, the negative component was insensitive to performance (i.e. correct/incorrect) information. The results of our study suggest that Gehring and Willoughby's findings may reflect the nature of the feedback stimuli used in their experiment. We found that when the salient aspect of the feedback signaled the performance value of the chosen outcome, the observed negativity was sensitive to this aspect of the feedback. Of course, our results cannot definitively rule out the possibility that the MFN and ERN reflect the activity of two different systems: one system involved in the rapid compu-

tation of utility, and one system involved in the rapid detection of errors. The activity of these systems may happen to be expressed in scalp negativities that are similar in timing, appearance, scalp topography, and purported generator. Indeed, in general it is difficult to exclude the possibility that two seemingly identical phenomena may actually reflect the operation of related, but dissociable cognitive and neural systems. However, if the two systems for processing gain/loss and correct/error information are operating in parallel, one would predict that we should see effects of both gain/loss and correct/error in both experiments, and these effects should sum. For example, we should see an effect of gain/loss in the experiment in which correct/error information is emphasized (Experiment 2), since the gain/loss information is clearly and explicitly specified in the feedback (as '+' and '-' symbols). However, in neither of the experiments do we observe such a graded effect, nor do Gehring and Willoughby in their experiment.

Therefore, the data are perhaps more naturally explained by the view that the MFN and feedback ERN reflect the activity of a single monitoring system that rapidly evaluates outcomes along a good-bad dimension on the basis of the most salient evaluative information in the environment. This view is consistent with a recently proposed theory (Holroyd and Coles, 2002), according to which the ERN is elicited when the monitoring system first detects that ongoing events (e.g. the outcomes of our actions) are worse than expected. In the context of this 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 (Holroyd *et al.*, 2002).

We believe that our account is more parsimonious than the two-systems account. One way in which our account might be disconfirmed is by establishing that the neural source of the feedback ERN and MFN are different; our prediction is that they have the same neural generator. Our finding that the scalp topographies of the two components did not reliably differ is suggestive rather than conclusive in this regard. The gain/loss effect was relatively small and was not consistently observed in individual participants. Moreover, the scalp topographies showed variability across participants, particularly so for the participants showing a gain/loss effect. As a consequence of these factors, the scalp distribution of the average gain/loss effect was essentially flat – there was not much effect to distribute. Confining the scalp topography to a subgroup with the strongest gain/loss effects yielded a clearer pattern, but this result is necessarily less powerful. Thus, the observed small difference between the scalp topographies in Figure 2 may simply be due to factors such as sampling error and low signal-to-noise ratio. However, the observed difference may also reflect a real difference in scalp topography, consistent with a two-systems account. Discrimination between these possibilities may require different methods, such as high-density EEG recordings or functional magnetic resonance imaging. Yet, although the scalp topography results do not allow us to draw any strong conclusions, it is interesting that both the gain/loss effect and the correct/error effect showed a somewhat right-lateralized frontocentral scalp distribution. The lateralization of the ERN is rather unusual, and may suggest the contribution to the scalp distribution of other reward-related neural activity that is specific to the gambling task used here.

The difference in sensitivity of the ERN between the two experiments was brought about by only a small change in the feedback display: the meaning of the background color. Indeed, the background color provided information that was completely redundant; the utilitarian value and performance value of the chosen outcome could be derived from the numbers and their valence, as indicated by the '+' and '-' symbols. Yet, the effect of manipulating color was large: In both experiments, almost all of the variance in ERN amplitude could be explained in terms of the variable emphasized by the color of the feedback display. Thus, the results are consistent with our hypothesis that the system underlying the ERN uses the most salient information available in the feedback display. This seems an essential property of a system that has been argued to play an important role in rapid, online adjustments of behavior (Gehring *et al.*, 1993). This is not to say that participants were not interested in how much they won or lost. We know that participants cared about (or at least paid attention to) the absolute magnitude of the chosen outcome ( $\pm 25$  versus  $\pm 5$ ), because in both experiments P300 amplitude was highly sensitive to this variable (Sutton *et al.*, 1978; N. Yeung and A. Sanfey, submitted). An important question for future research is the way in which rapidly available evaluative information (as reflected in the ERN) is integrated with more complex information, such as absolute magnitude of the outcome (as reflected in the P300).

We have recently begun to explore the effects of *context* on the amplitude of the feedback ERN. By context we mean the set of experimental factors that affect how a particular feedback stimulus is evaluated. The function that determines how a subjective value (i.e. along a good-bad dimension) is attributed to actual feedback outcomes is not straightforward. For instance, it is well known that people's subjective value does not increase linearly as a function of amount of monetary gain (Von Neumann and Morgenstern, 1947). In our view, context affects the function relating actual outcomes to subjective value, and therefore, indirectly, ERN amplitude. In a recent paper (Holroyd *et al.*, 2004), we have shown that ERN amplitude to a specific outcome is substantially modulated by the range of other possible outcomes in the task (i.e. shows *context dependence*). For instance, feedback indicating neither gain nor loss elicited a larger ERN in a condition in which this outcome constituted the worst possible outcome, than in a condition in which it was the best possible outcome. The present study suggests another form of context dependence of the feedback ERN. The study suggests that the emphasis placed on utilitarian and performance aspects of the feedback can influence the relative weighing of these two factors in computing the subjective value associated with the feedback.

Finally, we note that our results do not speak to the question of whether the ERN is a direct manifestation of the cognitive system that we have outlined above, or instead is a manifestation of a neural circuit involved in assessing the *emotional impact* of outcomes (Luu *et al.*, 2000; Gehring and Willoughby, 2002a). The emotion hypothesis leaves open the question *how* negative outcomes are detected in the brain; this function may be carried out by a system for reinforcement learning, which in turn may provide the input for a system involved in emotional and motivational functioning. Indeed, cognitive aspects of performance monitoring may be so intricately linked to emotional consequences that it may be impossible to tease

apart cognitive and emotional correlates of performance monitoring. If that is the case, the debate whether the ERN reflects cognitive or emotional processing becomes moot.

## Notes

This research was supported by the Netherlands Organization for Scientific Research (S.N.) and by postdoctoral fellowship MH63550 from the National Institute of Mental Health (C.B.). We thank Dustin Reagan for technical assistance, and Richard Ridderinkhof for helpful comments on an earlier draft of the manuscript.

Address correspondence to Sander Nieuwenhuis, Department of Cognitive Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. Email: stn20@dds.nl.

## References

- Coles MG, Scheffers MK, Holroyd CB (2001) Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biol Psychol* 56:173-189.
- Falkenstein M, Hohnsbein J, Hoormann J, Blanke L (1991) Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalogr Clin Neurophysiol* 78:447-455.
- Falkenstein M, Hoormann J, Christ S, Hohnsbein J (2000) ERP components on reaction errors and their functional significance: a tutorial. *Biol Psychol* 51:87-107.
- Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E (1993) A neural system for error detection and compensation. *Psychol Sci* 4:385-390.
- Gehring, WJ and Willoughby, AR (2002a) The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295:2279-2282.
- Gehring WJ, Willoughby AR (2002b) Medial prefrontal cortex and error potentials – Reply to Holroyd, Coles, & Nieuwenhuis. *Science* 296:1610-1611.
- Gratton G, Coles MGH, Donchin E (1983) A new method for off-line removal of ocular artifact. *Electroencephalogr Clin Neurophysiol* 55:468-484.
- Holroyd, CB and Coles, MGH (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev* 109:679-709.
- Holroyd CB, Coles MGH, Nieuwenhuis S (2002) Medial prefrontal cortex and error potentials. *Science* 296:1610-1611.
- Holroyd CB, Larsen JT, Cohen JD (2004) Context dependence of the event-related brain potential to reward and punishment. *Psychophysiology* (in press).
- Luu P, Collins P, Tucker DM (2000) Mood, personality, and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *J Exp Psychol Gen* 129:43-60.
- Luu P, Tucker DM, Derryberry D, Reed M, Poulsen C (2003) Electrophysiological responses to errors and feedback in the process of action regulation. *Psychol Sci* 14:47-53.
- Miltner WHR, Braun CH, Coles MGH (1997) Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a generic neural system for error detection. *J Cogn Neurosci* 9:788-798.
- Nieuwenhuis S, Ridderinkhof KR, Talsma D, Coles MGH, Holroyd CB, Kok A, Van der Molen MW (2002) A computational account of altered error processing in older age: Dopamine and the error-related negativity. *Cognit Affect Behav Neurosci* 2:19-36.
- Ruchow 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. *Neurosci Lett* 325:203-206.
- Sutton S, Tueting P, Hammer M, Hakerem G (1978) Evoked potentials and feedback. In: *Multidisciplinary perspectives in event-related potential research* (Otto DA, ed.), pp. 184-188. Washington, DC: Government Printing Office.
- Ullsperger, M and von Cramon, DY (2003) Error monitoring using external feedback: specific roles of the habenular complex, the

reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *J Neurosci* 23: 4308–4314.

Von Neumann, J and Morgenstern, O (1947) *Theory of games and economic behavior*, 2nd rev. edn. Princeton, NJ: Princeton University Press.

Yeung N, Botvinick MM, Cohen JD (2004) The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychol Rev* (in press).