In: *Errors, Conflict, and the Brain. Current Opinions on Performance Monitoring* (2004).
Edited by M. Ullsperger & M. Falkenstein. Leipzig: MPI of Cognitive Neuroscience, pp. 129-134.

Chapter 5: Feedback processing and Observation of Errors

What if I told you: "You were wrong"? Brain potentials and behavioral adjustments elicited by feedback in a time-estimation task

Rogier B. Mars[1,2], Ellen R.A. de Bruijn[1], Wouter Hulstijn[1], Wolfgang H.R. Miltner[3], and Michael G.H. Coles[2]
[1]*Nijmegen Institute for Cognition and Information*
[2]*F.C. Donders Centre for Cognitive Neuroimaging*
[3]*Department of Biological and Clinical Psychology, Friedrich-Schiller-University Jena*

Recent theories have associated the error-related negativity (ERN/Ne) with the arrival of an error signal in the anterior cingulate cortex (Holroyd & Coles, 2002). This error signal is generated when negative events occur, particularly when they are unexpected, and the anterior cingulate uses the error signal to select among appropriate courses of action. We evaluated these ideas by replicating and extending previous studies of the ERN following performance feedback in which subjects receive feedback after making a time-production judgment. In three different conditions, subjects received (1) correct or incorrect feedback, (2) correct, incorrect-slow, or incorrect-fast feedback, and (3) the same as condition (2), but with the graded incorrect feedback as a function of the degree of error. Behavioral data indicated that subjects adjusted their time-estimation as a function of feedback: following incorrect feedback in condition (2), they shortened or lengthened their judgments, and in condition (3) the amount of adjustment was related to the suggested degree of error. An ERN following negative feedback was present in all three conditions, being largest in the first condition. However, we did not find any relationship between ERN amplitude and behavioral adjustments. These results are discussed in terms of current theories on error processing.

Introduction

In order to lead a safe and productive life, human beings have to adjust their behavior to suit any particular situation. In particular, following an error, or following error feedback, adjustments have to be made to assure that the likelihood of future errors is minimized. In the laboratory, this phenomenon is evident in a slowing of reaction time after incorrect responses in choice-response tasks (Rabbitt, 1966).

The study of errors has recently been facilitated by the discovery of a component of the event-related brain potential, the error negativity (Ne) or error-related negativity (ERN), that accompanies the production of errors in choice reaction time tasks (Falkenstein et al., 1990; Gehring et al., 1993). This component has a peak latency of approximately 80 ms following the erroneous response and appears to be generated in the anterior cingulate cortex (ACC) (Dehaene et al., 1994; Holroyd et al., 1998). A similar component, also originating in the ACC, can be observed when subjects receive feedback indicating that they have made an error (e.g. Miltner et al, 1997; Badgayan & Posner, 1998). Consequently, this component is termed the feedback ERN.

In several early studies (e.g. Gehring et al., 1993; Coles et al., 1995), a relationship was found between the amplitude of the ERN and various examples of behavioral adjustments. These 'remedial actions' were reflected in the tendency to correct an error, to slow down following an error, and in the force of the error itself. However, to date, no one has investigated the relationship

between behavioral adjustments and the feedback ERN. One aim of the present study was to evaluate this relationship.
The theoretical framework for the study was provided by the 'Reinforcement Learning Theory' (Holroyd & Coles, 2002), which proposes that the ERN is associated with learning-relevant signals that are carried by the mesencephalic dopamine system (MDS). This model is based on the finding that the MDS carries reward signals indicating that ongoing evens are 'better' or 'worse' than expected (cf. Schultz, 2002). A monitoring system in the basal ganglia evaluates internal information about self-generated behaviors and external information from the environment, and predicts the expected value, or 'reward', of ongoing events. When the system revises its predictions for the worse, either because of an internally detected inappropriate motor action or upon receiving negative feedback, a phasic decrease of mesencephalic dopamine activity disinhibits apical dendrites of neurons in the ACC, resulting in an ERN (see Holroyd & Coles, 2002 for a more detailed description). Recent experimental evidence has supported the notion of dopamine influence in processes underlying the ERN (e.g. Holroyd et al., in press).
According to the Reinforcement Learning Theory, the prediction error signals are carried by the MDS to various brain areas, including the ACC, where they are used to adjust behavior to the task at hand: "The anterior cingulate cortex is trained to recognize the appropriate [motor] controller, with reinforcement learning signals conveyed to it via the mesencephalic dopamine system. We [...] assume that some of the motor controllers may themselves use those same reinforcement learning signals to identify the appropriate response strategy required of them" (Holroyd & Coles, 2002, p. 685).
In the present study, we further investigated the properties of the feedback ERN. First, we manipulated the information value of negative feedback in order to investigate how the system underlying generation of the feedback ERN reacts. Second, we were interested in the relationship between the feedback ERN and remedial actions. Although the Reinforcement Learning Theory states that the system learns from errors to adjust future behavior to suit the task at hand, it does not make any direct predictions on how this is reflected at the behavioral level. We addressed these questions in the context of a time-estimation task that has been used previously by Miltner et al. (1997) and Lemke (2003).

Materials and methods
*Subjects.*Eight subjects (6 female), ranging in age from 20 to 23 (M = 22.0) participated in the experiment. All participants had normal or corrected-to-normal vision. All subjects were paid 6 euros per hour plus a bonus depending on their performance. All participants provided written consent according to the institutional guidelines of the local ethics committee (CMO region Arnhem-Nijmegen, Netherlands).
Task. Subjects sat comfortably about 50 cm in front of a computer screen in an electrically shielded room. On each trial of the task, subjects saw a white star (angle .8 degrees) appear in the center of the screen. Participants were instructed to press a button with the index finger of their right hand when they estimated the star had been on screen for one second. Immediately following the button press, the star disappeared, and a blank screen was presented for 500 ms. Following this, feedback was provided for 500 ms.
In the first condition (one block of 150 trials), subject received feedback indicating '+6 cents' on correct trials and '-6 cents' on incorrect trials. In the second condition (one block of 150 trials), subjects received feedback indicating '+6 cents' and a white square underneath the text on correct trials, and '-6 cents' with an arrow underneath the text on incorrect trials. The arrow was pointing to the right when subjects were too fast in their estimation (indicating that they should 'respond later in time' in the future) and to the left when subjects were too slow in their estimation (indicating that they should 'respond earlier in time'). In the third condition (three blocks of 150 trials each), feedback was the same as in the second condition, except that the 'punishment' on incorrect trials was minus 2, minus 6, or minus 10 cents. Subjects were instructed that the larger the

punishment, the farther off their estimation was. In reality, each punishment level was presented one third of the negative feedback trials. Subjects were told they started with fifty cents bonus money and the money won or lost each trial would be added to or subtracted from their bonus money. The blocks of different conditions were presented in the order 3-1-3-2-3 or 3-2-3-1-3 (counterbalanced across subjects).

To keep the amount of errors equal in all conditions, we used a sliding criterion to determine if a response was correct or incorrect (cf. Miltner et al., 1997; Lemke, 2003). All subjects started with a criterion +/- 200 ms. Following correct trials the criterion was decreased 10 ms, following incorrect trials the criterion was increased 10 ms.

Data acquisition and analysis. Brain electrical activity was recorded from 61 Ag/AgCl electrodes, arranged equally over the scalp, referenced to linked earlobe references. Vertical and horizontal elecrooculograms were recorded from sites above and below the left eye and 1 cm external to the outer canthus of each eye. The electrode common was placed on the sternum. All electrode impedances were kept below 5 kΩ. EEG data were amplified using BrainAmp amplifiers and digitized at 250 Hz.

Data were filtered off-line with a .03-15 Hz bandpass filter. For each feedback type in each condition, a 700 ms epoch of data (100 ms baseline) was extracted for analysis. Ocular artifact was corrected using the procedure by Gratton et al. (1983) and waveforms of each electrode were checked for amplifier artifacts. Trials containing amplifier artifacts were discarded.

Feedback-locked average waveforms were computed for correct and incorrect trials for each condition and for each punishment level in the third condition. The feedback ERN was defined as the most negative peak, between 200 and 400 ms after feedback onset, in the difference wave of incorrect minus correct trials.

Results

In this section we will first present evidence from behavioral measures, indicating that the experimental manipulations were effective. We then consider the effects of these manipulations on the ERN and the relationship between the ERN and behavioral adjustment.

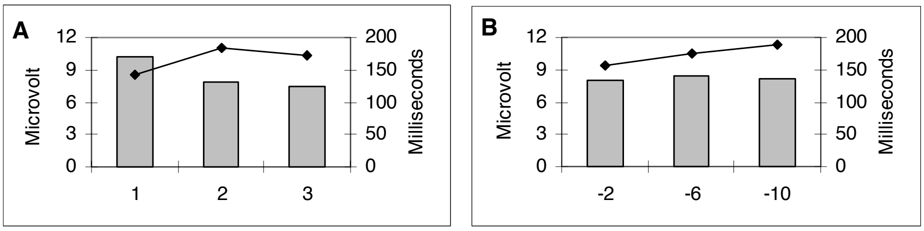

Figure 1.
Behavioral adjustments (lines) and absolute feedback ERN amplitudes (bars) across (A) different conditions and (B) different punishment levels.

Behavioral findings

A comparison of the error rates in the three different conditions showed that the sliding time window was effective. Subjects did not differ significantly in error rate in the different conditions ($F(2,14) = 1.576$, p = n.s.[1]). However, subjects were more accurate in their estimation of one second in the second and third conditions ($F(2,14) = 6.068$, $p = .034$), indicating that they made use of the feedback.

To study the use subjects made of the feedback, we looked at the absolute adjustment in time-estimation subjects made after incorrect feedback. As can be seen in Figure 1A, subjects made

[1] The Greenhouse-Geisser correction was applied when appropriate to correct for possible violations of the analysis of variance assumption of sphericity. The text lists corrected p values.

larger adjustments in the informative conditions ($F(2,14) = 11.563$, $p = .001$). Planned comparisons revealed that conditions two and three differed from the first conditions ($F(1,7) = 13.923$, $p = .007$ and $F(1,7) = 12.503$, $p = .010$), but not from each other ($F(1,7) = 2.376$, p = n.s.). Analysis of the data from condition three (see Figure 1B) indicated that subjects made larger adjustments when they received more punishment indicating they had made a larger error ($F(2,14) = 6.216$, $p = .012$). Follow-up analysis showed that a linear trend of increasing behavioral adjustment with increasing levels of punishment was significant ($F(1,7) = 8.217$, $p = .024$).
These results indicate that the subjects made use of the information provided by the feedback in adjusting their behavior from one trial to the next.

ERP findings
Figure 2 shows feedback-locked average waveforms for correct and incorrect trials in the different conditions. In all three conditions, a negative deviation is present for incorrect as compared to correct trials. This deviation reaches its maximum at electrode Cz, approximately 325 ms after feedback onset.
As shown in Figure 1A, the amplitude of the waveforms, as measured by the peak of the difference wave, differed among the conditions ($F(2,14) = 6.758$, $p = .009$)[2], being larger in the first condition then the other two, more informative, conditions ($F(1,7) = 8.958$, $p = .020$ and $F(1,7) = 7.501$, $p = .029$) Analysis of the ERN amplitude in condition three (Figure 1B) revealed no significant difference among the three punishment levels ($F(2,14) = .197$, p = n.s.).

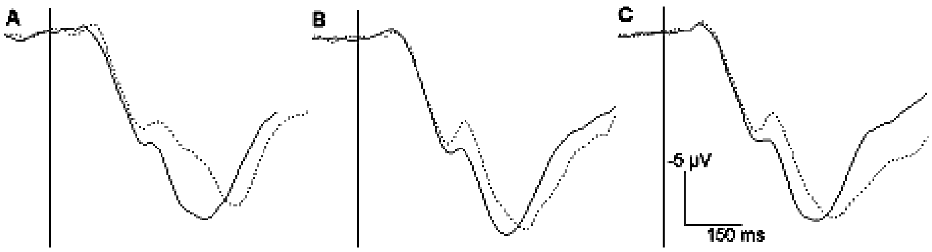

Figure 2.
Feedback-locked grand average waveforms for correct (solid lines) and incorrect (dotted lines) ERPs in the (A) first, (B) second, and (C) third condition. The vertical lines indicate feedback onset.

Relationship between ERN and behavioral adjustments
The relationship between ERN and behavioral adjustments was assessed directly in the following way. First, for each subject, all negative feedback trials in condition one were ordered as a function of the absolute adjustment in time-estimation on the following trial. Then four 'bins' were created representing four levels of adjustment. Third, the average ERP waveforms for each of the four bins were computed and the ERN amplitude for each bin was derived. Analysis of the relationship between bin and ERN amplitude failed to show any effect. There was no difference among the four bins in ERN amplitude ($F(3,21) = 1.498$, p = n.s.).
Another analysis examined the relationship between ERN amplitude and the quality of the subject's estimate on the trial following error feedback. Quality of estimate was defined as the absolute difference between the actual estimate and 1000 ms. As in the prior analysis, four bins

[2] It can be argued that the effects obtained are an artifact of the measure used, which is to measure feedback-ERN amplitude using difference waves (cf. Miltner et al., 1997; Lemke, 2003). To show that this is not the case, we also analyzed our data using a "base-to-peak" measure on the negative feedback waveform. The effects were in the same direction (largest ERN in the first condition, smallest ERN in the third condition), although the effect did not reach significance ($F(2,14) = 1.691$, p = n.s.).

were created representing different levels of estimate quality. Again, no significant differences in ERN amplitude among the bins were found ($F(3,21) = 1.040$, p = n.s.).

Discussion

The analyses of the behavioral data reveal that subjects used the information provided by the error feedback to adjust their behavior. Comparisons across the three experimental conditions indicate that more precise adjustments were evident following more versus less informative feedback. In addition, larger behavioral adjustments were seen when subjects received feedback that implied a large error than following feedback suggesting a small error. These results all support the inference that our manipulations were successful in influencing the processing of the error-feedback information. Subjects utilized the information to the extent that it could be used to guide their future behavior.

Analyses of the electrophysiological data confirmed that the component we identified as a feedback ERN was in fact a feedback ERN. The latency, shape of the waveforms, and scalp distribution are similar to those found in previous studies of the feedback ERN using this paradigm (Miltner et al., 1997; Lemke, 2003). ERN amplitude was smaller in the informative conditions (conditions two and three), and was not influenced by the degree of error, as indicated by the feedback in condition three. These results suggest a dissociation between the processes underlying generation of the ERN and the processes responsible for behavioral adjustments. The magnitude of behavioral adjustment was larger in the informative conditions and was influenced by the degree of error indicated by the feedback. Further evidence for a dissociation is provided by the analysis of the data in the first condition. Here the ERNs associated with different degrees of adjustment did not differ from each other.

These results are now considered in the context of prevailing theories of the ERN. As mentioned earlier, the Reinforcement Learning Theory of the ERN proposes that the mesencepahlic dopamine system (MDS) carries reward prediction errors to the ACC and other brain areas involved in selecting appropriate motor responses. This prediction error could be of (at least) two types. First, it could convey information about both an error in reward prediction and the kind of error that has been made, and could be used to guide a specific remedial action. In this case, one would expect a relationship between the ERN and remedial actions such as the behavioral adjustments measured in the present study. Second, the error signal carried by the MDS could act more as an ‘alerting’ signal, indicating that behavioral adjustments are needed, but leaving the decision as to how to adjust behavior to other brain systems. In this case, the additional information provided in the second and third conditions of the current experiment, would be used by systems other than the system that produced the ERN itself. The present data favor the second of these two conceptions of the nature of the error signal.

In his review of reward and the dopamine system, Schultz (2002) suggests that the reward prediction error signals carried by the dopamine reward system indicate the appetitive value of events relative to prediction, but do not discriminate between different types of reward. This may explain why we did not find any effects of the different punishment levels on the ERN. In terms of reward prediction, subjects presumably always make their best estimate and thus expect a positive reward of +6 cents. In all cases where negative feedback is given, the expected reward is not obtained. It is this observation that is reflected in the presence of the ERN. The magnitude of the difference between the expected and actual rewards, and the appropriate behavioral adjustments it indicates, might be of relevance to different brain systems, which are concerned with the remedial actions. In this perspective, the reward prediction error signals are said to convey a ‘scalar’ signal, signaling either ‘good’ or ‘bad’.

The proposal that the feedback ERN is the result of a simple ‘alerting’ signal is similar to proposals made by a different theory of the ERN, the so-called Conflict Theory (Carter et al, 1998; Botvinick et al, 2001). According to this theory, activity of the ACC is associated with the detection of response conflict in the system. Detection of conflict signals the need for cognitive control by the more frontal brain areas, for instance the dorsolateral prefrontal cortex[3]. The difference between the Reinforcement Learning Theory and the Conflict Theory is in the mechanism behind the activation of the ACC (and the presence of an ERN). However, according to both theories, the presence of ACC activation signals the need for action.

The observation that the ERN is smaller when more information is provided by the feedback replicates the observation made by Lemke (2003). In the more informative conditions, it is reasonable to assume that an alerting signal following error feedback is less important because decisions about remedial actions can be based on the information provided by the feedback. The computational process required to select an appropriate remedial action is also less complex than that required when the feedback merely indicates that an error has been made (cf. Lemke, 2003).

In summary, we conclude that the systems underlying generation of the feedback ERN are influenced by the degree of information presented by feedback stimuli (cf. Lemke, 2003), but not by the suggested degree of error as indicated by different levels of punishment. We did not find a direct relationship between feedback ERN amplitude and the degree of remedial action as indicated by behavioral adjustments, indicating that these measures may result from different, although related, neural processes. We have suggested that the ERN constitutes an 'alert' signal, related to the occurrence of an error, but does not give any more information concerning the type or error or any behavioral adjustments that should be made. This suggestion is consistent with current theories of reinforcement learning.

References

Badgaiyan RD, Posner MI (1998) Mapping the cingulate cortex in response selection and monitoring. Neuroimage 7:255-260.

Botvinick MM, Carter CS, Braver TS, Barch DM, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108:624-652.

Carter CS, Braver TS, Barch DM, Botvinick MM, Nol D, Cohen JD (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 280:747-749.

Coles MGH, Scheffers MK, Fournier L (1995) Where did you go wrong? Errors, partial errors, and the nature of human information processing. Acta Psychol 90:129-144.

Dehaene S, Posner MI, Tucker DM (1994) Localization of a neural system for error detection and compensation. Psychol Sci 5:303-305.

Falkenstein M, Hohnsbein J, Hoormann J, Blanke L (1990) Effects of errors in choice reaction tasks on the ERP under focused and divided attention. In: Psychophysiological brain research (Brunia CHM, Gaillard AWK, Kok A, eds), pp192-195. Tilburg: Tilburg University Press.

Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E (1993) A neural system for error detection and compensation. Psychol Sci 4:385-390.

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, Coles MGH (2002) The neural basis of error processing: Reinforcement learning, dopamine, and the error-related negativity. Psychol Rev 109:679-709.

Holroyd CB, Dien J, Coles MGH (1998) Error-related scalp potentials elicited by hand and foot movements: Evidence for an output-independent error-processing system in humans. Neurosci Lett 245:65-68.

Holroyd CB, Nieuwenhuis S, Yeung N, Cohen JD (in press) Errors in reward prediction are reflected in the event-related brain potential. Neuroreport.

Lemke U (2003) Hirnelektrische Prozesse der Fehlerverarbeitung. Unpublished Ph.D. thesis, Friedrich-Schiller-University Jena.

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.

Rabbitt PMA (1966) Errors and error correction in choice-response tasks. J Exp Psychol 71:246-272

Schultz W (2002) Getting formal with dopamine and reward. Neuron 36:241-263.

[3] It should be noted that the Conflict Theory has only been used to model internally detected errors, resulting in a response-locked ERN, while the current experiment deals with ERNs elicited by external negative feedback.