In: *Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring* (2004).
Edited by M. Ullsperger & M. Falkenstein. Leipzig: MPI of Cognitive Neuroscience, pp. 118-124.

Chapter 5: Feedback Processing and Observation of Errors

It wasn't me… or was it?
How false feedback affects performance

Ellen R.A. de Bruijn[1], Rogier B. Mars[1,2], Wouter Hulstijn[1]
[1]*Nijmegen Institute for Cognition and Information, Nijmegen, The Netherlands*
[2]*F.C. Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands*

The reinforcement learning model (Holroyd and Coles, 2002) argues that a feedback ERN is generated when the feedback is worse than expected. In the current experiment, we investigated whether the reinforcement learning system was also activated by unexpected feedback in a task in which feedback was not necessary to perform properly. A flankers task, in which false negative feedback was presented on four percent of the trials, was used. As expected, response-locked ERP data only showed an Ne/ERN after incorrect responses. Feedback-locked ERP data showed no feedback ERNs for the different feedback conditions. However, P300 amplitude increased significantly following false feedback. Performance adjustments, indexed by post-error slowing, were only present after incorrect responses. Interestingly, five out of nine participants also showed post-error slowing following false feedback. Participants were divided into two subgroups, on the basis of whether they showed post-error slowing following false feedback (PES) or not (No-PES). These subgroups revealed a difference between P300 amplitude and the presence of post-error slowing: the PES group had a larger P300 compared to the No-PES group, suggesting that the PES group attributed more meaning to false feedback. Overall, the results show that a dissociation between performance adjustments and the reinforcement learning system exists.

Introduction

Processing feedback may be essential for acquiring an error-free performance in certain tasks. Ever since the first reports of an event-related potential (ERP) component associated with the onset of error commission, an increasing number of experiments concerning this so-called error negativity (Ne: Falkenstein et al, 1990) or error-related negativity (ERN: Gehring et al, 1993) have been conducted. A variety of paradigms, for instance Eriksen flankers tasks, go-nogo tasks, time estimation tasks, gambling tasks, and learning tasks have been used. The latter three types of paradigms have led to the discovery of an ERN-like component elicited by negative feedback, the so-called feedback ERN (Miltner et al, 1997; Ruchsow et al, 2002; Gehring and Willoughby, 2002; Holroyd and Coles, 2002; Nieuwenhuis et al, 2002). Both the response- and feedback-locked ERNs have been shown to originate in the anterior cingulate cortex (ACC) (Dehaene et al., 1994; Miltner et al., 1997).

Recently, Holroyd and Coles (2002) have proposed a computational model that simulates the mechanisms by which response ERNs and feedback ERNs may interact. Their experiments showed that these two types of ERN are not independent ERP components, but that they are highly correlated. The model is based on reinforcement learning principles and accompanying reward expectancies. In brief, it argues that the ACC is trained to correctly select appropriate motor controllers through reinforcement learning. In a task in which participants had to learn correct stimulus-response mappings by making use of feedback, Holroyd and Coles showed that in the beginning of the learning task, a clear feedback ERN was present but no response ERN. After learning the correct stimulus-response mapping this pattern changed and the feedback ERN disappeared, while a response ERN became present. According to the model, the (feedback) ERN is elicited when the outcome is *worse then expected*. In the beginning of the task, participants have to rely completely on feedback, since they do not know the correct stimulus-response mappings. Consequently, they cannot detect their own errors at the moment of response onset and therefore, a feedback ERN is elicited upon delivery of the negative feedback. Near the end of the task, participants have learned the appropriate stimulus-response mappings, and can thus detect their own errors when they are committed (e.g., due to time pressure) leading to the generation of a response ERN. Since the participants have already detected the error, the following negative feedback will be expected and a feedback ERN will not be elicited.

Following this line of reasoning, one would expect the smallest feedback ERN on events that were highly expected and the largest feedback ERN on outcomes that were highly unexpected. As Holroyd and Coles (2002) showed, the response ERN is maximal when participants have finished learning the task. In such a learned situation, the feedback is no longer needed to perform the task, because participants detect their own errors and do not rely on feedback. Flankers and go-nogo tasks are examples of overlearned tasks in which learning is no longer necessary and feedback does not play a role. For this reason, feedback is typically not included in those types of experiments or otherwise ignored in the analyses. As a result, it is unknown whether these unexpected outcomes indeed elicit a feedback ERN in the learned situation of a flankers task.

Therefore, the aim of the current experiment was to investigate what happens when participants process unexpected performance feedback in a situation where they are not expecting to need the feedback to perform the task properly. We used a flanker task in which false negative feedback was presented in some trials to study if (1) the reinforcement learning system, as indexed by a feedback ERN, is activated by these unexpected outcomes and, since participants usually show an increase in reaction time following errors, if (2) these performance adjustments are also seen after false negative feedback.

Materials and Methods

Participants. Nine undergraduate students (five women) from the University of Nijmegen, ranging in the age from 22 to 27 years (mean 23.6), participated in this experiment. All participants were right-handed and had normal or corrected-to-normal vision. Participation was rewarded financially.

Design. We used a standard Flankers paradigm (Eriksen and Eriksen, 1974). Depending on the central letter (H/S) in a letter string (HHHHH, HHSHH, SSSSS, SSHSS) participants were instructed to respond as fast as possible with their left or right index finger by pushing a push button. Visual feedback, indicating whether the response was correct or incorrect, was given after each response and consisted of a triangle or a square. In four percent of the trials, participants received false feedback after a response to an incongruent trial. In those cases, the feedback would indicate that they made an error while in fact they responded correctly to the stimulus. This false feedback would appear quasi randomly, in such a way that at least 10 ‘normal’ trials would intermediate two following false feedback trials.

Procedure. Participants first received a practice block of 25 trials to let them get used to the relatively short stimulus presentation time. This presentation time of 50 ms was chosen on purpose to decrease the chance of participants discarding the false feedback as nonsense. The experimental phase consisted of 8 blocks of 100 trials. Participants first received a fixation point (100 ms) followed 300 ms later by the stimulus (50 ms). After the stimulus a blank screen (1050 ms) preceded the visual feedback stimulus (1000 ms) followed by an inter trial interval of 100 ms. One experimental session lasted about 1,5 hours including preparation and breaks.

Psychophysiological recording. The electroencephalogram (EEG) was recorded from 4 tin electrodes mounted in an elastic electrode cap (Electrocap international). The electrodes were located at the midline (Fz, FCz, Cz and Pz) and were referenced to the left mastoid. The vertical electro-oculogram (EOG) was recorded bipolarly from electrodes placed above and below the right eye. The horizontal EOG was also recorded bipolarly from electrodes lateral to each eye. All electrode impedances were below 5 kΩ. EEG signals were recorded in epochs starting 200 ms before stimulus onset and ending 2200 ms later. The EEG and EOG signals were amplified using a time-constant of 8 seconds and a low-pass filter of 15 Hz. All signals were digitized with a sample rate of 200 Hz using a 16 bit A/D converter.

*Analyses.*EOG artifact correction was carried out using the procedure proposed by Gratton et al. (1983). For both behavioral and ERP analyses, all responses faster than 200 ms and slower than 600 ms (5.58 %) were removed from the data sets. During the experiment, three different feedback conditions exist: two *true* feedback conditions (correct and incorrect) and one *false* feedback condition (incorrect while in fact correct). Epochs associated with correct and incorrect responses and correct responses followed by false feedback were averaged separately for each participant, time-locked to response onset, starting 100 ms before and ending 500 ms after response onset relative to a 100 ms pre-response baseline. Response ERN amplitude was defined in these averages as the most negative peak in the 0-150 ms time window after response onset at electrode FCz.

Epochs associated with correct, incorrect, and false feedback trials were also averaged separately for each participant, time-locked to feedback onset, starting 100 ms before and ending 800 ms after feedback onset relative to a 100 ms pre-feedback baseline. Feedback ERN amplitude on the feedback-locked ERPs was determined by subtracting the positive peak preceding the most negative peak in the 200-350 ms time window after stimulus onset at electrode FCz and P300 amplitude was defined as the most positive peak in the 300-800 ms time window after feedback onset at electrode Pz.

Individual averages for error rates, RTs, and amplitudes were entered in a General Linear Model (GLM) with repeated measures. Possible factors of the analyses in the following results section are congruency (2 levels: congruent vs incongruent), correctness (2 levels: correct vs incorrect), and feedback (3 levels: correct vs incorrect vs false). Greenhouse-Geisser corrections were applied when appropriate, but uncorrected degrees of freedom values are always given for purposes of interpretation.

Behavioral results

Reaction times and error rates (see Table 1)

Incongruent trials (401 ms) were responded to slower than congruent trials [365 ms; F (1, 8) = 25.09, p = .001]. The error rate for incongruent trials was higher (7.5 %) compared to the error rate for congruent trials [2.3 %; F (1, 8) = 35.56, p < .001] and incorrect responses were faster (362 ms) than correct responses [404 ms; F (1, 8) = 17.37, p = .003].

Table 1. Mean reaction times for correct and incorrect responses of congruent and incongruent stimulus type. Standard errors are given in parentheses.

	Congruent	Incongruent
Correct	380 (11)	428 (12)
Incorrect	350 (18)	374 (19)

Table 2. Mean RTs of the performance adjustments for the overall group, for participants who showed post-error slowing following false feedback (PES), and for participants who did not show post-error slowing following false feedback (No-PES). Standard errors are given in parentheses.

	Overall (N = 9)	PES (N = 5)	No-PES (N = 4)
Post-correct	397 (11)	380 (12)	419 (14)
Post-error	423 (11)	408 (12)	443 (14)
Post False	406 (11)	406 (15)	407 (17)

Performance adjustments (see Table 2)

Post-error slowing (Rabbitt, 1966) is characterized by a slowing of RTs after an error compared to RTs after a correct response. The analyses showed a main effect of feedback condition [F (2, 7) =

8.15, p = .015]. Simple contrasts referenced to the correct feedback condition showed that post-error slowing was only present after incorrect responses [26 ms; F (1, 8) = 16.71, p =.003], and not after correct responses followed by false feedback [9 ms; F (1, 8) = 1.14, p = .317].

ERP results

For the ERP analyses, average ERP amplitudes for individual participants were entered into a GLM with repeated measures on the three different feedback conditions. To investigate whether previously reported response ERN effects were also present in the current experiment, we first analyzed the three different feedback conditions at the moment of response onset. Second, we looked whether the amplitude of the feedback ERN was affected by the different feedback conditions, and third, because P300 amplitude is also known to be affected by expectancy (see e.g. Horst et al., 1980), we checked whether the amplitude of this component was affected by the different feedback conditions.

Response ERN (see Figure 1)

A main effect of feedback condition was present [F (2, 7) = 27.48, p < .001]. Simple contrasts referenced to the correct feedback condition revealed that a response ERN was only present after an incorrect choice of hand [-9.57 μV; F (1, 8) = 26.82, p = .001], peaking around 82 ms. No difference was found between correct responses followed by correct feedback (-1.63 μV) and correct responses followed by false feedback [-1.62 μV; F < 1].

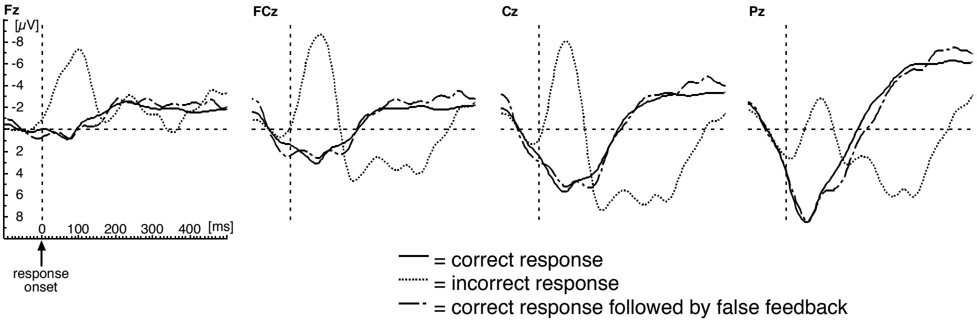

Figure 1.
Grand average ERPs time-locked to response onset for correct responses, incorrect responses and correct responses that were followed by false feedback.

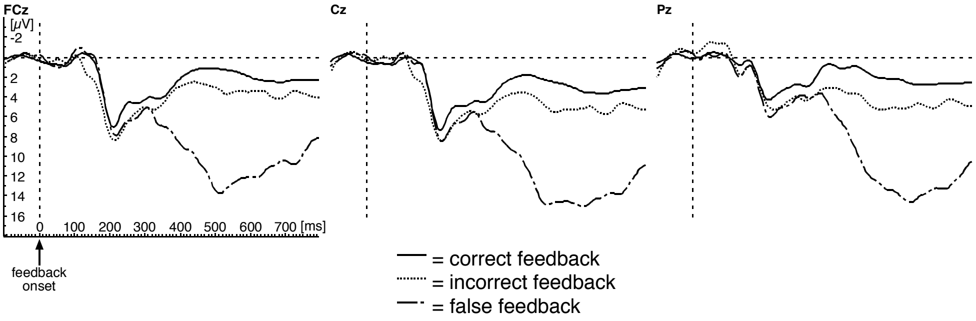

Figure 2.
Grand average ERPs time-locked to feedback onset for correct feedback, incorrect feedback and false feedback.

Feedback ERN (see Figure 2)

There was no main effect of feedback condition present [F (2, 7) = 2.86, p = .123], indicating that a feedback ERN was absent in the different feedback conditions.

Feedback-locked P300 (see Figure 2)

A main effect of feedback condition was present [F (2, 7) = 19.71, p < .001] and simple contrasts referenced to correct feedback revealed that the P300 amplitude was significantly larger after false feedback [16.59 μV; F (1, 8) = 25.61, p = .001]. There was no difference between correct (5.35 μV) and incorrect feedback [7.53 μV; F (1, 8) = 3.54, p = .097].

Individual differences

Although no effect of post-error slowing after false feedback was present in the overall analyses, large differences were visible between the participants. To investigate this further (see Table 2), we divided participants in two groups on the basis of whether post-error slowing was present (PES) following false feedback or not (No-PES). Five participants were assigned to the PES group (26 ms) and four participants to the No-PES group [-12 ms; t = 3.86, p = .006].

Following t-tests revealed that P300 amplitude following false feedback was significantly larger for the PES group (21.11 μV) than for the No-PES group [10.95 μV; t = 3.36, p = .012]. These group differences were neither present for the feedback ERN (t = -.29, t = .778) nor for the response ERN (t = 1.81, p = .113). The relation between P300 amplitude and the amount of post-error slowing following false feedback is depicted in Figure 3.

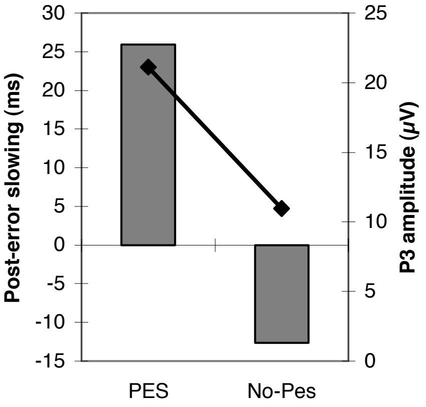

Figure 3.
Mean P300 amplitudes for the false feedback condition (line) depicted for the group of participants who did show post-error slowing following false feedback (PES) and for the group of participants who did not (No-PES).

Discussion

The aim of the current study was to investigate whether the reinforcement learning system would be activated (as reflected in the presence of a feedback ERN) by false performance feedback in a learned flankers task.

As expected from previous studies, the response-locked data show a clear response ERN after incorrect responses and no ERN activity after correct responses, irrespective of whether that response is followed by correct or by false feedback. Also the behavioural results were in accordance with previous studies using the flankers paradigm (see e.g. Gehring et al., 1993).

For the feedback-locked data, feedback ERNs are absent in the different feedback conditions and P300 amplitude does not differ between the correct and incorrect feedback conditions. However, the P300 is increased after false feedback compared to the other two conditions. Furthermore, the overall analyses show that performance adjustments, reflected in post-error slowing, are present after incorrect responses, but not after false feedback. The absence of these performance adjustments following false feedback suggests that participants do not use this feedback to 'improve' their performance. Taken together, these results are a strong indication that false feedback does not activate the reinforcement learning system.

The only component that is clearly affected by false feedback is the P300. According to Horst et al. (1980) this increased P300 amplitude after false feedback could be explained by the unexpectedness of the feedback. They showed that outcomes that did not confirm participants' expectations elicited a larger P300 than outcomes that did confirm their expectations, irrespective of whether that outcome was positive or negative. In addition, it should be noted that the false feedback events are rare compared to correct and incorrect feedbacks and that these infrequent events are also known to increase P300 amplitude. However, frequency alone cannot explain the increased P300 after false feedback. An increased P300 would then also be expected following incorrect feedback, an event that is much more infrequent than correct feedback. To investigate whether P300 amplitude is also related to the presence or absence of performance adjustments, we assigned the participants to two different groups, based upon them showing performance adjustments following the false feedback or not.

Participants who did not show post-error slowing after false feedback stimuli (No-PES) showed smaller P300s compared to participants who did show post-error slowing after false feedback (PES). This group effect was exclusively present for P300 amplitude following false feedback and thus suggests that the P300, a component not often considered in error monitoring research, is clearly related to performance adjustments. We can only speculate about why these participants show large differences with respect to post-error slowing, but it is reasonable to assume that individual strategies are underlying these effects. Although the current data cannot statistically support this assumption (group difference for correct RTs, $p = .075$), it may very well be possible that the No-PES group had a less impulsive response style. Adapting this more cautious strategy could have improved performance monitoring of the No-PES group compared to the PES group, in such a way that the No-PES group never regarded the false feedback stimulus as an event that called for a change of strategy. Instead of interpreting it as an error of their own, they may have thought of the false feedback being a software error. The PES group did try to improve their performance by making use of the false feedback. In this case, they interpreted the false feedback as an error they made themselves. As a result, the false feedback had much more meaning for the PES group than for the No-PES group, leading to increased P300 amplitude (see e.g. Johnson, 1986).

Previously, we have shown that in a force production task in which participants relied on the feedback on force performance to improve their behavior on a trial-to-trial basis, a large P300 was also elicited by negative feedback, but no feedback ERN (De Bruijn et al., 2003). One thing to consider is the possibility that the large P300 may be able to cancel out any feedback ERN activity. The force production task could be, unlike the current experiment, considered as a learning task, since participants learned to improve their force productions during the experiment by making use of the feedback. The main difference with learning discrete stimulus-response mappings (see Holroyd and Coles, 2002) is that force production involves learning on a more continuous dimension. Apparently, the largest effects on feedback ERNs in learning tasks are present when discrete stimulus-response mappings have to be learned. In our opinion, future research should also focus on the relationship between the presence of a feedback ERN, the presence of an increased P300, and the use of different types of learning.

Overall, we conclude that in an overlearned flankers task, unexpected false feedback does not activate the reinforcement learning system. The unexpected outcome does elicit an increased P300 that is related to the presence or absence of performance adjustments. Individuals who show performance adjustments following false feedback attribute more meaning to this type of feedback, leading to an increased P300 compared to individuals who do not show these performance adjustments. Therefore, the current data strongly suggest a dissociation between performance adjustments as indexed by post-error slowing and the reinforcement learning system as indexed by the presence of a feedback ERN.

References

De Bruijn ERA, Hulstijn W, Meulenbroek RGJ, Van Galen GP (2003) Action monitoring in motor control: ERPs following selection and execution errors in a force production task. Psychophysiology 40:786-795.

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

Eriksen BA, Eriksen CW (1974) Effects of noise letters upon the identification of a target letter in a non-search task. Percept Psychophys 16:143-149.

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.

Gehring WJ, Willoughby AR (2002) The medial frontal cortex and the rapid processing of monetary gains and losses. Science 295:2279-82.

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.

Horst RL, Johnson R Jr, Donchin E (1980) Event-related brain potentials and subjective probability in a learning task. Mem Cognit 8:476-488.

Johnson R Jr (1986) A triarchic model of P300 amplitude. Psychophysiology 23:367-384.

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. Cogn Affect Behav Neurosci 2:19-36.

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

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. Neurosci Lett 325:203-206.