Available online at www.sciencedirect.com
SCIENCE DIRECT
Neuroscience and Biobehavioral Reviews 28 (2004) 441–448
NEUROSCIENCE AND BIOBEHAVIORAL REVIEWS
www.elsevier.com/locate/neubiorev

Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance

Sander Nieuwenhuis[a,*], Clay B. Holroyd[b], Nisan Mol[a], Michael G. H. Coles[c]

[a]*Department of Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands*
[b]*Department of Psychology, Princeton University, Princeton, NJ 08544, USA*
[c]*FC Donders Centre for Cognitive Neuroimaging, Adelbertusplein 1, 6525 EK, Nijmegen, The Netherlands*

Received 15 March 2004; revised 26 May 2004; accepted 27 May 2004

Abstract

The development of the field of cognitive neuroscience has inspired a revival of interest in the brain mechanisms involved in the processing of rewards, punishments, and abstract performance feedback. One fruitful line of research in this area was initiated by the report of an electrophysiological brain potential in humans that was differentially sensitive to negative and positive performance feedback [J. Cogn. Neurosci. 9 (1997) 788]. Here we review current knowledge regarding the neural basis and functional significance of this feedback-evoked 'error-related negativity' (ERN). Our review is organized around a set of predictions derived from a recent theory, which holds that the ERN is associated with the arrival of a negative reward prediction error signal in anterior cingulate cortex.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: Anterior cingulate cortex; Error-related negativity; ERP; Feedback; Prediction error; Reward

Humans (and other animals) use feedback to learn how to behave. Such learning depends crucially on the ability of the organism to discriminate between positive feedback, indicating that the behavior was appropriate, and negative feedback, indicating that the behavior was in some way inappropriate. It is not surprising, therefore, that there is evidence from a variety of sources that indicates that the brain responds differentially to positive and negative feedback.

This differential neural response is evident in measures of the event-related brain potential, derived from recordings of the electroencephalogram (see Fig. 1). For example, in a study by Miltner and colleagues, human participants were required to estimate the duration of a 1-s interval [1]. Following a warning cue, they pressed a button when they believed that 1 s had elapsed. This response was followed 600 ms later by a feedback stimulus indicating whether their estimate was correct (positive feedback) or incorrect (negative feedback). A time window around 1 s was used to determine response accuracy and this window was adaptively adjusted so that the probabilities of positive and negative feedback stimuli were both 0.5. In different conditions, feedback was provided in auditory, visual and somatosensory modalities.

Analysis of the event-related brain potential responses (ERPs) following the feedback stimulus revealed that, following negative feedback, the potential became more negative. This negativity was isolated by subtracting the response to positive feedback from the response to negative feedback. The resulting waveform had an average amplitude of between 5 and 10 μV and a peak latency of between 230 and 270 ms, with the somatosensory and visual modalities being associated with the shortest and longest latencies, respectively. Several other studies had previously also reported a similar negative ERP under conditions of negative feedback [2–6] (for more references see Ref. [7]). However, these earlier studies either did not elaborate on this finding or did not control for the relative probabilities of positive and negative feedback.

When the source of the negative scalp potential was estimated using equivalent dipole analysis procedures, a generator in or near the anterior cingulate cortex (ACC) was suggested [1]. In this respect, the negativity closely resembled another event-related brain potential, the error-related negativity (ERN or Ne), which had previously been

* Corresponding author. Tel.: +31-20-4448807; fax: +31-20-4448971.
E-mail address: stn20@dds.nl (S. Nieuwenhuis).

0149-7634/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neubiorev.2004.05.003

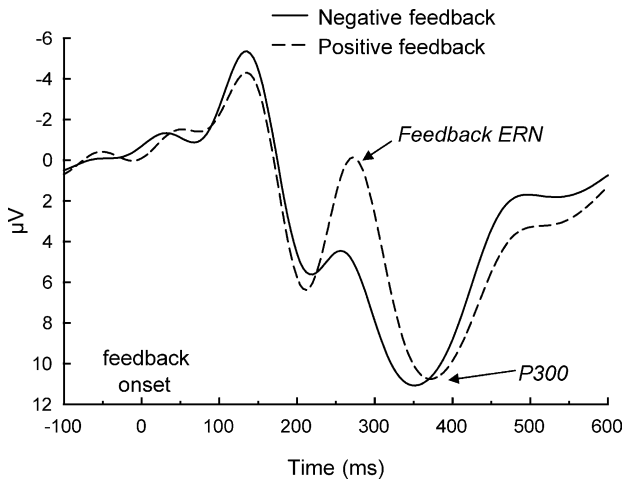

Fig. 1. Typical example of event-related brain potentials associated with negative and positive feedback (adapted from Ref. [25]). Negative is plotted up by convention. Waveforms were recorded from electrode Cz. Arrows indicate the peak of the feedback ERN and the P300 components in the waveform associated with negative feedback. Note that although the P300 reaches maximum amplitude over posterior parts of the scalp, the component is also visible over frontal regions as seen here.

identified in reaction time tasks (reviewed in Refs. [8,9]). This response ERN peaks within 100 ms of an error response and also appears to be generated by neural activity in the ACC [10,11]. This resemblance led to the proposal that the negativities following response errors (the response ERN) and negative feedback (the feedback ERN) were associated with the same neural and cognitive error-detection process [1].

1. Reinforcement learning theory of the ERN

A recent theory has extended the hypothesis of Miltner et al. [1] by proposing that both the response ERN and the feedback ERN are produced by a dopamine system for reinforcement learning [12]. Details regarding the neurophysiological motivation for this theory are reviewed elsewhere [12,13]. Briefly, the theory is predicated on previous research implicating the basal ganglia and midbrain dopamine system in reward prediction and reinforcement learning. According to this previous research [14,15] (for review, see Ref. [13]), the basal ganglia evaluate ongoing events and predict whether the events will end in success or failure. When the basal ganglia revise their predictions for the better, they induce a phasic increase in the activity of midbrain dopaminergic neurons, and when the basal ganglia revise their predictions for the worse, they induce a phasic decrease in the activity of midbrain dopaminergic neurons. These phasic increases and decreases in dopamine activity indicate that ongoing events are 'better than expected' and 'worse than expected,' respectively, and are used by the basal ganglia to update its predictions, such that the system gradually learns the earliest predictor of reward or punishment. Furthermore, the dopamine signals are also conveyed to the frontal cortex where they are used as reinforcement learning signals, serving the adaptive modification of behavior. The reinforcement learning theory of the ERN extends this theoretical framework by proposing that the impact of the dopamine signals on ACC modulates the amplitude of the ERN, such that phasic decreases in dopamine activity (indicating that ongoing events are worse than expected) are associated with large ERNs, and phasic increases in dopamine activity (indicating that ongoing events are better than expected) are associated with small ERNs [12,16]. According to this position, the dopamine signals are used by the ACC to improve performance on the task at hand.

Since Miltner et al.'s report in 1997, much progress has been made in understanding the neural basis and functional significance of the feedback ERN. Many of the empirical studies that have led to this progress have been inspired by the reinforcement learning theory of the ERN (hereafter called the 'RL-ERN theory'). Below, we will review the current knowledge about the feedback ERN. Our review will be organized around four core predictions of the RL-ERN theory: (i) The feedback ERN reflects a good/bad evaluation; (ii) feedback ERN amplitude depends on the relation between actual vs. expected outcome; (iii) feedback ERN amplitude varies inversely with response ERN amplitude as a function of learning; (iv) the feedback ERN is generated in ACC. We will discuss each of these predictions in turn, and evaluate the existing literature in the light of these predictions. Following this overview, we will discuss outstanding questions regarding the feedback ERN, including the possibility that the ERN reflects the emotional impact of a negative expectation violation. We note that aside from this emotion hypothesis, the RL-ERN theory is currently the only theory that attempts to explain the functional significance of the feedback ERN. For example, the conflict monitoring theory, while providing a powerful explanation of the response ERN and brain-activity associated with high-conflict correct trials, does not in its present form address the feedback ERN [9,17].

2. Four predictions of the reinforcement learning theory

2.1. The feedback ERN reflects a good–bad evaluation

The RL-ERN theory holds that the ERN reflects the outcome of an evaluation of events along a good–bad dimension, suggesting that the ERN should be sensitive to any performance-related feedback information indicating favorable or unfavorable outcomes. This notion is supported by the finding that an apparent ERN is observed following feedback indicating a loss [18,19] and following feedback indicating an incorrect response [1,20].

Although the negative ERP components elicited by losses and by feedback indicating an incorrect response have a similar morphology, timing with respect to

the feedback, and mediofrontal scalp distribution, there has been a debate regarding whether or not they are the same phenomenon [18,21,22]. To investigate this issue systematically, we conducted two gambling experiments in which the feedback stimuli on each trial conveyed information along two different dimensions: a 'gain/loss' dimension indicating whether the subject's choice led to a gain or loss of money, and a 'correct/error' dimension indicating whether the subject's choice was better or worse than the alternative choice that the subject could have made [23]. Furthermore, the background color of the feedback display (green or red) correlated with one of the two dimensions, emphasizing either the gain/loss aspect (in one experiment) or the correct/error value of the feedback (in the other experiment). The results demonstrated that the frontocentral negativity elicited by the feedback stimuli was sensitive to both the gain/loss information and the correct/error information conveyed by the feedback stimulus, depending on which dimension of the feedback was made most salient to the subjects. Gehring and Willoughby [18], using a similar gambling paradigm, found a negative component following the feedback that was sensitive to the gain/loss dimension but not to the correct/error dimension of the feedback. In this study, the most salient information in the feedback display was the gain/loss aspect of the chosen outcome, and the observed negativity was duly sensitive to this aspect.

Together, these studies seem to indicate that the ERN reflects a rapid evaluation of ongoing events along an abstract good–bad dimension, rather than in terms of correctness or gain/loss. Note that the RL-ERN theory is non-specific as to what constitutes a good or bad outcome. According to the theory, the ERN system can base its good–bad evaluations on different sources of information, and the choice of source can be determined by the context in which the information is provided.

At first glance, the results from a gambling study by Yeung and Sanfey [19] appear to pose a problem for the view, held by the RL-ERN theory, that the ERN scales with the goodness of ongoing events. On each trial in that study, subjects gambled on one of two possible response options and were then told the outcome of their choice. This could be a large gain of money (32–40¢), a small gain (7–11¢), a small loss (6–10¢), or a large loss (32–40¢). Yeung and Sanfey found that the feedback ERN was larger on trials that involved a loss than on trials that involved a gain of money. However, the amplitude of the ERN was not affected by the magnitude of the reward. In contrast, the amplitude of a later component of the ERP, the P300 (see Fig. 1), appeared to be selectively sensitive to the absolute magnitude of the reward, irrespective of the valence of the outcome. These results could be taken to suggest that the evaluation process indexed by the ERN is binary, simply coding whether events are good or bad regardless of the magnitude of reward or penalty. In contrast, the RL-ERN theory claims that ERN amplitude is sensitive to the size of the reward prediction error, and thus, would appear to predict a larger negativity for large negative outcomes.

However, the RL-ERN theory can accommodate the results of Yeung and Sanfey [19] in the following way. It should be noted that subjects in Yeung and Sanfey's experiment knew whether, on a particular trial, they gambled on a small (i.e. safe) or large (i.e. risky) outcome. It is possible that the monitoring system scales the variance of possible outcomes so that the extreme outcomes are weighted equally irrespective of their absolute magnitude [7]. For instance, the system may treat losing 10¢ when this represents the maximum loss in a similar way as losing 1¢ when this represents the maximum loss. If this hypothesis is correct, then an interesting case would be to present subjects with a range of possible outcomes, and to compare ERN amplitude for the intermediate outcomes relative to ERN amplitude for the extreme outcomes. Even if the monitoring system adjusts its sensitivity to the extreme outcomes, the RL-ERN theory predicts that intermediate outcomes should be associated with intermediate-sized ERN amplitudes. On this account, the ERN shows a graded but normalized (i.e. with respect to the experienced range of outcomes) sensitivity to outcome values. The currently available evidence regarding this issue is mixed [7,24], suggesting a need for additional research.

2.2. Feedback ERN amplitude depends on the relation between actual vs. expected outcome

According to the RL-ERN theory, the ERN reflects a negative reward prediction error, a signal elicited when the monitoring system has to revise its reward expectations for the worse. The amplitude of the ERN is proportional to the size of the prediction error. From this follows the prediction that the amplitude of the feedback ERN should be dependent on the difference between the actual outcome of a trial and the expected outcome of that trial. In a probabilistic learning study, described in more detail in Section 3, we found initial evidence in line with this prediction [12,25]. Subjects were required to learn a set of stimulus–response mappings on the basis of trial-to-trial performance feedback. In one task condition, the response was 80% predictive of the value of the feedback stimulus [25]. The behavioral results suggested that subjects gradually learned to select the response with the highest probability of reward. Nevertheless, on 20% of the trials this response led to unexpected penalty. The ERN associated with this unpredicted turn of events had an amplitude of more than 10 μV—probably the largest average feedback ERN reported in the literature. Furthermore, in a condition in which the correct stimulus–response mapping was randomly determined on each trial, the size of the ERN was larger if the correct mapping was different from the mapping that applied when the subject previously encountered the same stimulus [12]. This result suggests that ERN amplitude tracks the prediction error on a trial-to-trial basis.

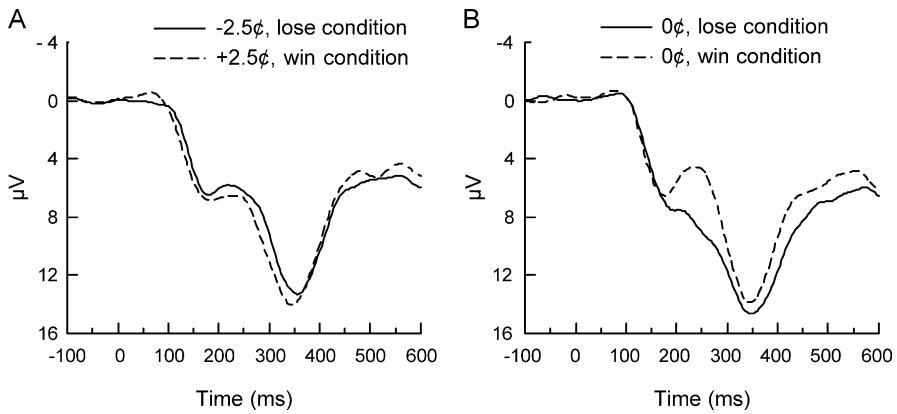

Fig. 2. Illustration of context sensitivity of the feedback ERN (adapted from Ref. [7]). Waveforms were recorded from electrode FCz. Time = 0 ms indicates the onset of the feedback stimulus. See text for details.

In line with this, Butterfield and Mangels [26] found that feedback-ERN amplitude was larger for semantic retrieval errors endorsed as correct with high confidence than for low-confidence errors.

In two recent feedback ERN studies, both using a gambling task, we have studied the effect of the relation between actual vs. expected outcome in more detail. In one study, we manipulated the expected value of the feedback by changing, between different task blocks, the probability of reward, while the size of the reward was held constant [27]. In this study, subjects were required, on each trial, to pick one of four balloons presented on the computer screen. The subjects were told that only one of the balloons contained money, and that their goal was to earn as much money as possible throughout the experiment. The outcome of each trial was communicated by means of a feedback stimulus. Unbeknown to the subjects, the location of the money was under control of the experimenter. In one task condition, subjects won money on 25% of the trials and picked an empty balloon on 75% of the trials. In another task condition, these probabilities were reversed. According to the RL-ERN theory, the negative prediction error associated with non-rewards (i.e. an empty balloon) should be larger when the system comes to expect rewards, and smaller when it comes to expect nothing. The results were consistent with this theoretical prediction: the feedback ERN was larger in the condition in which rewards were frequent than in the condition in which rewards were infrequent. A similar explanation may apply to the finding that the amplitude of the response ERN is inversely proportional to the frequency of response errors [12,28].

In another study [7], we manipulated the expected outcome by varying the range of possible outcomes communicated by the feedback stimulus, while keeping constant the probability of the various outcomes. Subjects performed a similar gambling task to that described above. However, instead of two, there were three, equiprobable possible outcomes on each trial. In a so-called 'win' condition, these were $+5¢$, $+2.5¢$, and $0¢$. In a 'lose' condition, involving separate task blocks, the possible outcomes were $-5¢$, $-2.5¢$, and $0¢$. Note that in both conditions, the intermediate outcome corresponded with the objective expected outcome. As predicted by the RL-ERN theory, the feedback ERN associated with these two outcomes did not differ in amplitude, even though one outcome entailed a loss and the other outcome entailed a gain of money (see Fig. 2A). Another interesting result concerned the ERPs elicited by the '0' outcomes in each condition. A large ERN was elicited in the win condition, in which '0' was the worst possible outcome (see Fig. 2B). In contrast, the feedback ERN was virtually absent in the lose condition, in which '0' was the best possible outcome. These and other results from this study suggest that the feedback ERN is sensitive not to the absolute magnitude of the reward, but rather to deviations from the expected value of the reward. Thus, the feedback ERN behaves as if it reflects a reward prediction error.

2.3. Feedback ERN amplitude varies inversely with response ERN amplitude as a function of learning

A central claim of the RL-ERN theory is that the ERN, like phasic activity of the midbrain dopamine system, is elicited following the earliest predictor of negative outcome [13]. This claim is consistent with the timing of the response ERN and the feedback ERN. In typical choice RT tasks, in which subjects are aware of the stimulus–response mappings, the response is the earliest predictor of the outcome of a trial. Indeed, subjects can often efficiently regulate their performance without the use of trial-to-trial feedback. Hence, in choice RT tasks, the ERN occurs immediately following an erroneous response. In contrast, in tasks such as the time estimation task used by Miltner et al. [1], subjects rely on trial-to-trial feedback for evaluation of their performance. As a consequence, the feedback stimulus itself is the earliest predictor of the outcome of the trial, and the ERN is observed following a negative feedback stimulus.

In a recent experiment, we have investigated the timing of the ERN more systematically [12]. Subjects performed a probabilistic learning task in which they were instructed to

produce speeded two-choice responses to a series of stimuli. Subjects were not informed about the stimulus–response mappings, but were instead required to determine these on the basis of trial-to-trial feedback. Some of the stimuli were consistently mapped to one or the other response, and choice accuracy associated with these stimuli showed a gradual increase from 50% to approximately 80% over the course of 50 presentations of each of the stimuli; the use of a stringent response deadline assured that accuracy did not reach perfection. The critical results in this experiment concerned the effect of learning of the stimulus–response mappings on the relative size of the response ERN and the feedback ERN. As predicted by the RL-ERN theory, during the initial stages of learning, the ERN was large following the feedback and absent following the response. However, as subjects learned the stimulus–response mappings, this pattern gradually reversed: The ERN slowly 'propagated back' from the feedback to the response as the predictive value of the response was learned. This pattern of results was not found for a separate set of stimuli for which the mapping to the response was randomly determined on each trial, and hence could not be learned. In this task condition, the ERN remained invariably high following the feedback, and did not propagate back to the response.

In a follow-up study, using a variant of the probabilistic learning task, we replicated these results [25]. In that study, we also established more clearly that the relative size of the response ERN and feedback ERN is highly sensitive to the degree to which a response is predictive of the value of the feedback. A stepwise increase of this predictive power (20–50–80–100%) led to a monotonic increase in response ERN amplitude and to a corresponding decrease in feedback ERN amplitude. Together, these results underline the intimate relationship between learning and the time of occurrence of the ERN.

2.4. The feedback ERN is generated in anterior cingulate cortex

The RL-ERN theory provides a precise account of how the ERN is generated: negative and positive reward prediction errors are coded as phasic decreases and increases in activity of the midbrain dopamine system, respectively. These phasic dopamine signals are then conveyed to several cortical brain regions including a part of the ACC associated with the cognitive control of motor behavior [29,30]. The negative and positive dopamine signals, respectively, disinhibit or inhibit the apical dendrites of motor neurons there, giving rise to differential activity of this area between correct trials and error trials, which is manifested at the scalp as the ERN. Thus, although the reward prediction error signals are coded by the midbrain dopamine system, the electrophysiological correlate of these signals, the ERN, is generated in ACC (see also Ref. [16]).

In line with this view, equivalent dipole source modeling studies have generally indicated the ACC as the most likely source of the feedback ERN [1,18,20]. One study has suggested a feedback ERN generator in a more caudal and dorsal region of medial frontal cortex, but this dipole model was not based on an exploratory fitting procedure, leaving open the possibility that a dipole in ACC provided a better fit of the data [31]. In any case, source modeling results must be interpreted with caution because the dipole source localization problem is underdetermined (the so-called 'inverse problem'). Another, more specific problem in modeling the source of the feedback ERN concerns the overlap of this component with the P300 (see Fig. 1). This complicates the source localization modeling, and may require the addition of extra dipoles [1], which increases the risk of finding a statistically appropriate but false source model solution. Nevertheless, the source modeling studies discussed above receive indirect support from neurophysiological recordings in monkeys, indicating that activity of ACC motor neurons is modulated by the absence of expected rewards [32–34].

Functional neuroimaging studies have also investigated the impact on ACC activity of positive and negative reinforcers. Although not observed in each study [35], some studies have found increased ACC activation in response to financial penalty [36–38]. Similarly, Bush and colleagues [39] found that dorsal ACC was activated by unexpected decreases in monetary reward. Other studies have investigated the neural response to abstract performance feedback. Ullsperger and von Cramon [40] found that activity in the rostral cingulate motor area was increased following negative compared to positive feedback in a dynamically adaptive motion prediction task. Monchi and colleagues reported a rostral ACC area that was reliably activated by negative feedback in the Wisconsin card sorting test [41].

In a recent functional magnetic resonance imaging study, using a probabilistic learning task, we demonstrated that a single area in dorsal ACC is sensitive to both error responses and negative feedback [42]. Importantly, the magnitude of the observed ACC activations mirrored the amplitude of the response ERN and feedback ERN in electrophysiological studies involving the same task [12,25], being largest when the reward prediction error is also largest. This presents compelling support for the RL-ERN theory, which claims that the response ERN and feedback ERN are generated in the same area of ACC. An intriguing question for future research is how the findings reviewed above are related to the well-known role of ACC in coding the negative affect associated with pain, a primary negative reinforcer [43].

3. Discussion

We have reviewed the current knowledge regarding the neural basis and functional significance of the feedback ERN, an electrophysiological brain potential in humans that is differentially sensitive to negative and positive performance feedback [1]. Studying the feedback ERN can provide

important clues about the nature of the neural mechanisms contributing to error processing, decision making, and reinforcement learning. The results from neurophysiological and functional imaging studies provide converging evidence for the view that the feedback ERN is generated in ACC, and electrophysiological studies appear to indicate that the ERN reflects an evaluation process that monitors for unexpected favorable or unfavorable events. The amplitude of the feedback ERN is sensitive to deviations between the actual and expected value of outcomes, rather than to the absolute magnitude of these outcomes. Furthermore, if subjects are able to learn the actions that lead to these bad outcomes, then the ERN propagates back to these actions, which come to serve as the earliest predictors of the negative outcomes. All of these findings are consistent with the recently proposed RL-ERN theory of the ERN, which suggests that the ERN reflects the impact on ACC of a negative reward prediction error signal, conveyed by the midbrain dopamine system, that is generated when ongoing events are suddenly worse than expected [12].

Despite the progress in understanding the feedback ERN signal, several outstanding questions remain. A first question is, what does the brain do with the reward prediction error signal thought to be indexed by the ERN? According to the RL-ERN theory, the ACC uses this signal to positively (negatively) reinforce the behaviors and motor systems involved in arriving at the positive (negative) outcome. This view is consistent with the observation that ERN amplitude varies as a function of learning. However, although there is indirect evidence from neurophysiological research and computational modeling studies [12], direct empirical evidence that the ERN reflects a reinforcement learning signal is still lacking. One possible avenue for exploring this issue is by studying how dopaminergic pharmacological agents affect ERN amplitude and simple associative learning [44,45].

A second outstanding question concerns the relation between the feedback ERN and immediately preceding action. According to the RL-ERN theory, the ERN specifically indicates when the *consequences of a response* are worse than expected. This notion refers to a fundamental principle of operant conditioning, according to which learning should occur only when the reward or punishment is contingent on the animal's behavior. This issue was addressed to some extent in the probabilistic learning experiment mentioned earlier [12]. One of the stimuli in this experiment was always associated with a negative trial outcome, irrespective of the subject's response, and hence was the earliest predictor of negative outcome. Because the negative outcomes associated with this stimulus were not the consequence of the subjects' responding, it was predicted that by the end of learning no ERN should be observed on these trials. As expected, in the course of learning, the amplitudes of the response ERN and feedback ERN gradually decreased. However, they did not diminish to zero—even after many presentations of the stimulus, there remained a small but clear ERN following response and feedback that appeared resistant to further learning. This result can perhaps be explained by making the assumption that an additional cognitive process was continuing the search for an appropriate response strategy even after the simple reward-prediction process posited by the RL-ERN theory had given up [12]. Thus, better tests seem needed to evaluate the relation between the ERN, previous responding, and cues that predict feedback irrespective of the response. For instance, an important question is whether ERNs will be observed when the subjects' task is to simply look at stimuli informing them about monetary rewards and punishments, in the absence of responding.

A third challenge for future research is to determine the relationship between the ERN and the amount or complexity of evaluative information in the feedback. Results from a study by Mars et al. [24] suggest that if the information conveyed by the feedback display is complex, then this may draw away attention from the simple good-or-bad character of the feedback. These authors studied the effect on the ERN of different types of feedback in the time estimation task. In one condition, the feedback had a binary character, indicating whether the time estimation was adequate or not. In this condition, the authors found a substantial ERN following negative feedback, replicating the findings of Miltner et al. [1]. In a second condition, the feedback could take on three different values, indicating whether the estimation was too short, appropriate, or too long. The feedback ERN in this condition was significantly smaller than in the first condition. Presumably, the increased amount of information in the feedback reduced the impact of the valence dimension of the feedback. The RL-ERN theory makes no prediction regarding the effect of feedback complexity on the ERN. According to the theory, the ERN reflects 'scalar' (good/bad) error signals, and not 'vector' signals that tell you what you should have done in addition to whether you were right or wrong. Indeed, this is a fundamental property of reinforcement learning theory [46]. Further research is needed to test this aspect of the RL-ERN theory and to extend the findings of Mars et al.

As a final issue, it has been proposed that the response ERN [31,47] and the feedback ERN [18] may reflect an emotional reaction to errors. Although future research is needed to investigate this issue, the RL-ERN theory appears to be compatible with the idea that the ERN is associated with emotional processing [48]. In particular, 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. One possibility is that the RL-ERN theory constitutes a formal instantiation of the somatic marker hypothesis [27,49]. According to this hypothesis, decision making is biased by emotionally induced somatic states, for instance through the influence

of particular neurotransmitter systems. The phasic dopamine signals hypothesized to underlie ERN generation may serve as a type of somatic marker. Another possibility is that the ERN is directly related to the error-detection process itself, and that the emotional reaction to errors is simply sensitive to the same variables as the ERN, yielding a correlation in the absence of a direct causal relation between the two phenomena.

Acknowledgements

The research of S.N. was supported by the Netherlands Organization for Scientific Research (NWO). The research of C.B.H. was supported by National Institute of Mental Health postdoctoral fellowship MH63550.

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: 788–98.
[2] Sutton S, Tueting P, Hammer M, Hakerem G. Evoked potentials and feedback. In: Otto D, editor. Multidisciplinary perspectives in even-related brain potential research. Publication No. EPA-600/9-77-043. Washington, DC: US Government Printing Office; 1978. p. 184–8.
[3] Campbell KB, Courchesne E, Picton TW, Squires KC. Evoked potential correlates of human information processing. Biol Psychol 1979;8:45–68.
[4] Johnston VS. Stimuli with biological significance. In: Begleiter H, editor. Evoked brain potentials and behavior. New York: Plenum Press; 1979. p. 1–12.
[5] Ruchkin DS, Sutton S, Munson R, Silver K, Macar F. P300 and feedback provided by the absence of the stimulus. Psychophysiology 1981;18:271–82.
[6] Takasawa N, Takino R, Yamazaki K. Event-related potentials derived from feedback tones during motor learning. Jpn J Physiol Psychol Psychophysiol 1990;8:95–101.
[7] Holroyd CB, Larsen JT, Cohen JD. Context dependence of the event-related brain potential to reward and punishment. Psychophysiology 2004;41:245–53.
[8] Holroyd CB, Nieuwenhuis S, Mars RB, Coles MGH. Anterior cingulate cortex, selection for action, and error processing. In: Posner MI, editor. Cognitive neuroscience of attention. New York: Guilford Press; 2004. in press.
[9] Yeung N, Botvinick MM, Cohen JD. The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol Rev 2004; in press.
[10] Dehaene S, Posner MI, Tucker DM. Localization of a neural system for error detection and compensation. Psychol Sci 1994;5:303–5.
[11] Holroyd CB, Dien J, Coles MGH. Error-related scalp potentials elicited by hand and foot movements: evidence for an output-independent error-processing system in humans. Neurosci Lett 1998;242:65–8.
[12] Holroyd CB, Coles MGH. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Rev 2002;109:679–709.
[13] Schultz W. Getting formal with dopamine and reward. Neuron 2002; 36:241–63.
[14] Barto AG. Adaptive critics and the basal ganglia. In: Houk J, Davis J, Beiser D, editors. Models of information processing in the basal ganglia. Cambridge, MA: MIT Press; 1995. p. 215–32.
[15] Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 1996;16:1936–47.
[16] Holroyd CB. A note on the oddball N200 and the feedback ERN. In: Ullsperger M, Falkenstein M, editors. Errors, conflicts, and the brain: current opinions on performance monitoring. Leipzig: MPI of Cognitive Neuroscience; 2004. p. 211–8.
[17] Van Veen V, Carter CS. The timing of action-monitoring processes in the anterior cingulate cortex. J Cogn Neurosci 2002;14:593–602.
[18] Gehring WJ, Willoughby AR. The medial frontal cortex and the rapid processing of monetary gains and losses. Science 2002;295:2279–82.
[19] Yeung N, Sanfey A. Independent coding of magnitude and valence in the human brain. J Neurosci 2004; in press.
[20] Ruchsow M, Grothe J, Spitzer M, Kiefer M. Human anterior cingulate cortex is activated by negative feedback: evidence from event-related potentials in a guessing task. Neurosci Lett 2002;325:203–6.
[21] Gehring WJ, Willoughby AR. Medial prefrontal cortex and error potentials—reply to Holroyd, Coles, & Nieuwenhuis. Science 2002; 296:1610–1.
[22] Holroyd CB, Coles MGH, Nieuwenhuis S. Medial prefrontal cortex and error potentials. Science 2002;296:1610–1.
[23] 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;14:741–7.
[24] Mars RB, De Bruijn ERA, Hulstijn W, Miltner WHR, Coles MGH. What if I told you: 'You were wrong'? Brain potentials and behavioral adjustments elicited by performance feedback in a time-estimation task. In: Ullsperger M, Falkenstein M, editors. Errors, conflicts, and the brain. Current opinions on performance monitoring. Leipzig: MPI of Cognitive Neuroscience; 2004. p. 129–34.
[25] Nieuwenhuis S, Ridderinkhof KR, Talsma D, Coles MGH, Holroyd CB, Kok A, Van der Molen MW. A computational account of altered error processing in older age: dopamine and the error-related negativity. Cogn Affect Behav Neurosci 2002;2:19–36.
[26] Butterfield B, Mangels JA. Neural correlates of error detection and correction in a semantic retrieval task. Brain Res Cogn Brain Res 2003;17:793–817.
[27] Holroyd CB, Nieuwenhuis S, Yeung N, Cohen JD. Errors in reward prediction are reflected in the event-related brain potential. Neuroreport 2003;14:2481–4.
[28] Falkenstein M, Hoorman J, Christ S, Hohnsbein J. ERP components on reaction errors and their functional significance: a tutorial. Biol Psychol 2000;51:87–107.
[29] Picard N, Strick PL. Motor areas of the medial wall: a review of their location and functional activation. Cereb Cortex 1996;6:342–53.
[30] Paus T. Primate anterior cingulate cortex: where motor control, drive, and cognition interface. Nat Rev Neurosci 2001;2:417–24.
[31] Luu P, Tucker DM, Derryberry D, Reed M, Poulsen C. Electrophysiological responses to errors and feedback in the process of action regulation. Psychol Sci 2003;14:47–53.
[32] Niki H, Watanabe M. Prefrontal and cingulate unit activity during timing behavior in the monkey. Brain Res 1979;171:213–24.
[33] Shima K, Tanji J. Role for cingulate motor area cells in voluntary movement selection based on reward. Science 1998;282:1335–8.
[34] Ito S, Stuphorn V, Brown JW, Schall JD. Performance monitoring by the anterior cingulate cortex during saccade countermanding. Science 2003;302:120–2.
[35] Elliott R, Newman JL, Longe OA, Deakin JFW. Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. J Neurosci 2003;23:303–7.
[36] Knutson B, Westdorp A, Kaiser E, Hommer D. FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 2000;12:20–7.
[37] O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci 2001;4:95–102.

[38] Delgado MR, Locke HM, Stenger VA, Fiez JA. Dorsal striatum responses to reward and punishment: effects of valence and magnitude manipulations. Cogn Affect Behav Neurosci 2003;3: 27–38.
[39] Bush G, Vogt BA, Holmes J, Dale AM, Greve D, Jenike MA, Rosen BR. Dorsal anterior cingulate cortex: a role in reward-based decision making. Proc Natl Acad Sci USA 2002;99:523–8.
[40] Ullsperger M, von Cramon DY. 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 2003;23:4308–14.
[41] Monchi O, Petrides M, Petre V, Worsley K, Dagher A. Wisconsin card sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. J Neurosci 2001;21:7733–41.
[42] Holroyd CB, Nieuwenhuis S, Yeung N, Nystrom LE, Mars RB, Coles MGH, Cohen JD. Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. Nat Neurosci 2004;7: 497–8.
[43] Rainville P, Duncan GH, Price DD, Carrier B, Bushnell MC. Pain affect encoded in human anterior cingulate but not somatosensory cortex. Science 1997;277:968–71.
[44] De Bruijn ERA, Hulstijn W, Verkes RJ, Ruigt GSF, Sabbe BGC. Drug-induced stimulation and suppression of action monitoring in healthy volunteers. Psychopharmacology 2004; in press.
[45] Holroyd CB, Yeung N. Alcohol and error processing. Trends Neurosci 2003;26:402–4.
[46] Sutton RS, Barto AG. Reinforcement learning: an introduction. Cambridge, MA: MIT Press; 1998.
[47] Pailing PE, Segalowitz SJ, Dywan J, Davies PL. Error negativity and response control. Psychophysiology 2002;39:198–206.
[48] Yeung N. Relating cognitive and affective theories of the error-related negativity. In: Ullsperger M, Falkenstein M, editors. Errors, conflicts, and the brain: current opinions on performance monitoring. Leipzig: MPI of Cognitive Neuroscience; 2004. p. 63–70.
[49] Damasio AR. The somatic marker hypothesis and the possible functions of the prefrontal cortex. Phil Trans R Soc Lond B Biol Sci 1996;351:1413–20.