

Performance monitoring and the causal attribution of errors

Marco Steinhauser · Andrea Kiesel

Published online: 14 May 2011
© Psychonomic Society, Inc. 2011

Abstract The present study investigated the role of causal attribution for performance monitoring in the medial frontal cortex. To this end, we compared internally and externally-caused errors in a selective attention task with respect to error-related EEG activity and error-induced adjustments of speed and attentional selectivity. Both error types evoked early negativities and later positivities in the response-locked event-related potential. However, whereas internally-caused errors caused posterror slowing, externally-caused errors were followed by reduced attentional selectivity. Moreover, the amount of reduced attentional selectivity was related to the amplitude of the early negativity on externally-caused errors. This suggests that posterror adjustments are initiated on the basis of perceived causality and, thus, causal attribution of errors.

Keywords Attention · Cognitive control · ERP

Introduction

The evaluation of action outcomes is crucial for self-regulation. Successful goal achievement requires that unintended action outcomes, such as errors, are detected and that behavior is adjusted accordingly. Different disciplines have developed complementary approaches to investigate these adjustment processes. Cognitive neurosci-

ence mainly has examined *how* behavioral adjustments are initiated upon the *detection* of an error. Research in this field has provided evidence for a performance-monitoring system in the medial prefrontal cortex (MFC; for an overview, see Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). Motivational psychologists, on the other hand, have been interested in *which* behavioral adjustment is chosen on the basis of the *evaluation* of action outcomes. For instance, attribution theory (Heider, 1958; Weiner, 1985) assumes that affective and behavioral reactions to failure are determined by perceived causality of action outcomes. In the present study, we make a first step toward an integration of these approaches. We investigated whether performance monitoring in the MFC involves causal attribution—that is, the evaluation of an error’s causality. To achieve this, we compared behavioral adjustments and brain activity following errors that are caused by oneself (internally-caused errors) and errors that are caused by external sources such as technical failure (externally-caused errors).

A viable method for investigating the dynamics of performance monitoring in the MFC is to consider event-related potentials (ERPs). Errors in speeded choice tasks elicit a negative deflection in ERPs that reaches its maximum about 50 ms after the error response on frontocentral channels. The source of this error negativity (Ne; Falkenstein, Hohnsbein, Hoormann, & Blanke, 1990) or error-related negativity (ERN; Gehring, Goss, Coles, Meyer, & Donchin, 1993) is assumed to be the dorsal anterior cingulate cortex (dACC; Yeung, Botvinick, & Cohen, 2004), a structure in the MFC. The Ne/ERN is typically followed by an error positivity (Pe; Falkenstein et al., 1990) with a centroparietal distribution that reaches a maximum between 250 and 400 ms after the error response (see also Overbeek, Nieuwenhuis, & Ridderinkhof, 2005). Whereas the Pe has been attributed to conscious error detection (Nieuwenhuis, Ridderinkhof, Blom, Band, &

M. Steinhauser (✉)
Fachbereich Psychologie, University of Konstanz,
Fach D29,
78457 Konstanz, Germany
e-mail: Marco.Steinhauser@uni-konstanz.de

A. Kiesel
Department of Psychology, University of Würzburg,
Röntgenring 11,
97070 Würzburg, Germany

Kok, 2001; Steinhauser & Yeung, 2010) or the affective-motivational significance of an error (Falkenstein et al., 1990), several theories have been proposed that assume a direct or indirect relation between the Ne/ERN and behavioral adjustments following errors (Holroyd & Coles, 2002; Maier, Steinhauser, & Hübner, 2008; Yeung et al., 2004).

Evidence for such a relation has been provided by a number of studies. In the following, we consider two types of behavioral adjustments following errors. The first and most frequently reported type of behavioral adjustment is posterror slowing, reflected by generally increased response times (RTs) following errors (Laming, 1979; Rabbitt, 1966), which has been shown to be positively correlated with the Ne/ERN amplitude in some studies (Debener et al., 2005; Gehring et al., 1993; Holroyd, Yeung, Coles, & Cohen, 2005), but not in others (e.g., Gehring & Fencsik, 2001; Nunez Castellar, Kühn, Fias, & Notebaert, 2010; van Meel, Heslenfeld, Oosterlaan, & Sergeant, 2007). The second behavioral adjustment is an increase in attentional selectivity in conflict tasks, in which participants classify a target feature while ignoring a distractor feature. In these tasks, the influence of the distractor on performance is measured by comparing trials on which the target and distractor lead to the same response (congruent stimuli) with trials on which both lead to a different response (incongruent stimuli). Using this method, some studies have shown reduced congruency effects following errors, indicating behavioral adjustments toward increased attentional selectivity (King, Korb, von Cramon, & Ullsperger, 2010; Maier, Yeung, & Steinhauser, *in press*; Ridderinkhof, 2002).

Recently, evidence has been provided that the MFC becomes active not only following internally-caused errors, but also following errors that are caused externally (e.g., Gentsch, Ullsperger, & Ullsperger, 2009; Yeung, Holroyd, & Cohen, 2005). This suggests that the MFC monitors for all events indicating failure of goal achievement. For instance, if a piano player fails to strike the right note, this might be registered by the MFC irrespective of whether this occurred because the wrong key was pressed (internally-caused) or because the piano is out-of-tune (externally-caused). The question emerges whether MFC processing distinguishes between internally and externally-caused errors—that is, whether there is causal attribution already at this early stage of action evaluation. One way to examine this is to consider behavioral adjustments following both error types. If MFC activity were associated with different behavioral adjustments following externally-caused errors than following internally-caused errors, this would imply that the underlying monitoring system evaluates errors with respect to causality.

Indeed, there are good reasons to expect different behavioral adjustments following internally and externally-caused errors. Posterror slowing and increased attentional selectivity are effective strategies for preventing internally-

caused errors, but they are less appropriate in the context of externally-caused errors. Instead of increasing effort to prevent internally-caused errors, it may even make sense to reduce effort following externally-caused errors in order to save resources. This hypothesis can be derived from studies showing that exposing humans or animals to uncontrollable action outcomes leads to reduced attention to task-relevant features (Brandstädter and Rothermund 2002; Mikulincer, 1989; Mikulincer, Kedem, & Zilkha-Segal, 1989; Reed & Antonova, 2007). For instance, using the same conflict task as in the present study, Reed and Antonova showed that attentional selectivity was reduced when participants received performance-independent feedback in a preceding categorization task.

On the basis of these considerations, we investigated whether error-related brain activity for externally-caused errors would be associated with subsequent behavioral adjustments that differed from those following internally-caused errors. We measured ERPs in a flanker paradigm (Eriksen & Eriksen, 1974) in which participants had to respond to a target letter while ignoring distractor letters that were either congruent or incongruent. Whereas internally-caused errors were induced by the presence of distractor letters and time pressure, externally-caused errors were generated by simulating a technical malfunction of the response keys on 10% of the trials. On these trials, goal achievement was prevented by blocking response execution. Accordingly, externally-caused errors (and thus, uncontrollable action outcomes) were not induced by manipulating external feedback (Gentsch et al., 2009; Reed & Antonova, 2007) but, rather, by preventing the correct response from being executed. This allowed us to focus on error monitoring, rather than feedback monitoring.

We expected that both internally and externally-caused errors would be associated with error-related brain activity. However, behavioral adjustments were hypothesized to differ for both error types. Internally-caused errors should lead to more cautious responding (i.e., posterror slowing) and/or an increased attentional selectivity, indicating a strategy change toward more effort in preventing errors. In contrast, externally-caused errors should lead to reduced attentional selectivity, indicating effort reduction. Such a pattern would suggest a role of causal attribution during performance monitoring.

Method

Participants

Twenty-eight right-handed participants (21 female) between 19 and 28 years of age ($M = 22.3$) with normal or corrected-to-normal vision participated in the study. Partic-

ipants were recruited at the University of Konstanz and received 5€ per hour. The study was conducted in accordance with institutional guidelines, and informed consent was acquired from all participants.

Task and procedure

Participants responded to a target letter (“H” or “S”), flanked by three identical flanker letters (“H” or “S”) on each side. Either the flankers were the same letters as the target (congruent stimuli; HHHHHHH or SSSSSSS), or they were different (incongruent stimuli; SSSHSS or HHSHHH). Letters were presented in white color on a black screen and subtended a visual angle of 7.5° width and 1.1° height at a viewing distance of 50 cm. Participants had to classify the target letter by giving a left-hand response or a right-hand response, respectively. The assignment of hands to target letters was counterbalanced across participants. Each trial started with a white fixation cross centrally presented for 600 ms. Then the target and flanker appeared and remained on the screen until a response was given. A new trial started 1,100 ms after the response.

Responses were given by pressing one of two response keys. Each key consisted of a lever (width, 11 mm; length, 85 mm) with the rear end mounted on a hinge and the front end resting on a spring. During presentation of the fixation cross (and unnoticed by the participants), the keys were switched between two modes. In normal mode, the lever could be pressed by 2 mm. In failure mode, the levers of both response keys were blocked by a wedge invisible to the participants and, therefore, could be pressed by 0.2 mm only. Response measurement occurred when the lever reached the stop. Note that the same mode (normal or failure) was applied to both keys at any time.

The whole experiment comprised ten blocks, and each block consisted of 72 trials. On 64 trials (32 congruent trials and 32 incongruent trials), the keys were in normal mode. On the remaining 8 trials (4 congruent trials and 4 incongruent trials), the keys were in failure mode. A relative frequency of 11% of externally-caused error was used because this is the rate of internally-caused errors typically obtained for incongruent stimuli without applying strong speed pressure in the present paradigm. Immediate repetitions of externally-caused errors were possible but were rather infrequent (about 1.2%). At the beginning of each session, participants signed an informed consent sheet and were fitted with the electrode cap. Then they read through written instructions containing information about the task, as well as the goal of the study. To explain the usage of the unusual and malfunctioning response keys, participants were told that the goal of the study was to measure the acceleration of keypresses during their responses. Moreover, to induce internally-caused errors,

participants were encouraged to respond as quickly as possible. After reading the written instructions, the experimenter additionally explained to the participants that the device measuring keypress acceleration would produce occasional malfunctions, in which case the key would be blocked and no response would be possible, but that these malfunctions should be ignored. No further instructions about the malfunction were provided. At the end of the experiment, participants were debriefed regarding the true nature of the malfunctions. Moreover, they were asked whether they had realized during the experiment that the malfunctions were part of the experiment. None of the participants reported that this had been the case.

Data acquisition and analysis

The electroencephalogram (EEG) was recorded using a BIOSEMI Active-Two system (BioSemi, Amsterdam, The Netherlands) with 64 Ag-AgCl electrodes from channels Fp1, AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C1, C3, C5, T7, TP7, CP5, CP3, CP1, P1, P3, P5, P7, P9, PO7, PO3, O1, Iz, Oz, POz, Pz, CPz, Fpz, Fp2, AF8, AF4, AFz, Fz, F2, F4, F6, F8, FT8, FC6, FC4, FC2, FCz, Cz, C2, C4, C6, T8, TP8, CP6, CP4, CP2, P2, P4, P6, P8, P10, PO8, PO4, and O2, as well as the left and right mastoid. The CMS (common mode sense) and DRL (driven right leg) electrodes were used as reference and ground electrodes. Vertical and horizontal electrooculograms (EOGs) were recorded from electrodes above and below the right eye and on the outer canthi of both eyes. All electrodes were offline rereferenced to averaged mastoids. EEG and EOG were continuously recorded at a sampling rate of 1024 Hz.

To separate the effects of internally and externally-caused errors, all trials on which an incorrect response had occurred in failure mode were excluded (2.3% of all the trials). Note that the frequency of such an incorrect response was higher in failure mode (21.0%) than in normal mode (9.0%), which was due to the small lift of key in the failure mode causing even slight finger movements to lead to a response. The remaining trials were assigned to one of three trial types. *Correct trials* were trials with a correct response and with keys being in normal mode. *Internally-caused errors* were trials with an incorrect response and with keys being in normal mode. *Externally-caused errors* were trials with a correct response and with keys being in failure mode. In this way, correct trials served as a common baseline for both error types. Internally-caused errors and correct trials differed only in the correctness of response selection, but not in key mode (failure or normal). Externally-caused errors and correct trials differed only in key mode, but not in the correctness of response selection. In addition, trials with multiple corrections (<1%) in normal mode were also excluded. For the analysis of RTs only, we additionally excluded trials with RTs less than or greater than

four standard deviations below or above the mean computed for each condition and each participant (<1%).

EEG data were analyzed using EEGLAB v6.01 (Delorme & Makeig, 2004) and custom routines written in MATLAB 7.0.4 (The Mathworks, Natick, MA). The data were band-pass filtered excluding activity below 1 Hz and above 30 Hz. Epochs were extracted ranging from 500 ms before to 1,000 ms after the response. Large artifacts were identified by computing the joint probability of each epoch and excluding epochs that deviated more than five standard deviations from the distribution mean. Ocular artifacts were corrected using a regression approach. Baseline activity was removed by subtracting the average voltage in an interval from 150 to 50 ms before the response. This baseline was chosen because it precedes the onset of the Ne/ERN.

Only incongruent trials of both error types were used for analyzing ERP data. The Ne/ERN was quantified as the difference amplitude at channel Cz between error trials and correct trials in a time window of 27–89 ms for internally-caused errors and 97–159 ms for externally-caused errors. The Pe was quantified as the difference amplitude at channel Cz between error trials and correct trials in a time window of 269–327 ms for internally-caused errors and 339–398 ms for externally-caused errors. These time windows were chosen because they captured the peak activity in each difference wave. All analyses were conducted at channel Cz because all the components showed strong activity at this channel, but similar results were obtained for channels at adjacent positions.

To examine differences in scalp topographies between error types, electrodes were combined into 11 clusters (see Fig. 3c), and scores were normalized by dividing the score from each cluster by the square root of summed squares from all clusters, separately for each condition and participant (McCarthy & Wood, 1985). The normalized scores were subjected to repeated measurement ANOVAs with the variables of cluster and error type. In addition, the similarities between scalp topographies were evaluated by calculating correlations across clusters separately for each participant. In all analyses, Huynh–Feldt corrections were applied whenever appropriate to compensate for violations of sphericity, and corrected *p*-values (but uncorrected degrees of freedom) are reported.

In a further step, we examined the relation between error-related brain activity and subsequent behavioral adjustment. Recent studies achieved this by correlating single-trial amplitudes of components of interest with RTs on subsequent trials (e.g., Debener et al., 2005). In the present study, such a method was difficult to apply because the congruency effect was a difference measure and, thus, could not be determined for single trials. Because of this, we chose a simpler method and compared behavioral adjustment following trials with high and low error-related brain activity. Trials were classified on the basis of a median split of error-related

brain activity, and this was done separately for the early (Ne/ERN) and late (Pe) error-related components and for internally and externally-caused errors.

Because single-trial amplitudes are typically very noisy, we improved classification validity by applying a linear integration method (Parra et al., 2002; Parra, Spence, Gerson, & Sajda, 2005). Using a logistic regression classifier, mean activity from all channels within a constant time period was integrated to discriminate between correct trials and error trials. The predicted value for a given trial can be viewed as a single-trial measure of error-related brain activity on this trial. To further reduce the influence of noise on the discrimination performance, we used a leave-one-out procedure (Stanislaw & Todorov, 1999). That is, the prediction value for a given trial was determined by using a classifier trained on the remaining trials. To take temporal variability of error-related brain activity into account, we constructed classifiers for three overlapping time windows for each component and error type (resulting from a moving window located around the peak of each component with a step size of 31 ms and a width of 62 ms: Ne/ERN–internally-caused errors, -5–58, 27–89, and 58–120 ms; Ne/ERN–externally-caused errors, 62–124, 93–155, and 124–187 ms; Pe–internally-caused errors, 238–300, 269–331, and 300–363 ms; Pe–externally-caused errors, 308–370, 339–402, and 370–433 ms). For each participant, component, and error type, we chose the time window in which the best discrimination performance was achieved. Discrimination performance was estimated by computing the area under the receiver operating characteristic curve, *Az*, which ranges between 0.5 (chance level) and 1 (perfect discrimination). On the basis of a median split of the prediction values, error trials were classified into low-amplitude trials and high-amplitude trials, and behavioral data following trials from both categories were compared to examine the relation between error-related activity and behavior adjustment.

Results

Behavioral data

First, we conducted an analysis on RTs of correct responses and on the rate of internally-caused errors depending on flanker congruency. As was expected, congruency had a significant effect on RTs, $F(1, 27) = 262.4$, $p < .001$, $\eta_p^2 = .91$, as well as on the error rates, $F(1, 27) = 79.2$, $p < .001$, $\eta_p^2 = .75$. RTs and error rates were larger for incongruent stimuli (441 ms, 9.5%) than for congruent stimuli (397 ms, 3.0%). Note that the rate of internally-caused errors for incongruent stimuli was comparable to the rate of externally-caused errors.

In a second step, we analyzed the effects of internally-caused errors, externally-caused errors, and correct responses on subsequent behavior. Because of the low number of errors on congruent trials and to avoid a confound between error type and congruency on trial $n-1$, we included only trials following an incongruent trial in this analysis. Furthermore, we considered only trials without key failure on the current trial. RTs of correct responses and error rates were entered into a two-way ANOVA with repeated measurement on the variables of congruency (congruent, incongruent) and $n-1$ trial (correct trial, internally-caused error, externally-caused error). Only effects involving the $n-1$ trial variable are presented.

The data are depicted in Fig. 1. For the RTs, only a main effect of $n-1$ trial was obtained. RTs were highest following internally-caused errors (433 ms), intermediate following correct responses (420 ms), and lowest following externally-caused errors (413 ms), $F(2, 54) = 5.52, p < .01, \eta_p^2 = .17$. The difference between externally-caused errors and correct responses was not significant, $F(1, 27) = 2.72, p = .11, \eta_p^2 = .09$, whereas the difference between internally-caused errors and correct responses reached significance, $F(1, 27) = 4.25, p < .05, \eta_p^2 = .14$, indicating posterror slowing following internally-caused errors. No significant interaction between congruency and $n-1$ trial was obtained, $F(2, 54) = 1.17, p = .32, \eta_p^2 = .04$. For the error rates, the interaction between $n-1$ trial and congruency was significant, $F(2, 54) = 4.70, p < .02, \eta_p^2 = .15$. The congruency effect was smallest following internally-caused errors (3.9%), intermediate following correct responses (5.5%), and largest following externally-caused errors (9.2%). The congruency effect differed significantly between externally-caused errors and correct responses, $F(1, 27) = 5.13, p < .04, \eta_p^2 = .16$, as well as between externally-caused errors and internally-caused errors, $F(1, 27) = 7.42, p < .02, \eta_p^2 = .22$, while it did not differ significantly between internally-caused errors and correct responses, $F(1, 27) = 0.89, p = .35, \eta_p^2 = .03$.

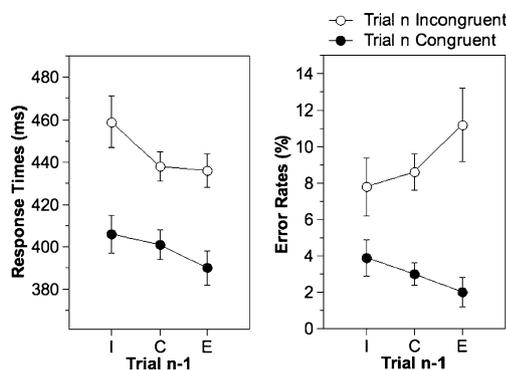


Fig. 1 Response times (left panel) and error rates (right panel) for congruent and incongruent trials as a function of previous trial type. Error bars represent standard errors of the means. I, internally caused error; C, correct; E, externally caused error

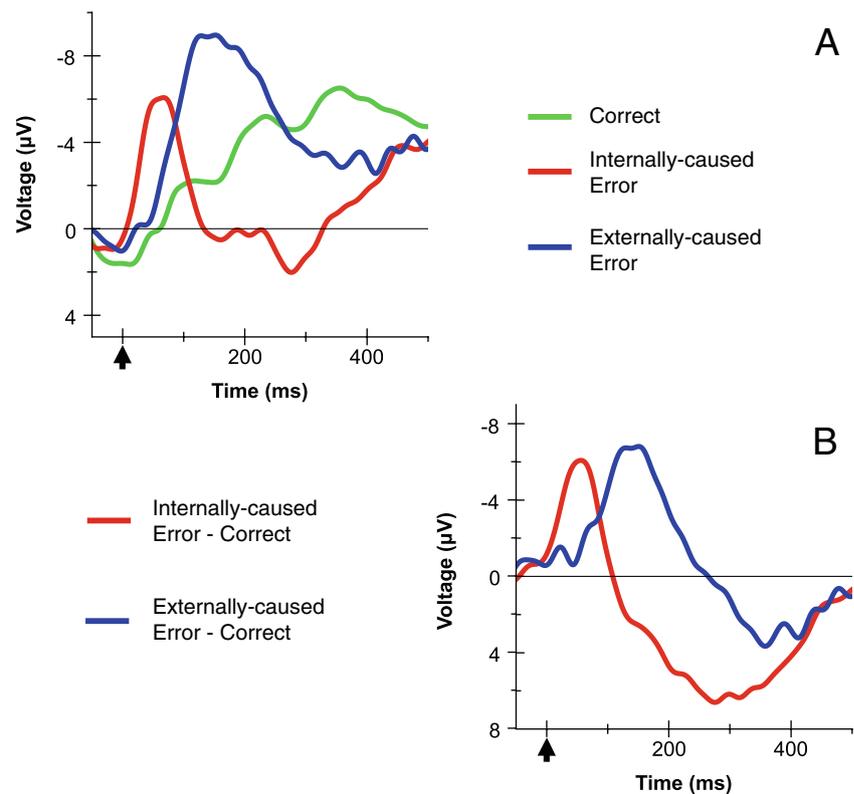
Taken together, we observed different behavioral adjustments following internally and externally-caused errors. Externally-caused errors were followed by an increased congruency effect in the error rates. In contrast, there was the typical posterror slowing in the RTs following internally-caused errors.

ERP data

In a second stage, we compared response-locked ERPs for internally-caused errors, externally-caused errors, and correct responses. Due to the small number of internally-caused errors on congruent trials, only incongruent trials were considered for all three trial types. Figures 2 and 3 show raw waveforms as well as difference waves and topographies representing the difference between internally-caused errors and correct trials and the difference between externally-caused errors and correct trials. In contrast to other studies using the flanker task, the waveform of correct responses showed a negative deflection following response execution. This could have been due to the rather long RTs in the present study, which implies that the peak of the stimulus-locked P300 occurred earlier relative to the response, than in studies with smaller RTs (e.g., Gehring et al., 1993). Another source of this negativity could be the so-called movement-monitoring potential—a negativity related to the monitoring of ongoing movements (De Bruijn, Hulstijn, Meulenbroek, & Van Galen, 2003; Foit, Grozinger, & Kornhuber, 1982)—which might be enhanced when response execution is potentially blocked. Because this potential should be similarly involved in correct and error responses, it should not affect difference waveforms between both.

As was expected, internally-caused errors produced an Ne/ERN peaking at about 60 ms following the response on frontocentral electrodes and a Pe peaking at about 300 ms following the response on centroparietal electrodes. Statistical analyses show that the Ne/ERN, $F(1, 27) = 26.3, p < .001, \eta_p^2 = .49$, and the Pe, $F(1, 27) = 37.2, p < .001, \eta_p^2 = .58$, represent significant deviations of the waveform on internally-caused error trials from that on correct trials. More interesting, however, is the observation that error-related brain activity emerges also for externally-caused errors. Externally-caused errors produce a significant negative deflection, $F(1, 27) = 31.0, p < .001, \eta_p^2 = .53$, peaking at about 130 ms after the response on centroparietal channels. Moreover, we also obtained a significant positive deflection, $F(1, 27) = 21.6, p < .001, \eta_p^2 = .44$, peaking at about 370 ms on central electrodes. Statistical analyses of the scalp topographies in Fig. 3 revealed a significant difference between the spatial distributions of early negativities in internally and externally-caused errors, as indicated by a significant interaction between electrode cluster and error type, $F(10, 270) = 5.28, p < .01, \eta_p^2 = .16$. Similarly, a significant difference was obtained for the late positivity, $F(10, 270) = 3.12, p < .02, \eta_p^2 = .10$. Taken together, externally-caused errors elicit a cascade of negative

Fig. 2 **a** Response-locked waveforms for correct trials, internally caused errors, and externally caused errors at channel Cz. **b** Difference waveforms representing internally caused errors minus correct trials and externally caused errors minus correct trials

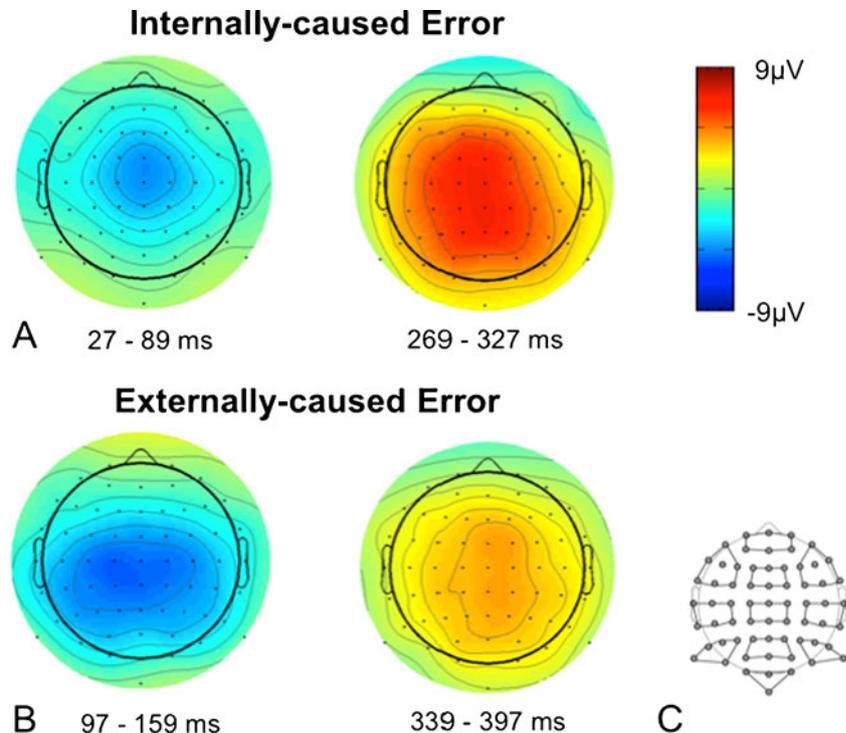


and positive components that occurs about 70 ms later than that elicited by internally-caused errors. However, the spatial distribution of activity associated with these components differs significantly for the early negativities, as well as for the late positivities.

Single-trial analysis

A further analysis aimed at examining the relation between the amount of error-related brain activity and subsequent behavioral adjustments. To this end, we distinguished

Fig. 3 **a** Scalp topographies for difference waveforms representing internally caused errors minus correct trials. **b** Scalp topographies for difference waveforms representing externally caused errors minus correct trials. Time intervals were chosen in which the early negativity (left panels) and the late positivity (right panels) were maximal. **c** Map of clusters used for testing differences between scalp topographies

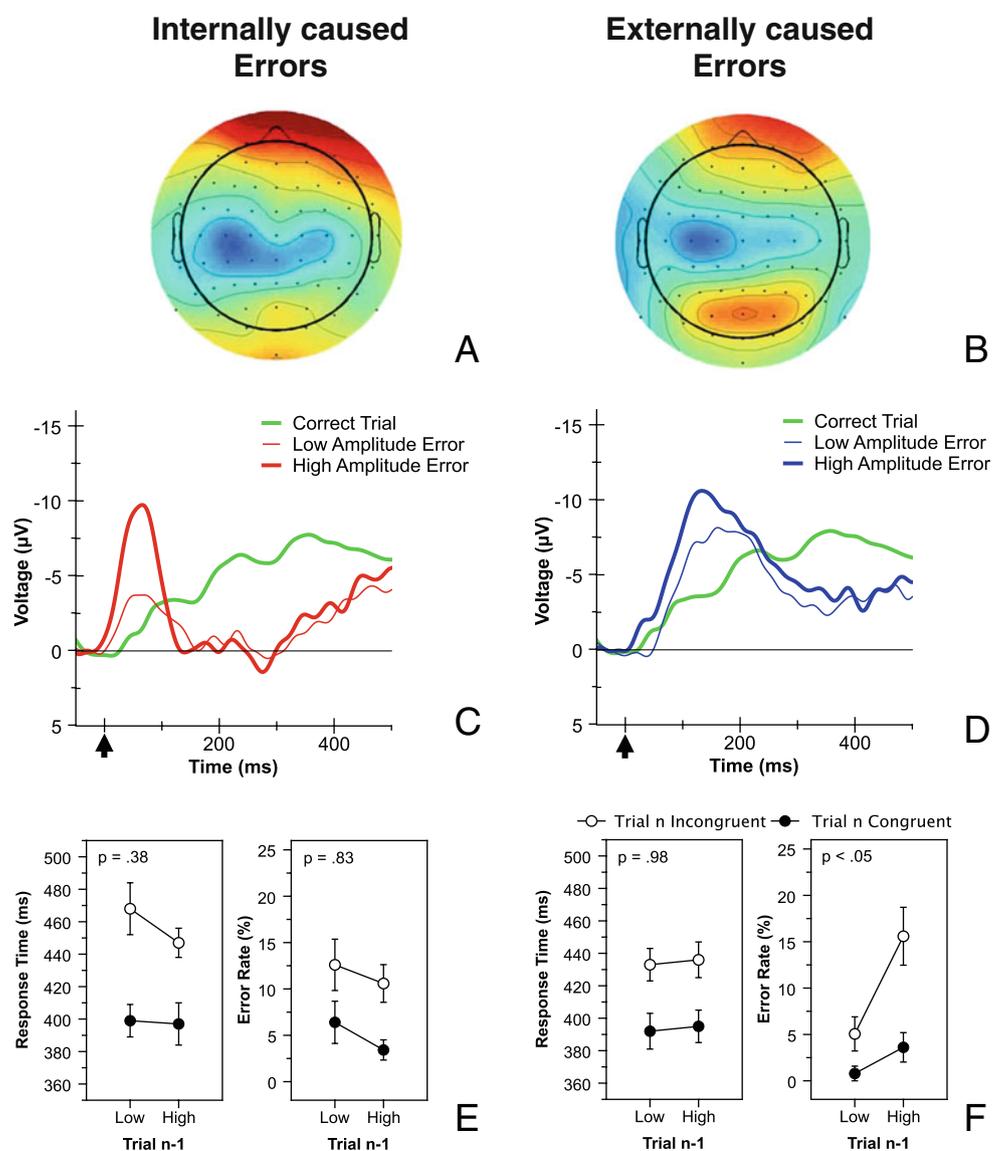


between errors with strong activity and errors with weak activity separately for each component (early negativity, late positivity) and each error type (externally-caused, internally-caused) according to the procedure described above.

We analyzed the early negativity separately for each error type. Discrimination performance of the logistic regression classifier was significantly above chance for internally-caused errors [$Az = 0.75$; $t(1, 25) = 10.5$, $p < .001$, $\eta_p^2 = .82$], as well as for externally-caused errors [$Az = 0.80$; $t(1, 27) = 23.0$, $p < .001$, $\eta_p^2 = .95$]. Figure 4a and b depict the predicted component activity of the classifier, which refers to the mean activity for each electrode multiplied with the classifier's normalized weight matrix. These topographies represent the spatial distribution of the discriminating component—that is, that component of EEG activity according to which we can best discriminate between correct trials and error trials. Interestingly, and in contrast to the raw waveforms, the spatial

distribution of the discriminating component was rather similar for internally and externally-caused errors. Statistical analysis still revealed a significant interaction between cluster and error type, $F(10, 250) = 3.58$, $p < .03$, $\eta_p^2 = .13$, suggesting that there were still significant differences between the two topographies. However, whereas the average correlation between spatial distributions of internally and externally-caused errors (calculated across clusters) was intermediate for the raw waveforms ($r = .31$), it was much larger for the classifier topographies ($r = .72$), $F(1, 25) = 11.2$, $p < .01$, $\eta_p^2 = .31$. Finally, and most important, in contrast to the raw waveforms, the distribution of the discriminating components peaked over the same central left-hemispheric electrodes. Altogether, this suggests that, despite the differences in raw waveforms, early error-related brain activity elicited by externally-caused errors resembles that elicited by internally-caused errors (Gentsch et al., 2009).

Fig. 4 Results of the single-trial analyses conducted independently for the early negativity of internally caused errors (a, c, e) and externally caused errors (b, d, f). **a, b** Spatial distributions of the component that best discriminates between errors and correct trials. **c, d** Response-locked waveforms for correct trials, high-amplitude errors, and low-amplitude errors at channel Cz. **e, f** Response times and error rates for congruent and incongruent trials following errors as a function of the estimated amplitude (low, high) of the early negativity on the preceding error trial. Values in the upper left corner represent the p -values of the interaction between congruency on trial n and amplitude on trial $n-1$. Error bars represent standard errors of the means



In a next step, the classifier was used to categorize trials as low-amplitude trials and high-amplitude trials. The waveforms for both trial types (Fig. 4c and d) indicate that we can validly distinguish between trials with low and high early negativity amplitudes. For both error types, the differences between high-amplitude and low-amplitude trials peaked in the same time windows in which the components also peaked. Most interesting, however, are the effects of amplitude on performance on the subsequent trial. To analyze this, RTs and error rates were entered into two-way ANOVAs with repeated measurement on the variables of congruency on trial n (congruent, incongruent) and amplitude on trial $n-1$ (low, high). For the analysis of internally-caused errors, data from 2 participants had to be excluded because, due to EEG artifact removal and an already low rate of internally-caused errors, fewer than five trials per cell were available for analysis. Only effects involving the amplitude variable on trial $n-1$ are reported.

A significant relation between the amplitude of the early negativity and the subsequent behavioral adjustment was obtained only following externally-caused errors (Fig. 4f). For the error rates, we obtained a significant main effect of amplitude on trial $n-1$, $F(1, 27) = 15.8$, $p < .001$, $\eta_p^2 = .37$, which was qualified by a significant interaction between amplitude on trial $n-1$ and congruency on trial n , $F(1, 27) = 4.25$, $p < .05$, $\eta_p^2 = .14$, indicating that a larger negativity was associated with a significantly larger congruency effect on the subsequent trial. For the RTs, there was a trend toward an increased congruency effect following errors with larger negativities, although both the effect of amplitude on trial $n-1$, $F(1, 27) = 0.11$, $p = .74$, $\eta_p^2 < .01$, and the interaction between both variables, $F(1, 27) < 0.001$, $p = .98$, $\eta_p^2 < .001$, were far from reaching significance. In contrast, no significant effect at all was found for internally-caused errors (Fig. 4e). The congruency effect was unaffected by amplitude on trial $n-1$ in the error rates, $F(1, 25) = 0.05$, $p = .83$, $\eta_p^2 < .01$, and the RTs, $F(1, 25) = 0.80$, $p = .38$, $\eta_p^2 = .03$. Mean RT was even slightly smaller following errors with high Ne/ERN amplitude—a difference that reached marginal significance, $F(1, 25) = 3.31$, $p < .09$, $\eta_p^2 = .12$ —suggesting no relation between Ne/ERN amplitude and posterror slowing in the present data.¹

¹ It is important to note that RTs for internally-caused errors varied with Ne/ERN amplitude. Mean RTs were increased for low-amplitude errors (378 ms), relative to high-amplitude errors (360 ms), $F(1, 25) = 10.7$, $p < .003$, $\eta_p^2 = .30$. As a consequence, when we defined posterror slowing as the RT increase on posterror trials, relative to immediately preceding error trials, we obtained a marginally significant stronger posterror slowing following high-amplitude errors (62 ms) than following low-amplitude errors (46 ms), $F(1, 25) = 3.21$, $p < .09$, $\eta_p^2 = .11$. However, it is unclear whether this effect really reflects an effect of Ne/ERN on posterror slowing, or whether it simply reflects the relation between error RT and Ne/ERN amplitude. Therefore, one should be cautious in interpreting this result.

The same analysis was applied to the late positivity for each error type, but, even though discrimination performance of the logistic regression classifier was significantly above chance for internally-caused errors [$Az = 0.73$; $t(1, 25) = 8.88$, $p < .001$, $\eta_p^2 < .76$] and for externally-caused errors [$Az = 0.71$; $t(1, 27) = 10.5$, $p < .001$, $\eta_p^2 < .80$], trials with low and high amplitudes of the late positivity did not differ significantly with respect to subsequent behavioral adjustments.

Discussion

Recent research suggested that the MFC is part of a performance-monitoring system that detects errors and triggers behavioral adjustments (for an overview, see Ridderinkhof et al., 2004). The aim of the present study was to investigate whether this system involves causal attribution of errors—that is, whether performance monitoring distinguishes between internally and externally-caused errors to determine optimal behavioral adjustment. To achieve this, we examined whether externally-caused errors, induced by simulated technical malfunctions, evoke different adjustments than do internally-caused errors. Moreover, we tested whether these adjustments are linked to activity of the performance-monitoring system in the MFC, as measured by error-related activity in the response-locked ERP.

Our results suggest that behavioral adjustments following internally and externally-caused errors differ considerably. Following internally-caused errors, we observed the usually observed posterror slowing. These results have typically been viewed as reflecting the system's attempt to avoid further internally-caused errors (but see Jentsch & Dudschig, 2009; Notebaert et al., 2009). However, this effect did not correlate with the amplitude of the ERN (at least when preceding error RT was not taken into account). This is in accord with the results of some studies (Gehring & Fencsik, 2001; Nunez Castellar et al., 2010; van Meel et al., 2007), but not with others (Debener et al., 2005; Gehring et al., 1993; Holroyd et al., 2005). One reason for the failure to find such a relation could be that internally-caused errors in the flanker paradigm represent a mixture of different error types (e.g., speed errors and attention errors; cf. Maier et al., 2008; Maier et al., in press). These error types might be associated with different behavioral adjustments being more or less correlated with the Ne/ERN amplitude. If this mixture varies across studies, this could also explain why no consistent relation between the Ne/ERN and these adjustments has been found.

In contrast, we obtained a reliable but different behavioral adjustment following externally-caused errors. Here, congruency effects in the error rates were increased, relative to trials following correct responses (and trials following

internally-caused errors). The reduced attentional selectivity reflected by the increased congruency effect could be interpreted in terms of a transient state of reduced effort or disengagement. Confronting participants with uncontrollable action outcomes has frequently been shown to cause reduced attention to task-relevant stimuli and responses (Brandstädter & Rothermund, 2002; Mikulincer, 1989; Mikulincer et al., 1989; Reed & Antonova, 2007). In contrast to adjustments following internally-caused errors, these adjustments do not serve to prevent further errors but, rather, aim at saving resources when the organism is confronted with an uncontrollable situation. Moreover, this shift in attentional strategy could also be viewed as a shift from an exploitation-oriented processing mode to an exploration-oriented processing mode, in which attention is broadened to disengage from blocked goals and to seek alternative courses of action (Brandstädter & Rothermund, 2002). Note that participants in the present experiment were instructed to ignore the occasional key malfunction, and when asked, no participant reported having been suspicious about the true nature of these malfunctions. Therefore, one could speculate that adjustments following externally-caused errors reflect an automatic, rather than an intended, process. This would also explain why these adjustments are transient. In the same way as adjustments following internally-caused errors fade in the course of subsequent correct trials, effort and attentional selectivity might recover when trials with externally-caused errors are followed by trials without externally-caused errors.

To examine whether the obtained behavioral adjustment following externally-caused errors is related to activity of the performance-monitoring system in the MFC, we considered error-related brain activity. First of all, we found externally-caused errors to be followed by an early negativity and a later positivity resembling the Ne/ERN and the Pe for internally-caused errors. Error-related activity for externally-caused errors was delayed by about 70 ms, relative to that for internally-caused errors, which presumably reflects the fact that externally-caused errors are not detectable before sensory information from the effectors indicates that the response key is blocked.

These findings mirror those of a recent study by Gentsch et al. (2009), in which externally-caused errors were induced by omitting a normally occurring positive feedback on correct responses. In this study, participants were told that the absence of feedback was due to a technical malfunction and that they had to repeat their response in such a case. In this task, an externally induced error did not imply that the correct response itself was prevented but, rather, that the positive consequence of this response was prevented. Moreover, in contrast to the present study, externally-caused errors were not entirely uncontrollable, because participants could correct these errors immediately.

Despite these differences, the results were rather similar to those in the present study. The absence of a feedback signal elicited a negativity similar to the Ne/ERN even when the response was correct, and the amplitude of this negativity was correlated with the speed of the correction response. As in our data, the negativity following externally-caused errors had a slightly more posterior spatial distribution, as compared with the Ne/ERN. Using independent component analysis (ICA), however, Gentsch et al. showed that both negativities can be attributed to the same source in the MFC, suggesting that both error types are similarly processed by the MFC.

In the present study, we obtained comparable results using a logistic regression classifier that was constructed to discriminate between correct and error trials (Parra et al., 2002; Parra et al., 2005). Although we applied this analysis independently to internally and externally-caused errors, the discriminating component extracted by the classifier (Fig. 4a and b) was very similar for both error types, suggesting that the early negativities for internally and externally-caused errors indeed share a common component. Whereas the spatial distribution of raw waveforms was rather different, presumably reflecting differences in tactile and force feedback between blocked and unblocked keypresses, the extracted discriminating components showed a uniform left-lateralized scalp distribution. On the basis of this similarity, we concluded that the early negativities for internally and externally-caused errors both reflect a similar underlying mechanism related to error processing. It is tempting to assume that the extracted component is identical to that obtained by Gentsch et al. (2009). However, whereas our component has a more left-lateralized spatial distribution, Gentsch et al. extracted a component with a frontocentral distribution. This difference could reflect the different methods used to extract these components. Whereas Gentsch et al. used an ICA that extracts independent components, our classifier-based method accumulates various sources in order to maximize discrimination between correct and error trials. Accordingly, our component presumably reflects different sources involved in error processing.

The observation of a left-lateralized early negativity resembles a finding from a study by Mathalon, Colrain, Gray, and Ford (2003), who induced errors in a go/no-go task by manipulating participants' finger movements. They found that these finger movements evoked a negativity with a left-lateralized distribution similar to that of our extracted component. Because this negativity was also obtained in a passive condition in which finger movements were induced without a task, Mathalon et al. assumed this effect not to be related to error processing but, rather, to somatosensory processes. This conclusion is plausible given that responses in their study were finger movements of the contralateral

right hand. However, our study revealed a similar result with a task in which both hands were used equally, and this result was found for externally-caused errors, as well as for internally-caused errors—at least when we considered componential activity that best discriminated between errors and correct responses. This suggests that the results of Mathalon et al. could, indeed, have reflected error processing. Future research should reveal whether this left-lateralized component represents the same source in MFC that is typically found for the Ne/ERN.

In the present study, we used the extracted component to predict subsequent behavioral adjustments. Whereas some earlier studies showed that the Ne/ERN amplitude predicts posterror slowing following internally-caused error (Debenner et al., 2005; Gehring et al., 1993; Holroyd et al., 2005; Kerns et al., 2004), we obtained a relation between the early negativity on externally-caused errors and attentional selectivity on the subsequent trial. Larger amplitudes of the early negativity for externally-caused errors were associated more with increased congruency effects—and thus, reduced attentional selectivity—than were smaller amplitudes. This implies that the adjustment of attentional selectivity following an externally-caused error depends on activity in the MFC. Provided that this activity depends on whether the type of adjustment is already determined (see the discussion below), this might suggest that the MFC is either directly involved in causal attribution or receives input from a process involved in causal attribution of error sources and that causal attribution occurs already at a very early stage of performance monitoring.

The question emerges as to how, exactly, causal attribution is achieved. One account of MFC function is the conflict-monitoring theory (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Yeung et al., 2004), which assumes that the dACC in the MFC monitors conflicts between opposite response tendencies. According to this idea, the Ne/ERN reflects posterror conflict between the still activated error response and an upcoming correction response (Yeung et al., 2004). Although it is unclear whether the early negativity following externally-caused errors is also due to dACC activity, one can speculate as to how this component could be explained by a conflict-monitoring account. For instance, the theory could be extended by assuming that the dACC also monitors other types of conflict, such as, for example, that between the intended action effect (keypress) and the actual action effect (no keypress, due to blocking). Moreover, to explain the present findings, one could further assume that performance monitoring distinguishes between both types of conflict and then triggers the respective behavioral adjustment. This theory would not necessarily imply that causal attribution already takes place *before* the early negativity is generated. Rather, one could assume that, first, the strength of behavioral adjustment is determined

according to the strength of conflict detected in the dACC and, later, the type of adjustment is determined on the basis of further evaluation of this conflict.

In contrast, the reinforcement-learning account of the Ne/ERN (Holroyd & Coles, 2002) suggests that a system located in the basal ganglia continuously updates the expectancy of action outcomes. Whenever an event implies that an action outcome will be worse than expected, a reinforcement signal is propagated to the dACC, which triggers behavioral adjustment. Other, related accounts assume that the dACC itself detects deviations from expectancy (e.g., Oliveira, McDonald, & Goodman, 2007). Basically, these theories can account for dACC activity on both error types, because internally-caused errors, as well as externally-caused errors, imply an unexpected negative action outcome. Within the framework of the reinforcement-learning account, evaluation of error causality should occur in the basal ganglia, because this is the anatomical locus of error detection and the generation of reinforcement signals (Holroyd & Coles, 2002). This theory presumably has to assume that the evaluation of error causality takes place before the early negativity is generated, because it seems plausible that, in order to adjust behavior effectively, the underlying mechanism has to determine what kind of behavioral adjustment is necessary.

Irrespective of which framework is appropriate for interpreting the early negativity following externally-caused errors, none of these accounts would assume that the early negativity is a direct correlate of the evaluation of causality or controllability. Rather, both would assume that the early negativities following internally and externally-caused errors reflect specific subprocesses of performance monitoring, such as response conflict or reinforcement. However, given that these negativities are associated with different behavioral adjustments depending on the error type, one can conclude that behavioral adjustments are initiated on the basis of an evaluation of causality or, alternatively, of controllability of action outcomes. Moreover, depending on the specific theory, one can infer that this evaluation takes place before or after these negativities are generated.

Taken together, the present results suggest that performance monitoring implies more than the mere detection of errors. In order to initiate appropriate behavioral adjustments, it is necessary that errors are causally attributed before adjustments are triggered. Provided that the Ne/ERN represents a direct correlate of behavioral adjustment (Holroyd & Coles, 2002), our results further imply that causal attribution occurs at a very early stage of performance monitoring. Further research will be required to show how different behavioral adjustments are accomplished on the basis of different error types.

This research was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG: STE 1708/1-1). We are grateful to Georg Schüssler for constructing the response keys, Lorena Deuker and Lilly Stamatescu for assistance in conducting the experiment, and Anja Achtziger and Martin Maier for helpful comments on an earlier version of the manuscript. Correspondence concerning this article should be addressed to Marco Steinhauser, Universität Konstanz, Fachbereich Psychologie, Fach D29, D-78457 Konstanz, Germany. E-mail: Marco.Steinhauser@uni-konstanz.de.

References

- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652.
- Brandstädter, J., & Rothermund, K. (2002). Intentional self-development: Exploring the interfaces between development, intentionality, and the self. In L. J. Crockett (Ed.), *Agency, motivation, and the life course: Nebraska Symposium on Motivation* (Vol. 48, pp. 31–75). Lincoln: University of Nebraska Press.
- De Bruijn, E. A., Hulstijn, W., Meulenbroek, R. G. J., & Van Galen, G. P. (2003). Action monitoring in motor control: ERPs following selection and execution errors in a force production task. *Psychophysiology*, *40*, 786–795.
- Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic imaging identifies the dynamics of performance monitoring. *The Journal of Neuroscience*, *25*, 11730–11737.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, *134*, 9–21.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *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 C. H. M. Brunia, A. W. K. Gaillard, & A. Kok (Eds.), *Psychophysiological brain research* (pp. 192–195). Tilburg, The Netherlands: Tilburg University Press.
- Foit, D., Grozinger, B., & Kornhuber, H. H. (1982). Brain potential differences related to programming, monitoring and outcome of aimed, nonaimed, fast and slow movements to a visual target: The movement monitoring potential (MMP) and the task outcome evaluation potential (TEP). *Neuroscience*, *7*, 571–577.
- Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *The Journal of Neuroscience*, *21*, 9430–9437.
- Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–390.
- Gentsch, A., Ullsperger, P., & Ullsperger, M. (2009). Dissociable medial frontal negativities from a common monitoring system for self- and externally caused failure of goal achievement. *NeuroImage*, *47*, 2023–2030.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709.
- Holroyd, C. B., Yeung, N., Coles, M. G. H., & Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General*, *134*, 163–191.
- Jentzsch, I., & Dudschig, C. (2009). Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. *The Quarterly Journal of Experimental Psychology*, *62*, 209–218.
- Kerns, J. G., Cohen, J. D., MacDonald, A. W. I., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026.
- King, J. A., Korb, F. M., von Cramon, D. Y., & Ullsperger, M. (2010). Post-error behavioral adjustments are facilitated by activation and suppression of task-relevant and task-irrelevant information processing. *The Journal of Neuroscience*, *30*, 12759–12769.
- Laming, D. (1979). Choice reaction performance following an error. *Acta Psychologica*, *43*, 199–224.
- Maier, M., Steinhauser, M., & Hübner, R. (2008). Is the error-related negativity amplitude related to error detectability? Evidence from effects of different error types. *Journal of Cognitive Neuroscience*, *20*, 2263–2273.
- Maier, M. E., Yeung, N., & Steinhauser, M. (in press). Error-related brain activity and adjustments of selective attention following errors. *NeuroImage*.
- Mathalon, D. H., Colrain, I., Gray, M., & Ford, J. M. (2003). It's not my fault: ERPs to induced errors. In M. Ullsperger & M. Falkenstein (Eds.), *Errors, conflicts, and the brain. Current opinions on performance monitoring* (pp. 27–35). Leipzig: MPI of Cognitive Neuroscience.
- McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *62*, 203–208.
- Mikulincer, M. (1989). Cognitive interference and learned helplessness: The effects of off-task cognitions on performance following unsolvable problems. *Journal of Personality and Social Psychology*, *57*, 129–135.
- Mikulincer, M., Kedem, P., & Zilkha-Segal, H. (1989). Learned helplessness, reactance, and cue-utilization. *Journal of Research in Personality*, *23*, 235–247.
- Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*, 752–760.
- Notebaert, W., Houtman, F., Van Opstal, F., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: An orienting account. *Cognition*, *111*, 275–279.
- Nunez Castellar, E., Kühn, S., Fias, W., & Notebaert, W. (2010). Outcome expectancy and not accuracy determines posterror slowing: ERP support. *Cognitive, Affective, & Behavioral Neuroscience*, *10*, 270–278.
- Oliveira, F. T. P., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: Expectancy deviation and the representation of action–outcome associations. *Journal of Cognitive Neuroscience*, *19*, 1994–2004.
- Overbeek, T. J. M., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-a-vis the ERN/Ne. *Journal of Psychophysiology*, *19*, 319–329.
- Parra, L., Alvino, C., Tang, A., Pearlmuter, B., Yeung, N., Osman, A., et al. (2002). Linear spatial integration for single-trial detection in encephalography. *NeuroImage*, *17*, 223–230.
- Parra, L., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage*, *28*, 326–341.

- Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, *71*, 264–272.
- Reed, P., & Antonova, M. (2007). Interference with judgments of control and attentional shift as a result of prior exposure to controllable and uncontrollable feedback. *Learning and Motivation*, *38*, 229–241.
- Ridderinkhof, K. R. (2002). Micro- and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research*, *66*, 312–323.
- Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, *306*, 443–447.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137–149.
- Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. *The Journal of Neuroscience*, *30*, 15643–15653.
- van Meel, C. S., Heslenfeld, D. J., Oosterlaan, J., & Sergeant, J. A. (2007). Adaptive control deficits in attention-deficit/hyperactivity disorder (ADHD): The role of error processing. *Psychiatry Research*, *151*, 211–220.
- Weiner, B. (1985). An attribution theory of achievement motivation and emotion. *Psychological Review*, *92*, 548–573.
- Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 939–959.
- Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex*, *15*, 535–544.