

Is the Error-related Negativity Amplitude Related to Error Detectability? Evidence from Effects of Different Error Types

Martin Maier*, Marco Steinhauser*, and Ronald Hübner

Abstract

■ The present study tested error detection theories of the error-related negativity (ERN) by investigating the relation between ERN amplitude and error detectability. To this end, ERN amplitudes were compared with a behavioral measure of error detectability across two different error types in a four-choice flanker task. If an erroneous response was associated with the flankers, it was considered a flanker error, otherwise it was considered a nonflanker error. Two experiments re-

vealed that, whereas detectability was better for nonflanker errors than for flanker errors, ERN amplitudes were larger for flanker errors than for nonflanker errors. Moreover, undetected errors led to strongly reduced ERN amplitudes relative to detected errors. These results suggest that, although error detection is necessary for an ERN to occur, the ERN amplitude is not related to error detectability but rather to error significance. ■

INTRODUCTION

An important prerequisite for the efficient adaptation of behavior is the detection of errors. One approach to investigate corresponding mechanisms is to consider a medio-frontal negative component of the event-related brain potential (ERP), emerging shortly after error commission in speeded response tasks. This *error negativity* (Ne; Falkenstein, Hohnsbein, Hoormann, & Blanke, 1990, 1991), or *error-related negativity* (ERN; Gehring, Goss, Coles, Meyer, & Donchin, 1993) is probably generated in the rostral anterior cingulate cortex (ACC; van Veen & Carter, 2002b; Ullsperger & von Cramon, 2001; Carter et al., 1998; Dehaene, Posner, & Tucker, 1994). This brain region is thought to be critically involved in error monitoring and behavioral adaptation (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001).

Whereas some accounts considered the ERN to be a correlate of error evaluation or behavioral adjustment that succeeds the detection of an error (e.g., Holroyd & Coles, 2002), other theories viewed the ERN to be a correlate of error detection itself. More specifically, these latter accounts assumed that the ERN represents the amount of mismatch between the actual response and the intended correct response (Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Bernstein, Scheffers, & Coles, 1995; Gehring et al., 1993; Falkenstein et al., 1990, 1991), or the amount of post-error response con-

flict, which arises when the intended correct response becomes activated during continued stimulus processing after an error (Yeung, Botvinick, & Cohen, 2004; Rodriguez-Fornells, Kurzbuch, & Münte, 2002; Botvinick et al., 2001). In both cases, the ERN would carry the information necessary for error detection because a mismatch or a conflict between the executed response and the intended correct response indicates the presence of an error. Therefore, we term these theories *error detection* theories of the ERN. From these theories, very precise predictions can be made regarding the relationship between error detection and the ERN amplitude. The greater the amount of mismatch or conflict, the greater should be the probability that the error is detected. Because the ERN is thought to represent the amount of mismatch or conflict, its amplitude should reflect the detectability of errors.

A possibility to test this prediction would be to compare ERN amplitudes between errors that were consciously detected and those that remained undetected. According to error detection theories, ERN amplitudes of detected errors should be larger than those of undetected errors. Conscious error detection can be measured by instructing participants to indicate each error by giving error signaling responses (Steinhauser, Maier, & Hübner, 2008; Steinhauser & Hübner, 2006; Ullsperger & von Cramon, 2006; Rabbitt, 1967, 1968, 2002). Unfortunately, however, studies using this method yielded inconclusive results. Whereas one study found that ERN amplitudes were reduced for undetected errors (Scheffers & Coles, 2000), other studies did not find such an effect

University of Konstanz, Germany

*The first and second authors contributed equally to this article.

(Klein et al., 2007; Endrass, Franke, & Kathmann, 2005; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001).

Another method to investigate the relation between error detectability and ERN amplitude is to compare ERN amplitudes across error types differing in detectability. Specifically, errors that are detected more efficiently should show larger ERN amplitudes than errors that are less detectable. Precisely, this prediction can be derived, for instance, from a computational model of error detection by Yeung et al. (2004), in which the ERN represents the post-error conflict between the correct and the executed response. In this model, an error is detected whenever the cumulated conflict exceeds a criterion. Because both the ERN and error detection are assumed to depend on the amount of post-error conflict, ERN amplitudes and error detection rates should be correlated across errors differing in post-error conflict. By comparing ERN amplitudes and error detection rates across these errors, one can examine whether increased detectability is accompanied by increased ERN amplitudes.

In the present study, we investigated the relation between error detectability and the ERN amplitude using this second method. We compared ERN amplitudes for two error types, which presumably differ in their detectability. We used a four-choice flanker task (Eriksen & Eriksen, 1974), in which participants had to categorize a central target letter while ignoring several irrelevant flanker letters. For stimuli in which target and flankers were associated with different responses, we distinguished between *flanker errors*, if the response associated with the flankers was executed, and *nonflanker errors*, if one of the remaining erroneous responses was performed (see Figure 1). Detectability of the error types was measured by signaling responses.

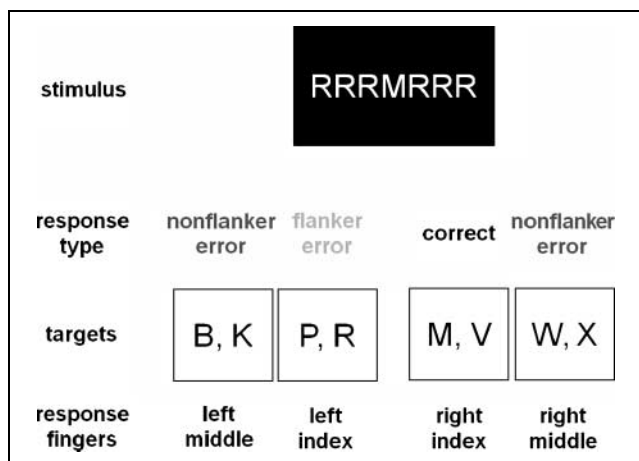


Figure 1. Stimulus–response mappings in the four-choice flanker task. Two targets were associated with each of the responses. If the target and the flankers were associated with different responses (incongruent stimulus), two types of errors could occur: If the response associated with the flankers was selected, this resulted in a flanker error. If one of the remaining alternatives was selected, this resulted in a nonflanker error.

Flanker errors and nonflanker errors are well suited for the present purpose because they are likely to differ in their detectability. In contrast to nonflanker errors, flanker errors should occur mainly on trials on which too much attention is accidentally allocated to the flankers, that is, when selective attention is suboptimal.¹ Suboptimal selective attention, however, implies that the stimulus activates the correct response only weakly. This should impair the detectability of flanker errors because error detection requires that the correct response is activated during continued stimulus processing after an error (Steinhauser et al., 2008; Yeung et al., 2004; Rabbitt, 2002). Therefore, we hypothesize that flanker errors are detected less frequently than nonflanker errors.

However, irrespective of which error type is easier to detect, the crucial question for the present study is whether the error type with better detectability also shows larger ERN amplitudes. This would support error detection theories of the ERN, because, according to this view, the ERN represents the information necessary for error detection. If, however, ERN amplitudes were smaller for the error type with better detectability, we could conclude that ERN amplitude and error detectability are unrelated. This would argue against error detection theories.

EXPERIMENT 1

In Experiment 1, a standard four-choice flanker paradigm was used to investigate whether error detectability and ERN amplitudes were different for flanker errors and nonflanker errors. In half of the blocks, the participants were instructed to signal their errors immediately by means of an error signaling response. In the other half of the blocks, no such signaling instruction was given. This condition served as control for the possibility that variations in ERN amplitudes are only due to the fact that signaling responses are given.

Methods

Participants

Fourteen participants (10 women) between 19 and 38 years old (mean = 23.2 years), with normal or corrected-to-normal vision, participated in the study. They 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.

Apparatus

Stimuli were presented on a 21-in. color monitor, and a PC-controlled stimulus presentation and response registration.

Stimuli

Stimuli were composed of the letters B, K, P, R, M, V, W, or X, and of the neutral symbols §, \$, %, &, #, or ? taken from the Arial font. Each character was resized to a visual angle of 0.64° height and 0.48° width, at a viewing distance of 72 cm. Each stimulus array consisted of a central target letter flanked by three identical distractors on each side. The whole array subtended a visual angle of 4.2° width. All letters were used as targets. Four letter pairs (B and K, P and R, M and V, W and X) were each assigned to one of four possible responses. For each of the eight possible target letters, six incongruent stimuli were constructed by combining the target letter with one of the six letters that were assigned to a different response than the target. Furthermore, for each of the eight possible target letters, six neutral stimuli were constructed by combining the target letter with one of the six neutral symbols. Together, this resulted in 48 incongruent stimuli and 48 neutral stimuli.

Procedure

Participants were instructed to respond to the identity of the target and to ignore the flankers. Responses had to be given by pressing the “W,” “S,” “L,” and “P” keys of a standard computer keyboard with the left middle finger, the left index finger, the right index finger, and the right middle finger, respectively. On some trials, participants were instructed to give an error signaling response immediately whenever they detected an error by simultaneously pressing the “Alt” and “Alt-Gr” keys with both thumbs.

Each trial started with the presentation of a fixation cross for 250 msec. Then, the stimulus array was presented for 150 msec followed by a blank screen. A new trial started after 1200 msec. If further responses occurred during this interval (e.g., an error signaling response or a spontaneous error correction), the interval was restarted.

In half of the blocks, participants were instructed to signal their errors. In the other half of the blocks, no error signaling instruction was given. Blocks with signaling instruction and blocks without signaling instruction alternated throughout the test sessions. Half of the participants began with a block with signaling instruction and the other half with a block without signaling instruction.

Each block consisted of 96 trials, one for each possible stimulus. Sixteen test blocks totaling 1536 trials were administered and were distributed over two 1-hr test sessions. In a preliminary practice session, participants had to perform six practice blocks without signaling instruction, during which they practiced the stimulus–response mapping. Furthermore, these blocks were used to adjust the error rate. Whenever the average error rate in a block fell below 15%, participants were

instructed to respond faster at the beginning of the next block. After six blocks without signaling instruction, subjects had to perform two blocks with signaling instruction so that the practice session totaled 8 blocks.

Psychophysiological Recording

Participants were seated comfortably in a dimly lit, electrically shielded room. The electroencephalogram (EEG) was recorded with Ag/AgCl electrodes from three electrode sites: Fz (frontal), FCz (fronto-central), and Cz (central). The right mastoid was recorded as an additional channel. Electrodes were referenced to the left mastoid and off-line re-referenced to linked mastoids. Electrode impedances were kept below 5 k Ω . Vertical and horizontal electrooculogram (EOG) was recorded from above and below the left eye and from the outer canthi of both eyes, respectively. EEG and EOG were continuously recorded at a sampling rate of 200 Hz and a high-pass filter of 0.1 Hz, and off-line filtered with a low-pass filter of 30 Hz.

Results

Error trials on which both signaling keys were pressed were classified as trials with valid signaling responses. Error trials on which only one signaling key was pressed were classified as trials with invalid signaling responses. The latency of a valid signaling response was calculated as the mean interval between the erroneous response and each of the two signaling keys.

To control for outliers, trials were excluded on which the response time of the choice response was two standard deviations above the condition mean (3.3%). Furthermore, error trials with invalid signaling responses and correct trials with false alarms (both <1%) were excluded. Finally, trials were excluded on which a spontaneous error correction occurred (2.5%). These trials cannot be classified as detected error trials because no signaling response occurred, nor can they be classified as undetected errors because spontaneous corrections share many properties with signaling responses and therefore are also regarded a form of error detection (e.g., Rabbitt, 2002).

Behavioral Data

Stimulus congruency. Error rates and response times of correct responses were subjected to two-way analyses of variance (ANOVAs) with repeated measurement on the variables congruency (neutral, incongruent) and block type (with signaling instruction, without signaling instruction). Whereas the error rate was higher for incongruent stimuli (22.5%) than for neutral stimuli (17.8%) [$F(1, 13) = 44.1, p < .001$], no effect of block type was revealed ($F < 1$). Similarly, mean response time

was higher for incongruent stimuli (563 msec) than for neutral stimuli (547 msec) [$F(1, 13) = 63.8, p < .001$]. Moreover, it was higher (560 msec) in blocks with signaling instruction than in blocks without signaling instruction (550 msec) [$F(1, 13) = 9.2, p < .01$].

The frequencies and latencies of signaling responses were entered into one-way ANOVAs with repeated measurement on the variable congruency (neutral, incongruent). Signaling responses were slightly more frequent but slower for errors on incongruent stimuli (96%, 449 msec) than for errors on neutral stimuli (95.0%, 446 msec). However, none of these effects reached significance ($F_s < 1$).

Error types. Flanker errors and nonflanker errors were compared by considering trials with incongruent stimuli only. A total of 43.4% of the errors on these trials were flanker errors. The response times of error responses were subjected to a two-way ANOVA with repeated measurement on the variables error type (flanker errors, nonflanker errors), and block type (with signaling instruction, without signaling instruction). It revealed an interaction between the variables block type and error type [$F(1, 13) = 5.04, p < .05$]. In blocks with signaling instruction, flanker errors were faster (509 msec) than nonflanker errors (541 msec), whereas in blocks without signaling instruction, this difference was smaller (511 and 523 msec, respectively).

The frequencies and latencies of valid signaling responses were subjected to one-way ANOVAs with repeated measurement on the variable error type (flanker errors, nonflanker errors). As expected, nonflanker errors were signaled more often (96.7%) than flanker errors (94.7%) [$F(1, 13) = 5.75, p < .05$; see Figure 2B]. No significant effect for the latency of signaling responses was found ($F < 1$).

In the preceding analyses, trials with spontaneous correction responses were excluded. However, error correction has often been viewed as another measure of error detection. Therefore, we were interested in whether the frequency of error correction shows similar effects as the frequency of error signaling. To analyze this, we entered the frequencies of spontaneous error corrections on trials with incongruent stimuli into a two-way ANOVA with repeated measurement on the variables error type (flanker errors, nonflanker errors) and block type (with signaling instruction, without signaling instruction). It revealed a trend toward more error corrections for nonflanker errors (11.7%) than for flanker errors (9.6%) which, however, was only marginally reliable [$F(1, 13) = 3.43, p < .10$]. The latencies of spontaneous corrections were not analyzed because of the low number of correction trials.

ERP Data

Segments of 500 msec before and after the first response were extracted from the continuous EEG. The average

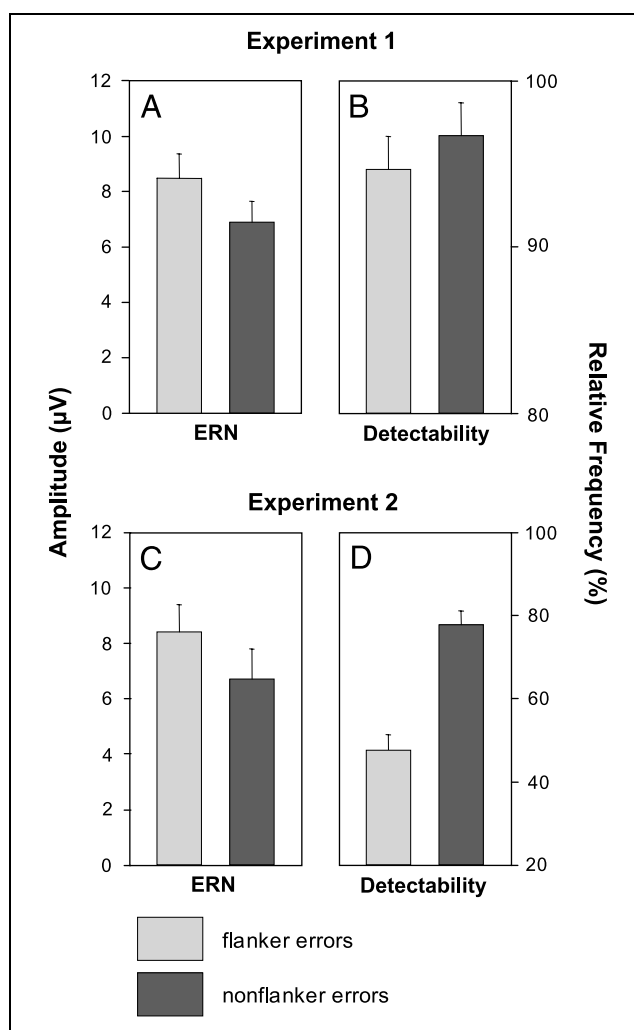


Figure 2. (A) ERN amplitudes in microvolts (μV) for flanker errors and for nonflanker errors in the incongruent condition of Experiment 1. (B) Relative frequency of error signaling responses after flanker errors and nonflanker errors in the incongruent condition in blocks with signaling instruction of Experiment 1. (C) ERN amplitudes in microvolts (μV) for signaled flanker errors and signaled nonflanker errors in the incongruent condition of Experiment 2. (D) Relative frequency of error signaling responses after flanker errors and nonflanker errors in the incongruent condition of Experiment 2. ERN = error-related negativity. Error bars represent standard errors of the mean.

voltage in the 100 msec preceding stimulus onset served as baseline.² Trials contaminated with oculomotor artifacts were rejected on the basis of vertical and horizontal EOG. The remaining ERP signals, locked to the response, were then averaged. In all conditions, a clear ERN at electrode FCz was observable, that is, the waveforms were more negative following error responses than following correct responses. The peak of this negativity was about 35 msec after the response. As a consequence, we defined the ERN amplitude as the mean difference between the ERP for correct trials and the ERP for error trials in a time window from 0 msec to 70 msec after the response at electrode FCz.

Stimulus congruency. To compare trials with neutral and incongruent stimuli, ERN amplitudes were subjected to a two-way ANOVA with repeated measurement on the variables congruency (neutral, incongruent) and block type (with signaling instruction, without signaling instruction). The mean ERN amplitude was 7.52 μV [$F(1, 13) = 79.5, p < .001$]. Moreover, it was slightly increased for incongruent stimuli (7.68 μV) as compared to neutral stimuli (7.35 μV), and for blocks with signaling instruction (7.67 μV) as compared to blocks without signaling instruction (7.36 μV). However, none of these effects reached significance ($F_s < 1$).

Error types. The averaged waveforms at electrode FCz from trials with incongruent stimuli are depicted in Figure 3A. Obviously, the waveforms for flanker errors were more negative than the waveforms for nonflanker errors. A t test confirmed that ERN amplitudes were increased for flanker errors (8.58 μV) than for nonflanker errors (7.03 μV ; see Figure 2A) [$t(13) = 2.28, p < .05$]. The same picture was revealed when we used block type as additional variable, and therefore, averaged the waveforms separately for each block type. A two-way ANOVA with repeated measurement on the variables block type and error type confirmed that ERN amplitudes were larger for flanker errors (8.48 μV) than for nonflanker errors (6.89 μV) [$F(1, 13) = 4.87, p < .05$], whereas neither the main effect of block type (with signaling instruction: 7.80 μV ; without signaling instruction: 7.57 μV) nor the interaction between both variables reached significance ($F_s < 1$).

As can be seen from further inspection of Figure 3A, the response-locked ERP for flanker errors was more negative than that for nonflanker errors in a time win-

dow from 200 to 50 msec before the response. To investigate this deflection, we computed its amplitude as the mean voltage in this time window. The amplitudes were subjected to a two-way ANOVA with repeated measurement on the variables block type (with signaling instruction, without signaling instruction) and error type (flanker errors, nonflanker errors). The ANOVA confirmed that the amplitude was larger for flanker errors ($-1.84 \mu\text{V}$) than for nonflanker errors ($-0.28 \mu\text{V}$) [$F(1, 13) = 4.84, p < .05$]. No further effects were found (all $F < 1$).

Discussion

In Experiment 1, we tested whether flanker errors and nonflanker errors differed in detectability and ERN amplitudes. As predicted, nonflanker errors were signaled more often than flanker errors (Figure 2B). Interestingly, although not reliably so, also the frequency of error corrections was larger for nonflanker errors. In contrast, however, ERN amplitudes were larger for flanker errors than for nonflanker errors (Figure 2A). This is surprising because it is often assumed that the ERN represents the information necessary for error detection. The finding that ERN amplitudes were larger for the error type with lower detectability is not compatible with such a view because it states that ERN amplitudes should be correlated with error detectability.

We argued that detectability of flanker errors is reduced because these errors occur particularly on trials on which too much attention is allocated to the flankers, which, in turn, impairs error detection. Although this interpretation is rather speculative, it receives support from an observation in our data. We obtained a negative component in the pre-response phase that was larger for flanker errors than for nonflanker errors. The timing of this component is consistent with a frontal N2, which is often interpreted as the correlate of a pre-response conflict (e.g., Yeung et al., 2004; van Veen & Carter, 2002a). This supports our interpretation because an increased pre-response conflict on flanker errors is consistent with the idea that flanker errors are due to an increased allocation of attention to the flankers.

Taken together, the results of this experiment contradict the prediction of error detection accounts of the ERN that an increased detectability of errors should be accompanied by an increased ERN. However, the question emerges whether our results are generalizable. One problem with our experiment was that the frequency of signaling responses was generally very high. Inspection of the individual signaling rates revealed that half of the participants signaled all errors in the incongruent condition. This led to rather small differences in detectability between the two error types.³ Therefore, we aimed at replicating the results under conditions in which error detectability is reduced.

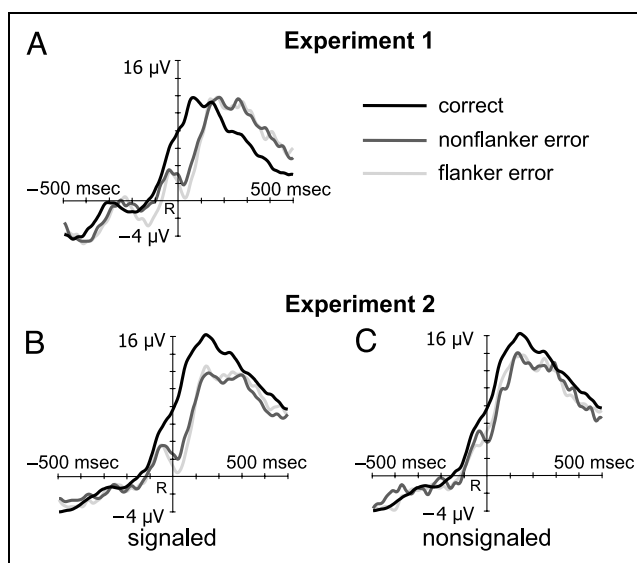


Figure 3. Grand-average response-locked waveforms at electrode FCz in the incongruent condition of Experiment 1 (A) and Experiment 2 (B and C). msec = milliseconds; μV = microvolts.

EXPERIMENT 2

In Experiment 2, a masking procedure was used to impede error detection. Masks consisting of a random feature pattern followed the letter stimuli by a short masking interval. In this way, stimulus processing should be interrupted. Moreover, if error detection relies on the continued processing of the stimulus after the error (Steinhauser et al., 2008; Yeung et al., 2004; Rabbitt, 2002), then interrupting stimulus processing should impair error detection considerably.

The main question was whether we can replicate the results of Experiment 1 under these conditions. Again, our main interest was the comparison of error detectability and ERN amplitudes between flanker errors and nonflanker errors. Because our masking procedure should produce a large number of nonsignaled errors, we additionally compared ERN amplitudes for signaled and nonsignaled errors. As mentioned earlier, this provides a further test of the error detection theories of the ERN. These theories would predict that nonsignaled errors show reduced ERN amplitudes (Scheffers & Coles, 2000).

Methods

Fourteen participants (9 women) between 20 and 29 years old (mean = 23.3 years) participated in the experiment. Except otherwise noted, the methods were identical to Experiment 1. Six different masking stimuli, one from each of the six letter stimuli, were constructed in the following way: First, each letter was cut into six equal-sized rectangles (corresponding to a two-by-three matrix). Then, each stimulus was reassembled by randomly changing the positions and orientations of the rectangles, resulting in a random feature pattern that consisted roughly of the same features as the original letter stimulus. A masking stimulus consisted of seven identical feature patterns arranged in a row.

Participants were always instructed to signal their errors. Because Experiment 1 showed that ERN amplitudes for the two error types did not differ regardless of whether error signaling was required or not, no control condition without signaling instruction was included. On each trial, the letter stimulus was followed by one of the masking stimuli with a fixed but individually adjusted stimulus-mask interval (SMI). The duration of stimulus presentation and mask presentation together was kept constant at 250 msec. Each letter stimulus was combined with each of the six masking stimuli, and each resulting combination was presented three times. As a consequence, eighteen blocks totaling 1728 trials were administered, which were distributed over two sessions of approximately 1 hr.

The practice session was organized as follows. First, the participants practiced the stimulus-response mapping during four blocks without signaling instruction.

At the end of each block, participants were instructed to respond faster whenever the error rate dropped below 15%. Then, the SMI was adjusted during four blocks with signaling instruction. The initial SMI was set to 83 msec. At the end of each block, the SMI was adjusted in a way that the error rate did not rise above 30%. The SMI of the fourth block with signaling instruction was taken as the SMI for the rest of the experiment. The resulting SMIs ranged from 33 to 67 msec (mean = 45 msec).

Results

Again, trials were excluded on which the response time of the choice response was two standard deviations above the condition mean (3.5%). Furthermore, error trials with invalid signaling responses (<1%), correct trials with false alarms (<1%), and trials with spontaneous corrections (1.3%) were excluded.

Behavioral Data

Stimulus congruency. Error rates and response times of correct responses were subjected to one-way ANOVAs with repeated measurement on the variable congruency (neutral, incongruent). Error rates were higher for incongruent stimuli (31.3%) than for neutral stimuli (19.5%) [$F(1, 13) = 70.7, p < .001$], and response times were higher for incongruent stimuli (602 msec) than for neutral stimuli (577 msec) [$F(1, 13) = 89.0, p < .001$].

The frequencies and latencies of signaling responses were subjected to one-way ANOVAs with repeated measurement on the variable congruency (neutral, incongruent). Significantly more errors were signaled for neutral stimuli (80.1%) than for incongruent stimuli (61.1%) [$F(1, 13) = 83.7, p < .001$]. In contrast, the mean latency of signaling responses was slightly but not significantly higher for incongruent stimuli (498 msec) than for neutral stimuli (491 msec, $F < 2.7$).

Error types. Again, only trials with incongruent stimuli were used to compare flanker errors and nonflanker errors. 53.6% of the errors on these trials were flanker errors. Response times of error responses were subjected to a one-way ANOVA with repeated measurement on the variable error type (flanker errors, nonflanker errors). Response times of flanker errors were shorter (571 msec) than those of nonflanker errors (580 msec), which, however, did not reach significance ($F < 1.2$).

The frequencies and latencies of valid signaling responses were subjected to one-way ANOVAs with repeated measurement on the variable error type (flanker errors, nonflanker errors). This revealed that nonflanker errors were signaled considerably more often (77.8%) than flanker errors (47.5%) [$F(1, 13) = 59.0, p < .001$];

see Figure 2D). No significant effect was obtained for the latencies of signaling responses ($F < 1$).

Again, we analyzed the frequency of spontaneous error corrections, which were subjected to one-way ANOVAs with repeated measurement on the variable error type (flanker errors, nonflanker errors). Nonflanker errors were corrected more frequently (6.5%) than flanker errors (2.9%) [$F(1, 13) = 9.11, p < .01$]. Due to the low number of correction trials, correction latencies were not analyzed.

ERP Data

Segmenting, artifact control, and filtering were performed as in Experiment 1. The average voltage in the 100-msec preceding the stimulus served as baseline. Again, the waveforms were more negative following error responses than following correct responses. The peak of this negativity was about 25 msec after the response at electrode FCz. As a consequence, we defined the ERN amplitude as the mean difference between the ERP for correct trials and the ERP for error trials in a time window between 10 msec before the response and 60 msec after the response at electrode FCz. Because this experiment yielded a sufficient number of nonsignaled errors, we now used the variable signaling (signaled, nonsignaled) as an additional variable in all ANOVAs.

Stimulus congruency. ERN amplitudes were subjected to a two-way ANOVA with repeated measurement on the variables congruency (neutral, incongruent) and signaling (signaled, nonsignaled). The mean ERN amplitude was 5.43 μV [$F(1, 13) = 31.9, p < .001$]. ERN amplitudes were larger for signaled (7.68 μV) than for nonsignaled errors (3.18 μV) [$F(1, 13) = 13.9, p < .01$]. Moreover, they were slightly but nonsignificantly larger for neutral stimuli (5.65 μV) than for incongruent stimuli (5.20 $\mu\text{V}, F < 1$).

Error types. The averaged ERP waveforms at electrode FCz in the incongruent condition are depicted in Figure 3B and C. For signaled errors, a clear ERN can be identified (Figure 3B). Moreover, the waveforms for signaled flanker errors were clearly more negative than the waveforms for signaled nonflanker errors. In contrast, ERN amplitudes were generally reduced for nonsignaled errors, and no clear difference between flanker errors and nonflanker errors is observable in this case (Figure 3C). These observations were supported by a two-way ANOVA with repeated measurement on the variables error type (flanker errors, nonflanker errors) and signaling (signaled, nonsignaled). A main effect of signaling indicated that ERN amplitudes were larger for signaled (7.55 μV) than for nonsignaled errors (3.35 μV) [$F(1, 13) = 19.7, p < .001$]. Furthermore, a significant interaction between the variables signaling and error

type was revealed [$F(1, 13) = 25.8, p < .001$]. Planned contrasts showed that ERN amplitudes were larger for signaled flanker errors (8.41 μV) than for signaled nonflanker errors (6.70 μV) [$F(1, 13) = 4.99, p < .05$; see Figure 2C]. However, no significant difference was found between nonsignaled flanker errors (2.79 μV) and nonsignaled nonflanker errors (3.91 $\mu\text{V}, F < 2.0$).⁴

Discussion

The objective of Experiment 2 was to replicate the results of Experiment 1 under conditions in which error detectability was impaired. As expected, our masking procedure successfully reduced the rate of signaled errors for all error types. Most important, however, our main results from Experiment 1 were fully replicated under these conditions. Whereas error detectability was higher on nonflanker errors (Figure 2D), ERN amplitudes were larger for flanker errors, at least when only signaled errors were considered (Figure 2C). Because stimulus masking impaired error detectability considerably, we obtained a sufficient number of nonsignaled errors for separate analyses. It turned out that ERN amplitudes were generally smaller for nonsignaled errors than for signaled errors (cf. Figure 3B and C). Moreover, nonsignaled errors showed no significant difference between flanker errors and nonflanker errors.

One result from Experiment 1 was not replicated. We obtained no N2-like difference between flanker errors and nonflanker errors in the pre-response phase. This could be a consequence of the masking procedure. When stimuli are masked, potentials evoked by the mask could overlap with stimulus-evoked potentials which might deteriorate stimulus-evoked effects (Verleger & Jaskowski, 2007). Alternatively, it is possible that interrupting stimulus processing implies that pre-response conflicts are attenuated to a degree not sufficient for a manifestation in the ERP, but which still allows flanker errors to occur.

Finally, in contrast to Experiment 1, error detectability was considerably increased for neutral stimuli (80.1%) as compared to incongruent stimuli (61.1%). Indeed, such a prediction can be derived from models of error detection (Steinhauser et al., 2008; Yeung et al., 2004). Interestingly, however, ERN amplitudes did not significantly differ between the two stimulus types. Basically, this could be viewed as additional support for the idea that the ERN is not related to error detection. If such a relation exists, a difference in error detectability of nearly 20% should be accompanied by a stable effect in ERN amplitude. However, one should be cautious with interpreting null results. Moreover, it has been reported that congruent stimuli, for which target and flankers were associated with the same response, led to larger ERN amplitudes than incongruent stimuli, (e.g., Scheffers & Coles, 2000). Therefore, the present result

might reflect the reduced power to detect such an effect when using neutral stimuli.

GENERAL DISCUSSION

The present study addressed the question whether the ERN is related to error detectability in two types of errors. An error was considered a flanker error when the erroneous response was associated with the flankers, and as a nonflanker error otherwise. In a first experiment, flanker errors showed lower detectability but larger ERN amplitudes than nonflanker errors. However, because the difference in detectability between the error types was rather small due to a nearly perfect detection performance, a second experiment was conducted in which lower detection rates were induced by means of a masking procedure. Even in this case, however, the detectability of flanker errors was considerably lower than that of nonflanker errors. Again, ERN amplitudes were larger for flanker errors, although this held for detected errors only. For undetected errors, ERN amplitudes were generally reduced relative to detected errors and no significant difference between the error types was found.

The finding that error detectability is impaired for flanker errors is consistent with current theories of error detection (Steinhauser et al., 2008; Yeung et al., 2004; Rabbitt & Vyas, 1981). Consider, for instance, the conflict monitoring model of error detection (Yeung et al., 2004), in which a conflict emerges between the executed erroneous response and the correct response because the correct response becomes activated after an error. In this model, an error is detected when the cumulated post-error conflict exceeds a criterion. Therefore, error detectability should be the better, the greater the amount of conflict after the error. It is further assumed that the proportion of attention allocated to the target and the flankers fluctuates across trials (Botvinick et al., 2001). In contrast to nonflanker errors, flanker errors should be particularly likely to occur on trials, on which a relatively large proportion of attention is allocated to the flankers. Due to the same reason, however, the correct response should become activated less strongly after the flanker error, and the resulting post-error conflict should be smaller. Consequently, the detectability of flanker errors should be impaired relative to that of nonflanker errors. Using a similar reasoning, impaired detectability of flanker errors can be predicted also from other error detection theories (Steinhauser et al., 2008; Rabbitt & Vyas, 1981).

The interpretation of a reduced detectability of flanker errors due to a suboptimal allocation of attention is supported by two further observations. First, flanker errors were not only less detectable, they also produced fewer spontaneous corrections. This provides more direct evidence that the correct response became less activated

after the error on trials with a flanker error. Second, we found an increased negativity in the range of the response-locked N2 for flanker errors in Experiment 1, which could be a marker of an increased pre-response conflict (Yeung et al., 2004; van Veen & Carter, 2002a). This is consistent with our assumption that flanker errors occur on trials on which too much attention is allocated to the flankers, because these trials should produce an increased pre-response conflict. However, this finding should be interpreted with caution because we did not replicate it in Experiment 2. One reason could be that mask-related potentials interfere with the N2-like effect, or that the pre-response conflict is attenuated by the mask.

Whereas our behavioral data alone are consistent with error detection theories, this is not the case when considered together with our ERN data. Error detection theories of the ERN state that the ERN amplitude represents the same phenomenon that constitutes the basis for error detection: an increased post-response conflict following an error (Yeung et al., 2004), or a mismatch between the executed and the intended correct responses (Bernstein et al., 1995; Falkenstein et al., 1990, 1991). Accordingly, these theories predict that increased ERN amplitudes should be accompanied by better error detectability. However, in both experiments, we observed just the opposite. ERN amplitudes were larger for errors with poorer detectability, and smaller for errors with better detectability. Thus, the phenomenon that gives rise to error detection cannot be the same phenomenon that determines the ERN amplitude.

The question arises what else is reflected by the ERN amplitude. One possibility is that the ERN amplitude reflects the evaluation of errors in terms of their significance for ongoing behavior (e.g., Hajcak, Moser, Yeung, & Simons, 2005). Accordingly, flanker errors could show larger ERN amplitudes because of their increased significance relative to nonflanker errors. Indeed, this is plausible given that the participant's goal is (a) to respond quickly but omit errors, and (b) to ignore the flankers. Whereas nonflanker errors indicate that the first goal has not been met, a flanker error implies that both goals were violated. Moreover, flanker errors indicate that more attention has subsequently been directed to the target. Thus, flanker errors are more informative than nonflanker errors with respect to necessary adaptive changes.

Regarding this latter idea, our results are also compatible with another account of the ERN, the reinforcement learning theory (Holroyd & Coles, 2002). According to this theory, the ERN is a correlate of a negative reinforcement signal, which is elicited whenever an event is worse than expected (as is the case when an error occurs instead of the expected correct response), and which serves for dynamically adjusting behavior (Holroyd & Coles, 2002). There are several ways on how this could

account for the increased ERN amplitude on flanker errors. On the one hand, flanker errors could represent a stronger violation of current expectations because they violate the current goals more strongly (as discussed above). On the other hand, the occurrence of a flanker error could be worse than the occurrence of a nonflanker error because flanker errors are less expected. Given the fact that participants detect more nonflanker errors than flanker errors, flanker errors might be less expected. As a consequence, detecting the less frequent flanker error represents a stronger violation of current expectations, which implies larger ERN amplitudes (Holroyd & Coles, 2002; see also Oliveira, McDonald, & Goodman, 2007).

If our conclusion is correct that the ERN amplitude is unrelated to error detectability, how can we explain the greatly reduced ERN amplitudes for nonsignaled errors in Experiment 2 (see also Scheffers & Coles, 2000)? One possibility is that, whereas the ERN amplitude reflects the significance of an error, the detection of this error is a necessary condition for the generation of an ERN. This is plausible because error detection should precede the evaluation of error significance, or the initiation of a reinforcement signal. From this perspective, the ERN does not reflect a process that is necessary for error detection (for instance, as implied by the conflict monitoring theory), it rather follows error detection.

This idea is not contradicted by the observation that we obtained a small but substantial ERN even on nonsignaled errors because nonsignaled errors were not necessarily undetected. For instance, Steinhauser et al. (2008) proposed that some internally detected errors remain nonsignaled because the execution of a signaling response can occasionally fail.⁵ As a consequence, the finding that nonsignaled errors showed an ERN at all could be due to the fact that some nonsignaled errors reflect errors for which error detection was successful but error signaling failed. If this interpretation is valid, it is not surprising that ERN amplitudes for nonsignaled errors did not differentiate between our error types. The mean ERN for nonsignaled errors should result from a mixture of many undetected trials on which no ERN was generated and some detected trials on which an ERN was generated. Because only detected trials can produce a difference between flanker errors and nonflanker errors, the resulting effect might be too small to be observable in the context of noise.

The idea that a failure of error signaling can produce detected but nonsignaled errors could also explain why other studies did not find any differences in ERN amplitudes between signaled and nonsignaled errors. For instance, Nieuwenhuis et al. (2001) used an antisaccade task. With respect to error detection, their results differed in two ways from typical results obtained with the flanker task. On the one hand, in the antisaccade task, many errors were not signaled. On the other hand, almost all of these nonsignaled errors were spontane-

ously corrected. Most theories of error detection would assume that corrected errors are implicitly detected even if no error signaling occurs (see, e.g., Steinhauser et al., 2008). In this case, however, signaled and nonsignaled (but corrected) errors should both produce an ERN.

Taken together, the present study provided evidence for the idea that the amplitude of the ERN is related to the significance of an error rather than to the detectability of an error, although error detection seems to be a necessary condition for the generation of the ERN. In this way, our results support theories which view the ERN as reflecting a mechanism that serves for the evaluation of errors (Magno, Foxe, Molholm, Robertson, & Garavan, 2006; Hajcak et al., 2005; Hajcak, McDonald, & Simons, 2004; Holroyd & Coles, 2002; Luu, Flaisch, & Tucker, 2000). Finally, our results are consistent with the general view that the medial prefrontal cortex, which is assumed to be the source of the ERN, is not only involved in detecting conflicts and errors (e.g., Botvinick et al., 2001; Ullsperger & von Cramon, 2001; Carter et al., 1998), but also in evaluating the relevance of internal and external events for the adaptive adjustment of behavior (e.g., Ullsperger & von Cramon, 2003; for an overview, see, Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004).

Acknowledgments

This research was supported by a grant to the third author from the Deutsche Forschungsgemeinschaft (DFG: Hu 432/8-1).

Reprint requests should be sent to Martin Maier, Universität Konstanz, Fachbereich Psychologie, Fach D29, D-78457 Konstanz, Germany, or via e-mail: martin.maier@uni-konstanz.de.

Notes

1. Indeed, it is a frequent assumption in models of the flanker task that the proportion of attention allocated to the target and the flankers fluctuates across trials (e.g., Botvinick et al., 2001).
2. Usually, a pre-response baseline is applied for analyses of the ERN. However, this period was subject to large amplitude differences between the conditions of interest (Figure 3A). Therefore, we chose a prestimulus baseline interval that aligned the voltages of the conditions of interest in the pre-responses phase.
3. However, we probably even underestimated the differences between the two error types in the present study. Whereas flanker errors should occur due to evidence collected from the flanker elements of the stimulus, nonflanker errors are selected due to the influence of noise in the system. Theoretically, however, the response associated with the flankers can also be selected due to noise without considering evidence from the stimulus. Therefore, some errors classified as flanker errors could, in fact, be nonflanker errors. Consequently, we might actually be comparing flanker errors mixed up with nonflanker errors in one condition to nonflanker errors in another condition, what would lead to an underestimation of differences between the two conditions.

4. Because we did not obtain a pre-response effect as in Experiment 1, and to ensure that our results are not due to the prestimulus baseline, we additionally analyzed the data using a pre-response baseline ranging from 250 to 50 msec before the response. This period was chosen because it preceded the point where the waveforms of correct and error responses diverge. The results were the same as that of the initial analysis. We obtained a main effect of signaling indicating that ERN amplitudes were larger for signaled (6.60 μV) than for nonsignaled errors (2.57 μV) [$F(1, 13) = 15.3, p < .01$]. Moreover, there was a significant interaction between the variables signaling and error type [$F(1, 13) = 10.2, p < .01$]. ERN amplitudes were larger for signaled flanker errors (7.40 μV) than for signaled nonflanker errors (5.81 μV) [$F(1, 13) = 6.53, p < .05$]. Again, no significant difference was obtained between nonsignaled flanker errors (2.0 μV) and nonsignaled nonflanker errors (3.13 μV , $F < 2.0$).

5. Steinhauser et al. (2008) investigated the mechanisms underlying error signaling and error correction. Their results suggested that both types of error detection responses are based on the detection of internal error corrections (which occurs when a second response is selected by exceeding a threshold after another response has already been executed). However, whereas error correction simply requires that the already selected correction response is overtly executed, error signaling involves a response that is not part of the response set of the initial choice response. Accordingly, error signaling requires switching to a new response set as well as the selection, preparation, and execution of this response. These additional processes could be responsible that error signaling is susceptible to distraction and failure (see also Rabbitt, 2002).

REFERENCES

- Bernstein, P. S., Scheffers, M. K., & Coles, M. G. (1995). "Where did I go wrong?" A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1312–1322.
- 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.
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, *280*, 747–749.
- Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science*, *5*, 303–305.
- Endrass, T., Franke, C., & Kathmann, N. (2005). Error awareness in a saccade countermanding task. *Psychophysiology*, *19*, 275–280.
- 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* (Vol. 1, pp. 192–195). Tilburg: Tilburg University Press.
- Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components: II. Error processing in choice reaction tasks. *Electroencephalography & Clinical Neurophysiology*, *78*, 447–455.
- Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology*, *51*, 87–107.
- Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–390.
- Hajcak, G., McDonald, N., & Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain and Cognition*, *56*, 189–197.
- Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology*, *42*, 151–160.
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709.
- Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D. Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *Neuroimage*, *34*, 1774–1781.
- Luu, P., Flaisch, T., & Tucker, D. M. (2000). Medial frontal cortex in action monitoring. *Journal of Neuroscience*, *20*, 464–469.
- Magno, E., Foxe, J. J., Molholm, S., Robertson, I. H., & Garavan, H. (2006). The anterior cingulate and error avoidance. *Journal of Neuroscience*, *26*, 4769–4773.
- 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.
- 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.
- Rabbitt, P. M. A. (1967). Time to detect errors as a function of factors affecting choice–response time. *Acta Psychologica*, *27*, 131–142.
- Rabbitt, P. M. A. (1968). Three kinds of error-signalling responses in a serial choice task. *Quarterly Journal of Experimental Psychology*, *20*, 179–188.
- Rabbitt, P. M. A. (2002). Consciousness is slower than you think. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *55*, 1081–1092.
- Rabbitt, P. M. A., & Vyas, S. (1981). Processing a display even after you make a response to it. How perceptual errors can be corrected. *Quarterly Journal of Experimental Psychology*, *33A*, 223–239.
- 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.
- Rodriguez-Fornells, A., Kurzbuch, A. R., & Münte, T. F. (2002). Time course of error detection and correction in humans: Neurophysiological evidence. *Journal of Neuroscience*, *22*, 9990–9996.
- Scheffers, M. K., & Coles, M. G. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 141–151.
- Steinhauser, M., & Hübner, R. (2006). Response-based strengthening in task-shifting: Evidence from shift effects produced by errors. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 517–534.
- Steinhauser, M., Maier, M., & Hübner, R. (2008). Modeling behavioral measures of error detection in

- choice tasks: Response monitoring versus conflict monitoring. *Journal of Experimental Psychology: Human, Perception and Performance*, *34*, 158–176.
- Ullsperger, M., & von Cramon, D. Y. (2001). Subprocesses of performance monitoring: A dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *Neuroimage*, *14*, 1387–1401.
- Ullsperger, M., & von Cramon, D. Y. (2003). Error monitoring using external feedback: Specific roles of the habenular complex, the reward system, and the cingulated motor area revealed by functional magnetic resonance imaging. *Journal of Neuroscience*, *233*, 4308–4314.
- Ullsperger, M., & von Cramon, D. Y. (2006). How does error correction differ from error signaling? An event-related potential study. *Brain Research*, *1105*, 102–109.
- van Veen, V., & Carter, C. S. (2002a). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & Behavior*, *77*, 477–482.
- van Veen, V., & Carter, C. S. (2002b). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, *14*, 593–602.
- Verleger, R., & Jaskowski, P. (2007). Disentangling neural processing of masked and masking stimulus by means of event-related contralateral–ipsilateral differences of EEG potentials. *Advances in Cognitive Psychology*, *3*, 193–210.
- 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*, 931–959.