The stability of error-related brain activity with increasing trials

DOREEN M. OLVET AND GREG HAJCAK

Department of Psychology, Stony Brook University, Stony Brook, New York, USA

Abstract

The error-related negativity (ERN) and error positivity (Pe) are increasingly being examined as neural correlates of response monitoring. The minimum number of error trials included in grand averages varies across studies; indeed, there has not been a systematic investigation on the number of trials required to obtain a stable ERN and Pe. In the current study, the ERN and Pe were quantified as two random trials were added to participants' ($N = 53$) ERP averages. Adding trials increased the correlation with the grand average ERN and Pe; however, high correlations ($rs > .80$) were obtained with only 6 trials. Internal reliability of the ERN and Pe reached moderate levels after 6 and 2 trials and the signal-to-noise ratio of the ERN and Pe did not change after 8 and 4 trials, respectively. Combined, these data suggest that the ERN and Pe can be quantified using a minimum of between 6 and 8 error trials.

Descriptors: Cognition, EEG/ERP, Normal volunteers

The error-related negativity (ERN) is an event-related potential (ERP) that presents as a negative deflection approximately 50 ms following an erroneous response at fronto-central midline recording sites (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). Functionally, the ERN appears to reflect relatively early error processing in the medial prefrontal cortex (cf. Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004), and source localization studies suggest that the ERN is generated in the anterior cingulate cortex (ACC; Dehaene, Posner, & Tucker, 1994; Holroyd, Dien, & Coles, 1998; van Veen & Carter, 2002). Computational models propose that the ERN represents a reinforcement learning signal (Holroyd & Coles, 2002) or that it represents response conflict (Yeung, Cohen, & Botvinick, 2004).

Although much less studied, the error positivity (Pe) is another ERP component observed on error trials. The Pe follows the ERN as a positive deflection 200–400 ms after the commission of an error (Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001; Overbeek, Nieuwenhuis, & Ridderinkhof, 2005) and has a more posterior midline scalp distribution than the ERN (Falkenstein et al., 2000). It has been suggested that the Pe reflects error awareness (Leuthold & Sommer, 1999; Nieuwenhuis et al., 2001), the emotional assessment of the error (Falkenstein et al., 2000), or a P300-like orienting response to errors (Davies, Segalowitz, Dywan, & Pailing, 2001; Hajcak, McDonald, & Simons, 2003).

The ERN and Pe have been investigated in a variety of within-subject (Hajcak & Foti, 2008; Hajcak, Moser, Yeung, & Simons, 2005; Nieuwenhuis et al., 2001; Pailing & Segalowitz, 2004; Ullsperger & von Cramon, 2006) and between-subject designs (Band & Kok, 2000; Franken, van Strien, Franzek, & van de Wetering, 2007; Hajcak, Franklin, Foa, & Simons, 2008; Luu, Collins, & Tucker, 2000; Mathalon et al., 2002). In these experiments, participants typically perform between 250 and 1,500 trials of a speeded reaction time task in relatively rapid succession. Errors tend to be rare, however, resulting in a relatively low number of trials in some ERP averages. In fact, the minimum number of error trials for conditions and participants varies greatly, ranging from 5 to 300 (e.g., Amodio et al., 2004; Franken et al., 2007; Hajcak & Simons, 2008; Morris, Yee, & Nuechterlein, 2006; Ullsperger & von Cramon, 2006; Zirnheld et al., 2004). It has been suggested that only a few error trials are required for a stable ERN (see Hajcak & Simons, 2008); however, guidance on the actual number of trials required to obtain a stable ERN based on empirical work is lacking (for similar work on the P130, cf. Cohen & Polich, 1997; Polich, 1986).

In the present study, we set forth to systematically assess the stability of the ERN and the Pe as an increasing number of errors trials were examined. We used methods similar to that reported by Polich and colleagues (Cohen & Polich, 1997; Polich, 1986). First, we measured the mean and standard deviation of the ERN and Pe as random pairs of error trials were added to participant averages (i.e., 2, 4, 6, 8, 10, 12, and 14 error trials). Additionally, we calculated the correlation between these averages and the ERN/Pe for all error trials (i.e., grand average). Internal reliability of the ERN and Pe as a function of increasing trial numbers was quantified with Cronbach's alpha; finally, the

signal-to-noise ratio (SNR) of the ERPs was also assessed. By using a multitude of analytic methods, we sought to determine how adding error trials would influence the quantification of error-related brain activity.

Method

Participants

Seventy undergraduate students (43 female) participated in the current study. Data from 2 participants were not used due to excessive artifacts. Only participants who made at least 14 errors ($N = 53$; 33 female) were included. Additionally, for the Pe analysis, 1 participant did not have usable data from the electrode of interest (Pz). No participants discontinued their participation in the experiment once procedures had begun, and all participants received course credit for their participation.

Task

The task was an arrow version of the flanker task (cf. Hajcak et al., 2005). On each trial, five horizontally aligned arrowheads were presented, and participants had to respond to the direction of the central arrowhead by pressing the left or right mouse button. Half of all trials were compatible ("$< < < < <$" or "$> > > > >$") and half were incompatible ("$< < > < <$" or "$> > < > >$"); all stimuli were presented for 200 ms with an intertrial interval that varied randomly from 500 to 1000 ms.

Procedure

Following a brief description of the experiment, electroencephalographic (EEG) sensors were attached and the participant was given detailed task instructions. Participants performed a practice block consisting of 30 trials and were told to try to be as accurate and fast as possible. The actual experiment consisted of eight blocks of 30 trials. To encourage both fast and accurate responding, participants received feedback based on their performance at the end of each block. If performance was 75% correct or lower, the message "Please try to be more accurate" was displayed; performance above 90% correct was followed by "Please try to respond faster"; otherwise, the message "You're doing a great job" was displayed.

Psychophysiological Recording, Data Reduction, and Analysis

All EEG recording, filtering, and eye movement correction parameters were identical to those reported in Hajcak and Foti (2008). Briefly, continuous EEG activity was sampled at 512 Hz using an ActiveTwo head cap and the ActiveTwo BioSemi system (BioSemi, Amsterdam, The Netherlands). Recordings were taken from 64 scalp electrodes based on the 10–20 system, as well as from 2 electrodes placed on the left and right mastoids. Four electrodes recorded the electrooculogram generated from eyeblinks and eye movements: Vertical eye movements and blinks were measured with 2 electrodes placed approximately 1 cm above and below the right eye; horizontal eye movements were measured with 2 electrodes placed approximately 1 cm beyond the outer edge of each eye. Off-line analysis was performed using Brain Vision Analyzer software (BVA, Brain Products, Gilching, Germany). EEG data were re-referenced to the numeric mean of the mastoids and bandpass filtered with cutoffs of 0.1 and 30 Hz. Eyeblink and ocular corrections were made using the method developed by Gratton, Coles, and Donchin (1983). Specific intervals for individual channels were rejected in each trial using a

semiautomated procedure, with physiological artifacts identified by the following criteria: a voltage step of more than 50.0 μV between sample points, a voltage difference of more than 300.0 μV within a trial, and a maximum voltage difference of less than 0.50 μV within a 100-ms interval. All trials were also visually inspected for other artifacts.

Response-locked ERPs were computed for error trials. The ERN was evaluated as the average activity in a 0–100-ms window relative to response onset at FCz; the Pe was evaluated as the average activity from 200 to 400 ms following response onset at Pz. A 200-ms window prior to the response (-400 to -200 ms) served as the baseline. For all analyses, the ERN and Pe were quantified based on a random subset of 2, 4, 6, 8, 10, 12, and 14 errors from each participant; the grand average ERN and Pe were quantified based on all error trials. Area measures of the ERN and Pe were chosen for two reasons: First, peak measures might be especially sensitive to noise (see Luck, 2005); second, within-subject measures such as SNR and Cronbach's alpha necessitated the use of an area measure, and it was preferable to use the same ERN/Pe metric across all analyses.

In all cases, the ERN and Pe were statistically evaluated using SPSS (Version 16.0) General Linear Model software; Greenhouse–Geisser correction was applied to p values associated with multiple degrees of freedom, repeated measures comparisons, when appropriate. Paired-sample t tests were used to compare the ERN and the Pe for each randomly chosen error trial averages (i.e., 2, 4, 6, 8, 10, 12, and 14) as 2 trials were increasingly added to the average. Pearson's correlation coefficient was also used to examine the relationship between smaller error trial averages and the grand average ERN and Pe. Internal reliability of the ERN and Pe as a function of increasing trial numbers was quantified with Cronbach's alpha. SNR at FCz and Pz was estimated using a process available in BVA. First, noise is estimated by summing the squares of the difference between each data point and the average EEG value; this sum is then divided by the number of data points minus one. Second, average total power is estimated by taking the average of the squared values of each data point. Average power of the signal then equals the average total power minus the average noise power. SNR is then calculated as the average signal power divided by average noise power. SNR was assessed using a repeated measures ANOVA using number of trials (2, 4, 6, 8, 10, 12, 14, and grand average) as the within-subjects factor. Post hoc analyses were performed using paired-sample t tests, and significance levels were adjusted with Bonferroni's correction for multiple comparisons ($p < .007$).

Results

Between-Subject Comparisons

On average, participants made 27.47 ($SD = 8.15$) errors while performing the task. Figure 1 presents the average and standard deviation of the ERN and Pe for random trial averages and for the grand averages. ERPs for the ERN and the Pe from a representative subject for random trial averages and the grand average are presented in Figure 2. Paired-sample t-tests were performed on both the ERN and the Pe area measures to examine differences between smaller trial averages as 2 trials were increasingly added to the averages. There were no significant differences when comparing increasing number of trials (2 vs. 4 trials, 4 vs. 6 trials, 6 vs. 8 trials, 8 vs. 10 trials, 10 vs. 12 trials,
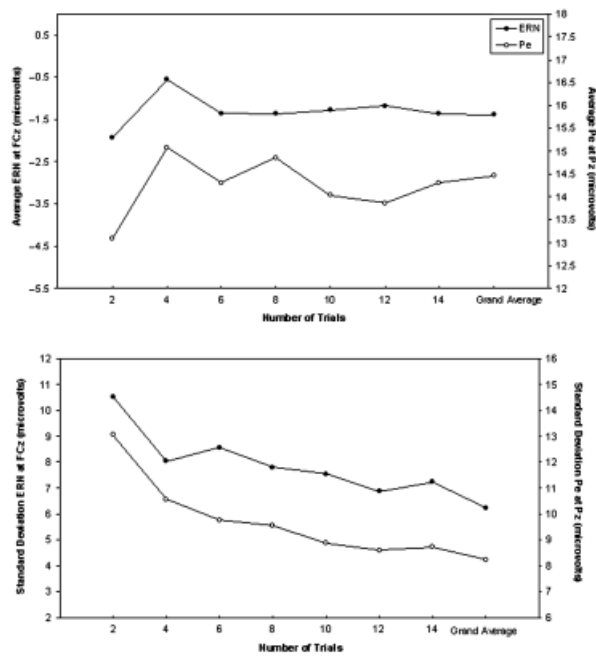
Figure 1. The average (top) and standard deviation (bottom) of the ERN at FCz (left ordinate) and the Pe at Pz (right ordinate) as a result of adding two random trials to the average and for the grand average. Please note that the left and right ordinates have different scales.

12 vs. 14 trials, and 14 vs. grand average; all *p*s > .05). The Pe for all trials also did not differ (*p*s > .05). Figure 1 suggests that both the average of the ERN and the Pe appear to stabilize after six errors are included in the average. The standard deviation for both components appears to decrease as more trials are added, although the standard deviation appears fairly similar after approximately 10 trials.

Additionally, we explored the relationship between each trial average and the ERN/Pe grand average using the Pearson correlation coefficient. Figure 3 presents the correlation coefficient between the grand average ERN/Pe and the ERN/Pe based on fewer trials. All pairs were highly significant (*p* < .001). Moreover, the ERN and Pe averages based on just six trials were highly correlated with the grand average ERN and Pe (*r*s > .80).

Within-Subject Comparisons

Hinton, Brownlow, McMurray, and Cozens (2004) have suggested that Cronbach's alpha exceeding .90 indicates excellent internal reliability, between .70 and .90 indicates high internal reliability, from .50 to .70 indicates moderate internal reliability, and below .50 is low. Figure 4 presents Cronbach's alpha for the ERN as progressively more trials were considered. Consistent with the impression from Figure 4, moderate and high internal reliabilities for the ERN were obtained with 6 and 10 errors, respectively. Cronbach's alpha for the Pe as a function of trial number is also presented in Figure 4. For the Pe, moderate and high internal reliabilities were obtained with just 2 and 6 trials, respectively. Thus, the ERN and Pe both demonstrated moderate internal reliability with just 6 errors and high internal reliability with 10 errors.

Estimates of the SNR for the ERN and Pe were also examined. SNR scores for the ERN starting with at least six error trials ranged from .25 to .63. For the ERN, a repeated measures ANOVA confirmed that there was a significant difference in the SNR across averages, $F(7,364) = 15.06$, $p < .001$. Post hoc paired-sample *t* tests indicated that there was a significant difference when comparing the ERN for 2 versus 4 trials, $t(52) = 3.61$, $p < .001$, 6 versus 8 trials, $t(52) = 2.86$, $p < .001$, and 14 versus grand average, $t(52) = 3.17$, $p < .003$; however, there was no difference between the ERN for 4 versus 6 trials, $t(52) = -0.04$, $p > .05$, 8 versus 10 trials, $t(52) = 1.56$, $p > .05$, 10 versus 12 trials, $t(52) = -0.32$, $p > .05$, and 12 versus 14 trials, $t(52) = 0.20$, $p > .05$. Thus, the SNR for the ERN did not change substantially after 8 trials were added to the average.

SNR scores for the Pe starting with at least 6 error trials ranged from 1.74 to 2.35. For the Pe, the SNR also differed as more trials were added to the averages, $F(7,357) = 19.25$, $p < .001$. Post hoc paired *t* tests showed that there was a significant difference when comparing the Pe for 2 versus 4 trials, $t(51) = 3.58$, $p < .001$; however, there was no significant difference for 4 versus 6 trials, $t(51) = 1.29$, $p > .05$, 6 versus 8 trials, $t(51) = 0.02$, $p > .05$, 8 versus 10 trials, $t(51) = 2.47$, $p > .01$, 10 versus 12 trials, $t(51) = 0.28$, $p > .05$, 12 versus 14 trials, $t(51) = -0.57$, $p > .05$, and 14 versus grand average, $t(51) = 2.34$, $p > .02$. Therefore, the SNR for the Pe did not significantly change once 4 trials were included in the average.
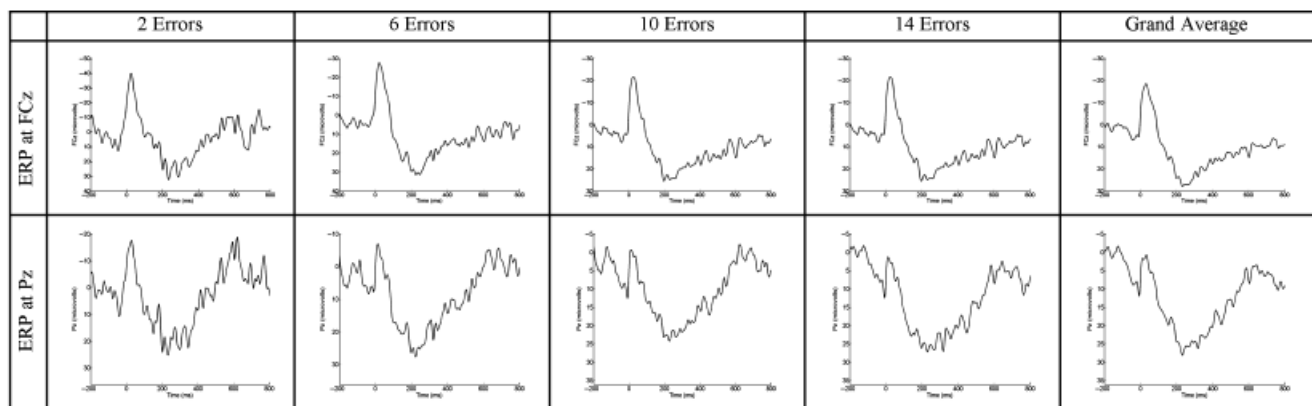

Figure 2. ERPs from a representative single subject for trials containing the average of 2, 6, 10, and 14 randomly selected trials and the grand average (i.e., 21 errors). ERP averages at FCz (top) and Pz (bottom) following error responses. Please note that the ERP figures have different scales.
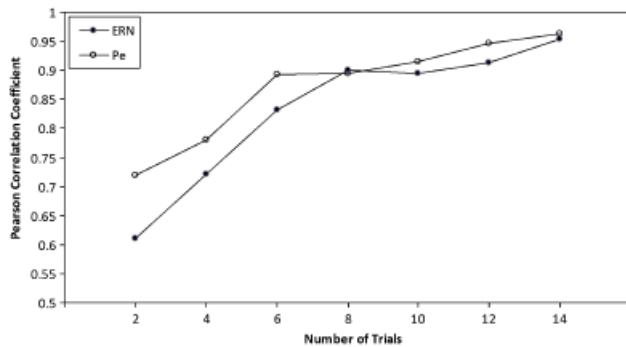
Figure 3. The Pearson correlation coefficient between the ERN at FCz and Pe at Pz derived from averages based on increasing numbers of trials with the grand average ERN and the grand average Pe, respectively.

Figure 4. Cronbach's alpha for the ERN at FCz (left ordinate) and the Pe at Pz (right ordinate) as increasing number of trials were examined. Please note that the left and right ordinates have different scales.

Discussion

The present study examined ERPs among 53 individuals who made approximately 27 errors on average—each made at least 14 errors—to examine how measures of the ERN and Pe performed when calculated based on an increasing number of randomly chosen error trials. Analyses suggested that the ERN and Pe became fairly stable once between six and eight trials were included, depending on the metric examined. For instance, the average of the ERN and Pe appeared to stabilize after about six trials per participant were included in averages. Moderate internal reliability was also achieved with 6 errors. Moreover, in ERP averages based on six trials per participant, the ERN and Pe were highly correlated ($r > .80$) with the grand average ERN and Pe, respectively. In addition, the SNR for a smaller number of trials did not significantly differ much from the SNR for the grand average after eight and four trials were added to the average for the ERN and the Pe, respectively. Although it is possible that the lack of significant differences is due to the small number of trials in each average, we used multiple measures that all converged on a common result.

In the present study, the ERN was elicited in a relatively brief speeded reaction time task comprised of only 240 trials—a task that lasted less than 10 min. Out of a total of 70 participants, 64 made at least eight errors and 66 participants made at least six errors. Thus, if one uses six to eight error trials as a minimum, less than 10% of the original sample would need to be excluded. Collectively, these data suggest that brief tasks can be used to elicit a sufficient number of errors for quantifying the ERN and Pe. This is particularly re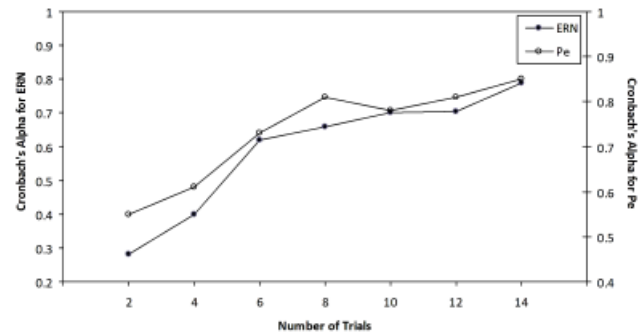levant in light of a growing literature relating the ERN to certain psychiatric disorders. For instance, evidence suggests that internalizing psychopathologies such as anxiety (Gehring, Himle, & Nisenson, 2000; Hajcak et al., 2008; Johannes et al., 2001; Ladouceur, Dahl, Birmaher, Axelson, & Ryan, 2006) and depression (Chiu & Deldin, 2007; Holmes & Pizzagalli, 2008) are characterized by increased ERNs, whereas externalizing psychopathology is related to reduced ERNs (cf. Olvet & Hajcak, 2008, for a review).

The current data suggest that six to eight trials are adequate to reliably assess error-related brain activity, which might aid in brief ERN assessments in clinical contexts. Additionally, it seems reasonable to include participants who make a relatively small number of mistakes. These data further suggest that it is possible to examine more infrequent error-related phenomena (i.e., double errors; cf. Hajcak & Simons, 2008). In light of the suggestion that an increased ERN may reflect a stable trait-like marker for internalizing psychopathology (Olvet & Hajcak, 2008), it will be important for future research to also examine the test–retest reliability of error-related brain activity.

In the present study, we used a number of statistical approaches to determine the stability of the ERN and the Pe. Future studies might further examine this issue using more sophisticated statistical approaches (i.e., hierarchical linear modeling) and consider whether the current results generalize to other tasks (e.g., Stroop) that are used in some studies to elicit the ERN. Moreover, it will be important to independently examine reliability measures in children to determine whether more trials are required to obtain a stable ERN and Pe earlier in development.

REFERENCES

Amodio, D. M., Harmon-Jones, E., Devine, P. G., Curtin, J. J., Hartley, S. L., & Covert, A. E. (2004). Neural signals for the detection of unintentional race bias. *Psychological Science*, *15*, 88–93.

Band, G. P., & Kok, A. (2000). Age effects on response monitoring in a mental-rotation task. *Biological Psychology*, *51*, 201–221.

Chiu, P. H., & Deldin, P. J. (2007). Neural evidence for enhanced error detection in major depressive disorder. *American Journal of Psychiatry*, *164*, 608–616.

Cohen, J., & Polich, J. (1997). On the number of trials needed for P300. *International Journal of Psychophysiology*, *25*, 249–255.

Davies, P. L., Segalowitz, S. J., Dywan, J., & Pailing, P. E. (2001). Error-negativity and positivity as they relate to other ERP indices of attentional control and stimulus processing. *Biological Psychology*, *56*, 191–206.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science*, *5*, 303–305.

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 and 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.

Franken, I. H., van Strien, J. W., Franzek, E. J., & van de Wetering, B. J. (2007). Error-processing deficits in patients with cocaine dependence. *Biological Psychology*, *75*, 45–51.

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.

Gehring, W. J., Himle, J., & Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive-compulsive disorder. *Psychological Science*, *11*, 1–6.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.

Hajcak, G., & Foti, D. (2008). Errors are aversive: Defensive motivation and the error-related negativity. *Psychological Science*, *19*, 103–108.

Hajcak, G., Franklin, M. E., Foa, E. B., & Simons, R. F. (2008). Increased error-related brain activity in pediatric obsessive-compulsive disorder before and after treatment. *American Journal of Psychiatry*, *165*, 116–123.

Hajcak, G., McDonald, N., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology*, *40*, 895–903.

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology*, *42*, 151–160.

Hajcak, G., & Simons, R. F. (2008). Oops!. . . I did it again: An ERP and behavioral study of double-errors. *Brain and Cognition*, *68*, 15–21.

Hinton, P. R., Brownlow, C., McMurray, I., & Cozens, B. (2004). *SPSS explained*. East Sussex, UK: Routledge.

Holmes, A. J., & Pizzagalli, D. A. (2008). Spatiotemporal dynamics of error processing dysfunctions in major depressive disorder. *Archives of General Psychiatry*, *65*, 179–188.

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., Dien, J., & Coles, M. G. (1998). Error-related scalp potentials elicited by hand and foot movements: Evidence for an output-independent error-processing system in humans. *Neuroscience Letters*, *242*, 65–68.

Johannes, S., Wieringa, B. M., Nager, W., Rada, D., Dengler, R., Emrich, H. M., et al. (2001). Discrepant target detection and action monitoring in obsessive-compulsive disorder. *Psychiatry Research*, *108*, 101–110.

Ladouceur, C. D., Dahl, R. E., Birmaher, B., Axelson, D. A., & Ryan, N. D. (2006). Increased error-related negativity (ERN) in childhood anxiety disorders: ERP and source localization. *Journal of Child Psychology and Psychiatry*, *47*, 1073–1082.

Leuthold, H., & Sommer, W. (1999). ERP correlates of error processing in spatial S-R compatibility tasks. *Clinical Neurophysiology*, *110*, 342–357.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Luu, P., Collins, P., & Tucker, D. M. (2000). Mood, personality, and self-monitoring: Negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *Journal of Experimental Psychology, General*, *129*, 43–60.

Mathalon, D. H., Fedor, M., Faustman, W. O., Gray, M., Askari, N., & Ford, J. M. (2002). Response-monitoring dysfunction in schizophrenia: An event-related brain potential study. *Journal of Abnormal Psychology*, *111*, 22–41.

Morris, S. E., Yee, C. M., & Nuechterlein, K. H. (2006). Electrophysiological analysis of error monitoring in schizophrenia. *Journal of Abnormal Psychology*, *115*, 239–250.

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*, 752–760.

Olvet, D. M., & Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: Toward an endophenotype. *Clinical Psychology Review*, *28*, 1343–1354.

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.

Pailing, P. E., & Segalowitz, S. J. (2004). The effects of uncertainty in error monitoring on associated ERPs. *Brain and Cognition*, *56*, 215–233.

Polich, J. (1986). P300 development from auditory stimuli. *Psychophysiology*, *23*, 590–597.

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.

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. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, *14*, 593–602.

Yeung, N., Cohen, J. D., & Botvinick, M. M. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 931–959.

Zirnheld, P. J., Carroll, C. A., Kieffaber, P. D., O'Donnell, B. F., Shekhar, A., & Hetrick, W. P. (2004). Haloperidol impairs learning and error-related negativity in humans. *Journal of Cognitive Neuroscience*, *16*, 1098–1112.