Psychophysiology, 48 (2011), 1420–1425. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2011.01206.x

PSYCHOPHYSIOLOGY

BRIEF REPORT

Longer term test–retest reliability of error-related brain activity

ANNA WEINBERG AND GREG HAJCAK
Department of Psychology, Stony Brook University, Stony Brook, New York, USA

Abstract

The error-related negativity (ERN) is a negative deflection in the event-related potential (ERP) following an erroneous response and is thought to reflect activity of the anterior cingulate cortex. There is accumulating evidence that the component has trait-like properties; prior evidence further suggests test–retest reliability estimates ranging from .40 to .82 over a period of 2 to 6 weeks. The present study examined temporal stability over a longer time period. Error-related brain activity was recorded from 26 subjects during an arrow version of the flankers task on two occasions separated by 1.5 to 2.5 years. Depending on the scoring method, test–retest reliability of the ERN ranged from .56 to .67. These data are consistent with previous suggestions that the ERN is a moderately stable, trait-like neural measure.

Neural correlates of error monitoring are increasingly studied using the error-related negativity (ERN), a component of the event-related potential (ERP) observed immediately following the commission of an error. The ERN is a frontocentrally maximal response-locked negative deflection in the ERP that peaks approximately 50 ms following errors. A wealth of evidence suggests that the ERN is generated by the anterior cingulate cortex (Brázdil, Roman, Daniel, & Rektor, 2005; Holroyd, Dien, & Coles, 1998; Luu, Tucker, & Makeig, 2004; Miltner et al., 2003), an area of the medial prefrontal cortex associated with the integration of affective and cognitive information (Bush, Luu, & Posner, 2000). Because the ERN appears to reflect the functioning of an executive system concerned with identifying the discrepancy between intended and actual behavior and generating adaptive responses (Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Hajcak & Simons, 2008; Holroyd & Coles, 2002), interest in the component has flourished in recent years.

Variability in error-related brain activity has been related to multiple psychiatric disorders (for a review, see Olvet & Hajcak, 2008) as well as a number of stable trait-like characteristics (Amodio, Master, Yee, & Taylor, 2008; Boksem, Tops, Wester, Meijman, & Lorist, 2006; Hajcak, McDonald, & Simons, 2004). There is also increasing evidence that the ERN exhibits trait-like qualities of its own in that it is state independent (Hajcak, Franklin, Foa, & Simons, 2008; Segalowitz et al., 2010), and resistant to symptom provocation (Moser, Hajcak, & Simons, 2005); additionally, stable measurements of error monitoring can be obtained from as few as six trials (Olvet & Hajcak, 2009b; Pontifex et al., 2010). The component has also demonstrated substantial heritability (Anokhin, Golosheykin, & Heath, 2008; Riesel, Endrass, Kaufmann, & Kathmann, 2011), supporting assertions that the ERN may represent a trait-like neural marker of executive control during response monitoring.

Previous work has demonstrated excellent short-term (i.e., 2 weeks to 6 week) reliability of error-related brain activity (Olvet & Hajcak, 2009a; Segalowitz et al., 2010). The present study sought to extend previous work on the reliability of the ERN by examining error-related brain activity across a longer period. Other ERP components associated with error responding were also examined. These include a small ERN-like component, called the correct response negativity (CRN), which is observed following correct responses (Ford, 1999; Gehring & Knight, 2000; Scheffers & Coles, 2000) and which may index the degree of engagement of response monitoring processes on correct trials. In addition, a parietal positive-going deflection in the waveform following the ERN, called the error positivity (Pe; Falkenstein et al., 2000) was examined. Relative to the ERN, the Pe is less well understood (Overbeek, Nieuwenhuis, & Ridderinkhof, 2005), but increasing evidence suggests the component may be related to error awareness (Endrass, Reuter, & Kathmann, 2007; Hughes & Yeung, 2010; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). However, research regarding variation linked to individual differences has been mixed (Endrass, Klawohn, Schuster, & Kathmann, 2008; Hajcak et al., 2008; Overbeek et al., 2005; Ridderinkhof, Ramautar, & Wijnen, 2009). To examine these components, ERP data were collected from 26 participants who performed an arrow version of the flanker task in two sessions separated by 1.5 to 2.5 years.

Method

Participants

A total of 26 Stony Brook University undergraduates (9 female), all of whom had previously participated in studies involving the

Address correspondence to: Greg Hajcak, Ph.D., Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500, USA. E-mail: greg.hajcak@stonybrook.edu

flanker task, were recruited to return to the laboratory for a second visit as a part of a larger study. Participants returned between 1.5 and 2.5 years after their initial visit, and the mean time between visits was 1.9 years ($SD = 0.23$). The mean age was 21.12 years ($SD = 1.48$), and 46% of the sample was Caucasian/European, 15.4% was African American, 30.8% was Asian, 3.8% was Hispanic, and 7.7% reported "other." All participants were paid $20.00 an hour for their participation in the return visit of the study.

Task and Materials

The procedures at Time 1 and Time 2 were identical. At each laboratory visit, an arrow version of the flanker task (Eriksen & Eriksen, 1974) was administered on a Pentium D class computer, using Presentation software (Neurobehavioral Systems, Inc., Albany, CA). Each stimulus was displayed on a 19-in. (48.3 cm) monitor. On each trial, five horizontally aligned arrowheads were presented. Half the trials were compatible ("< < < < <" or "> > > > >") and half were incompatible ("< < > < <" or "> > < > >"); the order of compatible and incompatible trials was random. Each set of arrowheads occupied approximately 1.3° of visual angle vertically and 9.2° horizontally. All stimuli were presented for 200 ms followed by an intertrial interval that varied randomly between 2300 and 2800 ms.

Procedure

Participants were seated at a viewing distance of approximately 24 in. (61 cm) and were instructed to press the right mouse button if the center arrow pointed right and to press the left mouse button if the center arrow pointed left. Participants performed a practice block containing 30 trials during which they were instructed to be both as accurate and fast as possible. The actual task consisted of 11 blocks of 30 trials (330 trials total), with each block initiated by the participant. To encourage both fast and accurate responding, participants received performance-based feedback 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

Continuous electroencephalogram (EEG) recordings were collected using an elastic cap and the ActiveTwo BioSemi system (BioSemi, Amsterdam, Netherlands). Sixty-four electrodes were used, based on the 10/20 system, as well as two electrodes on the right and left mastoids. Electrooculogram (EOG) generated from eye movements and eyeblinks was recorded using four facial electrodes: horizontal eye movements were measured via two electrodes located approximately 1 cm outside the outer edge of the right and left eyes. Vertical eye movements and blinks were measured via two electrodes placed approximately 1 cm above and below the right eye. The EEG signal was preamplified at the electrode to improve the signal-to-noise ratio with a gain of 1 by a BioSemi ActiveTwo system (BioSemi, Amsterdam). The data were digitized at 24 bit resolution with a sampling rate of 512 Hz using a low-pass fifth order sinc filter with a half-power cutoff of 102.4 Hz. Each active electrode was measured online with respect to a common mode sense active electrode producing a monopolar (nondifferential) channel. Off-line, all data were referenced to the average of the left and right mastoids and bandpass filtered with low and high cutoffs of 0.1 and 30 Hz, respectively; eyeblink and ocular corrections were conducted per Gratton, Coles, and Donchin (1983).

A semiautomatic procedure was employed to detect and reject artifacts. Data from individual channels were rejected if a voltage step of more than 50.0 μV between sample points or a voltage difference of 300.0 μV within a trial existed. In addition, data were identified as artifacts if a voltage difference of less than 0.50 μV within 100-ms intervals was present. Visual inspection of the data was then conducted to detect and reject any remaining artifacts.

The EEG was segmented for each trial beginning 500 ms before each response onset and continuing for 1500 ms (i.e., for 1000 ms following the response); a 200-ms window from −500 to −300 ms prior to response onset served as the baseline. The ERN and CRN were evaluated as the average activity on error and correct trials, respectively, from 0 to 100 ms after response at FCz; the reliability of the difference between the ERN and CRN (i.e., ΔERN) was also evaluated. In addition to area measures, the ERN and CRN were scored as the largest negative peak in the 0–100-ms window after error and correct responses, respectively. The negative peak was also scored on the difference wave between error and correct trials (ΔERN peak). Finally, a 50-ms window consisting of the area around the peak (i.e., 25 ms before and after the peak) was calculated for both incorrect (ERN) and correct (CRN) trials, as was a difference score (area around the ΔERN peak). Peak latency of the CRN and ERN were also evaluated between Times 1 and 2. Consistent with previous work, the Pe was scored as the average activity from 200 to 400 ms at Pz on error trials; the corresponding activity in this time window was also evaluated on correct trials. All ERP measures were evaluated with a 2 (trial type: correct vs. incorrect) × 2 (time point: Time 1 vs. Time 2) repeated-measures analysis of variance (ANOVA) to assess differences between error and correct trials, testing session, and their potential interaction. Paired samples t tests were performed for follow-up post hoc comparisons.

Reaction times were evaluated with a 2 (trial type: correct vs. incorrect) × 2 (time point: Time 1 vs. Time 2) repeated-measures ANOVA. Accuracy, expressed as a percentage of correct trials, was compared between Times 1 and 2 using a paired samples t test. Trials were removed from the analysis if reaction times were faster than 200 ms or slower than 800 ms.

All data were screened for multivariate outliers using mahalanobis distances; none were detected. Statistical analyses were conducted using SPSS (version 17.0). Test–retest reliability of behavioral and ERP measures at the initial testing (Time 1) and 1.5 to 2.5 years later (Time 2) was examined in terms of both intersubject stability (i.e., Pearson's r) and score agreement (i.e., intraclass correlation, ICC).

Results

Comparing Time 1 to Time 2

Behavioral data from Times 1 and 2 are presented in Table 1. Participants were faster on error compared to correct trials, $F(1,24) = 188.80$, $p < .001$, $\eta p^2 = .89$, but did not differ between Times 1 and 2, $F(1,24) = 2.30$, $p = .14$, $\eta p^2 = .09$. However, the

Table 1. *Means, Standard Deviations, and Test–Retest Reliability for Behavioral and ERP Measures at Time 1 and Time 2*

	Time 1 Mean (*SD*)	Time 2 Mean (*SD*)	Time 1 to Time 2	
			r	ICC
Accuracy (% correct)	88.03 (4.98)	88.08 (4.23)	.40	.41*
Correct RT (ms)	387.84 (38.81)	387.22 (41.03)	.65**	.66**
Incorrect RT (ms)	317.98 (30.07)	338.80 (49.43)	.48*	.39*
ERN area (μV)	1.64 (6.92)	− 0.85 (6.26)	.67**	.62**
CRN area (μV)	10.64 (6.08)	9.48 (4.65)	.75**	.72**
ΔERN area (μV)	− 9.00 (5.42)	− 10.32 (5.99)	.69**	.66**
Pe area (μV)	17.05 (7.21)	13.40 (6.63)	.68**	.68**
ERN peak (μV)	− 2.67 (7.81)	− 5.41 (7.71)	.65**	.62**
CRN peak (μV)	8.84 (6.62)	8.09 (5.06)	.58**	.55**
ΔERN peak (μV)	− 11.56 (6.07)	− 13.50 (7.53)	.63**	.60**
Area around ERN peak (μV)	1.52 (7.33)	− 1.61 (6.29)	.66**	.66**
Area around CRN peak (μV)	9.96 (6.40)	8.92 (4.71)	.66**	.66**
Area around Δpeak (μV)	− 8.44 (5.24)	− 10.53 (6.31)	.56**	.56**
ERN peak latency (ms)	46.48 (20.33)	43.21 (17.43)	.29	.29
CRN peak latency (ms)	30.35 (28.48)	33.28 (30.07)	− .08	− .08
ΔERN peak latency (ms)	17.25 (33.36)	9.92 (33.71)	− .13	− .14

Note. ICC = intraclass correlations.
*$p < .05$, **$p < .01$.

effect of response type was qualified by time, $F(1,24) = 6.01$, $p < .05$, $\eta p^2 = .21$, such that participants were slower on error trials at the second testing session, $t(25) = 2.33$, $p < .05$, though there were no significant differences in correct reaction times between time points, $t(25) = 0.09$, $p = .93$. Accuracy did not differ from Time 1 to Time 2, $t(25) = 0.05$, $p = .96$.

Average ERP values from Times 1 and 2 are presented in Table 1; Figure 1 presents the grand-average stimulus-locked ERPs at FCz and Pz for error and correct trials at Time 1 and Time 2. Consistent with previous literature, area measures of the ERN were more negative than area measures of the CRN, $F(1,24) = 79.89$, $p < .001$, $\eta p^2 = .78$. There was no significant difference between Times 1 and 2, $F(1,24) = 3.78$, $p = .08$, $\eta p^2 = .15$, nor did the effect of trial type vary according to time, $F(1,24) = 2.25$, $p = .15$ $\eta p^2 = .09$.[1]

Peak measures of the ERN were also significantly more negative than peak measures of the CRN, $F(1,24) = 86.65$, $p < .001$, $\eta p^2 = .81$. Peak measures did not differ significantly between Times 1 and 2, $F(1,24) = 2.66$, $p = .12$, $\eta p^2 = .11$, nor did the effect of trial type vary according to time, $F(1,24) = 1.92$, $p = .18$, $\eta p^2 = .08$. Likewise, area around the ERN peak was significantly more negative than the area around the CRN peak, $F(1,24) = 90.87$, $p < .001$, $\eta p^2 = .79$, although area around the peak measures did not differ significantly between Times 1 and 2, $F(1,24) = 3.81$, $p = .07$, $\eta p^2 = .17$, nor did time interact with response type to determine area around the peak, $F(1,24) = 2.67$, $p = .16$, $\eta p^2 = .10$. In terms of latency, the ERN peaked significantly later than the CRN, $F(1,24) = 10.26$, $p < .01$, $\eta p^2 = .33$, though there were no significant differences in peak latency from Time 1 to Time 2, $F(1,24) < 1$, $p = .72$, $\eta p^2 = .01$. The effect of time did not interact with the effect of trial type, $F(1,24) < 1$, $p = .60$, $\eta p^2 = .01$.

[1]Because there was a trend toward a main effect of time in the time window of the ERN/CRN, we conducted two post hoc *t* tests to explore whether either the ERN or the CRN differed from Time 1 to Time 2. The CRN did not differ significantly from Time 1 to Time 2, $t(21) = 1.07$, $p = .29$; however, at Time 2, the ERN was larger than at Time 1, $t(21) = 2.13$, $p = .05$.

In the time window of the Pe, responses on incorrect trials were significantly more positive than responses on correct trials, $F(1,24) = 140.09$, $p < .001$, $\eta p^2 = .86$. In the time window of the Pe, ERP values were more positive on both error and correct trials at Time 1 compared to Time 2, $F(1,24) = 4.71$, $p < .05$, $\eta p^2 = .17$. However, the effect of trial type did not vary according to time, $F(1,24) = 3.54$, $p = .07$, $\eta p^2 = .13$.

Test–Retest Reliability

Test–retest reliability indices for behavioral and ERP measures are presented in Table 1; individual subject scores for the area measures at Time 1 and Time 2 are plotted in Figure 2. As indicated, the stability—both in terms of intersubject stability (*r*) and score agreement (ICC)—was high across most measures of error-related brain activity. In addition, behavioral measures, area ERP measures, and peak ERP measures were characterized by comparable stability for correct and incorrect trials. Exceptions include peak latency measures for both the ERN and CRN.

Because the CRN consists of a relatively minor deflection riding on top of a much greater positivity (the P300 following the imperative stimulus), partial correlation analyses were conducted to examine the relationship between the CRN at Time 1 and Time 2, controlling for the P3 following correct responses at each time point. After partialling out the variance contributed by the P3, test–retest reliability for the CRN remained high ($r = .81$, $p < .001$).

Discussion

Consistent with previous work suggesting that the ERN is a remarkably stable and reliable signal (Olvet & Hajcak, 2009a, 2009b; Pontifex et al., 2010; Segalowitz et al., 2010), the present study demonstrated test–retest reliability estimates between .56 and .75 for indices of error-related brain activity—regardless of whether peak or area measures were used—over a period of approximately 2 years. This is comparable to retest reliability scores of other major ERP components (e.g., P1, N1, P2, N2, P300; Fallgatter et al., 2001; Segalowitz & Barnes, 1993; Walhovd & Fjell, 2002), which have ranged from .37 to .91 in sessions

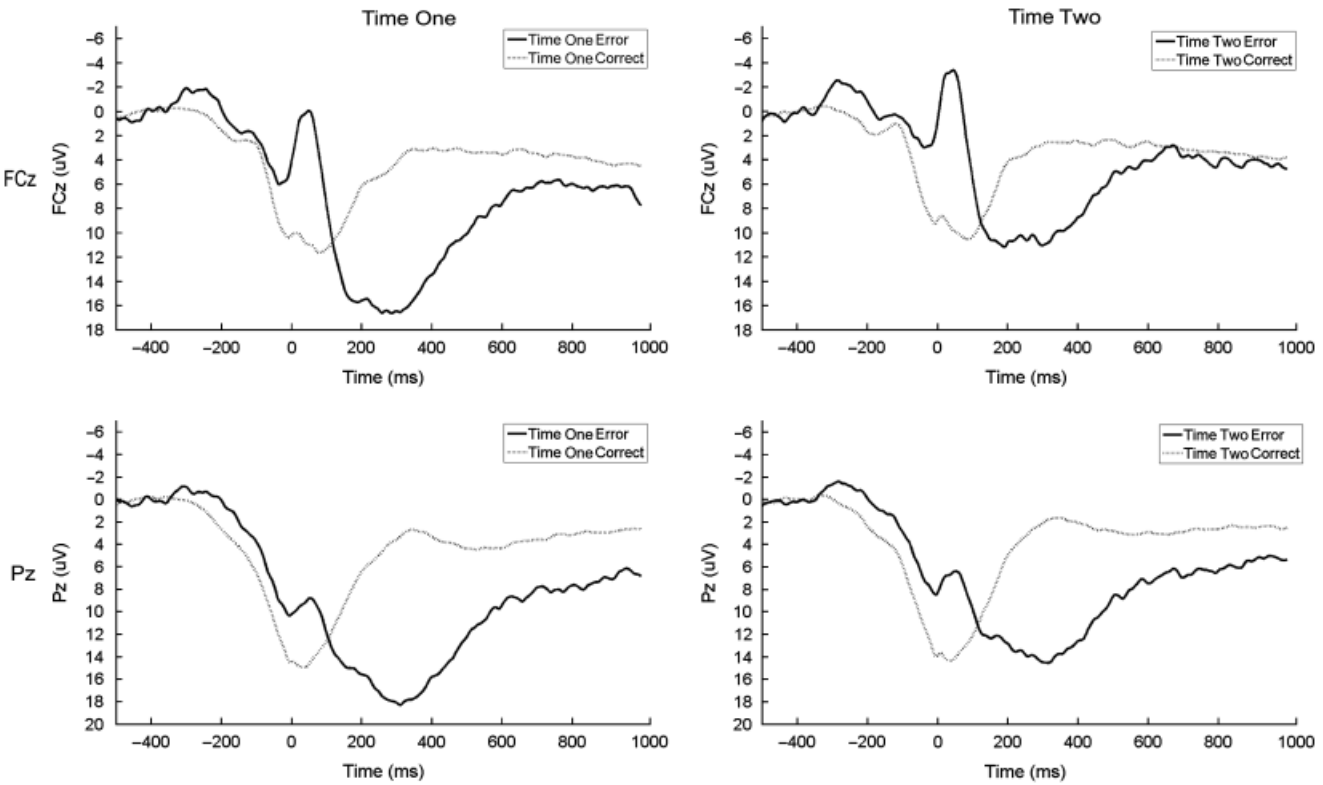

Figure 1. Response-locked ERP waveforms recorded at Time 1 (left) and Time 2 (right) from FCz (top) and Pz (bottom), comparing correct and error trial waveforms. Response onset occurred at 0 ms and negative is plotted up.

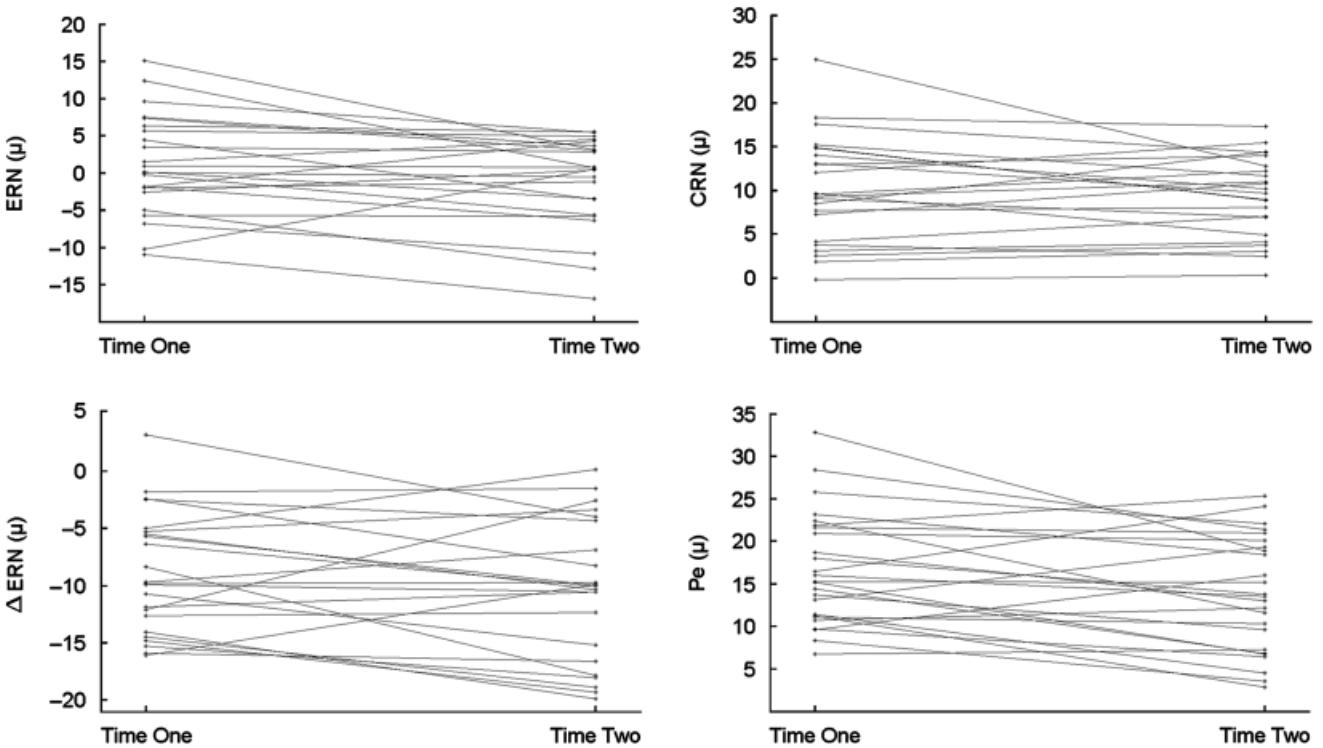

Figure 2. Individual subject scores at Time 1 and Time 2 for area measures of the ERN (top left), CRN (top right), ΔERN (bottom left), and Pe (bottom right).

measured 1–3 years apart, often in larger sample sizes (Sandman & Patterson, 2000; Segalowitz & Barnes, 1993; Walhovd & Fjell, 2002). Previous reliability estimates for the ERN have been derived from sessions separated by anywhere from 20–30 min (Fallgatter et al., 2001; Segalowitz et al., 2010) to 2–6 weeks (Olvet & Hajcak, 2009a; Segalowitz et al., 2010). The present study demonstrates similar levels of test–retest reliability as previous research on the ERN across a substantially longer period of time.

Peak latency measures, however, were not reliable from Time 1 to Time 2, consistent with an earlier report on the 2-week reliability of error-related brain activity (Olvet & Hajcak, 2009a). Latencies of ERPs are subject to a number of influences, including individual differences in response time and component processing time; variability in either area could contribute to the observed latency, which might decrease the 2-year reliability of these measures.

Although there is evidence that the magnitude of the ERN may be impacted by state-related manipulations (Bartholow et al., 2005; Chiu & Deldin, 2007; Endrass et al., 2010; Hajcak, Moser, Yeung, & Simons, 2005; Ridderinkhof et al., 2002), we believe it to be the case that the majority of variation in the ERN is trait linked (Anokhin et al., 2008; Hajcak et al., 2008; Olvet & Hajcak, 2009a; Riesel et al., 2011). However, there is also increasing evidence that the impact of state-related manipulations may interact with trait variables to determine the amplitude of the ERN (e.g., Endrass et al., 2010; Olvet & Hajcak, 2011). Future studies should investigate the stability of ERN response to such manipulations as well as how individual difference variables might interact with these manipulations over time.

Of late, proposals to integrate neurobiological markers into psychometric systems for diagnosing psychological disorders have become more influential (e.g., Patrick & Bernat, 2010). Evidence from the present study demonstrating long-term test–retest reliability of the ERN is useful in establishing the component as a viable neurobiological marker, insofar as the validity of any measure cannot exceed that measure's reliability (e.g., Segalowitz & Barnes, 1993). However, though the reliability estimates demonstrated here appear to be sufficiently robust for group studies (i.e., greater than .50; Helmstadter, 1964), they are likely not yet sufficient to establish these neural markers of error processing as clinically diagnostic *on an individual level* (e.g., cutoffs of .94 or above; Helmstadter, 1964; Segalowitz & Barnes, 1993). Nonetheless, many self-report and interview batteries that are frequently used in clinical diagnosis also fall well below this cutoff (e.g., Sheehan et al., 1997; Williams et al., 1992; Zanarini et al., 1987). Combined with other studies suggesting its substantial heritability (Anokhin et al., 2008; Riesel et al., 2011), the results of the present study further suggest that the ERN is a viable neural marker with which to assess group-level trait characteristics and individual differences in ACC functioning and executive control.

References

Amodio, D., Master, S., Yee, C., & Taylor, S. (2008). Neurocognitive components of the behavioral inhibition and activation systems: Implications for theories of self-regulation. *Psychophysiology*, *45*, 11–19.

Anokhin, A., Golosheykin, S., & Heath, A. (2008). Heritability of frontal brain function related to action monitoring. *Psychophysiology*, *45*, 524–534.

Bartholow, B., Pearson, M., Dickter, C., Sher, K., Fabiani, M., & Gratton, G. (2005). Strategic control and medial frontal negativity: Beyond errors and response conflict. *Psychophysiology*, *42*, 33–42.

Boksem, M. A., Tops, M., Wester, A. E., Meijman, T. F., & Lorist, M. M. (2006). Error-related ERP components and individual differences in punishment and reward sensitivity. *Brain Research*, *1101*, 92–101.

Brázdil, M., Roman, R., Daniel, P., & Rektor, I. (2005). Intracerebral error-related negativity in a simple go/nogo task. *Journal of Psychophysiology*, *19*, 244–255.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciencse*, *4*, 215–222.

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

Endrass, T., Klawohn, J., Schuster, F., & Kathmann, N. (2008). Overactive performance monitoring in obsessive-compulsive disorder: ERP evidence from correct and erroneous reactions. *Neuropsychologia*, *46*, 1877–1887.

Endrass, T., Reuter, B., & Kathmann, N. (2007). ERP correlates of conscious error recognition: Aware and unaware errors in an antisaccade task. *European Journal of Neuroscience*, *26*, 1714–1720.

Endrass, T., Schuermann, B., Kaufmann, C., Spielberg, R., Kniesche, R., & Kathmann, N. (2010). Performance monitoring and error significance in patients with obsessive-compulsive disorder. *Biological Psychology*, *84*, 257–263.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters on the identification of target letters in a non-search task. *Perception and Psychophysics*, *16*, 143–149.

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.

Fallgatter, A., Bartsch, A., Strik, W., Mueller, T., Eisenack, S., Neuhauser, B., et al. (2001). Test–retest reliability of electrophysiological parameters related to cognitive motor control. *Clinical Neurophysiology*, *112*, 198–204.

Ford, J. M. (1999). Schizophrenia: The broken P300 and beyond. *Psychophysiology*, *36*, 667–682.

Gehring, W. J., & Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. *Nature Neuroscience*, *3*, 516–520.

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

Hajcak, G., Franklin, M., Foa, E., & Simons, R. (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. (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.

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

Helmstadter, G. (1964). *Principles of psychological measurement*. New York: Appleton-Century-Crofts, Inc.

Holroyd, C., & Coles, M. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709.

Holroyd, C., Dien, J., & Coles, M. (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.

Hughes, G., & Yeung, N. (2010). Dissociable correlates of response conflict and error awareness in error-related brain activity. *Neuropsychologia*, *49*, 405–415.

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation. *Clinical Neurophysiology*, *115*, 1821–1835.

Miltner, W., Lemke, U., Weiss, T., Holroyd, C., Scheffers, M., & Coles, M. (2003). Implementation of error-processing in the human anterior

cingulate cortex: A source analysis of the magnetic equivalent of the error-related negativity. *Biological Psychology*, *64*, 157–166.

Moser, J. S., Hajcak, G., & Simons, R. F. (2005). The effects of fear on performance monitoring and attentional allocation. *Psychophysiology*, *42*, 261–268.

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

Olvet, D., & Hajcak, G. (2009a). Reliability of error-related brain activity. *Brain Research*, *1284*, 89–99.

Olvet, D. M., & Hajcak, G. (2009b). The stability of error-related brain activity with increasing trials. *Psychophysiology*, *46*, 957–961.

Olvet, D., & Hajcak, G. (2011). The error-related negativity (ERN) relates to sadness following mood induction among individuals with high neuroticism. *Social Cognitive and Affective Neuroscience*. Advance online publication. doi: 10.1093/scan/nsr007

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.

Patrick, C. J., & Bernat, E. M. (2010). Neuroscientific foundations of psychopatholoy. In T. Millon, R. F. Krueger, & E. Simonsen (Eds.), *Contemporary directions in psychopathology: Scientific foundations of the DSM-V and ICD-II* (pp. 519–452). New York: Guilford.

Pontifex, M., Scudder, M., Brown, M., O'Leary, K., Wu, C., Themanson, J., et al. (2010). On the number of trials necessary for stabilization of error related brain activity across the life span. *Psychophysiology*, *47*, 767–773.

Ridderinkhof, K., de Vlugt, Y., Bramlage, A., Spaan, M., Elton, M., Snel, J., et al. (2002). Alcohol consumption impairs detection of performance errors in mediofrontal cortex. *Science*, *298*, 2209–2211.

Ridderinkhof, K. R., Ramautar, J. R., & Wijnen, J. G. (2009). To P(E) or not to P(E): A P3-like ERP component reflecting the processing of response errors. *Psychophysiology*, *38*, 752–760.

Riesel, A., Endrass, T., Kaufmann, C., & Kathmann, N. (2011). Overactive error-related brain activity as an endophenotype for obsessive-compulsive disorder: Evidence from unaffected first-degree relatives. *Biological Psychiatry*, *168*, 317–324.

Sandman, C., & Patterson, J. (2000). The auditory event-related potential is a stable and reliable measure in elderly subjects over a 3 year period. *Clinical Neurophysiology*, *111*, 1427–1437.

Scheffers, M., & Coles, M. (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.

Segalowitz, S., & Barnes, K. (1993). The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology*, *30*, 451–459.

Segalowitz, S., Santesso, D., Murphy, T., Homan, D., Chantziantoniou, D., & Khan, S. (2010). Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology*, *47*, 260–270.

Sheehan, D., Lecrubier, Y., Harnett Sheehan, K., Janavs, J., Weiller, E., Keskiner, A., et al. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, *12*, 232–241.

Walhovd, K., & Fjell, A. (2002). One-year test–retest reliability of auditory ERPs in young and old adults. *International Journal of Psychophysiology*, *46*, 29–40.

Williams, J. B., Gibbon, M., First, M. B., Spitzer, R. L., Davies, M., Borus, J., et al. (1992). The structured clinical interview for DSM-III-R (SCID). II. Multisite test–retest reliability. *Archives of General Psychiatry*, *49*, 630–636.

Zanarini, M., Frankenburg, F., Chauncey, D., & Gunderson, J. (1987). The Diagnostic Interview for Personality Disorders: Interrater and test–retest reliability. *Comprehensive Psychiatry*, *28*, 467–480.

(Received November 7, 2010; Accepted February 25, 2011)