BRAIN RESEARCH 1284 (2009) 89–99

available at www.sciencedirect.com

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Reliability of error-related brain activity

*Doreen M. Olvet, Greg Hajcak**

Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500, USA

ARTICLE INFO

Article history:
Accepted 23 May 2009
Available online 6 June 2009

Keywords:
Error-related negativity
Error positivity
Correct response negativity
Reliability
ERN
CRN
Pe

ABSTRACT

Recent studies that have examined neural correlates of action monitoring with event-related potentials (ERPs) have focused on the error-related negativity (ERN) and error positivity (Pe) on error trials, as well as the correct response negativity (CRN) on correct trials. Moreover, the ERN has been assessed in relation to a number of personality traits and psychiatric disorders. However, no study to date has assessed the reliability of the ERN, Pe, and CRN. We measured these ERPs in 45 undergraduates at baseline and 2 weeks later. For split-half and test–retest reliabilities, both the intersubject stability and score agreement were high for the ERN, CRN, and Pe. These data demonstrate excellent reliability of ERPs elicited during response monitoring, and further suggest that these ERPs are well-suited to assess trait characteristics and individual differences.

1. Introduction

The ability to detect errors is crucial for goal-directed behavior and learning (cf. Holroyd and Coles, 2002). In laboratory tasks, studies suggest that errors are followed by compensatory behavioral adjustments (Hajcak and Simons, 2008; Rabbitt and Rodgers, 1977) and signal a need to increase cognitive control (Holroyd and Coles, 2002; Kerns et al., 2005; Yeung et al., 2004). From a motivational perspective, errors can place organisms in danger and threaten safety; consistent with this notion, recent work indicates that errors increase defensive reflexes (Hajcak and Foti, 2008). Moreover, certain forms of psychopathology appear characterized by excessive concern over mistakes (for review see Shafran and Mansell, 2001). Thus, errors have become a phenomenon of interest within cognitive, social, and clinical psychology.

Interest in errors has been fueled by neuroscientific investigations on the relatively rapid neural correlates of error processing. In particular, studies that utilize event-related potentials (ERPs) have reported on the error-related negativity (ERN)—a negative deflection at frontal-central midline recording sites that occurs approximately 50 ms following an erroneous response (Falkenstein, Hohnsbein et al., 1991; Gehring et al., 1993). Functionally, the ERN appears to reflect early error processing in the medial prefrontal cortex (cf. Ridderinkhof et al., 2004) and source localization studies suggest that the ERN is generated in the anterior cingulate cortex (ACC; Brazdil et al., 2005; Brazdil et al., 2002; Dehaene et al., 1994; Holroyd et al., 1998; van Veen and Carter, 2002).

The reinforcement learning theory of the ERN (RL-ERN; Holroyd and Coles, 2002) suggests that the ERN is a learning signal that results from a phasic decrease in midbrain dopamine activity, which in turn disinhibits neurons in the ACC; this signal is thought to reflect the evaluation of on-going events as worse than expected. Another prominent theory is the conflict monitoring theory, which purports that the ERN reflects conflict between error and error-correcting responses (Yeung et al., 2004). More recently, it has been suggested that a theory which incorporates aspects of both the RL-ERN and conflict monitoring theory would be optimal in explaining the

---

* *Corresponding author*. Fax: +1 631 632 7876.
E-mail address: greg.hajcak@stonybrook.edu (G. Hajcak).

0006-8993/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2009.05.079

functional significance of the ERN (Botvinick, 2007; Holroyd et al., 2005).

Almost immediately following the ERN, the ERP on error trials is characterized by a positive deflection referred to as the error positivity (Pe). The Pe is maximal 200–400 ms after the commission of an error (Falkenstein et al., 2003; Nieuwenhuis et al., 2001; Overbeek et al., 2005) and has a more posterior midline scalp distribution relative to the ERN (Falkenstein et al., 2000). It has been suggested that the Pe reflects error awareness (Leuthold and Sommer, 1999; Nieuwenhuis et al., 2001); it stands to reason that the Pe is a P300-like orienting response following error commission (Davies et al., 2001; Hajcak et al., 2003b; Ridderinkhof et al., 2009). Relative to the ERN, the Pe has received much less attention in ERP studies of error processing (cf. Overbeek et al., 2005).

A small ERN-like component can also occur on correct trials; it is referred to as the correct response negativity (CRN; Falkenstein et al., 2000; Ford, 1999; Gehring and Knight, 2000; Scheffers and Coles, 2000; Vidal, Hasbroucq et al., 2000). There has been debate regarding the meaning of the CRN: it has been suggested that the CRN reflects a response comparison process (Falkenstein et al., 2000; Vidal et al., 2000), an emotional reaction (Luu et al., 2000a), uncertainty of a correct response (Coles et al., 2001; Pailing et al., 2002), or the co-activation of correct and incorrect responses (Luu et al., 2000a; Scheffers et al., 1996; Vidal et al., 2000). Bartholow et al. present compelling data that the CRN may reflect the evaluation of adaptive versus maladaptive response strategies on correct trials (Bartholow et al., 2005). Consistent with the role of the CRN in cognitive control, studies have also suggested that error-preceding trials are characterized by a reduced CRN (Allain et al., 2004; Hajcak et al., 2005b).

In addition to interest regarding the functional characterization of the ERN, Pe, and CRN, a number of studies have also explored these components with respect to individual difference variables. In particular, the ERN has been studied extensively in a variety of psychiatric disorders, including obsessive–compulsive disorder (Gehring et al., 2000; Hajcak et al., 2008) and major depressive disorder (Chiu and Deldin, 2007; Holmes and Pizzagalli, 2008). Other studies have linked the ERN to related personality traits. For example, individuals who score high in negative affect (Hajcak et al., 2004b; Luu et al., 2000b), neuroticism (Pailing and Segalowitz, 2004), worry (Hajcak et al., 2003a), obsessive–compulsive symptoms (Hajcak and Simons, 2002) and punishment sensitivity (Amodio et al., 2008; Boksem et al., 2006; Frank et al., 2005) have all been associated with an increased ERN. On the other hand, studies show that individuals who score high on measures of impulsivity (Potts et al., 2006; Ruchsow et al., 2005), externalizing (Hall et al., 2007) or low on socializing (Dikman and Allen, 2000) have decreased ERN amplitudes in response to errors. Some studies have found that these individual difference variables relate to both the ERN and CRN (Hajcak and Simons, 2002)—as well as variation in the Pe (Boksem et al., 2006; Hajcaket al., 2004b; Ruchsow et al., 2005).

Although the ERN has been consistently related to stable trait-like characteristics, no study to date has examined whether the ERN *itself* is reliable over time. Moreover, the degree to which the ERN, Pe, and CRN reflect valid metrics of information-processing will depend on the reliability of these measures (Helmstader, 1964). While there has been extensive research indicating high test–retest reliability of the P300 component of the ERP (Fabiani et al., 1987; Fallgatter, et al., 2001; Sandman and Patterson, 2000; Segalowitz and Barnes, 1993; Walhovd and Fjell, 2002; Williams et al., 2005), no such study on the reliability of the ERN, Pe, and CRN have been published to date. Therefore, we set forth to examine the reliability of the ERN and Pe, as well as the CRN, and behavioral measures both within and across two testing sessions. To this end, ERP data was analyzed from 45 participants who performed the arrow version of the flanker task in two sessions separated by exactly 2 weeks.

2. Results

2.1. *Comparing Session 1 to Session 2*

Behavioral data from Sessions 1 and 2 are presented in Table 1. There was no significant difference between the accuracy at Sessions 1 and 2 ($t(42) = -1.22$, $p > 0.05$). Participants responded faster on error than correct trials ($F(1,42) = 307.67$, $p < 0.001$) and were also slower in Session 1 than 2 ($F(1,42) = 4.29$, $p < 0.05$). In terms of speed-accuracy trade-off, these data collectively suggest that participants' performance improved from Session 1 to 2: they were able to respond faster without making more mistakes in Session 2.

Fig. 1 displays grand averages for correct- and error-trial ERPs from Sessions 1 and 2; mean (and standard deviation) of ERP area measures are shown in Table 1. Fig. 2 displays individual averages for correct- and error-trial ERPs from 6 randomly selected participants. The average number of epochs included in the ERP averages for correct trials was 213.20 (SD = 16.22) at Session 1 and 211.12 (SD = 11.44) at Session 2 and for error trials was 26.49 (SD = 8.56) at Session 1 and 28.27 (SD = 11.03) at Session 2. In terms of area measures, the ERN was more negative than the CRN ($F(1,44) = 167.96$, $p < 0.001$). There was also a significant interaction between Trial Type and Session ($F(1,44) = 13.03$, $p < 0.001$), but no effect of Session alone ($F(1,44) = 0.75$, $p > 0.05$). Post-hoc paired samples' *t*-tests confirmed that the CRN was significantly larger at Session 2 than 1 ($t(44) = -3.26$, $p < 0.01$), whereas the ERN was comparable across testing sessions ($t(44) = 1.08$, $p > 0.05$).[1] The ERN–CRN area was also larger at Session 2 compared to Session 1 ($t(44) =\ 3.61$, $p < 0.001$). On the other hand, there was no difference in Pe area between the two sessions ($t(44) = -0.74$, $p > 0.05$).

When examining peak measures, the ERN was more negative than the CRN ($F(1,44) = 210.95$, $p < 0.001$). This was

---

[1] When we analyzed the ERP data using filtering in the theta range (4–7Hz), we found no significant difference in the CRN between sessions for the area ($t = -1.68$, $p > 0.05$) and peak measures ($t = -0.78$, $p > 0.05$). This differs from the original analysis in that the CRN area was larger at Session 2 compared to Session 1 ($t = -3.26$, $p < 0.01$). These data suggest that variance in the CRN may in fact have been due to the P3. The results for the ERN, though, are consistent with our original peak data. There was a significant increase in the ERN at Session 2 using peak measures ($t = 2.26$, $p < 0.05$) and, in addition, this is confirmed using the area measure ($t = 2.33$, $p < 0.05$).

Table 1 – Mean (and standard deviation), as well as split-half reliability metrics, of behavioral and ERP measures at Session 1 (left) and Session 2 (right).

Measure	Session 1			Session 2		
	Mean (SD)	r	ICC	Mean (SD)	r	ICC
Reaction time: correct trials (ms)	401[a] (43)	0.99[**]	0.98[**]	387[a] (41)	0.99[**]	0.98[**]
Reaction time: error trials (ms)	335[a] (40)	0.73[**]	0.57[**]	331[a] (42)	0.82[**]	0.69[**]
Accuracy (%)	88.66 (3.89)	n/a	n/a	87.58 (5.94)	n/a	n/a
CRN area measure (μV)	7.50[b] (5.84)	0.98[**]	0.96[**]	9.27[b] (6.11)	0.98[**]	0.95[**]
ERN area measure (μV)	0.14 (6.73)	0.88[**]	0.75[**]	−0.67 (6.33)	0.84[**]	0.72[**]
Difference area measure (μV)	−7.36 (5.40)	0.76[**]	0.57[**]	−9.94 (4.74)	0.66[**]	0.48[**]
Pe area measure (μV)	14.35 (7.79)	0.91[**]	0.84[**]	14.95 (7.38)	0.83[**]	0.71[**]
CRN peak measure (μV)	6.78 (5.81)	0.97[**]	0.93[**]	8.72 (6.30)	0.98[**]	0.95[**]
ERN peak measure (μV)	−2.96[b] (6.80)	0.88[**]	0.76[**]	−4.59[b] (6.46)	0.83[**]	0.71[**]
Difference peak measure (μV)	−10.45 (5.98)	0.84[**]	0.65[**]	−13.39 (6.10)	0.76[**]	0.60[**]
CRN peak latency (μV)	32.29 (23.52)	0.88[**]	0.78[**]	35.72 (25.64)	0.83[**]	0.67[**]
ERN peak latency (μV)	42.19 (23.12)	0.65[**]	0.48[**]	36.72 (25.20)	0.47[*]	0.30[*]
Difference peak latency (μV)	42.80 (27.74)	0.77[**]	0.62[**]	46.66 (20.57)	0.65[**]	0.49[**]

[a] Participants were significantly faster at Session 2 compared to Session 1 on both correct and error trials ($p < 0.001$).
[b] Both CRN area and ERN peak measures were significantly larger at Session 2 compared to Session 1 ($p < 0.01$).
[*] $p < 0.05$.
[**] $p < 0.001$.

qualified by a significant interaction between Trial Type and Session ($F(1,44) = 143.95$, $p < 0.001$), but no effect of Session alone ($F(1,44) = 0.06$, $p > 0.05$). Post-hoc t-tests showed that the ERN was larger at Session 2 than 1 ($t(44) = -2.97$, $p < 0.01$), whereas there was no difference in the CRN ($t(44) = 1.83$, $p > 0.05$). The ERN–CRN peak was also larger at Session 2 compared to Session 1 ($t(44) = 3.61$, $p < 0.001$).

The peak latency of the ERN and CRN were similar ($F(1,44) = 2.42$, $p > 0.05$), and did not differ between the two testing sessions ($F(1,44) = 0.09$, $p > 0.05$); similarly, the ERN–CRN peak latency did not vary between Sessions 1 and 2 ($t(44) = -0.86$, $p > 0.05$). Collectively, these results confirm the impression from Fig. 1 and Table 1 that the ERN, as well as the difference between the ERN and CRN, was larger at the second testing session.

2.2. *Test–retest reliability*

Table 2 contains test–retest reliability indices for behavioral and ERP measures, both in terms of intersubject stability (r) and score agreement (ICC) for all subjects (top); when comparing male and female participants separately (middle), and when comparing participants who made relatively few versus many errors (bottom).

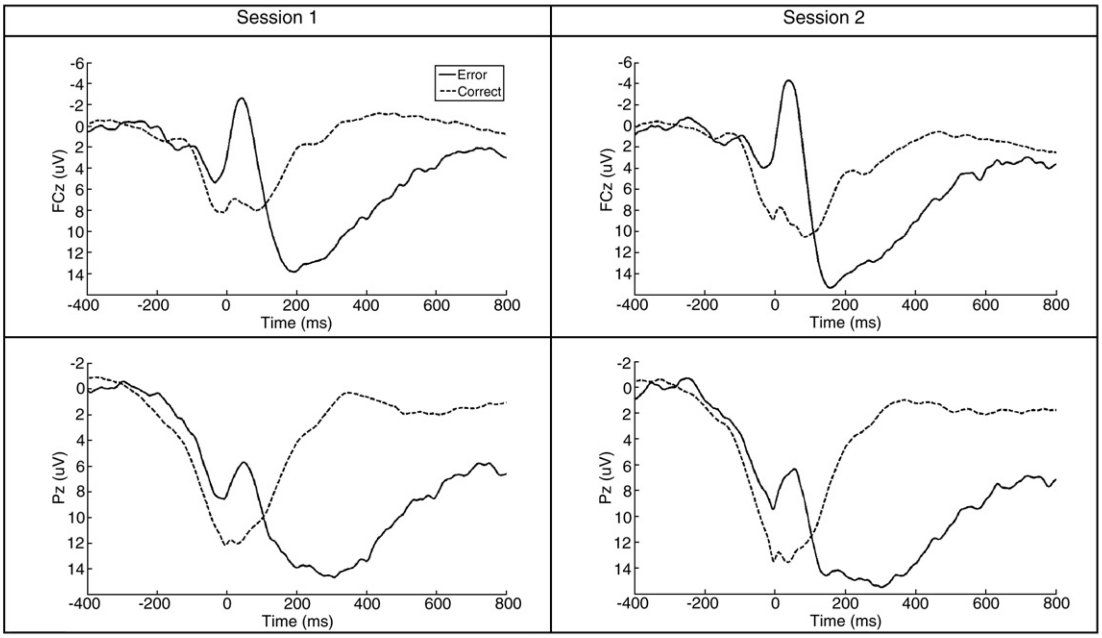

Fig. 1 – The response-locked ERPs for error and correct trials at FCz, where the ERN was maximal for Session 1 (top left) and Session 2 (top right) and at Pz where the Pe was maximal for Session 1 (bottom left) and Session 2 (bottom right). Response onset occurred at 0 ms and negative is plotted up.

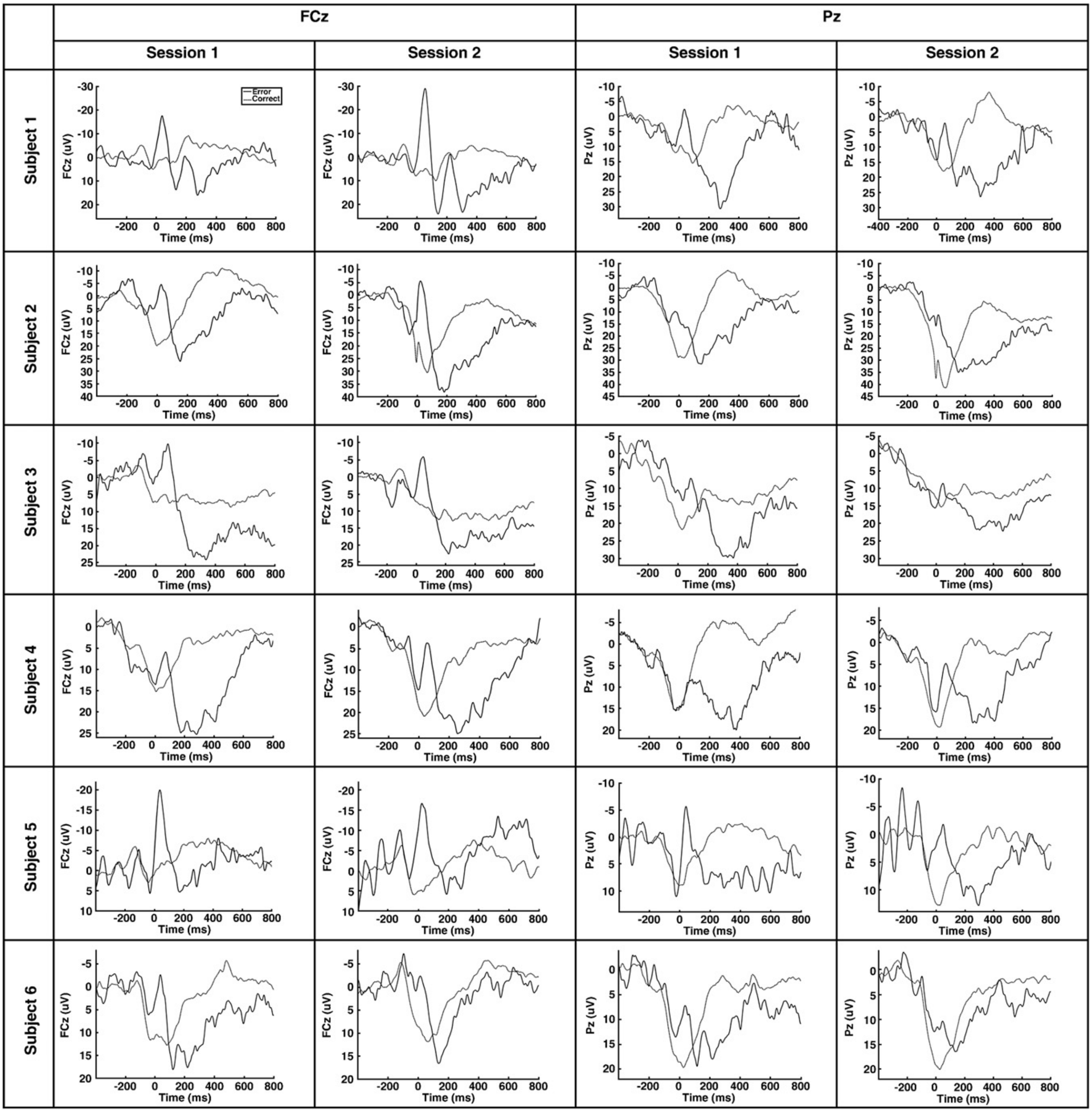

Fig. 2 – Response-locked ERPs for error and correct trials at FCz (left columns) and Pz (right columns) for Session 1 and Session 2 in 6 randomly selected subjects. Response onset occurred at 0 ms and negative is plotted up; note that the ordinate differs across subjects.

2.2.1. *Behavioral measures*

There was a significant correlation between the reaction time at Session 1 and Session 2 on correct ($r=0.50$, $p<0.001$, ICC=0.51) and error trials ($r=0.80$, $p<0.001$, ICC=0.75). There was also a significant correlation between the number of errors committed at Sessions 1 and 2 ($r=0.38$, $p<0.01$, ICC=0.36).

2.2.2. *Area measures*

Scatterplots depicting ERP area measures at both Sessions 1 and 2 are presented in Fig. 3. For all individual ERP area measures, reliability estimates were high (CRN: $r=0.82$, $p<0.001$, ICC=0.78; ERN: $r=0.70$, $p<0.001$, ICC=0.70; Pe: $r=0.75$, $p<0.001$, ICC=0.75); in fact, even the ERN–CRN difference score demonstrated moderate reliability ($r=0.56$, $p<0.001$, ICC=0.47).

2.2.3. *Peak measures*

Fig. 4 presents scatterplots for ERP peak measures at Sessions 1 and 2. Consistent with reliability estimates obtained on area measures, all peak measures were highly reliable between Sessions 1 and 2 (CRN: $r=0.59$, $p<0.001$, ICI=0.58; ERN: $r=0.74$,

Table 2 – Test-retest reliability metrics for behavioral and ERP measures for all participants (top), when evaluated for male and female participants separately (middle), and for participants who made a low versus high number of errors (bottom).

	Behavior				Area measures								Peak measures											
	# of errors		Reaction time (correct/error)		CRN area		ERN area		ERN–CRN area		Pe area		CRN peak		ERN peak		Difference peak		CRN peak latency		ERN peak latency		ERN–CRN peak latency	
	r	ICC	*r*	ICC	*r*	ICC	*r*	ICC	*r*	ICC	*r*	ICC	*r*	ICC	*r*	ICC	*r*	ICC	*r*	ICC	*r*	ICC	*r*	ICC
Overall (*n* =45)	0.38**	0.36**	0.50*** 0.80***	0.51*** 0.75***	0.82***	0.78***	0.70***	0.70***	0.56***	0.47***	0.75***	0.75***	0.59***	0.58***	0.74***	0.70***	0.59***	0.51***	−0.02	−0.02	0.43**	0.42**	0.25	0.24
Gender																								
Male (*n* =23)	0.69***	0.69***	0.39 0.80***	0.40* 0.75***	0.85***	0.84***	0.77***	0.74***	0.65***	0.58***	0.82***	0.81***	0.59**	0.55**	0.89***	0.89***	0.70***	0.65***	−0.08	−0.06	0.38	0.37*	0.15	0.15
Female (*n* =22)	0.15	0.13	0.57** 0.74***	0.58** 0.68***	0.80***	0.71***	0.68***	0.68***	0.43*	0.32	0.73***	0.72***	0.64***	0.62***	0.64***	0.54**	0.47*	0.35*	0.07	0.06	0.56**	0.52**	0.36	0.36*
Number of errors																								
Low (*n* =22)	0.11	0.07	0.32 0.79***	0.32 0.79***	0.71***	0.71***	0.64***	0.64***	0.61***	0.61***	0.79***	0.78***	0.67***	0.66***	0.63**	0.63**	0.64***	0.64***	−0.05	−0.03	0.34	0.34	−0.10	−0.10
High (*n* =21)	0.17	0.13	0.62** 0.80***	0.61** 0.78***	0.89***	0.89***	0.72***	0.72***	0.40	0.40*	0.74***	0.70***	0.46*	0.46*	0.82***	0.82***	0.46*	0.46*	0.03	0.03	0.46*	0.46*	0.53*	0.46*

* Significant correlations at the $p < 0.05$ level.

** Significant correlations at the $p < 0.01$ level.

*** Significant correlations at the $p < 0.001$ level.

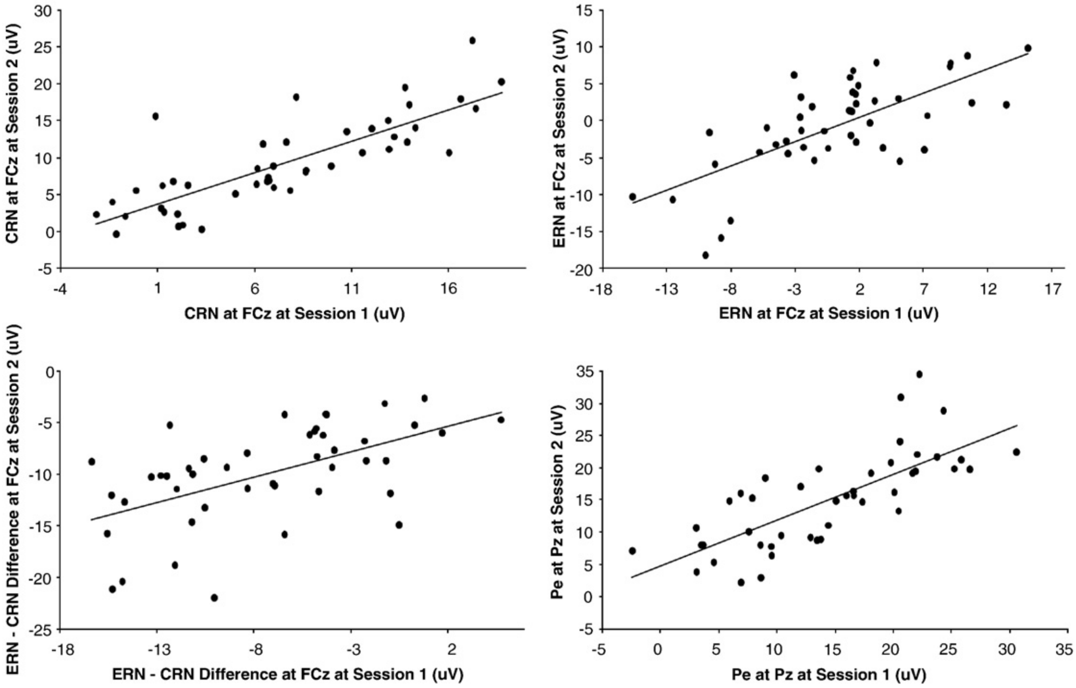

Fig. 3 – Scatterplots depicting the Pearson correlation between Session 1 and Session 2 for area measures of the CRN at FCz (top left), the ERN at FCz (top right), the ERN–CRN at FCz (bottom left) and Pe at Pz (bottom right).

$p<0.001$, ICI = 0.70); again, the ERN–CRN difference peak also demonstrated moderate reliability ($r=0.59$, $p<0.001$, ICI = 0.51). On the other hand, peak latency measures were characterized by low to moderate test–retest reliability (ERN: $r=0.43$, $p<0.01$, ICI = 0.42; CRN: $r=-0.02$, $p>0.05$, ICI = −0.02; ERN − CRN: $r=0.25$, $p>0.05$, ICI = 0.24).

2.2.4. *Analyses by gender and number of error trials*

In general, test–retest reliability metrics were similar for males and females—although several were higher for males than females. Additionally, we divided participants into two groups based on the number of errors they made at Session 1. The median number of error trials at Session 1 was 29, therefore those who made fewer than 29 errors were categorized as making a low number of errors ($n=22$) and those who made more than 29 errors were categorized as making a high number of errors ($n=21$). At Session 2, 13 subjects who made a high number of errors at Session 1 also made a high number of errors at Session 2 (median number of errors at Session 2 = 28). As in the male/female comparison, test–retest reliability estimates were similar for low and high error-committing groups, except for the ERN–CRN latency which was reliable in subjects who made a high number of errors ($r=0.53$, $p<0.05$, ICC = 0.46), but not in subjects who made a low number of errors ($r=-0.10$, $p>0.05$, ICC = −0.10).

2.3. *Split-half reliability*

Table 1 presents split-half reliability measures (intersubject stability (r) and score agreement (ICC)) at both Sessions 1 (left) and 2 (right).

2.3.1. *Behavioral measures*

Split-half reliability measures were high for reaction time measures within Session 1 (correct: $r=0.99$, $p<0.001$, ICC = 0.98; error: $r=0.73$, $p<0.001$, ICC = 0.57) and Session 2 (correct: $r=0.99$, $p<0.001$, ICC = 0.98; error: $r=0.82$, $p<0.001$, ICC = 0.69).

2.3.2. *Area measures*

For all area measures, split-half reliability was high at Session 1 (CRN: $r=0.98$, $p<0.001$, ICC = 0.96; ERN: $r=0.88$, $p<0.001$, ICC = 0.75; ERN − CRN: $r=0.76$, $p<0.001$, ICC = 0.57; Pe: $r=0.91$, $p<0.001$, ICC = 0.84) and Session 2 (CRN: $r=0.98$, $p<0.001$, ICC = 0.95; ERN: $r=0.84$, $p<0.001$, ICC = 0.72; ERN − CRN: $r=0.66$, $p<0.001$, ICC = 0.48; Pe: $r=0.83$, $p<0.001$, ICC = 0.71).

2.3.3. *Peak measures*

Consistent with the area measures, split-half reliability measures were high at Session 1 (CRN: $r=0.97$, $p<0.001$, ICC = 0.93; ERN: $r=0.88$, $p<0.001$, ICC = 0.76; ERN − CRN: $r=0.84$, $p<0.001$, ICC = 0.65) and Session 2 (CRN: $r=0.98$, $p<0.001$, ICC = 0.95; ERN: $r=0.83$, $p<0.001$, ICC = 0.71; ERN − CRN: $r=0.76$, $p<0.001$, ICC = 0.60). Peak latency measures were also reliable at Session 1 (CRN: $r=0.88$, $p<0.001$, ICC = 0.78; ERN: $r=0.65$, $p<0.001$, ICC = 0.48; ERN − CRN: $r=0.77$, $p<0.001$, ICC = 0.62) and Session 2 (CRN: $r=0.83$, $p<0.001$, ICC = 0.67; ERN: $r=0.47$, $p<0.05$, ICC = 0.30; ERN − CRN: $r=0.65$, $p<0.001$, ICC = 0.49).

3. Discussion

The present study analyzed ERPs from 45 individuals who performed the arrow version of the flanker task at two time-points to examine the split-half and two week test–retest

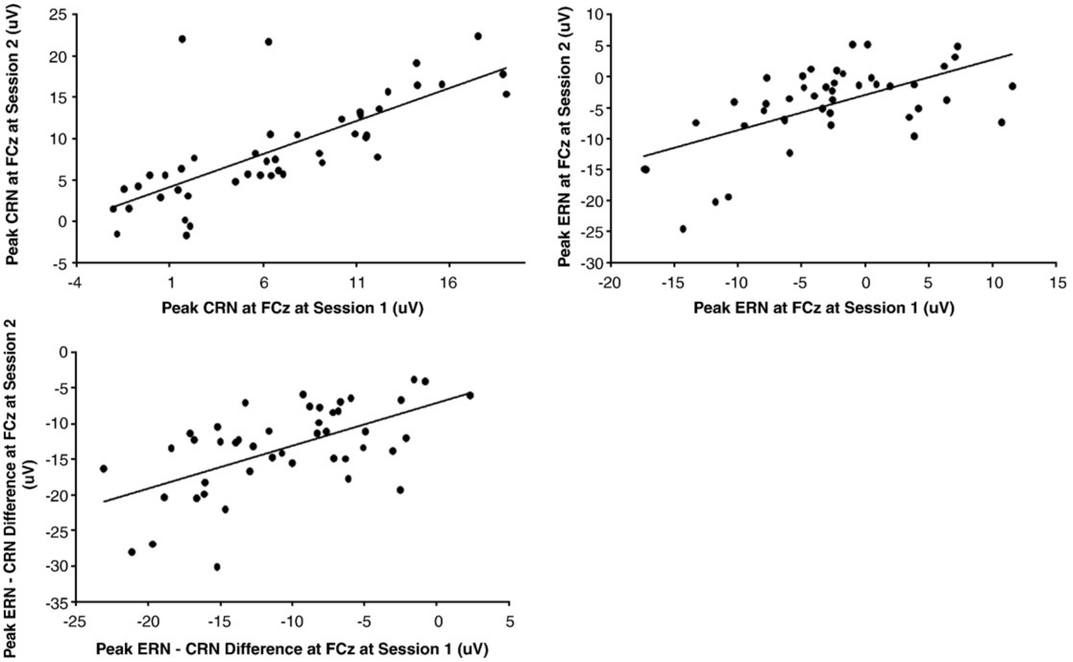

Fig. 4 – Scatterplots depicting the Pearson correlation between Session 1 and Session 2 for peak measures of the CRN at FCz (top left), the ERN at FCz (top right), and the ERN–CRN at FCz (bottom left).

reliability of the ERN, Pe and CRN. Although the focus of the current study was on error-related brain activity, behavioral measures were also evaluated in terms of reliability. Overall, ERP measures were characterized by significant stability and score agreement over 2 weeks—and comparable reliability estimates were obtained even when using different baselines, filter settings, and at multiple midline recording sites. Collectively, these data suggest relatively robust stability in measures of the ERN, CRN, and Pe.

Although the test–retest reliabilities of the ERN, CRN, and Pe were generally on par with their respective split-half reliability measures, split-half reliabilities were generally higher than the test–retest reliabilities. This is likely due to many state factors that are consistent within the split-half analyses; for example, the electrodes have not been removed and replaced, which occurs between sessions. Additionally, arousal level of the subject is likely quite uniform when comparing odd to even trials within a session, whereas the subjects' arousal level might vary considerably between testing sessions. Along these lines, Segalowitz and Barnes (1993) suggested that high split-half reliability indicates that an ERP component might be a stable index of state-related variables, and that high test–retest reliability would suggest more of a trait-like characteristic of ERP components. In this context, both split-half and test–retest reliabilities of these ERP components were high, and many of the split-half metrics were on par with the two week test–retest reliability, together suggesting strong trait-like qualities (cf. Segalowitz and Barnes, 1993).

Area measures are typically more conservative than peak measures; in particular, peak measures may be especially sensitive to noise (Luck, 2005). In the error-related brain activity literature, however, the Pe is most often scored using an area measure (i.e., Hajcak, et al., 2004b; Nieuwenhuis et al., 2001), whereas studies score the ERN either in terms of area (i.e., Hajcak and Foti, 2008; Hajcak et al., 2005a) or peak measurements (i.e., Compton et al., 2008; Holmes and Pizzagalli, 2008). In fact, the current study found quite similar ERP reliabilities using both the area and peak measures. Two exceptions were higher split-half reliability of the ERN–CRN difference peak measurement, and somewhat higher test–retest reliability of the of the CRN area measure. Thus, both the intersubject stability and score agreement over a two week period of the ERN and CRN were quite good—regardless of whether peak or area measures were used. In comparing the error-related components, the ERN and Pe area measures were characterized by similarly high reliability estimates both within- and between-sessions.

Consistent findings emerged when examining male and female subjects separately, and when comparing individuals who made a low versus high number of errors separately. These analyses further suggest robust reliability of the ERN, Pe, and CRN, and are consistent with a recent report from our lab that demonstrated that a stable ERN and Pe could be measured with just 6 error trials (Olvet and Hajcak, in press). The present analyses that examined participants who made high versus low number of errors further suggest that reliability estimates of the ERN, CRN, and Pe are not much affected by the number of error trials—that is, even though the high group made an average of 36 error trials, whereas the low group made an average of 20 error trials, test–retest reliability estimates were quite similar.

It is worth comparing the reliability estimates obtained for the ERN, CRN, and Pe in the current study to previous work on other ERP components. A number of previous studies have examined split-half and test–retest reliabilities of the P300, and reliability estimates in the current study were comparable or greater than reliability estimates of the P300 (Fabiani et al., 1987; Fallgatter et al., 2001; Sandman and Patterson, 2000;

Segalowitz and Barnes, 1993; Walhovd and Fjell, 2002; Williams et al., 2005). Moreover, the current reliability estimates are even comparable to measures on self-report measures of anxiety (Foa et al., 2002; Hajcak et al., 2004a).

In terms of the practical implications of these results, reliability places an upper limit on the validity of a measure (cf. Sechrest, 1984; Segalowitz and Barnes, 1993). Thus, the present data bode well for studies that examine the ERN, Pe, and CRN as stable metrics of information-processing (i.e. Hajcak et al., 2004b; Pailing and Segalowitz, 2004). Recently, we suggested that the ERN might serve as a useful endophenotype for internalizing disorders (for review see Olvet and Hajcak, 2008). This would be impossible if error-related components were not reliable. Therefore, the reliability of the ERN is especially important in light of a number of recent studies that have examined the ERN in clinical populations (Gehring et al., 2000; Holmes and Pizzagalli, 2008), before and after treatment (Hajcak et al., 2008; Ladouceur et al., 2007), and in personality traits that are stable over time (Hajcak et al., 2004b; Pailing and Segalowitz, 2004). The current data are certainly consistent with the possibility that the ERN reflects a trait-like measure of neural activity following errors.

Unlike magnitude metrics, however, peak latency reliabilities were fairly low—especially for the CRN and ERN–CRN (difference) peaks. It is important to note that the peak latencies of these ERPs are influenced not only by variability in the component processing time, but also variability in the time it takes to initiate a response. It is difficult to determine which factor is contributing to the measured latency—and this confound may play a role in the low test–retest reliability of latency measures.

It is also worth noting overall differences between the two testing sessions. Behaviorally, participants were faster and equally as accurate at the second testing session; these data suggest that participants' performance improved from the first to second testing session. Moreover, the CRN area and the ERN peak measures were larger at the second than first testing session. Insofar as the CRN difference was eliminated when the data was filtered in the theta range (i.e., 4–7 Hz), CRN differences across sessions may have been driven by larger P300 activity in the response-locked average (compare the post-ERN/CRN positivities in Fig. 1). However, the between-session difference in ERN remained even when the data was filtered in the theta range. One possibility is that the larger ERN at the second testing session relates to behavioral changes across testing sessions: improved performance ought to relate to larger ERNs based on contemporary computational models of the ERN (Holroyd and Coles, 2002; Yeung et al., 2004). Future studies might further examine other between-session changes in response to errors, including both self-reported effort and other peripheral psychophysiological measures (cf. Hajcak et al., 2003b).

The current reliability estimates were based on data from 45 participants performing an arrow version of the flanker task at two time-points. Although this number of subjects is consistent with some studies that have assessed the reliability of other ERP components (Sandman and Patterson, 2000, $N = 46$), most studies include only about 20 subjects (Fallgatter et al., 2001, $N = 23$; Segalowitz and Barnes, 1993, $N = 19$; Williams et al., 2005, $N = 21$). Thus, the current data is relatively large. However, it might be important for future studies to examine reliability of the ERN, Pe, and CRN in other task contexts. Moreover, future studies might investigate the reliability of these error-related components over longer periods of time and in individuals at different ages.

4. Experimental procedures

Fifty-seven undergraduates (27 male, 30 female) participated in the current study. Seven subjects were not included in the data analysis due to excessive artifacts ($n = 1$) or failing to return for the second testing session ($n = 6$). Participants who made fewer than 6 errors were excluded (Olvet and Hajcak, in press); the final sample consisted of 45 participants (22 female). For the Pe analysis, one additional subject did not have useable data from the electrode of interest (Pz). For the behavioral data, two subjects were not included due to a computer error. Data from 26 subjects at Session 1 were published in another manuscript on within-session properties of the ERN (Olvet and Hajcak, in press). No participant discontinued their participation in the experiment once procedures had begun, and all participants received course credit for their participation.

4.1. *Task*

The task was an arrow version of the flanker task (cf. Eriksen and Eriksen, 1974; Moser 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 followed by a blank screen for 1800 ms with an ITI that varied randomly from 500 to 1000 ms.

4.2. *Procedure*

The present task was administered on a Pentium D class computer, using Presentation software (Neurobehavioral Systems, Inc., Albany, California, USA) to control the presentation and timing of all stimuli. Following a brief description of the experiment, EEG sensors were attached and the participant was given detailed task instructions. Participants performed a practice block containing 30 trials and were told to try to be as accurate and fast as possible. The actual experiment consisted of 8 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. Participants returned 2 weeks later and repeated the procedure.

4.3. *Psychophysiological recording, data reduction and analysis*

The continuous EEG activity was recorded using an elastic head cap and the ActiveTwo BioSemi system (BioSemi, Amsterdam, Netherlands). Recordings were taken from 64 scalp electrodes based on the 10/20 system, as well as two electrodes placed on the left and right mastoids. The electrooculogram (EOG) generated from blinks and eye movements were recorded from four facial electrodes: two approximately 1 cm above and below the participant's right eye, one approximately 1 cm to the left of the left eye, and one approximately 1 cm to the right of the right eye. As per BioSemi's design, the ground electrode during acquisition was formed by the Common Mode Sense active electrode and the Driven Right Leg passive electrode. All bioelectric signals were digitized on a laboratory microcomputer using ActiView software (BioSemi, Amsterdam, Netherlands).

Off-line analysis was performed using Brain Vision Analyzer software (Brain Products, Gilohing, Germany). EEG data were re-referenced to the numeric mean of the mastoids and band-pass filtered with cutoffs of 0.1 and 30 Hz.[2] The EEG was segmented for each trial, beginning 400 ms before each picture onset and continuing for 1000 ms. The EEG was corrected for blinks and eye movements using the method developed by Gratton et al. (1983). Specific intervals for individual channels were rejected in each trial using a semi-automated 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 300.0 μV within a trial, and a maximum voltage difference of less than 0.50 μV within 100 ms intervals.

Response-locked average ERPs were computed for correct and error trials. Because the ERN and CRN are often evaluated using different measures, these ERP components were analyzed both in terms of area and peak measures; because the Pe is a broader component, it was evaluated only using an area measure. In terms of the area measures, the ERN and CRN were quantified on error and correct trials, respectively, as the average activity in a 0–100 ms window relative to response onset at electrode site FCz[3]; in addition, the difference (ERN−CRN) between the ERN and CRN was also calculated in the same window. The Pe was evaluated on error trials as the average activity from 200 to 400 ms following response onset at electrode site Pz. For peak measures, the ERN and CRN were evaluated on error and correct trials, respectively, as the largest negative peak in the 0–100 ms window after response. The ERN−CRN peak was also computed on the difference wave between error and correct trials in the same window. A 200 ms window prior to the response (−400 to −200 ms) served as the baseline.[4]

In all cases, reaction time, number of errors, ERN, CRN, and Pe were statistically evaluated using SPSS (Version 14.0) General Linear Model software; Greenhouse–Geisser correction was applied to *p* values associated with multiple-*df*, repeated measures comparisons, when appropriate. A 2 (Trial Type)×2 (Session) ANOVA was used to detect differences between the two sessions and paired *t*-tests were performed for follow-up post-hoc testing. Both the Pearson (*r*) and intraclass correlation (ICC) coefficients were used to examine intersubject stability and score agreement, respectively, of the relationship between behavioral measures and ERP components at the initial testing (Session 1) and 2 weeks later (Session 2). Intraclass correlation coefficients reflect the consistency of a measure taking into account variance related to the time of testing (Shrout and Fleiss, 1979), whereas the Pearson's correlation coefficient reflects only intersubject stability in how subjects are ranked. If an ERP reflects a trait-like characteristic, then the actual agreement between measure values should be high, which would be indicated by high ICC estimates of reliability (cf. Segalowitz and Barnes, 1993). Split-half reliability was also reported in terms of both Pearson and intraclass correlation coefficients, and was performed within each session by comparing even and odd trial values. Because split-half reliability metrics were based on half of the trials, these measures were corrected using the Spearman–Brown prophecy formula (Helmstader, 1964).

REFERENCES

Allain, S., Carbonnell, L., Falkenstein, M., Burle, B., Vidal, F., 2004. The modulation of the Ne-like wave on correct responses foreshadows errors. Neurosci. Lett. 372 (1–2), 161–166.

Amodio, D.M., Master, S.L., Yee, C.M., Taylor, S.E., 2008. Neurocognitive components of the behavioral inhibition and activation systems: implications for theories of self-regulation. Psychophysiology 45, 11–19.

Bartholow, B.D., Pearson, M.A., Dickter, C.L., Sher, K.J., Fabiani, M., Gratton, G., 2005. Strategic control and medial frontal negativity: beyond errors and response conflict. Psychophysiology 42 (1), 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 Res. 1101 (1), 92–101.

Botvinick, M.M., 2007. Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. Cogn. Affect. Behav. Neurosci. 7 (4), 356–366.

---

[2] These filter settings are commonly used in the ERN literature, however, we also evaluated reliability of the ERN using a filter of 4–7Hz. Test–retest reliabilities were comparable to the 0.1–30Hz filtering for area measures (CRN: $r = 0.84$; ERN: $r = 0.58$; Pe: $r = 0.64$; ERN − CRN (difference): $r = 0.70$), peak measures (CRN: $r = 0.76$; ERN: $r = 0.60$; ERN − CRN (difference): $r = 0.73$), and peak latency measures (CRN: $r = 0.24$; ERN: $r = 0.47$; ERN − CRN (difference): $r = 0.32$). The split-half reliability estimates were also comparable.

[3] The paper is mainly intended as a resource to inform work on the ERN, which is typically scored at the FCz electrode. However, reliability estimates for the CRN, ERN, Pe and ERN − CRN (difference) were comparable at Cz, CPz, Pz, and POz (*r*'s ranging from 0.54 to 0.86), however reliability estimates for Fz were somewhat lower (CRN: $r = 0.50$; ERN: $r = 0.49$; Pe: $r = 0.46$, ERN − CRN (difference): $r = 0.45$).

[4] In order to examine the impact of a different baseline, ERP data was reanalyzed using a baseline of − 200 to 0ms prior to the response: test–retest reliabilities were comparable to the baseline of − 400 to 200ms for area measures (CRN: $r = 0.75$; ERN: $r = 0.74$; Pe: $r = 0.66$; ERN − CRN (difference): $r = 0.68$), peak measures (CRN: $r = 0.82$; ERN: $r = 0.62$; ERN − CRN (difference): $r = 0.67$), and peak latency measures (CRN: $r = 0.18$; ERN: $r = 0.52$; ERN − CRN (difference): $r = 0.32$). The split-half reliability estimates were also comparable.

Brazdil, M., Roman, R., Falkenstein, M., Daniel, P., Jurak, P., Rektor, I., 2002. Error processing—evidence from intracerebral ERP recordings. Exp. Brain Res. 146 (4), 460–466.

Brazdil, M., Roman, R., Daniel, P., Rektor, I., 2005. Intracerebral error-related negativity in a simply go/nogo task. J. Psychophysiol. 19, 244–255.

Chiu, P.H., Deldin, P.J., 2007. Neural evidence for enhanced error detection in major depressive disorder. Am. J. Psychiatry 164 (4), 608–616.

Coles, M.G., Scheffers, M.K., Holroyd, C.B., 2001. Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. Biol. Psychol. 56 (3), 173–189.

Compton, R.J., Lin, M., Vargas, G., Carp, J., Fineman, S.L., Quandt, L.C., 2008. Error detection and posterror behavior in depressed undergraduates. Emotion 8 (1), 58–67.

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. Biol. Psychol. 56 (3), 191–206.

Dehaene, S., Posner, M.I., Tucker, D.M., 1994. Localization of a neural system for error detection and compensation. Psychol. Sci. 5, 303–305.

Dikman, Z.V., Allen, J.J., 2000. Error monitoring during reward and avoidance learning in high- and low-socialized individuals. Psychophysiology 37 (1), 43–54.

Eriksen, B.A., Eriksen, C.W., 1974. Effects of noise letters on the identification of target letters in a non-search task. Percept. Psychophys. 16, 143–149.

Fabiani, M., Gratton, G., Karis, D., Donchin, E., 1987. The definition, identification, and reliability of measurement of the P3 component of the event-related potential. In: Ackles, P.K., Jennings, J.R., Coles, M.G. (Eds.), Advances in Psychophysiology (Vol. 2). JAI Press, Greenwich.

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. Electroencephalogr. Clin. Neurophysiol. 78 (6), 447–455.

Falkenstein, M., Hoormann, J., Christ, S., Hohnsbein, J., 2000. ERP components on reaction errors and their functional significance: a tutorial. Biol. Psychol. 51 (2–3), 87–107.

Falkenstein, M., Hoormann, J., Hohnsbein, J., Kleinsorge, T., 2003. Short-term mobilization of processing resources is revealed in the event-related potential. Psychophysiology 40 (6), 914–923.

Fallgatter, A.J., Bartsch, A.J., Strik, W.K., Mueller, T.J., Eisenack, S.S., Neuhauser, B., et al., 2001. Test–retest reliability of electrophysiological parameters related to cognitive motor control. Clin. Neurophysiol. 112 (1), 198–204.

Foa, E.B., Huppert, J.D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., et al., 2002. The Obsessive–compulsive inventory: development and validation of a short version. Psychol. Assess. 14 (4), 485–496.

Ford, J.M., 1999. Schizophrenia: the broken P300 and beyond. Psychophysiology 36 (6), 667–682.

Frank, M.J., Woroch, B.S., Curran, T., 2005. Error-related negativity predicts reinforcement learning and conflict biases. Neuron 47 (4), 495–501.

Gehring, W.J., Knight, R.T., 2000. Prefrontal–cingulate interactions in action monitoring. Nat. Neurosci. 3 (5), 516–520.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E., 1993. A neural system for error detection and compensation. Psychol. Sci. 4, 385–390.

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

Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artifact. Electroencephalogr. Clin. Neurophysiol. 55 (4), 468–484.

Hajcak, G., Foti, D., 2008. Errors are aversive: defensive motivation and the error-related negativity. Psychol. Sci. 19 (2), 103–108.

Hajcak, G., Simons, R.F., 2002. Error-related brain activity in obsessive–compulsive undergraduates. Psychiatry Res. 110 (1), 63–72.

Hajcak, G., Simons, R.F., 2008. Oops!.. I did it again: an ERP and behavioral study of double-errors. Brain Cogn. 68 (1), 15–21.

Hajcak, G., McDonald, N., Simons, R.F., 2003a. Anxiety and error-related brain activity. Biol. Psychol. 64 (1–2), 77–90.

Hajcak, G., McDonald, N., Simons, R.F., 2003b. To err is autonomic: error-related brain potentials, ANS activity, and post-error compensatory behavior. Psychophysiology 40 (6), 895–903.

Hajcak, G., Huppert, J.D., Simons, R.F., Foa, E.B., 2004a. Psychometric properties of the OCI-R in a college sample. Behav. Res. Ther. 42 (1), 115–123.

Hajcak, G., McDonald, N., Simons, R.F., 2004b. Error-related psychophysiology and negative affect. Brain Cogn. 56 (2), 189–197.

Hajcak, G., Moser, J.S., Yeung, N., Simons, R.F., 2005a. On the ERN and the significance of errors. Psychophysiology 42 (2), 151–160.

Hajcak, G., Nieuwenhuis, S., Ridderinkhof, K.R., Simons, R.F., 2005b. Error-preceding brain activity: robustness, temporal dynamics, and boundary conditions. Biol. Psychol. 70 (2), 67–78.

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. Am. J. Psychiatry 165 (1), 116–123.

Hall, J.R., Bernat, E.M., Patrick, C.J., 2007. Externalizing psychopathology and the error-related negativity. Psychol. Sci. 18 (4), 326–333.

Helmstader, G.C., 1964. Principles of Psychological Measurement. Appleton-Century-Crofts, New York.

Holmes, A.J., Pizzagalli, D.A., 2008. Spatiotemporal dynamics of error processing dysfunctions in major depressive disorder. Arch. Gen. Psychiatry 65 (2), 179–188.

Holroyd, C.B., Coles, M.G., 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol. Rev. 109 (4), 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. Neurosci. Lett. 242 (2), 65–68.

Holroyd, C.B., Yeung, N., Coles, M.G., Cohen, J.D., 2005. A mechanism for error detection in speeded response time tasks. J. Exp. Psychol. Gen. 134 (2), 163–191.

Kerns, J.G., Cohen, J.D., MacDonald III, A.W., Johnson, M.K., Stenger, V.A., Aizenstein, H., et al., 2005. Decreased conflict- and error-related activity in the anterior cingulate cortex in subjects with schizophrenia. Am. J. Psychiatry 162 (10), 1833–1839.

Ladouceur, C.D., Dahl, R.E., Birmaher, B., Axelson, D.A., Ryan, N.D., 2007. Decreased Pe, but not ERN, amplitude following treatment of children diagnosed with an anxiety disorder: preliminary results. Psychophysiology 44, s99.

Leuthold, H., Sommer, W., 1999. ERP correlates of error processing in spatial S-R compatibility tasks. Clin. Neurophysiol. 110 (2), 342–357.

Luck, S.J., 2005. An Introduction to the Event-Related Potential Technique. MIT Press, Cambridge, MA.

Luu, P., Collins, P., Tucker, D.M., 2000a. Mood, personality, and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. J. Exp. Psychol. Gen. 129 (1), 43–60.

Luu, P., Collins, P., Tucker, D.M., 2000b. Mood, personality, and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. J. Exp. Psychol. Gen. 129 (1), 43–60.

Moser, J.S., Hajcak, G., Simons, R.F., 2005. The effects of fear on performance monitoring and attentional allocation. Psychophysiology 42 (3), 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 (5), 752–760.

Olvet, D.M., Hajcak, G., 2008. The error-related negativity (ERN) and psychopathology: toward an endophenotype. Clin. Psychol. Rev. 28 (8), 1343–1354.

Olvet, D.M., & Hajcak, G., in press. The stability of error-related brain activity with increasing trials. Psychophysiology.

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. J. Psychophysiol. 19, 319–329.

Pailing, P.E., Segalowitz, S.J., 2004. The error-related negativity as a state and trait measure: motivation, personality, and ERPs in response to errors. Psychophysiology 41 (1), 84–95.

Pailing, P.E., Segalowitz, S.J., Dywan, J., Davies, P.L., 2002. Error negativity and response control. Psychophysiology 39 (2), 198–206.

Potts, G.F., George, M.R., Martin, L.E., Barratt, E.S., 2006. Reduced punishment sensitivity in neural systems of behavior monitoring in impulsive individuals. Neurosci. Lett. 397 (1–2), 130–134.

Rabbitt, P., Rodgers, B., 1977. What does a man do after he makes an error? An analysis of response programming. Q. J. Exp. Psychol. 29, 727–743.

Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., Nieuwenhuis, S., 2004. The role of the medial frontal cortex in cognitive control. Science 306 (5695), 443–447.

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 46 (3), 531–538.

Ruchsow, M., Spitzer, M., Gron, G., Grothe, J., Kiefer, M., 2005. Error processing and impulsiveness in normals: evidence from event-related potentials. Cogn. Brain Res. 24 (2), 317–325.

Sandman, C.A., Patterson, J.V., 2000. The auditory event-related potential is a stable and reliable measure in elderly subjects over a 3 year period. Clin. Neurophysiol. 111 (8), 1427–1437.

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. J. Exp. Psychol. Hum. Percept. Perform. 26 (1), 141–151.

Scheffers, M.K., Coles, M.G., Bernstein, P., Gehring, W.J., Donchin, E., 1996. Event-related brain potentials and error-related processing: an analysis of incorrect responses to go and no-go stimuli. Psychophysiology 33 (1), 42–53.

Sechrest, L., 1984. Reliability and validity. In: Bellack, A.S., Hersen, M. (Eds.), Research Methods in Clinical Psychology. Pergamon Press, New York.

Segalowitz, S.J., Barnes, K.L., 1993. The reliability of ERP components in the auditory oddball paradigm Psychophysiology 30 (5), 451–459.

Shafran, R., Mansell, W., 2001. Perfectionism and psychopathology: a review of research and treatment. Clin. Psychol. Rev. 21 (6), 879–906.

Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlation: uses in assessing rater reliability. Psychol. Bull. 86, 420–428.

van Veen, V., Carter, C.S., 2002. The timing of action-monitoring processes in the anterior cingulate cortex. J. Cogn. Neurosci. 14 (4), 593–602.

Vidal, F., Hasbroucq, T., Grapperon, J., Bonnet, M., 2000. Is the 'error negativity' specific to errors? Biol. Psychol. 51 (2–3), 109–128.

Walhovd, K.B., Fjell, A.M., 2002. One-year test–retest reliability of auditory ERPs in young and old adults. Int. J. Psychophysiol. 46 (1), 29–40.

Williams, L.M., Simms, E., Clark, C.R., Paul, R.H., Rowe, D., Gordon, E., 2005. The test–retest reliability of a standardized neurocognitive and neurophysiological test battery: "neuromarker". Int. J. Neurosci. 115 (12), 1605–1630.

Yeung, N., Cohen, J.D., Botvinick, M.M., 2004. The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol. Rev. 111 (4), 931–959.