Biological Psychology 93 (2013) 377–385

Contents lists available at SciVerse ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks

Anja Riesel[a,b,*], Anna Weinberg[b], Tanja Endrass[a], Alexandria Meyer[b], Greg Hajcak[b]

[a] *Institut für Psychologie, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany*
[b] *Department of Psychology, Stony Brook University, New York, USA*

ARTICLE INFO

Article history:
Received 11 April 2012
Accepted 9 April 2013
Available online 19 April 2013

Keywords:
Construct validity
Discriminant validity
Error-related negativity
Error positivity
ERN
Pe

ABSTRACT

Error-processing is increasingly examined using the error-related negativity (ERN) and error positivity (Pe) – event-related potentials (ERPs) that demonstrate trait-like properties and excellent reliability. The current study focuses on construct validity by applying a multitrait–multimethod approach, treating error-related ERPs (i.e., ERN, Pe and the difference between error minus correct, referred to as ΔERN and ΔPe, respectively) as traits measured across multiple tasks (i.e., Flanker, Stroop, and Go/NoGo). Results suggest convergent validity of these ERPs ranging between .62 and .64 for ΔERN. Values were somewhat smaller for ERN (range .33–.43), Pe (range .37–.49) and ΔPe (range .30–.37). Further, the correlations for ERN and Pe are higher within components across tasks than between different components suggesting discriminant validity. In conclusion, the present study revealed evidence for convergent and discriminant validity of error-related ERPs, further supporting the use of these components as psychophysiological trait markers.

1. Introduction

Adaptive behavior in a changing world requires a flexible system that monitors performance and detects errors. Psychophysiological research on performance monitoring has flourished since error-related event-related potentials (ERPs) were discovered 20 years ago (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). Studies of performance monitoring have focused in particular on the error-related negativity (ERN; Gehring et al., 1993) or error negativity (Ne; Falkenstein et al., 1991), a sharp negative deflection that appears shortly after the commission of an error over frontocentral electrodes. The anterior cingulate cortex (ACC) has been suggested as the primary generator of the ERN based on studies using both functional neuroimaging (Debener et al., 2005; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004) and source localization techniques (Dehaene, Posner, & Tucker, 1994).

In addition to the ERN, other response-related ERPs are used to examine action monitoring. Several studies report a smaller, but similar looking, negative-going component following correct responses, called the correct response negativity (CRN; Ford, 1999; Vidal, Hasbroucq, Grapperon, & Bonnet, 2000). The ERN is typically followed by the error positivity (Pe; Falkenstein et al., 1991). The Pe has a centroparietal distribution and occurs within 200–500 ms after incorrect responses.

Despite a considerable body of research, the functional significance of these electrophysiological measures of action monitoring is still debated. The Pe has been related to error awareness (Endrass, Reuter, & Kathmann, 2007; Hughes & Yeung, 2011; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001) and some suggest it represents a P3-like response to infrequent error commission (Arbel & Donchin, 2009, 2011; Overbeek, Nieuwenhuis, & Ridderinkhof, 2005; Ridderinkhof, Ramautar, & Wijnen, 2009).

The ERN on the other hand is assumed to signal the need to adjust behavior and to increase cognitive control to improve future performance (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Falkenstein et al., 1991; Gehring et al., 1993; Holroyd & Coles, 2002). More specifically, the ERN has been thought to reflect a neural correlate of conflict monitoring (Botvinick et al., 2001; Yeung, Botvinick, & Cohen, 2004), reinforcement learning (Holroyd & Coles, 2002) or error-likelihood (Brown & Braver, 2005). In addition, the ERN is related to motivational and individual difference variables and is thought to be a trait marker that reflects individual differences in the subjective value of errors based on context, personality, and learning history (Hajcak, 2012; Olvet & Hajcak, 2008; Weinberg, Riesel, & Hajcak, 2012).

Despite this increasing interest in the ERN as a potential trait-like biomarker, the psychometric properties of the ERN have not been thoroughly investigated. There is evidence that the internal consistency and temporal stability of error-related ERPs (ERN, CRN,

---

* Corresponding author at: Humboldt-Universität zu Berlin, Institut für Psychologie/Klinische Psychologie, Rudower Chaussee 18, 12489 Berlin, Germany.
Tel.: +49 30 2093 4918; fax: +49 30 2093 4859.
E-mail address: anja.riesel@hu-berlin.de (A. Riesel).

0301-0511/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.biopsycho.2013.04.007

Pe) as well as the difference between the ERN and CRN (referred to as ΔERN) are excellent (Olvet & Hajcak, 2009b; Segalowitz et al., 2010; Weinberg & Hajcak, 2011). Specifically, estimates of test–retest reliability range between .40 and .82 over a period of 2–6 weeks (Olvet & Hajcak, 2009b; Segalowitz et al., 2010) and are similar in size (ranging from .56 to .67) after as long as 2 years (Weinberg & Hajcak, 2011).

However, few studies investigated the degree to which the ERN is comparable *across tasks* within the same individuals. This is particularly important given that the ERN is measured in a variety of tasks, including the Flanker (e.g., Gehring et al., 1993), Go/NoGo (e.g., Bates, Kiehl, Laurens, & Liddle, 2002) and Stroop tasks (e.g., Hajcak & Simons, 2002). Initial evidence suggests that the ERN elicited in a Flanker and a Go/NoGo task is highly correlated within the same individuals (Segalowitz et al., 2010). Further, the ERN is often scored at FCz where it is maximal (e.g., Bates et al., 2002; Falkenstein, Hoormann, Christ, & Hohnsbein, 2000; Gentsch, Ullsperger, & Ullsperger, 2009; Luu, Tucker, Derryberry, Reed, & Poulsen, 2003; Ridderinkhof et al., 2002), but it is unknown where the ERN has the highest convergence across tasks.

Nonetheless, there is some evidence for task-dependent modulations of the ERN (Grundler, Cavanagh, Figueroa, Frank, & Allen, 2009; Mathews, Perez, Delucchi, & Mathalon, 2012; Nieuwenhuis, Nielen, Mol, Hajcak, & Veltman, 2005; Olvet & Hajcak, 2009a). For example, research suggests that enhanced ERN amplitudes in obsessive-compulsive disorder may only be found in response-conflict tasks and not in tasks with probabilistic stimulus–response mappings (Grundler et al., 2009; Mathews et al., 2012; Nieuwenhuis et al., 2005). It has even been proposed that different, albeit overlapping, neural systems can underlie the ERN, depending on the specific task (i.e., execution of an incorrect motor response vs. suboptimal choice; Cavanagh, Grundler, Frank, & Allen, 2010; Grundler et al., 2009). In short, it is not yet clear whether the ERN measured across common tasks reflects a unitary phenomenon or has task-dependent characteristics. Thus, there is a need to directly compare indices of error-related brain activity across different tasks to examine the construct validity of the ERN.

If correlations between ERNs measured across tasks were low, this would suggest that the ERN is not a singular entity, and that correlations with individual difference measures may depend heavily on the task used to elicit the ERN. In this case, it would be useful to specify results in terms of the task employed (e.g. Stroop-ERN). Given that the ERN is discussed as a promising biomarker, which could be useful for diagnostic or prognostic purposes, this question is of central importance in guiding such research efforts.

To this end, error-related ERPs (i.e., ERN, Pe) were assessed in the current study using three commonly employed speeded response tasks (i.e., Flanker, Stroop, Go/NoGo). An adjusted multitrait–multimethod matrix (MMTM; Campbell & Fiske, 1959) was applied to examine whether indices of error-related brain activity (i.e., traits) converge across different tasks (i.e., methods). This was done in order to examine how comparable (convergent validity) and distinguishable (discriminant validity) error-related components are across tasks.

2. Methods

2.1. Participants

Forty-seven undergraduate students (20 female) from Stony Brook University participated in this study. Two participants committed fewer than six errors and were therefore excluded from further analysis, since evidence suggests that between 6 and 8 error trials are needed to reliably quantify the ERN and Pe (Olvet & Hajcak, 2009c; Pontifex et al., 2010). Data from two subjects were excluded due to excessive EEG artifacts. The final sample consisted of 43 participants (19 female). All participants had normal or corrected-to-normal vision and reported no history of head trauma or neurological disease. The mean age was 19.14 years ($SD = 1.42$). 38.6% of the sample was Caucasian/European, 45.5% was Asian-American, 6.8% was Hispanic, 2.3% was African-American and 6.8% identified as “other.” All participants received verbal and written information about the aims and procedure of the study and written consent was obtained. All participants received course credit for their participation.

2.2. Task and procedure

The experiment consisted of three tasks: a modified Flanker task, a Go/NoGo task, and a Stroop task. The order of tasks was counterbalanced across participants. All tasks were administered using Presentation software (Neurobehavioral Systems, Inc., Albany, CA, USA). Prior to each task, the participants performed a practice block containing 20 trials. All three tasks consisted of 420 trials presented in 7 blocks of 60 trials. All stimuli were presented for 200 ms. An intertrial interval (ITI) that varied randomly from 600 to 1000 ms followed the response. Throughout all tasks, participants were encouraged to be both fast and accurate in their performance. To encourage both fast and accurate behavior, performance-based feedback was presented at the end of each block. If performance accuracy was below 75%, a message appeared instructing participants to respond more accurately. When performance was above 90%, participants were instructed to respond faster. Error rates between 10 and 25% were followed by the feedback “You’re doing a great job.” For the Go/NoGo task, the performance feedback was given with regard to error rates that were calculated for NoGo trials only (i.e., errors of commission). The total duration of the three tasks combined was approximately 60 min.

Flanker task: On each trial of the Flanker task (Eriksen & Eriksen, 1974; Kopp, Rist, & Mattler, 1996), five horizontally aligned white arrowheads were presented and participants were instructed to respond with the left or right mouse button in accordance with the direction of the central arrowhead. Half the trials were compatible (e.g. flanker arrows and target point in the same direction) and half were incompatible (e.g. flanker arrows and target point in opposite directions). The trials were displayed in a pseudorandomized order. At a viewing distance of approximately 65 cm, the set of arrows filled 2° of visual angle vertically and 10° horizontally.

Stroop task: On each trial, one of three color words (‘red’, ‘green’, ‘blue’) was shown, and was presented in either red or green font. Subjects were instructed to press the left mouse button if the color word was presented in red, and press right button if the color word was presented in green. Thus, 1/3 of trials were compatible (e.g. color word and font color require the same response, ‘red’ in red font, ‘green’ in green font), 1/3 were incompatible (e.g. color word and font color require different responses, ‘red’ in green font, ‘green’ in red font), and 1/3 were neutral (e.g. the color word, ‘blue’ in red or green font). At a viewing distance of approximately 65 cm, each word occupied between 2° and 3° of visual angle.

Go/NoGo task: In the Go/NoGo task, a green triangle was presented on each trial. Participants were instructed to press the right mouse button in response to an upright triangle, which occurred on 80% of the trials. Additionally, participants were told to withhold responses to slightly tilted triangles (10°), which occurred on 20% of the trials. At a viewing distance of approximately 65 cm, each triangle occupied $3^\circ \times 3^\circ$ of the visual angle.

2.3. Psychophysiological recording, data reduction and analysis

The continuous EEG was recorded using an elastic cap and the ActiveTwo BioSemi system (BioSemi, Amsterdam, Netherlands). Sixty-four electrode sites were used, based on the 10/20 system, as well as two electrodes on the right and left mastoids. All electrodes were sintered Ag–AgCl electrodes. The Electrooculogram (EOG) was recorded using four additional facial electrodes: two electrodes placed approximately 1 cm outside of the right and left eyes and two electrodes mounted approximately 1 cm above and below the right eye. To improve the signal-to-noise ratio, the EEG signal was pre-amplified at the electrode with a gain of 1× by a BioSemi ActiveTwo system (BioSemi, Amsterdam, Netherlands). The EEG was digitized with a sampling rate of 512 Hz using a low-pass fifth order sinc filter with a half-power cutoff of 102.4 Hz. A common mode sense (CMS) active electrode producing a monopolar (non-differential) channel was used as recording reference. The EEG was analyzed using Brain Vision Analyzer (Brain Products, Gilching, Germany).

2.4. ERP analysis

Offline, the data were referenced to the average of the left and right mastoids, and band-pass filtered with low and high cutoffs of 0.1 and 30 Hz, respectively. Eye movement artifacts were corrected using the algorithm developed by Gratton, Coles, and Donchin (1983). Response-locked epochs with a duration of 1200 ms, including a 400 ms prestimulus interval, were extracted. A semi-automatic procedure was used to detect and reject artifacts. Epochs containing a voltage step of more than 50 μV between sample points, a voltage difference of 300 μV within a segment, and a maximum voltage difference of less than 0.50 μV within 100 ms intervals were rejected. In addition, visual inspection of the data was conducted to detect and reject any remaining artifacts. Response-locked ERPs were averaged separately for each participant, each task, and for incorrect and correct responses. For all tasks, trials with response times below 100 ms and above 700 ms were excluded from averaging. Because the ERN can begin prior to the completion of the motor response, we used the 400–200 ms pre-response interval as the baseline in order to avoid subtracting out activity of interest (Weinberg, Olvet, & Hajcak, 2010). To quantify the ERN and

Table 1
Task performance and measures of error-processing (means and standard deviations) in the Flanker, Go/NoGo and Stroop task.

	Flanker task	Go/NoGo task	Stroop task	Task effect	
				F	*p*
Number of Errors	52.95 (25.38)	30.44 (10.06)	55.05 (31.14)	20.78	p < .001
Correct RT in ms	396 (50)	318 (42)	422 (68)	115.38	p < .001
Error RT in ms	323 (46)	268 (32)	394 (71)		
Post-Error Slowing in ms	9 (21)	26 (43)	23 (33)	3.95	p < .05
ΔERN	-7.66 (5.32)	-8.86 (5.06)	-6.16 (4.53)	7.86	p < .01
ERN	.84 (5.09)	-2.25 (5.01)	.98 (4.97)	6.93	p < .05
CRN	8.50 (5.05)	6.61 (5.18)	7.14 (6.44)		
ΔPe	10.56 (4.93)	10.50 (5.84)	6.61 (4.44)	10.73	p < .001
Pe	13.01 (6.36)	14.43 (7.63)	10.63 (5.53)	3.20	p < .05
Correct Positivity	2.45 (3.61)	3.93 (3.94)	4.02 (4.52)		

CRN, we calculated mean amplitudes for the time interval between 0 and 100 ms after responses at FCz, where error-related brain activity was maximal. A negative deflection is typically observable after both error and correct trials, therefore it is common to analyze not just ERN and CRN, but also the difference between them (ΔERN) in order to isolate activity unique to error-processing from activity more broadly related to response-monitoring (Simons, 2010). This difference score shows comparable reliability with ERN or CRN alone (Olvet & Hajcak, 2009b; Weinberg & Hajcak, 2011). The Pe was evaluated as the average activity from 200 to 400 ms at Pz, and again a difference score subtracting the positivity on error trials minus correct trials was calculated (i.e., ΔPe). Grand averages were filtered with a 15-Hz low-pass filter for visual presentation.

2.5. Statistical analysis

Statistical analyses were conducted using SPSS (Version 19.0). The significance level was $\alpha = .05$, two-tailed. The analyses are described in several parts below. First, a 2 (*response*: error, correct) × 3 (*task*: Flanker, Stroop, Go/NoGo) repeated-measurement analysis of variance (ANOVA) was used to analyze the reaction time and electrophysiological data. Paired *t*-tests were performed for follow-up, post hoc tests. When appropriate, the Greenhouse–Geisser correction was used for all comparisons with more than two within-subject levels and ε is reported. Second, correlational analyses (Pearson's *r*) were conducted to examine the relationship among measures of error-related brain activity (Tables 2 and 3) and correlations with behavioral indices. In addition, correlations were used to examine on which electrode position the ΔERN shows the highest associations (i.e., convergence) across tasks.

The interpretation and naming of the correlation matrix shown in Tables 2 and 3 is derived from a multitrait–multimethod matrix (MMTM, Campbell & Fiske, 1959), though ERP components are used as *traits*, and tasks reflect different *methods*.[1] Construct validity is evaluated by a rule-based examination of the observed correlation patterns (Campbell & Fiske, 1959). Validity can be assumed when the matrix follows the rules that are introduced below:

1. Reliability diagonal (monocomponent–monotask: black color): The main diagonal of the correlation matrix reflects the split-half reliability. The split-half reliability was calculated by using the odd-even method (i.e., correlation between even and odd trials). Coefficients in the reliability diagonal should consistently be the highest in the matrix, given that each measure should be more highly correlated with itself than other measures.
2. Validity diagonals (monocomponent–heterotask: red color): Correlations of the same measure assessed by different tasks. Each validity coefficient should be higher than values in its column and row in the same heterotask block.
3. Heterocomponent–monotask blocks (blue color): correlations between different measures assessed within a single task. A similar pattern of relationships between measures should be seen in all task triangles. Further, these correlations should be smaller than the correlations observed in the validity diagonals (i.e., the ERN in the flanker task should correlate more highly with the ERN in the Stroop task than with the Pe in the Flanker task). The latter criterion allows examining discriminant validity of these components.
4. Heterocomponent–heterotask blocks (gray color): Correlations between different measures assessed by different tasks. These coefficients should be the lowest in matrix.

3. Results

3.1. Behavioral results

Behavioral results for the three tasks are presented in Table 1. Reaction time varied significantly between the three tasks ($F(2, 82) = 115.38$, $p < 0.001$, $\varepsilon = .87$). The fastest reaction times were observed in the Go/NoGo task compared to the Flanker ($t(41) = 14.74$, $p < 0.001$) and Stroop task ($t(41) = 14.81$, $p < 0.001$). The Flanker task was characterized by faster reactions compared to the Stroop task ($t(42) = 4.76$, $p < 0.001$). In all tasks, reaction times were significantly faster for incorrect than correct responses ($F(1, 41) = 246.26$, $p < 0.001$). However, the difference in reaction times between incorrect and correct responses varied across tasks ($F(2, 82) = 36.07$, $p < 0.001$, $\varepsilon = .98$). This difference was most pronounced for the Flanker task (73 ms) compared to the Go/NoGo task (50 ms, $t(41) = 4.62$, $p < 0.001$) and Stroop task (28 ms, $t(42) = 8.21$, $p < 0.001$). In addition, the difference in reaction times in the Go/NoGo task was more pronounced compared to the Stroop task ($t(41) = 4.11$, $p < 0.001$).

Number of errors also differed between tasks ($F(2, 82) = 20.78$, $p < 0.001$, $\varepsilon = .86$), such that the Go/NoGo task elicited a smaller number of errors compared to the Flanker ($t(41) = 6.62$, $p < 0.001$) and Stroop tasks ($t(41) = 5.80$, $p < 0.001$). The Flanker and Stroop tasks did not differ from one another in terms of the number of errors they elicited ($t(42) = .44$, $p = 0.66$).

Post-error slowing was analyzed by comparing the difference between reaction times following correct trials and reaction times following errors. Post-error slowing was observable in all tasks (Flanker: $t(41) = 2.56$, $p < 0.05$, Stroop: $t(41) = 4.36$, $p < 0.001$, Go/NoGo: $t(41) = 3.76$, $p < 0.01$). Again, there was a main effect of task ($F(2, 82) = 3.95$, $p < 0.05$, $\varepsilon = .84$), such that both the Go/NoGo ($t(41) = 2.67$, $p < 0.05$) and Stroop tasks ($t(41) = 2.54$, $p < 0.05$) were characterized by a pronounced post-error slowing compared to the Flanker task. The slowing after an error did not differ between Go/NoGo and Stroop task ($t(41) = .63$, $p = 0.53$).

---

[1] These changes in naming were undertaken to indicate that (a) different methods have not been used but instead different tasks within one method (EEG) and (b) because it is still a matter of discussion whether these ERP components can be conceptualized as traits.

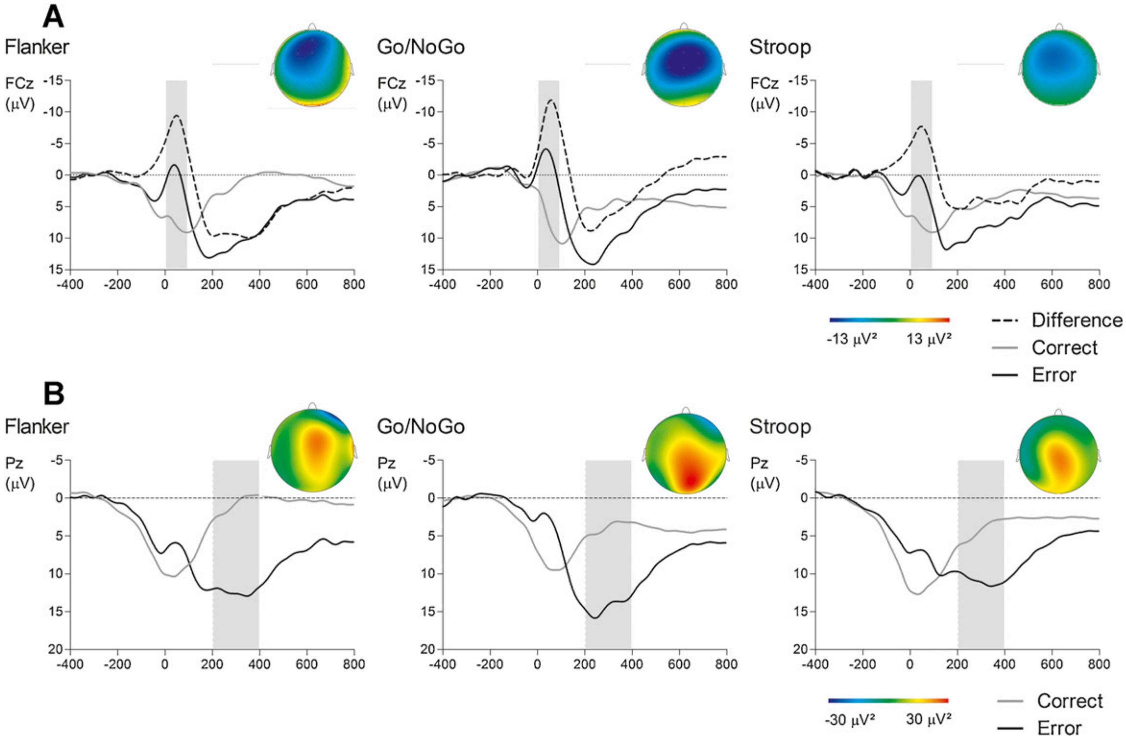

Fig. 1. (A) Grand average waveforms for correct and incorrect responses and the difference wave at electrode FCz for each of the three tasks, as well as each task's associated error minus correct scalp topography (0–100 ms, current source density). (B) Grand average waveforms for correct and incorrect responses at electrode Pz for each of the three tasks, as well as each task's associated topography for errors (200–400 ms, current source density). Note: the used interval for ERP and topography analyses is indicated with a gray bar.

3.2. ERP results

Fig. 1A presents the response-locked ERP waveforms for correct and error responses and topographies for ΔERN in the three tasks. Consistent with previous studies, the ERN was observed across tasks as a sharp frontocentral negative deflection that peaked shortly after the commission of an error and was more negative than the CRN ($F(1, 41) = 124.67$, $p < .001$). In addition, a main effect of task ($F(2, 82) = 6.93$, $p < .01$, $\varepsilon = .96$) and an interaction between task and response ($F(2, 82) = 7.89$, $p < .01$, $\varepsilon = .96$) was also observed, indicating that the difference between ERN and CRN (i.e., ΔERN) varied as a function of task. The ΔERN in the Go/NoGo task was more pronounced than the ΔERN in the Stroop task ($t(42) = 4.40$, $p < 0.001$) and was marginally larger than the ΔERN elicited by the Flanker task ($t(42) = 1.83$, $p = 0.07$). In addition, the ΔERN in the Flanker task was more negative compared to the ΔERN in the Stroop task ($t(42) = 2.3$, $p = 0.05$).

The positivity after errors for the three tasks is shown in Fig. 1B. The Pe was also enhanced following errors (compared to the positivity following correct responses, $F(1, 41) = 253.40$, $p < .001$). As above, a significant main effect of task ($F(2, 82) = 3.20$ $p < .05$, $\varepsilon = .96$) was qualified by an interaction with response type ($F(2, 82) = 12.43$, $p < .001$, $\varepsilon = .95$). The difference in the positivity between errors and correct responses was larger for both the Flanker ($t(42) = 4.46$, $p < 0.001$) and Go/NoGo tasks ($t(42) = 3.73$, $p < 0.01$) compared to the Stroop task. The ΔPe elicited by the Flanker and Go/NoGo tasks did not differ ($t(42) = .07$, $p = 0.95$).[2]

3.3. Construct validity

Fig. 2 displays the size of the correlation between the ΔERN across tasks as a function of electrode site. Maximal correlations across tasks can be observed at frontocentral electrode sites, consistent with the scalp maximum of the ΔERN. All subsequent correlations across tasks focus on scores at FCz, where the ΔERN was both maximal and maximally correlated across tasks. Tables 2 and 3 show the correlations between the different measures of error-monitoring both within and across tasks. Table 2 reports the results for difference scores isolating either error- or conflict-specific processes, whereas Table 3 reports measures derived from error trials only. Since the overall pattern of results is similar in both matrices, the results and discussion will focus on difference (i.e., Δ) measures.

Reliability diagonal (monocomponent-monotask, black color): The main diagonal of the correlation matrix reflects the split-half reliability. The split-half reliability coefficients were the highest correlations in the matrix, consistent with the notion that each measure should be more highly correlated with itself than other measures. However, the reliability of the error-related measures differed between tasks (.53–.82), with split-half reliability values highest in the Flanker task.

Validity diagonals (monocomponent-heterotask, red color): These diagonals represent correlations of the same ERP measure assessed by different tasks. The validity diagonals indicate significant correlations for each measure of error-related brain activity across tasks. The highest validity values were observed for ΔERN measures across tasks, with values above .60. The ERN, Pe and ΔPe also had reasonably high indices across tasks (range .30–.49). Importantly and as expected, the validity values within any given variable were higher than the correlations observed between that variable and another variable having neither component nor task in common (i.e., gray values), which were all low and non-significant.

---

[2] *Correlations to behavioral indices*: Correlations were used to explore the associations between error-related ERPs (ΔERN and Pe) and behavioral indices (post-error slowing and numbers of errors). Neither ΔERN nor Pe were associated with number of errors or post-error slowing (all $rs < .30$, $ps > .05$).

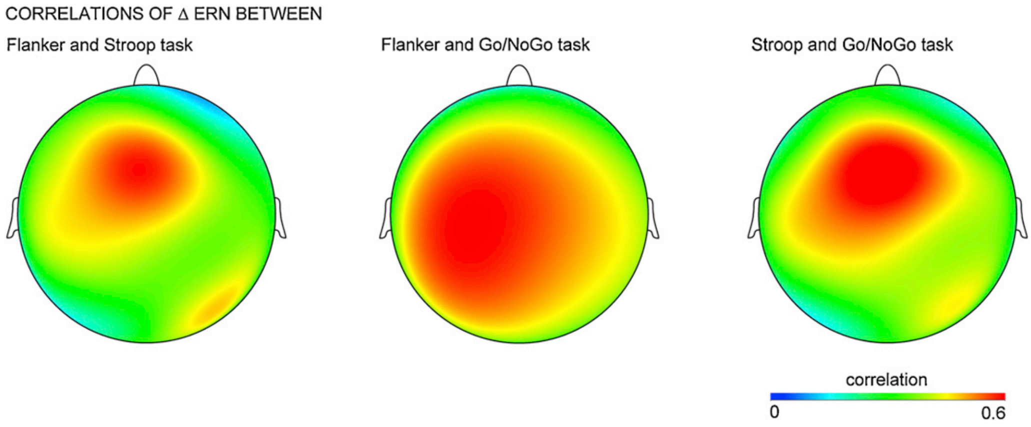

Fig. 2. Correlational headmaps for ΔERN at all electrodes depicting the strength of correlations between Flanker and Stroop task (left), Flanker and Go/NoGo task (middle) and Stroop and Go/NoGo task (right).

Table 2
Bivariate Pearson correlations between different Δ-scored ERPs across tasks.

		Flanker		Stroop		Go/NoGo	
		Δ ERN	Δ Pe	Δ ERN	Δ Pe	Δ ERN	Δ Pe
Flanker	Δ ERN	.76**					
	Δ Pe	-.07	.82**				
Stroop	Δ ERN	.65**	-.09	.69**			
	Δ Pe	-.08	.34*	.18	.53**		
Go/NoGo	Δ ERN	.65**	-.26	.66**	-.05	.64**	
	Δ Pe	-.07	.37**	-.15	.30*	.16	.56**

Note: the main diagonal (in black) shows split-half reliability indices, the validity diagonals are depicted in red, the blue values display heterocomponent–monotask correlations, the gray values show heterocomponent–heterotask correlations.

* $p < .05$.

** $p < .01$.

Table 3
Bivariate Pearson correlations between different error-related ERPs across tasks.

		Flanker		Stroop		Go/NoGo	
		ERN	Pe	ERN	Pe	ERN	Pe
Flanker	ERN	.81**					
	Pe	.37*	.87**				
Stroop	ERN	.37*	.28	.69**			
	Pe	-.03	.41**	.39*	.58**		
Go/NoGo	ERN	.43**	.10	.33*	.11	.60**	
	Pe	.12	.49**	.10	.37*	.58**	.73**

Note: the main diagonal (in black) shows split-half reliability indices, the validity diagonals are depicted in red, the blue values display heterocomponent–monotask correlations, the gray values show heterocomponent–heterotask correlations.

* $p < .05$.

** $p < .01$.

In addition, there is also evidence for task-dependent effects on error-related brain activity. For example, the construct validity of ΔERN and ΔPe (and also ERN and Pe) was somewhat lower when comparing Stroop and Go/NoGo tasks. Further, the amount of shared variance for one component assessed with different tasks suggests that distinct influences must exist (R^2 range 9–44%, see Tables 2 and 3). Further, each task appears to share *unique* variation in the ΔERN with each of the other tasks as reflected in partial correlations. For instance, the correlation between ΔERN in the Flanker and Go/NoGo task, controlling for Stroop task was $r = .40$, $p < .01$. Similarly, the correlation between Flanker and Stroop task, controlling for the Go/NoGo task was $r = .39$, $p < .05$. Finally, the correlation between ΔERN in the Go/NoGo and Stroop task, controlling for Flanker task was $r = .37$, $p < .05$. Thus, each measure of ΔERN indexes both unique and overlapping variation in error-related brain activity that is observed across different tasks.

Heterocomponent–monotask blocks (blue color): Correlations between different measures assessed within a single task. A similar pattern of component interrelationships in the heterocomponent–monotask blocks can be seen for each task: ΔERN and ΔPe were not correlated with one another (Table 2). However, ERN and Pe were significantly correlated (Table 3), such that more positive Pe amplitudes were associated with more positive (i.e., smaller) ERN amplitudes. These data suggest better discriminant validity between the ΔERN and ΔPe than the ERN and Pe, perhaps because the use of difference measures represent a quantification of error-processing that is more independent of overall task-related differences, interindividual differences in ERP size, and other potential confounds.

Heterocomponent–heterotask blocks (gray color): Correlations between different measures assessed by different tasks. As expected these correlations were the lowest and none reached significance.

4. Discussion

The aim of the present study was to examine the construct validity of error-related ERPs. The MMTM approach (Campbell & Fiske, 1959) was adjusted and applied to the present data set to examine how error-related measures, taken as traits, relate across three commonly used methods (i.e., tasks: Flanker, Stroop and Go/NoGo task). Establishing the psychometric properties of neural indicators of error-processing is important because different tasks have been utilized across studies to index the ERN as an individual difference variable (Weinberg et al., 2012). However, studies had not examined the overlap between ERNs elicited from multiple tasks.

4.1. ERP results

Consistent with previous work, the ERN was observed across tasks as a sharp negative deflection maximal at frontocentral electrodes; the ERN was followed by a centroparietal positivity (i.e., the Pe). However, both the ERN and Pe varied as a function of task, and were smallest in the Stroop task. The Stroop task was also associated with the highest number of errors and longest reaction times. Thus, the Stroop task may have been the most difficult – and the reduced ERN/Pe might have reflected increased task difficulty. Consistent with this possibility, previous findings indicate a decrease in the magnitude of the ERN with increasing task difficulty (Falkenstein, 2004; Hoffmann & Falkenstein, 2010; Johannes et al., 2002; Pailing & Segalowitz, 2004). The present data suggest that the Go/NoGo task was the easiest, whereas the Flanker task was moderately difficult among the examined paradigms. An identical pattern of results for ERN and Pe suggests that differences related to task difficulty may account for variation in these neural measures of error processing.

4.2. Construct validity

Reliability: Consistent with the notion that a component should be more highly correlated with itself than with other measures, split-half coefficients were the highest correlations observed. However, reliability values varied between tasks, and were largest for the Flanker task. Since reliability is a prerequisite for validity and provides an upper limit to correlations with other measures, differences in reliability between tasks may help explain some inconsistent results in previous individual differences studies on the ERN (e.g., in OCD research; Grundler et al., 2009; Mathews et al., 2012; Nieuwenhuis et al., 2005). Lower reliability of a task means that more of the observed score is error – which could produce increases in both Types I and II errors. One possible explanation for these observed differences in reliability relates to variation in task difficulty and possible variations in task engagement. A curvilinear relationship between task engagement and difficulty has been observed, such that higher task engagement is associated with moderately difficult tasks (Gendolla, 1999). Assuming that task engagement has positive effects on the quality of the collected data and will enhance reliability, it may be that moderately difficult tasks like the Flanker task lead to more reliable data.

Across all tasks, results of this study add support to a growing body of research indicating that the ERN and Pe can be assessed reliably (Olvet & Hajcak, 2009b; Segalowitz et al., 2010; Weinberg & Hajcak, 2011), and that this is true in a variety of speeded response tasks. However, the present study also suggests that differences in reliability between tasks should be considered in the planning and interpretation of studies.

Convergent validity: When assessing construct validity, the same measures of error-related brain activity assessed in the same individuals using different tasks should demonstrate convergence, indicating a common neural and cognitive substrate. And indeed, the examinations of the MTMM revealed evidence for convergent validity of the ERN, ΔERN, Pe and ΔPe as indicated by the significant correlations *within* each measure across tasks. This is in line with results by Segalowitz et al. (2010) indicating correlations of around .5 between the ERN derived from the Flanker and Go/NoGo task in adolescents. Therefore, it appears that different tasks can be used to assess a common neural and cognitive process reflected in the ERN and Pe. The highest validity values were observed for ΔERN amplitudes, suggesting that the ERN difference score more precisely isolates common variance in error processing across tasks, and the ΔERN has higher convergent validity compared to the neural responses to errors alone (i.e., the ERN). Importantly, the association between ΔERNs across tasks were high, with correlations around $r = .65$ – similar in size to the observed split-half reliabilities in this study and the observed retest reliability indices for the ERN in other studies (Olvet & Hajcak, 2009b; Weinberg & Hajcak, 2011). Moreover, maximal correlations between ΔERN across tasks were observed at frontocentral electrode sites, consistent with the usual scalp distribution of the ΔERN. This suggests that the ΔERN is not only *largest* at FCz, but also that the ΔERN measured at FCz best reflects common error-related brain activity across tasks. This validates an already-widespread tendency in studies to focus on ΔERN amplitudes measured at FCz. It is worth noting, however, that the ΔPe compared to the Pe does not show the same advantages as the ΔERN compared to the ERN. That is, higher validity values were observed for the Pe relative to ΔPe suggesting higher trait-like properties of the positivity when it is not scored as a difference measure; indeed, scoring the Pe only on error trials is a practice fairly common in existing studies (e.g., Aarts & Pourtois, 2010; Boksem, Tops, Wester, Meijman, & Lorist, 2006;

Falkenstein et al., 2000; Nieuwenhuis et al., 2001; Ridderinkhof et al., 2009).

Nonetheless, the results also reveal that a considerable amount of variance in error-related brain activity can be attributed to task-specific influences. Partial correlations indicate that the ERNs elicited by two tasks share variance even when controlling for the amount of variance that can also be explained by a third task. This demonstrates independent relationships between ERNs assessed with different tasks, and that each ERN involves task-specific variance. This is in line with studies suggesting at least partial task-related dissociations in ERN amplitude (Grundler et al., 2009; Mathews et al., 2012; Nieuwenhuis et al., 2005; Olvet & Hajcak, 2009a). These task-related modulations in error-processes may be caused by differences in difficulty or error-type (e.g., failure in response inhibition, as in the Go/NoGo task, compared to slips due to interference, as in the Flanker task) as well as stimulus–response mapping between tasks (e.g., fully-determined, as in the Flanker task, compared to probabilistic as in a reinforcement learning task; Grundler et al., 2009; Nieuwenhuis et al., 2005). Thus, the present study provides evidence for both common error-processes across tasks, as well as task-dependent and unique influences on error-related brain activity.

Discriminant validity: With regard to discriminant validity, the present study investigated the relations between the ERN and Pe with the aim of examining the degree to which these measures are related – or if they are at least partially dissociable indicators of error-processing. In the present study, the observed interrelations *within* components – even across multiple tasks – were stronger than the relationships *between* different components assessed with either the same or different tasks, providing evidence for discriminant validity and separable processes. The current results contribute to research suggesting that ERN and Pe are partly dissociable components of error processing (Overbeek et al., 2005). However, similarities between the ERN and Pe (i.e., comparable variations across tasks, correlations between ERN and Pe) also suggest that they are not completely independent, consistent with evidence that they may originate from a shared neural network (Brazdil, Roman, Daniel, & Rektor, 2005; Debener et al., 2005; Eichele, Juvodden, Ullsperger, & Eichele, 2010). Moreover, in a broader context it seems important that these components are related and work in concert in order for error-processing to function well. The two components appear to reflect both common and unique variance in the error-processing system – they reflect overlapping but non-redundant information. One possibility is that the ERN reflects the early evaluation of conflict and that the subsequent Pe might index a slower and somewhat more elaborated response to errors that, like the P3, may reflect further attentional processing (Polich, 2007). This increased processing might underlie awareness of errors often associated with the Pe (Endrass et al., 2007; Hughes & Yeung, 2011; Nieuwenhuis et al., 2001).

4.3. Limitations

The present study has potential limitations. First, it is difficult to definitely interpret multiple correlations presented in the MTMM (i.e., whether the correlations across tasks are high enough to suggest a singular component). The interpretation of an MTMM is driven by a qualitative rule-based examination of the whole of the interrelations (Campbell & Fiske, 1959). Overall, these results support the convergent and discriminant validity of error-related ERPs. Further according to Cohen (1992) a correlation of 0.5 is considered to be large, 0.3 is moderate, and 0.1 is small. In this sense, the observed correlations between ΔERN across tasks are high. The Pe, ΔPe and ERN are moderately correlated across tasks. However, even the observed high correlations of around .65 only share approximately 40% of variance. This indicates that error-related ERPs across tasks have both shared and unique variance and they are clearly not redundant. Future studies might recruit larger samples using multiple tasks to allow for the use of more effective statistical tools for the evaluation of convergent and discriminant validity such as confirmatory factor analysis (see Strauss & Smith, 2009).

The present results also indicate task-specific effects on error-processing as an important source of variation. This leads to questions of whether different tasks might exhibit differential relationships to individual difference measures. Future research might address these issues by recruiting larger samples – to assess the relationship between individual difference measures and error-related brain activity across tasks. However, an association between the ERN and obsessive-compulsive symptoms has been shown with different tasks like the Stroop (Hajcak & Simons, 2002), Flanker (Gehring et al., 1993), Go/NoGo (Ruchsow et al., 2005) and Simon task (Hajcak, Franklin, Foa, & Simons, 2008) which suggests that the ERN does also share functional characteristics across tasks.

Anatomical differences are considered to be the most significant source of variation in ERPs between subjects (Luck, 2005). Therefore, it is possible that the observed correlations across tasks may be partially caused by anatomical convergence in the neural networks underlying ERPs across tasks within individuals, which in fact supports convergent validity. Moreover, the use of difference score measures (i.e., ΔERN and ΔPe) controls for interindividual differences in ERP size. Accordingly, the observed pattern of results cannot be attributed to interindividual variation in anatomy.

Finally, this study focused on construct validity. Future studies should also demonstrate criterion validity of the ERN via the inclusion of a clinical or external outcome variable (e.g. clinical symptoms). This is of particular interest for studies interested in the ERN as a trait-like biomarker that may indicate risk for psychopathology (e.g., Olvet & Hajcak, 2008; Riesel, Endrass, Kaufmann, & Kathmann, 2011; Weinberg et al., 2012). The first step toward identifying the ERN as a trait marker is establishing that it can be reliably and validly assessed. However, validity alone is not sufficient and does not guarantee its utility as a trait marker. To qualify as a useful biomarker, the predictive power and the specificity of ERN in relation to individual differences in personality or psychiatric disorders should be examined. Existing data is promising in this regard, but few, if any, studies have directly assessed the relationship between personality or psychopathology and the ERN derived from multiple tasks.

5. Conclusion

Given the extensive resources committed thus far to research neurobiological ERP markers of psychopathology and personality, surprisingly little attention has been paid to their psychometric properties. Yet evidence for good psychometric properties will be critical to the continued use of these markers. This study revealed evidence for construct validity of the ERN and Pe, further supporting the use of these components as psychophysiological trait markers. Enhanced ERN amplitudes have been observed for a range of disorders and personality features characterized by anxiety and negative affect, and it has been suggested that the ERN relates to individual differences in defensive reactivity following errors (Weinberg et al., 2012). In contrast to the ERN, results regarding variations of the Pe as a function of personality or psychopathology are less clear (Overbeek et al., 2005). Refining our understanding of how these ERPs relate to personality and psychopathology will be a critical direction for future research. In addition to results indicating overlapping variation in error-related brain activity across different tasks, task-specific effects on error-processes were demonstrated and may represent an important source of variation between studies. Forming a composite scale of error-related indicators assessed with different tasks may reduce task specificity and further improve psychometric properties. Further, by examining more ERP measures and adding theoretically related trait measures assessed with different methods (e.g., psychophysiological, questionnaires, imaging) we can further increase our understanding of the functional significance of these measures.

References

Aarts, K., & Pourtois, G. (2010). Anxiety not only increases, but also alters early error-monitoring functions. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(4), 479–492. http://dx.doi.org/10.3758/CABN.10.4.479

Arbel, Y., & Donchin, E. (2009). Parsing the componential structure of post-error ERPs: A principal component analysis of ERPs following errors. *Psychophysiology*, *46*(6), 1179–1189. PSYP857 [pii].

Arbel, Y., & Donchin, E. (2011). How large the sin? A study of the event related potentials elicited by errors of varying magnitude. *Psychophysiology*, *48*(12), 1611–1620. http://dx.doi.org/10.1111/j.1469-8986.2011.01264.x

Bates, A. T., Kiehl, K. A., Laurens, K. R., & Liddle, P. F. (2002). Error-related negativity and correct response negativity in schizophrenia. *Clinical Neurophysiology*, *113*(9), 1454–1463. S1388245702001542 [pii].

Boksem, M. A., Tops, M., Wester, A. E., Meijman, T., & Lorist, M. (2006). Error-related ERP components and individual differences in punishment and reward sensitivity. *Brain Research*, *1101*(1), 92–101.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. http://dx.doi.org/10.1037/0033-295x.108.3.624

Brazdil, M., Roman, R., Daniel, P., & Rektor, I. (2005). Intracerebral error-related negativity in a simple Go/NoGo task. *Journal of Psychophysiology*, *19*(4), 244–255. http://dx.doi.org/10.1027/0269-8803.19.4.244

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science*, *307*(5712), 1118–1121. http://dx.doi.org/10.1126/science.1105783, 307/5712/1118 [pii]

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Cavanagh, J. F., Grundler, T. O., Frank, M. J., & Allen, J. J. (2010). Altered cingulate sub-region activation accounts for task-related dissociation in ERN amplitude as a function of obsessive-compulsive symptoms. *Neuropsychologia*, *48*(7), 2098–2109. http://dx.doi.org/10.1016/j.neuropsychologia.2010.03.031

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience*, *25*(50), 11730–11737. http://dx.doi.org/10.1523/JNEUROSCI.3286-05.2005

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

Eichele, H., Juvodden, H. T., Ullsperger, M., & Eichele, T. (2010). Mal-adaptation of event-related EEG responses preceding performance errors. *Frontiers in Human Neuroscience*, *4* http://dx.doi.org/10.3389/fnhum.2010.00065, 65 [pii]

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*(6), 1714–1720. http://dx.doi.org/10.1111/j.1460-9568.2007.05785.x

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. (2004). Errors, conflicts, and the brain: A review of the contributions to the error conference, Dortmund 2003. *Journal of Psychophysiology*, *18*(4), 153–163 (Print Electronic).

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*(6), 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*(2–3), 87–107. S0301051199000319 [pii].

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

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.

Gendolla, G. H. E. (1999). Self-relevance of performance, task difficulty, and task engagement assessed as cardiovascular response. *Motivation and Emotion*, *23*(1), 45–66.

Gentsch, A., Ullsperger, P., & Ullsperger, M. (2009). Dissociable medial frontal negativities from a common monitoring system for self- and externally caused failure of goal achievement. *Neuroimage*, *47*(4), 2023–2030. http://dx.doi.org/10.1016/j.neuroimage.2009.05.064. S1053-8119(09)00589-8 [pii]

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484.

Grundler, T. O., Cavanagh, J. F., Figueroa, C. M., Frank, M. J., & Allen, J. J. (2009). Task-related dissociation in ERN amplitude as a function of obsessive-compulsive symptoms. *Neuropsychologia*, *47*(8–9), 1978–1987. http://dx.doi.org/10.1016/j.neuropsychologia.2009.03.010

Hajcak, G. (2012). What we've learned from mistakes: Insights from error-related brain activity. *Current Directions in Psychological Science*, *21*(2), 101–106. http://dx.doi.org/10.1177/0963721412436809

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*(1), 116. http://dx.doi.org/10.1176/appi.ajp.2007.07010143

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

Hoffmann, S., & Falkenstein, M. (2010). Independent component analysis of erroneous and correct responses suggests online response control. *Human Brain Mapping*, *31*(9), 1305–1315. http://dx.doi.org/10.1002/hbm.20937

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. http://dx.doi.org/10.1037/0033.4.679-295X109

Hughes, G., & Yeung, N. (2011). Dissociable correlates of response conflict and error awareness in error-related brain activity. *Neuropsychologia*, *49*(3), 405–415. http://dx.doi.org/10.1016/j.neuropsychologia.2010.11.036

Johannes, S., Wieringa, B. M., Nager, W., Muller-Vahl, K. R., Dengler, R., & Munte, T. F. (2002). Excessive action monitoring in Tourette syndrome. *Journal of Neurology*, *249*(8), 961–966. http://dx.doi.org/10.1007/s00415-002-0657-9

Kopp, B., Rist, F., & Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*, *33*(3), 282–294.

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

Luu, P., Tucker, D., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science*, *14*(1), 47–53.

Mathews, C. A., Perez, V. B., Delucchi, K. L., & Mathalon, D. H. (2012). Error-related negativity in individuals with obsessive-compulsive symptoms: Toward an understanding of hoarding behaviors. *Biological Psychology*, http://dx.doi.org/10.1016/j.biopsycho.2011.12.018

Nieuwenhuis, S., Nielen, M. M., Mol, N., Hajcak, G., & Veltman, D. J. (2005). Performance monitoring in obsessive-compulsive disorder. *Psychiatry Research*, *134*(2), 111–122. http://dx.doi.org/10.1016/j.psychres.2005.02.005

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. *Clinical Psychology Review*, *28*(8), 1343–1354. http://dx.doi.org/10.1016/jcpr200807003

Olvet, D. M., & Hajcak, G. (2009a). The effect of trial-to-trial feedback on the error-related negativity and its relationship with anxiety. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(4), 427.

Olvet, D. M., & Hajcak, G. (2009b). Reliability of error-related brain activity. *Brain research*, *1284*, 89–99. http://dx.doi.org/10.1016/j.brainres.2009.05.079

Olvet, D. M., & Hajcak, G. (2009c). The stability of error-related brain activity with increasing trials. *Psychophysiology*, *46*(5), 957–961. http://dx.doi.org/10.1111/j.1469-8986.2009.00848.x

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*(2), 215–233 (Print Electronic).

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. http://dx.doi.org/10.1016/j.clinph.2007.04.019

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*(4), 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*(5601), 2209–2211. http://dx.doi.org/10.1126/science.1076929 [1076929] [pii]

Ridderinkhof, K., Ramautar, J., & Wijnen, J. (2009). To P(E) or not to P(E): A P3-like ERP component reflecting the processing of response errors. *Psychophysiology*, http://dx.doi.org/10.1111/j.1469-8986.2009.00790.x. PSYP790 [pii]

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. http://dx.doi.org/10.1126/science.1100301

Riesel, A., Endrass, T., Kaufmann, C., & Kathmann, N. (2011). Overactive error-related brain activity as a candidate endophenotype for obsessive-compulsive disorder: Evidence from unaffected first-degree relatives. *American Journal of Psychiatry*, *168*(3), 317–324. http://dx.doi.org/10.1176/appi.ajp.2010.10030416

Ruchsow, M., Gron, G., Reuter, K., Spitzer, M., Hermle, L., & Kiefer, M. (2005). Error-related brain activity in patients with obsessive-compulsive disorder and in healthy controls. *Journal of Psychophysiology*, *19*(4), 298.

Segalowitz, S. J., Santesso, D. L., Murphy, T. I., Homan, D., Chantziantoniou, D. K., & Khan, S. (2010). Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology*, *47*(2), 260–270. http://dx.doi.org/10.1111/j.1469-8986.2009.00942.x

Simons, R. F. (2010). The way of our errors: Theme and variations. *Psychophysiology*, *47*(1), 1–14. http://dx.doi.org/10.1111/j.1469-8986.2009.00929.x

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, *5*, 1–25. http://dx.doi.org/10.1146/annurev.clinpsy.032408.153639

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the 'error negativity' specific to errors? *Biological Psychology*, *51*(2–3), 109–128. S0301051199000320 [pii].

Weinberg, A., & Hajcak, G. (2011). Longer term test–retest reliability of error-related brain activity. *Psychophysiology*, *48*(10), 1420–1425. http://dx.doi.org/10.1111/j1469-8986201101206

Weinberg, A., Olvet, D. M., & Hajcak, G. (2010). Increased error-related brain activity in generalized anxiety disorder. *Biological Psychology*, *85*, 472–480. http://dx.doi.org/10.1016/j.biopsycho.2010.09.011

Weinberg, A., Riesel, A., & Hajcak, G. (2012). Integrating multiple perspectives on error-related brain activity: ERN as a neurobehavioral trait. *Motivation and Emotion*, *36*, 84–100. http://dx.doi.org/10.1007/s11031-011-9269-y

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959. http://dx.doi.org/10.1037/00334.931-295x111