Acta Neuropsychiatrica

cambridge.org/neu

Reliable and efficient recording of the error-related negativity with a speeded Eriksen Flanker Task

Franziska Suchan, Juliane Kopf, Heike Althen, Andreas Reif and Michael M. Plichta

Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Heinrich-Hoffmann-Straße 10, 60528 Frankfurt, Germany

Original Article

Cite this article: Suchan F, Kopf J, Althen H, Reif A, Plichta MM. (2019) Reliable and efficient recording of the error-related negativity with a speeded Eriksen Flanker Task. *Acta Neuropsychiatrica* 31:135–142. doi: 10.1017/neu.2018.36

Received: 23 July 2018
Revised: 17 October 2018
Accepted: 15 November 2018
First published online: 18 December 2018

Key words:
electroencephalography; evoked potentials; reproducibility of results

Author for correspondence:
Michael M. Plichta, Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Heinrich-Hoffmann-Straße 10, 60528 Frankfurt, Germany. Tel: +49 (0)69 6301 5897; Fax: +49 (0)69 6301 81671; E-mail: michael.plichta@kgu.de

Abstract

Objective: There is accumulating evidence that the error-related negativity (ERN), an event-related potential elicited after erroneous actions, is altered in different psychiatric disorders and may help to guide treatment options. Thus, the ERN is a promising candidate as a psychiatric biomarker. Basic methodological requirements for a biomarker are that their measurements are standardised and reliable. The aim of the present study was to establish ERN acquisition in a reliable, time-efficient and patient-friendly way for use in clinical practice. *Methods:* Healthy subjects performed a speeded Eriksen Flanker Task that increases the number of errors. In a test–retest design ($N = 14$) with two sessions separated by 28 days we assessed the reliability of the ERN. To ensure external validity, we aimed to replicate previously reported correlation patterns of ERN amplitude with (A) number of errors and (B) negative affect. In order to optimise the clinical use of the task, we determined to which extent the task can be shortened while keeping reliability >0.80. *Results:* We found excellent reliability of the ERN (intraclass correlation coefficients = 0.806–0.947) and replicated ERN correlation patterns. The task can be halved to a patient-friendly length of 200 trials (recorded in 8 min) keeping reliability >0.80. *Conclusions:* The modified task provides reliable and efficient recording of the ERN, facilitating its use as a psychiatric biomarker.

Significant outcomes

- The modified Eriksen Flanker Task provides an error-related negativity (ERN) with excellent reliability; the task can be halved to a patient-friendly length of 200 trials.

Limitations

- Instant feedback does not allow for analysing feedback-related potentials; sample size although sufficient for detecting the ERN does not allow sub-analyses (e.g. gender effects).

Introduction

Distinguishing error from correctness is an essential requirement for learning progress (1). In order to understand the function of error-related brain activity, an event-related potential (ERP) has been investigated in several electroencephalography (EEG) studies: the ERN, a negative deflection appearing within 100 ms after an erroneous response that peaks in fronto-central midline recording sites (2,3). To elicit the ERN the Eriksen Flanker Task (4) is broadly used (5–8) which involves discriminating a central target symbol (e.g. an arrow) from surrounding distracting 'flanker' symbols. There is strong evidence that the ERN is generated in the anterior cingulate cortex (9–11), an area of the medial prefrontal cortex responsible for the integration of affective and cognitive information (12).

A similar, but smaller negative ERP can arise also after correct responses in the same time window and at the same recording sites as the ERN: the correct-related negativity (CRN) (8,13–15). It has been discussed whether the same process (15) or two different processes (16,17) underlie the ERN and CRN.

The function of the ERN is described in different models with regard to an error detection system (3), reinforcement learning (1) or general conflict-detection process (18). Recently, by application of a forward model it has been discovered (19), that the ERN is likely to reflect an error-prediction. This is in line with the predicted response-outcome (PRO) model (20), which interprets the ERN as a surprise signal caused by non-occurrence of a predicted event.

Several factors have been shown to influence the ERN. Particularly important is the performance of the individual subject: The higher the error rate, the lower the ERN amplitude (21,22). In addition, the structure of the task and the instruction are relevant: (a) using congruent stimuli (i.e. target and flanker arrows point to the same direction) leads to increased ERN compared to incongruent stimuli (14); (b) task instruction focusing on accuracy over speed leads to increased ERN (2,8); (c) the ERN scales with the availability of sensory information and the task goal (23).

Moreover, negative affect (24,25) and several psychiatric disorders (26–29) are related to the ERN amplitude. Recently, it has been demonstrated that the ERN can (a) predict the onset of internalising disorder (30) such as anxiety disorder during the adolescence (31,32), (b) provide evidence for therapy responsiveness (26,27,33–35) and (c) help to guide treatment decisions (36).

Particularly the latter case emphasises the clinical relevance of the ERN and makes it a promising candidate as a biomarker for psychiatric disorders. A basic requirement for a biomarker is the reliable measurement. Only a few studies have investigated ERN reliability by using different Eriksen Flanker Task variants and found intraclass correlation coefficients (ICCs) between 0.62 and 0.74 (5,7,37–39).

With the present study we seek to investigate test–retest reliability of the ERN by using a modified Eriksen Flanker Task with an adaptive reaction time (RT) deadline (40,41) in two measurement sessions separated by 28 days. The application of an adaptive RT deadline is intended to maximise reliability due to higher error rate (39) while ERN is significantly different from CRN amplitude and a potential decrease of ERN amplitude (21,22) is negligible. At the behavioural level it is expected that the accuracy data are constant across sessions due to the adaptive RT deadline, whereas RT is predicted to be faster in session 2 because of training effects (7). To ensure the validity of the modified Eriksen Flanker Task, we attempt to replicate known correlation patterns:(1) positive correlation of ERN amplitude with number of errors (21,22) and (2) negative correlation of ERN amplitude with negative affect, measured by the Positive and Negative Affect Schedule (PANAS) questionnaire (24). In order to optimise a potential future clinical use of the task, we determine whether the task can be shortened without significant loss in reliability.

Aims of the study

To quantify test–retest reliability of the ERN evoked by a modified Eriksen Flanker Task with an adaptive RT deadline. We seek to determine whether the task can be shortened without significant loss in reliability.

Materials and methods

Participants

For the pilot study $N = 12$ healthy participants were recruited to adjust task parameters. Two subjects had to be excluded from analyses due to technical problems. Power estimation for the main study was calculated based on the pilot study results. Using G*Power 3.1.9.2 we calculated a required sample size of $N = 11$ subjects given a statistical power of 0.80, $\alpha = 0.05$ (one-tailed) and an effect size of 0.83 for a *t*-test with dependent means. To compensate for drop-outs a new sample of $N = 15$ subjects was recruited for the main study. One subject had to be excluded from the main study due to technical problems. Finally, test–retest data from $N = 14$ subjects (9 F/5 M; mean age = 23.5 years, SD = 2.07 years, range = 20–28 years) were included for main analyses. All participants were tested for mental health by the Mini International Neuropsychiatric Interview (M.I.N.I.), German Version 5.0.0 (42). Exclusion criteria included current or preceding psychiatric diagnoses. We documented consumption of cigarettes, caffeine (including coffee, coke, or caffeinated tea) and alcohol before the first testing and requested the subjects to appear in a comparable condition for the second testing.

All participants were compensated for their participation and gave written informed consent after detailed explanation of the experimental procedure. The study was approved by the Ethics Committee of the University of Frankfurt and is in accordance with the latest version of the Declaration of Helsinki.

PANAS

The German version of the PANAS is a self-report measuring instrument of affect adapted by Krohne et al. (43) from the English language questionnaire PANAS (44). The questionnaire consists of 20 adjectives describing different emotions (see Supplementary Material). Ten adjectives each cover the dimensions positive affect and negative affect. Every item can be rated on a Five-Point Likert-Type Scale ranging from 1 'not at all' to 5 'extremely'. Subjects responded on the basis of their present mood. The sum scores representing negative and positive affect have adequate internal consistency, test–retest reliability, and convergent and discriminant validity (44).

Subjects completed the PANAS before the Eriksen Flanker Task started at both sessions. To ensure validity by replicating known correlation patterns, state affect was correlated with the ERN amplitude.

Modified Eriksen Flanker Task

Subjects performed a modified arrow version of the Eriksen Flanker Task (4) two times (session *t*1 and session *t*2) separated by exactly 28 days (Fig. 1) in a dimly illuminated room (subjects of the pilot study finished only session *t*1). Presentation software Version 18.1 (Neurobehavioral Systems Inc.) was used. The whole task included 411 trials, 12 exercise trials and 399 experimental trials. In order to force many errors only incongruent stimuli were included. On each trial five horizontally aligned arrows were shown in the middle of the monitor ('« > «' or '> > < > >' or '> < > < >' or '< > < > <') for 125 ms followed by a white

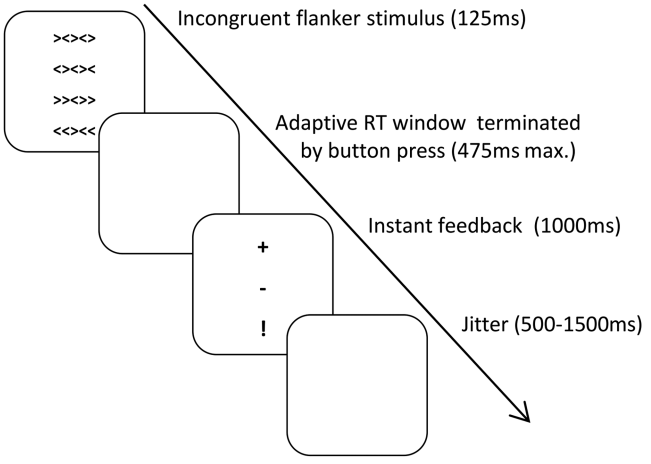

Fig. 1. Procedure of the Eriksen Flanker Task. RT, reaction time.

https://doi.org/10.1017/neu.2018.36 Published online by Cambridge University Press

screen during the RT deadline of maximal 475 ms. Each of the stimulus types was intended to be shown 100 times (due to a technical problem, '« > «' was only shown 99 times on both sessions and all subjects). The subject was instructed to respond as fast and accurate as possible with the right or left arrow key using his/her right index finger on a keyboard, congruent to the direction of the central arrow. Immediately after the button press, a feedback was presented: a plus (+) sign for correct answers, minus (−) for erroneous answers and exclamation point (!) was shown when the subject did not answer within the current RT deadline. In order to force quick answers, the RT deadline was adjusted after each trial by a reduction of 25 ms in case the subject reacted correctly within the current RT deadline or an extension by 25 ms in case the response took longer than the current RT deadline. Between each trial a white screen without fixation cross was shown for randomly 500–1500 ms.

EEG recording

The EEG was recorded using an elastic head cap with 64 scalp electrodes according to the international 10/20-System. Four additional electrodes were placed to record an electrooculogram, two close to each angulus oculi lateralis, one on the supercilium and one on the palpebrae inferioris. Ground electrode was placed between the FPz and Fz electrode, reference electrode between the Fz and Cz electrode. All signals were digitised with a 64-channel DC-amplifier and the software 'BrainVision Recorder' 2.0 (BrainProducts, Munich, Germany) with a sampling rate of 5000 Hz.

Data analysis

EEG data were analyzed using the software 'BrainVision analyzer' 2.0 (BrainProducts, Munich, Germany). First, electrode TP9 and TP10 were disabled, since they are placed on the mastoid and not used as reference. Data were band pass filtered with a low cutoff of 0.1 Hz, a high cutoff of 50 Hz and a notch filter of 50 Hz. Blinks and eye movements were corrected based on the method established by Gratton et al. (45). The algorithm corrects eye artefacts by subtracting the eye channel voltages multiplied by a channel-dependent corrective factor from the respective EEG channels.

Subsequently data were re-referenced on an average reference of all electrodes and the former reference was reused as channel FCz. The EEG was segmented response-locked with an entire length of 800 ms, with 400 ms pre- and post-response each. The automatic artifact rejection searched for values exceeding a difference of ±70 µV within 200 ms and excluded data 200 ms before and after the artefact. This procedure did not reveal any artefacts. Afterwards the segments were averaged separately into correct, error and missed trials and a window −400 to −200 ms before the response was used as baseline. The ERP components ERN and CRN were analyzed in terms of area and peak measures at electrode sites FCz and Cz. For area measures the mean activity in the interval 0–100 ms after response was calculated, for peak analysis automatic peak detection identified the largest negativity in the same interval.

In the process of our analysis it was necessary to evaluate the EEG data additionally stimulus-locked. The window −400 to −200 ms pre-stimulus was used as baseline and the average time course separated into correct and error as well as sessions *t1* and *t2* was calculated.

Statistical methods

For statistical calculations IBM SPSS statistics (version 22) and MATLAB R2017b (The Mathworks, Natick, MA, USA) was used.

In case of behavioural data we used Wilcoxon-test ($\alpha = 0.05$; two-tailed) due to non-normally distributed data as tested by Shapiro–Wilk tests. In order to analyse EEG data, we tested for Gaussian distribution by Shapiro–Wilk tests (all $ps > 0.42$) and calculated a 2×2 repeated measure ANOVA (analysis of variance) with factors (1) accuracy (CRN, ERN) and (2) sessions (*t1*, *t2*). *Posthoc* dependent *t*-tests ($\alpha = 0.05$; two-tailed) were performed in case of significant interaction effects.

Test–retest reliability was assessed by calculating ICC (1,2) for absolute agreement defined by Shrout and Fleiss (46) as:

$$\mathrm{ICC}(2,1) = \mathrm{BMS} - \mathrm{EMS} / (\mathrm{BMS} + (k-1) \times \mathrm{EMS} + k \times (\mathrm{JMS} - \mathrm{EMS}) / N)$$

BMS = between-subjects mean square; EMS = error mean square; JMS = session mean square (the original terminology of 'J' is 'Judge'); k = number of repeated sessions and N = number of subjects. Thus, in the current study, $k = 2$ and $N = 14$.

Following Shrout and Fleiss (46) we defined ICC values <0.4 as poor, 0.4–0.75 as fair to good and >0.75 as excellent. Negative ICC values were reset to 0 (47).

For correlation analyses of ERN amplitude, we calculated the correlation according to Spearman (one-tailed), since the scores of negative affect and number of errors were not Gaussian distributed.

Results

Behavioural results

Participants responded significantly faster in session *t2* compared to session *t1*, for both correct and error trials (Table 1). There was a significant effect on number of correct trials, but not on number of error and missed trials between sessions (Table 1). Across sessions the accuracy was consistent at a level of ~80% (see Supplementary Fig. 1).

Table 1. Performance data

	Session *t1*		Session *t2*		Session *t1*–*t2*		
	Median	IQR	Median	IQR	U	*p*	dz
RT correct trials (ms)	404	53	366	36	−3.296	<0.001	1.935
RT error trials (ms)	373	50	343	30	−3.233	<0.001	1.604
Number of correct trials	199	8	201	2	−2.280	0.021	0.565
Number of erroneous trials	68	53	81	55	−0.471	0.659	0.181
Number of missed trials	136	46	120	56	−0.031	0.985	0.017

IQR = interquartil range.

https://doi.org/10.1017/neu.2018.36 Published online by Cambridge University Press

Comparing CRN and ERN

Figure 2a shows response-locked ERPs for error and correct trials at FCz electrode averaged over all subjects and trials. As expected, there was a significant difference between CRN and ERN [peak amplitude measures: $F(1,13) = 16.673$, $p < 0.001$, $\eta p^2 = 0.562$; area measures: $F(1,13) = 10.008$, $p = 0.007$, $\eta p^2 = 0.435$] with more pronounced negativity for ERN versus CRN. For factor session, there was a significant effect [peak amplitude measures: $F(1,13) = 15.282$, $p = 0.002$, $\eta p^2 = 0.540$; area measures: $F(1,13) = 27.924$, $p < 0.001$, $\eta p^2 = 0.682$] with a more pronounced negativity for session 1 versus session 2. In addition, there was a significant interaction of accuracy and session for peak amplitude measures [$F(1,13) = 11.484$, $p = 0.005$; $\eta p^2 = 0.469$] but not for area measures. *Posthoc* *t*-tests revealed that the interaction resulted from a significant change in the CRN amplitude ($t = -4.270$, $p < 0.001$, $dz = 1.141$) while the ERN amplitude difference was not significant across sessions ($t = -1.841$, $p = 0.089$, $dz = 0.492$).

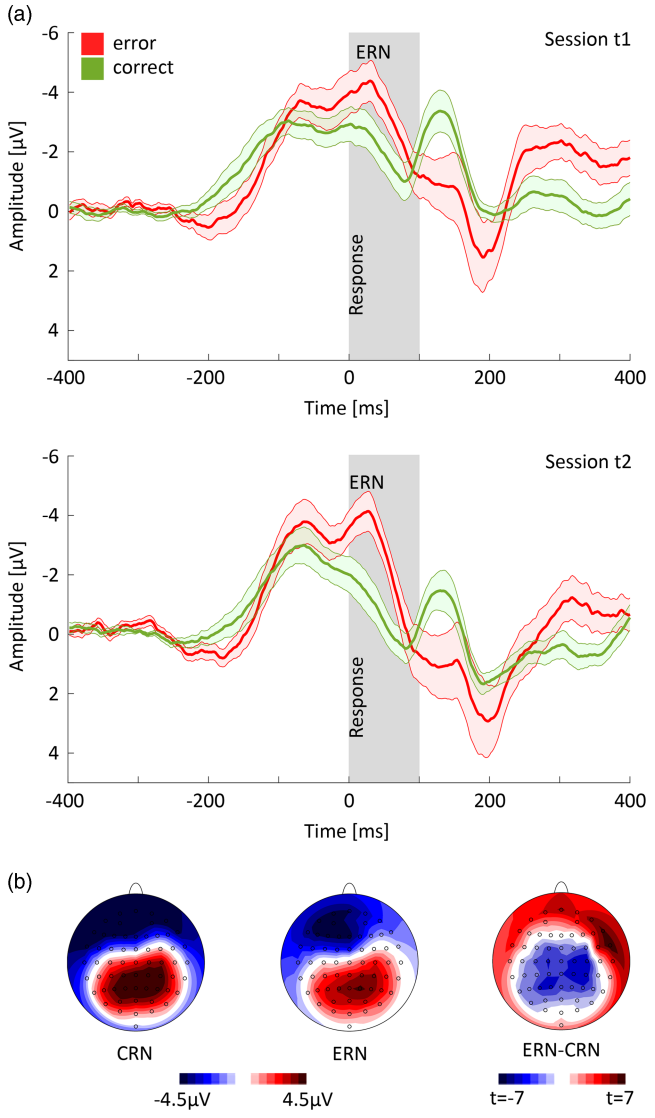

Fig. 2. (a) Response-locked time courses of correct and error trials at FCz electrode (± SE) for session *t*1 and *t*2. (b) Topographic mapping of correct-related negativity (CRN) and error-related negativity (ERN) and *t*-map of the difference ERN–CRN (area measure).

The topographies showed a more pronounced negativity in frontal areas for error compared to correct trials and the major difference between CRN and ERN in the central cortex (Fig. 2b).

Test–retest reliability

Table 2 shows test–retest reliability indices of ERP measures for error and correct trials at FCz and Cz electrode. Considering the FCz electrode, ICC_{ERN} was excellent (peak amplitude measures: ICC = 0.947, $p < 0.001$; area measures: ICC = 0.806, $p < 0.001$) and ICC_{CRN} was fair to good (peak amplitude measures: ICC = 0.747, $p < 0.001$; area measures: ICC = 0.675, $p < 0.001$). For peak amplitude measures the $ICC_{ERN-CRN}$ was excellent (ICC = 0.792, $p < 0.001$) and fair to good for area measures (ICC = 0.585, $p = 0.013$). On the contrary, peak latency measures were characterised by a low non-significant reliability of ERN (ICC = 0.143, $p = 0.290$) and CRN (ICC = 0.347, $p = 0.113$) but a moderate and significant reliability of ERN–CRN (ICC = 0.690, $p = 0.002$). For Cz electrode we found comparable results.

Validity

Spearman correlation for the ERN amplitude (FCz) with relative number of errors (Fig. 3a) revealed a trend to significance

Table 2. Test–retest reliability for error-related negativity (ERN) and correct-related negativity (CRN) at FCz and Cz electrode*

	ERP	ICC† (95% CI)	*p*
FCz electrode			
Peak amplitude measures	ERN	0.947 (0.832–0.983)	<0.001
	CRN	0.747 (0.030–0.930)	<0.001
	ERN–CRN	0.792 (0.226–0.939)	<0.001
Area measures	ERN	0.806 (0.052–0.950)	<0.001
	CRN	0.675 (0.032–0.899)	<0.001
	ERN–CRN	0.585 (0.092–0.846)	0.013
Peak latency measures	CRN	0.347 (0.000–0.736)	0.113
	ERN	0.143 (0.000–0.594)	0.290
	ERN–CRN	0.690 (0.295–0.887)	0.002
Cz electrode			
Peak amplitude measures	ERN	0.890 (0.679–0.964)	<0.001
	CRN	0.740 (0.281–0.914)	<0.001
	ERN-CRN	0.628 (0.195–0.861)	0.004
Area measures	ERN	0.826 (0.482–0.943)	<0.001
	CRN	0.800 (0.482–0.931)	<0.001
	ERN–CRN	0.743 (0.361–0.910)	<0.001
Peak latency measures	ERN	0.569 (0.055–0.840)	0.017
	CRN	0.572 (0.081–0.839)	0.014
	ERN–CRN	0.525 (0.012–0.819)	0.025

CI, confidence interval.
*Note that the intraclass correlation coefficients (ICCs) are comparable at C2 electrode where the difference between ERN and CRN was at maximum.
†ICC for absolute agreement.

https://doi.org/10.1017/neu.2018.36 Published online by Cambridge University Press

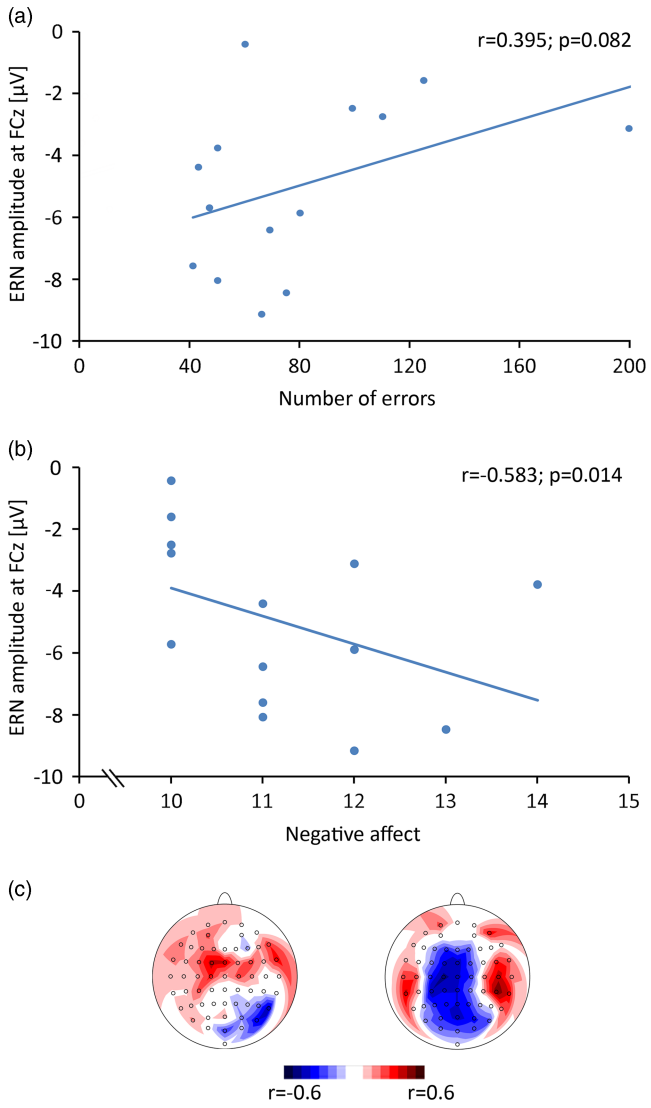

Fig. 3. (a) Correlation of error-related negativity (ERN) amplitude with absolute number of errors. (b) Correlation of ERN amplitude with negative affect, measured by the Positive and Negative Affect Schedule (PANAS) questionnaire. (c) Topographic mapping of correlation values of ERN amplitude with absolute number of errors (left) and negative affect (right).

($r = 0.394$; $p = 0.082$). The correlation of negative affect and ERN amplitude (Fig. 3b) reached significance ($r = -0.583$, $p = 0.014$). Topographic mappings of the correlations show that the absolute maxima were located at central electrodes (Fig. 3c).

A negative deflection preceding the ERN is noticeable in our response-locked time courses (Fig. 2a). To further examine this negative deflection, we analysed the EEG data stimulus-locked (see Supplementary Fig 2) and identified visual evoked potentials: the negative potential before the ERN and CRN is most likely the N200 which peaks at FCz electrode (correct: 292 ms ($t1$)/281 ms ($t2$); error: 291 ms ($t1$)/ 277 ms ($t2$) post-stimulus) (48).

Can the task be shortened?

Figure 4 shows ICC_{ERN} and ICC_{CRN} values with increasing number of included trials at FCz electrode. Analysing peak measures the ICC_{ERN} exceeded the threshold of >0.80 including 35 trials, for area measures 45 trials were required.

Analysing ICC_{CRN} values for peak measures at least 50 trials were required, for area measures the threshold was not exceeded.

Discussion

The overall objective of this study was to establish ERN acquisition in a reliable, time-efficient and patient-friendly way. Therefore, we used a modified Eriksen Flanker Task that increases the number of errors. To ensure external validity we aimed to replicate previously reported correlation patterns of ERN amplitude with number of errors and negative affect. In order to optimise the clinical use of the task, we determined to which extent the task can be shortened while keeping reliability >0.80. Overall, we (A) found excellent reliability of the ERN, which was >0.80 even when the task was reduced to halve of the trials and (B) ensured external validity of the ERN assessed by replicating previously reported correlation patterns with internal and external variables.

Reliability and effects of the adaptive RT deadline

Excellent reliability of the ERN was found. For peak measures, reliability is higher compared to other studies (5,7,37–39), with a 95% CI ranging from 0.832 to 0.983. A potential explanation for this is the adaptive RT deadline which produced about twice as many errors in comparison to other studies. For example, the subjects in the study of Larson et al. (39) made errors in 12% of the incongruent trials on average. In the studies of Weinberg and Hajcak (38) and Olvet and Hajcak (7) error rate was 11.97 and 11.34%. In our paradigm, however, error rate was 20%. (The data refers to the first session, but the second session is comparable.) An increasing number of error trials has been shown to increase the ERN reliability (39) and power (21,49).

In addition, the adaptive RT deadline counteracts a potential learning effect. According to the PRO model (20), the ERN amplitude changes with the likelihood of errors. When a subject performs the same task at two sessions, a learning effect arises and thereby a difference in likelihood for errors between the sessions. However, due to the adaptive RT deadline, the paradigm adapts to the performance level of the subject and the likelihood remains stable despite the learning effect. This may explain the excellent reliability of the ERN as found in our study.

A further advantage of the adaptive RT deadline is performance adjustment across groups. Several studies (21,22) demonstrated a negative relationship between number of errors and ERN amplitude. This can lead to biased results when comparing groups with different error rates (21). According to the PRO model (20), different performance levels, for example, in healthy controls and patients would lead to different subjects' expectations of making errors and thus may confound the ERN amplitudes. The adaptive RT deadline can reduce this potential bias because subjects would produce a comparable error rate.

However, there are also potential caveats: a high task performance can be defined not only by the error rate but also by the RT. According to the forward model (19,50) better task performance corresponds to more accurate forward model predictions about the performance outcome. This could lead to higher ERN amplitudes in subjects with faster RT. Therefore, differences in RT, for example, between patients and healthy controls might lead to biased ERN comparisons.

Finally, other studies generate sufficient number of errors by increasing task length [e.g. 900 trials (39)], while we achieved the high number of errors by a higher error rate. According to the

https://doi.org/10.1017/neu.2018.36 Published online by Cambridge University Press

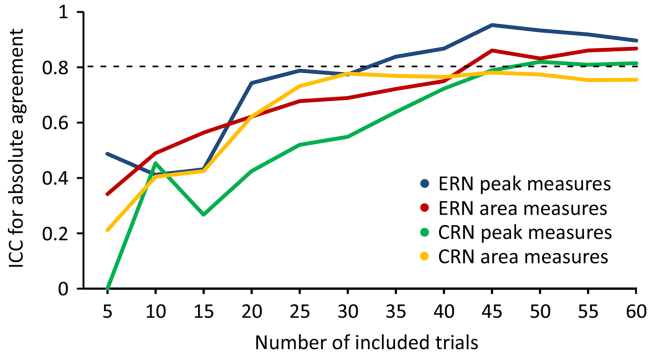

Fig. 4. Intraclass correlation coefficients (ICC) values of correct-related negativity (CRN) and error-related negativity (ERN) with increasing number of included trials at FCz electrode.

PRO model (20) and Fischer et al. (21) a smaller ERN amplitude is then expected. However, this ERN amplitude decrement seems to be negligible in our case since we have detected significant differences between ERN and CRN.

Validity

We found evidence for validity of the recorded ERN by partially replicating known correlation patterns: (A) a trend-wise positive correlation of ERN amplitude with number of errors (21,22) and (B) negative correlation of ERN amplitude with negative affect (24). In our study, the correlation between ERN amplitude and number of errors showed only a trend to significance. However, this is likely due to the small sample because our revealed effect size is in line with the reported values (21).

An additional aspect supporting the validity of the ERN is the topographic mapping: the ERN peaks in fronto-central midline recording sites as reported by previous studies (2,3). However, compared to other ERN studies a negative deflection preceding the ERN is noticeable in our response-locked time courses (Fig. 2a). To further examine this potential, we evaluated stimulus-locked time courses (see Supplementary Figure 2) and identified this negative deflection most likely as the N200. It has been shown in former studies that the N200 appears particularly on incongruent flanker stimuli (48).

Can the task be shortened?

To determine whether the task can be shortened without significant loss in reliability we analysed from which number of processed trials a reliability >0.80 (51) can be achieved. Our analyses showed that at least 35 error trials are necessary to achieve reliability >0.80 for peak amplitude measures of the ERN. For area measures, 45 error trials are required. A subject made 68 ($t1$) and 81 ($t2$) errors on average during the entire task. Therefore it can be concluded that a reduction of the paradigm to approximately half of the trials (=200) can equally ensure excellent reliability of ERN peak measures. Processing the whole task took on average 16.33 min. Thus, our paradigm can acquire highly reliable ERN within 8 min. This is advantageous in clinical practice as patients often have shorter concentration spans (52).

Limitations and recommendations for future studies

No comparison and reliability assessment of congruent versus incongruent trials could be conducted, because only incongruent stimuli were shown.

Moreover, our modifications of the Eriksen Flanker Task, that is using only incongruent stimuli and an adaptive RT deadline, might have influenced the ERN. For example, it has been shown that faster RTs are associated with larger ERNs (53) while higher error numbers (21,22) and incongruent stimuli (14) lead to reduced ERN amplitudes. In order to investigate these influences systematically, future studies should compare the ERN elicited by a flanker task variant with versus without these modifications.

The instant feedback does not allow for analysing feedback-related potentials (54). To achieve this, introducing a delay period between response and feedback would be necessary. Furthermore, contaminations of the response-locked ERP components by the visual feedback cannot be ruled out.

Although sufficient for detecting the ERN with a power >0.80, the current sample size does not allow any sub-analyses, for example, gender effects. Studies focusing on such effects should include larger sample sizes.

In order to use the ERN as a biomarker e.g. to control the course of an intervention (36) it is important to assess and interpret the ERN of a single subject (e.g. assignment into treatment type). However, measuring the ERN in single subjects usually is fairly difficult because of high variance due to diverse pre-analytical and analytical sources (55) that all have an potential impact on reliability. Future studies have to investigate further criteria for establishing the ERN (effects of sex, stress, age, pre-existing disease, medication effects, circadian rhythm, etc.) as a trans-diagnostic biomarker in particular within the Research Domain Criteria (56) matrix (57,58).

Finally, the task and its practicability should be evaluated in patients to examine feasibility and compare reliability.

Conclusion

The present study found an excellent reliability of the ERN acquired by a modified Eriksen Flanker Task with adaptive RT deadline with only 200 trials which is time-efficient and clinically feasible. Summarising, the present modified task provides a reliable and efficient recording of the ERN, which will facilitate its use in psychiatry.

Supplementary Material. To view supplementary material for this article, please visit https://doi.org/10.1017/neu.2018.36

Acknowledgements. The authors thank Lea Marie Schnetzler for her help. The authors gratefully acknowledge the subjects who participated. Authors' contributions: All authors (F.S., J.K., A.R., M.M.P.) substantially contributed to the conception and design of the study. F.S. recorded the data. F.S., H.A. and M.M.P. analysed and interpreted the data and wrote the manuscript. All authors critically revised the manuscript and gave their final approval of the version to be published.

Funding. This work was supported by the German Research Association (DFG) [SFB 1193 Z03]; BMBF BipoLife; EU Horizon 2020 [CoCA No 667302, MiND No 643051, Eat2BeNICE No 667302] and EU FP7 Aggressotype [No 602805].

Statement of Interest. The authors declare no financial, professional and personal relationships with the potential to bias the work.

Ethical Standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guides on the care and use of laboratory animals.

References

1. **Holroyd CB and Coles MGH** (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev* **109**, 679–709.
2. **Gehring G and Coles M** (1993) Donchin. A neural system for error detection and conpensation. *Psychol Sci* **4**, 385–390.
3. **Falkenstein H and Hoormann B** (1991) Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalogr Clin Neurophysiol* **78**, 447–455.
4. **Eriksen CW and Eriksen BA** (1979) Target redondancy in visual search: do repetitions of target within the display impair processing? *Percept Psychophys* **26**, 356–370.
5. **Cassidy SM, Robertson IH and O'Connell RG** (2012) Retest reliability of event-related potentials: evidence from a variety of paradigms. *Psychophysiology* **49**, 659–664.
6. **Ehlis AC, Herrmann MJ, Bernhard A and Fallgatter AJ** (2005) Monitoring of internal and external error signals. *J Psychophysiol* **19**, 263–269.
7. **Olvet DM and Hajcak G** (2009) Reliability of error-related brain activity. *Brain Res* **1284**, 89–99.
8. **Falkenstein H and Christ H** (2000) ERP components on reaction errors and their functional significance: a tutorial. *Biol Psychol* **51**, 87–107.
9. **Brázdil M, Roman R, Daniel P and Rektor I** (2005) Intracerebral error-related negativity in a simple Go/NoGo task. *J Psychophysiol* **19**, 244–255.
10. **Holroyd CB, Dien J and Coles MGH** (1998) Error-related scalp potentials elicited by hand and foot movements: evidence for an output-independent error-processing system in humans. *Neurosci Lett* **242**, 65–68.
11. **Luu P, Tucker DM and Makeig S** (2004) Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clin Neurophysiol* **115**, 1821–1835.
12. **Bush G, Luu P and Posner MI** (2000) Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn Sci* **4**, 215–222.
13. **Gehring WJ and Knight RT** (2000) Prefrontal-cingulate interactions in action monitoring. *Nat Neurosci* **3**, 516–520.
14. **Scheffers MK and Coles MGH** (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**, 141–151.
15. **Vidal F, Hasbroucq T, Grapperon J and Bonnet M** (2000) Is the 'error negativity' specific to errors? *Biol Psychol* **51**, 109–128.
16. **Coles MGH, Scheffers MK and Holroyd CB** (2001) Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biol Psychol* **56**, 173–189.
17. **Yordanova J, Falkenstein M, Hohnsbein J and Kolev V** (2004) Parallel systems of error processing in the brain. *Neuroimage.* **22**, 590–602.
18. **Yeung N, Botvinick MM and Cohen JD** (2004) The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol Rev* **111**, 931–959.
19. **Joch M, Hegele M, Maurer H, Müller H and Maurer LK** (2017) Brain negativity as an indicator of predictive error processing: the contribution of visual action effect monitoring. *J Neurophysiol* **118**, 486–495.
20. **Alexander WH and Brown JW** (2011) NIH public access. *Brain.* **14**, 1338–1344.
21. **Fischer AG, Klein TA and Ullsperger M** (2017) Comparing the error-related negativity across groups: the impact of error- and trial-number differences. *Psychophysiology* **54**, 998–1009.
22. **Hajcak G, McDonald N and Simons RF** (2003) To err is autonomic: error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology* **40**, 895–903.
23. **Brown JW and Braver TS** (2005) Learned predictions of error likelihood in the anterior cingulate cortex. *Science* **307**, 1118–1121.
24. **Hajcak G, McDonald N and Simons RF** (2004) Error-related psychophysiology and negative affect. *Brain Cogn* **56**, 189–197.
25. **Luu P, Collins P and Tucker DM** (2000) Mood, personality, and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *J Exp Psychol Gen* **129**, 43–60.
26. **Fissler M, Winnebeck E, Schroeter TA, Gummbersbach M, Huntenburg JM, Gärtner M and Barnhofer T** (2017) Brief training in mindfulness may normalize a blunted error-related negativity in chronically depressed patients. *Cogn Affect Behav Neurosci* **17**, 1164–1175.
27. **Rabella M, Grasa E, Corripio I, Romero S, Mañanas MÀ, Antonijoan RM, Münte TF, Pérez V and Riba J** (2016) Neurophysiological evidence of impaired self-monitoring in schizotypal personality disorder and its reversal by dopaminergic antagonism. *NeuroImage Clin* **11**, 770–779.
28. **Gehring WJ, Himle J and Nisenson LG** (2016) Action-monitoring dysfunction in obsessive-compulsive disorder. In: William JG, Joseph H and Laura GN. Sage Publications Inc. on behalf of the Association for Psychological Science Stable. Available at http://www11:p. 1–6.
29. **Meyer A, Hajcak G, Glenn CR, Kujawa AJ and Klein DN** (2017) Error-related brain activity is related to aversive potentiation of the startle response in children, but only the ern is associated with anxiety disorders. *Emotion* **17**, 487–496.
30. **Meyer A, Danielson CK, Danzig AP, Bhatia V, Black SR, Bromet E, Carlson G, Hajcak G, Kotov R and Klein DN** (2017) Neural biomarker and early temperament predict increased internalizing symptoms after a natural disaster. *J Am Acad Child Adolesc Psychiatry* **56**, 410–416.
31. **Meyer A** (2017) A biomarker of anxiety in children and adolescents: a review focusing on the error-related negativity (ERN) and anxiety across development. *Dev Cogn Neurosci* **27**, 58–68.
32. **Meyer A, Nelson B, Perlman G, Klein DN and Kotov R** (2018) A neural biomarker, the error-related negativity, predicts the first onset of generalized anxiety disorder in a large sample of adolescent females. *J Child Psychol Psychiatry* **59**, 1162–1170.
33. **Schroder HS, Moran TP and Moser JS** (2018) The effect of expressive writing on the error-related negativity among individuals with chronic worry. *Psychophysiology* **55**, e12990.
34. **Forster SE, Zirnheld P, Shekhar A, Steinhauer SR, O'Donnell BF and Hetrick WP** (2017) Event-related potentials reflect impaired temporal interval learning following haloperidol administration. *Psychopharmacology (Berl)* **234**, 2545–2562.
35. **Hobson NM, Bonk D and Inzlicht M** (2017) Rituals decrease the neural response to performance failure. *Peer J* **5**, e3363.
36. **Gorka SM, Burkhouse KL, Klumpp H, Kennedy AE, Afshar K, Francis J, Ajilore O, Mariouw S, Craske M, Langenecker S, Shankman SA and Luan Phan K** (2017) Error-related brain activity as a treatment moderator and index of symptom change during cognitive-behavioral therapy or selective serotonin reuptake inhibitors. *Neuropsychopharmacology* **43**, 1355–1363.
37. **Segalowitz SJ, Santesso DL, Murphy TI, Homan D, Chantziantoniou DK and Khan S** (2010) Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology* **47**, 260–270.
38. **Weinberg A and Hajcak G** (2011) Longer term test-retest reliability of error-related brain activity. *Psychophysiology.* **48**, 1420–1425.
39. **Larson MJ, Baldwin SA, Good DA and Fair JE** (2010) Temporal stability of the error-related negativity (ERN) and post-error positivity (Pe): the role of number of trials. *Psychophysiology* **47**, 1167–1171.
40. **Debener S** (2005) Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *J Neurosci* **25**, 11730–11737.
41. **Unger K, Heintz S and Kray J** (2012) Punishment sensitivity modulates the processing of negative feedback but not error-induced learning. *Front Hum Neurosci* **6**, 1–16.
42. **Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R and Dunbar GC** (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* **59**(Suppl. 20):22–33.
43. **Krohne H, Egloff B, Kohlmann C-W and Tausch A** (1996) Untersuchungen mit einer deutschen Version der 'Positive and Negative Affect Schedule' (PANAS). *Diagnostica* **42**, 139–156.
44. **Watson D, Clark LA and Tellegen A** (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Personal Soc Psychol* **54**, 1063–1070.
45. **Gratton G, Coles MGH and Donchin E** (1983) A new method for off-line removal of ocular artifact. *Electroencephalogr Clin Neurophysiol* **55**, 468–484.

https://doi.org/10.1017/neu.2018.36 Published online by Cambridge University Press

46. **Shrout PE and Fleiss JL** (1979) Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* **86**, 420–428.
47. **Bartko JJ** (1976) On various intraclass correlation reliability coefficients. *Psychol Bull* **83**, 762–765.
48. **Kopp B, Rist F and Mattler U** (1996) N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology* **33**, 282–294.
49. **Boudewyn MA, Luck SJ, Farrens JL and Kappenman ES** (2017) How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology* **55**, e13049.
50. **Joch M, Hegele M, Maurer H, Müller H and Maurer LK** (2018) Accuracy of motor error predictions for different sensory signals. *Front Psychol* **9**, 1–13.
51. **Rosaroso RC** (2015) Using reliability measures in test validation. *Eur Sci J* **11**, 1857–7881.
52. **American Psychiatric Association** (2013) Diagnostic and statistical manual of mental disorders. Arlington: APA, 991 pp.
53. **Quik EH** (2012) The somatotropic axis : effects on brain and cognitive functions. Netherlands: Uitgeverij BOXPress.
54. **Bismark AW, Hajcak G, Whitworth NM and Allen JJB** (2013) The role of outcome expectations in the generation of the feedback-related negativity. *Psychophysiology* **50**, 125–133.
55. **Micheel CM and Ball JR** (2010) Evaluation of biomarkers and surrogate endpoints in chronic disease.
56. **Carcone D and Ruocco AC** (2017) Six years of research on the National Institute of Mental Health's Research Domain Criteria (RDoC) initiative: a systematic review. *Front Cell Neurosci* **11**, 1–8.
57. **Ladouceur CD** (2016) The error-related negativity: a transdiagnostic marker of sustained threat? *Psychophysiology* **53**, 389–392.
58. **Weinberg A, Meyer A, Hale-Rude E, Perlman G, Kotov R, Klein DN and Hajcak G** (2016) Error-related negativity (ERN) and sustained threat: conceptual framework and empirical evaluation in an adolescent sample. *Psychophysiology* **53**, 372–385.

https://doi.org/10.1017/neu.2018.36 Published online by Cambridge University Press