Biological Psychology 156 (2020) 107969

Contents lists available at ScienceDirect

Biological Psychology

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

Neural correlates of error detection during complex response selection: Introduction of a novel eight-alternative response task

Jutta Stahl[a,*], André Mattes[a], Manuela Hundrieser[a], Kilian Kummer[a], Markus Mück[a], Eva Niessen[a,b,c], Elisa Porth[a], Yohana Siswandari[a], Peter Wolters[a], Sebastian Dummel[a]

[a] *Department of Individual Differences and Psychological Assessment, University of Cologne, Germany*
[b] *Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre Juelich, Germany*
[c] *Institute of Neuroscience & Psychology, University of Glasgow, United Kingdom*

ARTICLE INFO

Keywords:
Complex response task
Error processing
Response force
Memory errors
Impulsive errors

ABSTRACT

Error processing in complex decision tasks should be more difficult compared to a simple and commonly used two-choice task. We developed an eight-alternative response task (8ART), which allowed us to investigate different aspects of error detection. We analysed event-related potentials (ERP; $N = 30$). Interestingly, the response time moderated several findings. For example, only for fast responses, we observed the well-known effect of larger error negativity (N_e) in signalled and non-signalled errors compared to correct responses, but not for slow responses. We identified at least two different error sources due to post-experimental reports and certainty ratings: impulsive (fast) errors and (slow) memory errors. Interestingly, the participants were able to perform the task and to identify both, impulsive and memory errors successfully. Preliminary evidence indicated that early (N_e-related) error processing was not sensitive to memory errors but to impulsive errors, whereas the error positivity seemed to be sensitive to both error types.

1. Introduction

Instantaneous error detection is essential for all sorts of actions—-from a simple button press to complex action sequences—because it provides important information, for example, whether behavioural adaption is required or not. Since the early 1990s, when two error-related components of the event-related potential (ERP) were discovered, more than 1,200 studies investigated various research questions using these neural correlates. However, there is still a scientific debate about the functional meaning of the *error negativity* ($N_{e,}$ which peaks about 100 ms after error response onset and the N_e amplitude is usually larger than correct-response negativity, N_c; Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993) as well as the *error positivity* (P_e, which peaks about 300 ms after detected errors and the P_e amplitude is usually larger in detected compared to undetected errors as well as the amplitude of P_c in correct responses; Falkenstein et al., 1991). In brief, the N_e was discussed as the neural correlate reflecting a mismatch of response representations (Falkenstein et al., 1991), an ongoing response conflict (Yeung, Botvinick, & Cohen, 2004), a prediction error (Brown & Braver, 2005), response uncertainty (Scheffers & Coles, 2000), or reinforcement learning (Holroyd & Coles, 2002; Holroyd, Coles, & Nieuwenhuis, 2002). The P_e amplitude was supposed to be a correlate of error awareness (Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001), because the P_e was smaller for undetected errors than for detected errors. P_e also correlates with error evidence accumulation (Steinhauser & Yeung, 2010; 2012) and post-error slowing (Hajcak, McDonald, & Simons, 2003; Overbeek, Nieuwenhuis, & Ridderinkhof, 2005).

To reliably investigate error processing, a sufficient number of errors is essential; this has been achieved in different ways. The Flanker task is a commonly used paradigm in this research area; several hundred studies used either the original Flanker task with *letters* (e.g., HHSHH; Eriksen & Eriksen, 1974) or any modified version of it (*digits*, e.g., Stahl, 2010; *arrows*, Weinberg & Hajcak, 2011; *beverages cans*, Brion et al., 2018). Furthermore, go/no-go tasks (Niessen, Fink, Hoffmann, Weiss, & Stahl, 2017), Stroop tasks (Hajcak & Simons, 2002), and different kinds of perceptual discrimination tasks (e.g. Steinhauser & Yeung, 2010) were used. A key feature of these tasks is the unequivocal stimulus-response mapping, that is, for each stimulus there is only one correct response; this type of stimulus-response mapping is well suitable

---

[*] Corresponding author at: Department of Individual Differences and Psychological Assessment, University of Cologne, Pohligstr 1, 50969 Köln, Germany.
E-mail address: jutta.stahl@uni-koeln.de (J. Stahl).

https://doi.org/10.1016/j.biopsycho.2020.107969
Received 11 November 2019; Received in revised form 20 September 2020; Accepted 5 October 2020

for the study of error detection. When the stimulus-response mapping of a task is not unequivocal, however, it seems that successful instantaneous error detection might be well impossible. For example, using a probabilistic learning task with a varying stimulus-response mapping, (Kaczkurkin, 2013) observed no difference between the N_e and N_c (see also Holroyd & Coles, 2002). The importance of a clear and simple definition of a "correct" response for successful error detection was also demonstrated in tasks with varying stimulus visibility (Gibbons, Fritzsche, Bienert, Armbrecht, & Stahl, 2011) and in tasks with larger response-set sizes (Maier, Steinhauser, & Hübner, 2010). The N_e–N_c differences were smaller for a high task complexity (e.g., response set of eight) compared to the differences observed with lower task complexity (e.g., response set of two). Considering that in the more complex tasks the $N_{e/c}$-related process was not (or only marginally) sensitive to the actual response outcome, one might wonder whether this early process is part of successful error processing in more complex actions at all. Alternatively, the small or missing $N_{e/c}$ differences might be a result of a high response uncertainty (Scheffers & Coles, 2000) or an ongoing response conflict (Yeung et al., 2004) which may be present for both, errors and correct responses. Simultaneously ongoing processes might have affected both, the error detection process itself and the unequivocal assessment because of overlapping neural responses from different (neighbouring) neural sources (for a detailed discussion, see supplement). Despite the difficulty of finding clear evidence for neural correlates of fast error detection in studies using complex tasks, observation of everyday life actions with complex stimulus-response mappings—such as piano playing—suggests that immediate error detection should be possible. Interestingly, highly trained piano players showed instantaneous error detection even before the onset of an auditory signal of an incorrect tone (e.g., Ruiz, Jabusch, & Altenmüller, 2009). Thus, motor learning is an important aspect of error detection. For example, Beaulieu, Bourassa, Brisson, Jolicoeur, and de Beaumont (2014) demonstrated that the N_e was sensitive to learning of complex motor sequences. The authors reported a larger N_e-N_c difference in the last four blocks of the experiment as an indicator of learning, which was further supported by a larger N_e in trials of blocks with clear motor sequences compared to trials of blocks with random sequences (Rüsseler, Münte, & Wiswede, 2018).

1.1. Objective of the present study

The present study was designed to investigate error detection performance in a newly developed task, which required higher cognitive load, especially during response selection. We were interested in the following research questions: Is error detection possible in tasks with challenging response selection? And, if so, how does it affect the behavioural and neural error processing indicators? For a systematic investigation of these questions, we developed an eight-alternative response task (8ART). To provoke a sufficient number of error types, we relied on the general rationale of a Simon task (e.g. Simon and Rudell, 1967), which showed that spatial-identity incompatibility increased the number of errors (for details, see Method).

In the above-mentioned tasks (e.g. Flanker tasks, motor sequence tasks), error detection is often a more or less implicit process and sometimes requires external feedback to become aware (e.g. Van der Helden, Boksem, & Blom, 2010). However, to explicitly induce internal error detection, we asked participants to rate each response immediately after responding. This type of error detection assessment allowed differentiating errors that were signalled by participants as errors (signalled errors), and errors signalled as correct responses (non-signalled errors), as well as correct responses that were signalled as correct (signalled correct responses), and correct responses signalled as incorrect (non-signalled correct responses). It was often shown that the N_e amplitudes in signalled errors and non-signalled errors were larger compared to the N_c amplitude in signalled correct responses, whereas the two error types did not differ (e.g. Nieuwenhuis et al., 2001, but for review see also Wessel, 2012). In contrast, the P_e amplitude in non-signalled errors did not differ from the P_c amplitude in signalled correct responses, but the P_e amplitude in signalled errors was larger compared to the other two response types (e.g. Nieuwenhuis et al., 2001). Note that non-signalled correct responses are unfortunately rare events; hence, it is difficult to investigate them systematically (Wessel, 2012)

As mentioned above, a challenging aspect in tasks with a high load in response selection (for the participant and the investigation of error detection) might be the simultaneous activation of more than one motor command in a trial. This could be a result of fast error correction, response conflict, as well as uncertainty. Hitherto, many studies used computer keyboards to measure response time (RT), especially, when more than two response alternatives were investigated. In addition to the imprecise temporal assessment of these keyboards (from 11 to 73 ms delay; Shimizu, 2002), involuntary movements (e.g., side-slip from a key) and voluntary movements (e.g., error correction) cannot be controlled during response execution. To this end, we developed a new assessment tool—a keyboard with eight force-sensitive keys (for details see Method and Fig. 1)—which allows one to monitor the smallest intended or unintended movements of all eight fingers simultaneously in real time. In addition to the precise RT measurement, the keys further enabled us to measure the applied response force (RF) of each finger, which has been demonstrated to be an indicator of uncertainty (Mattes and Ulrich, 1997), time pressure (Jaśkowski et al., 2000), inhibition (Ko et al., 2012), and response conflict (Kantowitz, 1973). Bode and Stahl (2014) suggested that RF could be an indicator of error correction based on their finding of smaller RF on error trials compared to correct trials. We wanted to replicate this finding using the new task. The present approach has, in many aspects, a clear exploratory character. However, in a first step, we want to replicate some well-documented behavioural findings, $N_{e/c}$ findings (Gehring, Liu, Orr, & Carp, 2012), as well as $P_{e/c}$ findings (Wessel, 2012) using the new complex paradigm. Thus, we expected faster (Rabbitt, 1966) and less forceful (e.g., Bode & Stahl, 2014) responses for error trials compared to correct response trials. Furthermore, we predicted so-called post error slowing, that is slower responses in trials immediately following an error compared to responses following correct responses (Rabbitt, 1966). Finally, we predicted larger N_e for error trials (signalled and non-signalled) compared to N_c in signalled correct trials, and a larger P_e for signalled errors compared to P_e for non-signalled errors and P_c for correct responses.

2. Method

2.1. Participants

A power analysis suggested for a medium effect size of $\eta^2 = 0.06$, a Type-I error of .05 and an estimated power of .84 a sample of 30 participants. The data sets of 30 undergraduate students (17 females, 0 diverse, 13 males; age: 18–42 years; mean = 24.4 years, SD = 5.84 years) were used (participant recruitment via an online recruitment system; Elson & Bente, 2009). 27 participants were right-handed (mean handedness coefficient 65.3 ± 10.3, Oldfield, 1971). In total, 38 participants took part in the experiment; data from five participants were excluded from analyses because they did not show a sufficient number of non-signalled errors; three additional data sets were excluded due to technical problems (i.e., defect electrodes). The study was approved by the ethics committee of the German Psychological Association. Informed written consent was given by each participant.

2.2. Procedure

The participants were instructed to respond as fast and as accurate as possible to a stimulus presented on a computer screen (TFT, 22") by pressing one of eight force-sensitive keys (Fig. 1A). Eight signs (§, ⁋, &,?, #, %, @, +) served as target stimuli; each sign was assigned to the

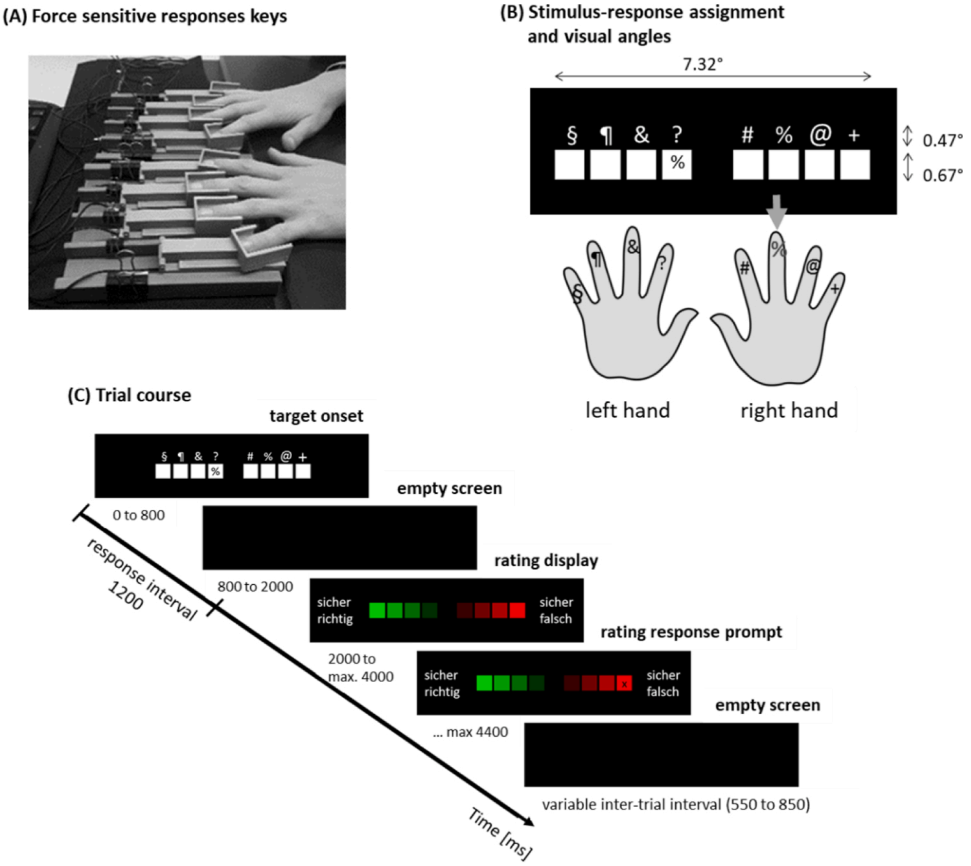

Fig. 1. (A) Photo of the adjustable force sensitive keys. (B) Stimulus-response assignment of the eight signs to the eight fingers and the stimuli's visual angles. (C) Trial course of the eight alternative response task (8ART) with an eight-point confidence rating.

response of one finger (thumbs excluded, see Fig. 1B for a detailed-finger assignment). A chinrest was used to maintain both, a constant distance to the screen (86 cm) and a stable posture of a participant while performing the task.

Each trial started with the presentation of the target display (see Fig. 1C and B for visual angles), showing eight white squares (representing the eight fingers) on a black background. To reduce memory load, the corresponding signs were depicted above the eight squares (see Fig. 1B). The target sign was presented within one of the eight white squares; participants were instructed to ignore the square position and to respond to the default sign position (e.g., '%' was assigned to the right middle finger, but presented on the position of the left index finger). The target display was presented for 800 ms followed by a black screen (presented for 2000 ms). The response interval ended 1200 ms after stimulus onset if RT exceeded 1200 ms, visual feedback was given by displaying the German word for "too slow" ("zu langsam") in red font and the trial was terminated since Stahl (2010) showed that RT errors could have an additional effect on $N_{e/c}$ amplitude. Following the response, an eight-point response evaluation rating (see Fig. 1C) was presented and participants were asked to indicate their subjective certainty about the accuracy of the given response. We used an 8-point scale with the poles "certainly correct" to "certainly incorrect"; as the main task was quite demanding, we neither wanted to distract the participants with a rating display which was very different from the target display nor confuse them with a very different way of giving the rating response. Thus, this scale was used to match the visual settings and motor requirements as similar to the main task as possible. The squares were coloured with different shades of green and red indicating the degree of certainty and the evaluation of accuracy. The scale orientation was randomly switched (i.e. in 50 % of the trials "certainly right" was presented at the left end and in 50 % at the right end) to avoid automatic replies without deeper processing of certainty. A cross indicated the participant's choice ("rating response prompt" in the rating display; cf. Fig. 1). The rating was not shown in trials with an RT > 1200 ms (i.e., too slow responses). The inter-trial interval was variable ranging from 550 ms to 850 ms (on average 750 ms).

The participants completed twelve blocks with 64 trials (each 5:30 min). Across trials, each target sign occurred equally often in each position; we presented target signs randomly to avoid anticipatory responses and effects of motor learning (e.g. Beaulieu et al., 2014; Rüsseler et al., 2018). The first block served as practice block and there was no time pressure; in this block, the target display remained on the screen until response onset, so that participants had enough time to practice the stimulus-response assignment. In case of less than 55 correct responses, a second practice block would have been performed (which was never required). A participant could start the next block after a 2 min-break.

2.3. Apparatus

Behavioural data were recorded using the eight custom-made force-sensitive keys. In each key, a force sensor (FCC221-0010-L, DigiKey MSP6948-ND) was implemented (with a range of 0–4448 cN). A VarioLab AD converter digitised the analogous response signal at a sampling rate of 1024 Hz with a resolution of 16 bits. Further attached to this converter is a photo sensor, which assessed the exact onset of a stimulus on the screen by change of brightness. To our knowledge, this is the first technical equipment which allows simultaneously assessing both stimuli and responses with eight different fingers in real time (i.e., without noticeable temporal delay or temporal jitter). The custom-made keys further allow adjusting the finger rests to individual hand and finger sizes as well as a comfortable distance between hands before the experiment starts. The forearms rest on a board in height of the keys in order to attain forearms and fingers on the same comfortable level aiming to reduce variations in hand position during the experiment. The keys are calibrated to the individual finger weight before the experiment. A response onset was registered when the applied force exceeded the threshold of 45 cN; thus, RT was defined as the interval between stimulus onset and this threshold. The median of the individual RTs was used as the within-subject RT distribution is positively skewed. RF was measured as the maximal applied force (i.e., peak RF). Behavioural adaptation was assessed by the pre-post-RT difference using the method suggested by Dutilh et al. (2012) and defined as

$RT_{pre\text{-}post} = RT_{n+1} – RT_{n\text{-}1}$.

RT differences between the pre-response RT ($RT_{n\text{-}1}$) and the post-response RT (RT_{n+1}) for each trial (n) were computed separately for signalled and non-signalled error trials as well as correct trials. Only correct trials before (n-1) and after (n+1) the responses of interest were included in the analyses.

2.4. *Electrophysiological data*

EEG signals were recorded from 63 scalp electrodes which were arranged according to the standard international 1020 system (Jasper, 1958) (FP1, FP2, AF7, AF3, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C3', C1, Cz, C2, C4, C4', C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, PO9, O1, Oz, O2, PO10). The active Ag/AgCl electrodes (Brain Products) were referenced against the two reference electrodes positioned at the mastoids (left active, right passive reference). The electrooculogram (EOG) signals were recorded using passive bipolar Ag/AgCl electrodes (ExG-Amplifier, Brain Products) above and below the left eye (vertical EOG), and at the left and right temples (horizontal EOG). The EEG signal was recorded continuously at a sampling rate of 500 Hz using a BrainAmp DC amplifier (Brain Products) with a filter from DC to 70 Hz.

For the pre-processing of the ERP data, an ocular correction was used to correct for blinks (Gratton, Coles, & Donchin, 1983). Then, the EEG data was divided into response-locked segments (100 ms before to 800 ms after response onset) separately for each condition. The 100 ms before stimulus onset served for baseline correction. An artefact rejection removed the trials in which the ERP waves exceeded ± 150 µV. We observed overlapping (positive and negative) activities from parietal areas in the classic ERPs (see supplement); thus, we performed a current source density (CSD) analysis to get EEG signals which were (a) independent from the references, and (b) cleared from overlapping activities from neighbouring electrode sites (Perrin et al., 1989). The analyses of the classic ERPs as well as a detailed discussion of the benefits and limitations of the CSD approach are presented in the supplement.

The post-response signalling was defined by the 8-point confidence rating. Because most participants used only the extreme ends of the scale, and in order to have a sufficient number of trials for the ERP analyses, we aggregated all trials with an 'error signalling' rating (i.e., four different red shades) and with a 'correct response' signalling rating (i.e., four different green shades). The number of trials differed between the conditions; however, we could show (see supplement) that this had no substantial effect on the shape of the components.

The $N_{e/c}$ peak was defined as the most negative amplitude at FCz in a time window relative to the response onset (0–150 ms); the $P_{e/c}$ peak was defined as the most positive amplitude at Cz (150–300 ms). The topographical maps (for details see Results) supported that for our new task, the local maximum of $N_{e/c}$ and $P_{e/c}$ were FCz and Cz, respectively.

2.5. *Statistical analyses*

In the first set of analyses, we performed several *t*-tests (response rates) and univariate ANOVAs with repeated measures for certainty ratings, RT, RF, $N_{e/c}$ amplitudes, and $P_{e/c}$ amplitudes to examine the main research question. The factor Response Type was three-fold (signalled error, non-signalled error, signalled correct). Unfortunately, the non-signalled correct response had to be excluded due to the small number of trials. Greenhouse-Geisser correction was applied when the sphericity assumption was violated. Effect sizes were generalized eta^2 (ANOVA) and Cohen's d. Tukey's HSD for within-comparison was used as post-hoc tests. Several exploratory two-way ANOVAs with repeated measures for the factors Response Type (signalled error, non-signalled error, signalled correct) and Speed (fast, slow; for a detailed explanation of this factor, see the Exploratory Analyses) were applied. Data and R script are online available: www.osf.io/b4cak.

3. Results

3.1. *Planned analyses*

The descriptive statistics (mean ± standard error of mean; SEM) are presented in Table 1 as a function of Response Type.

3.1.1. *Behavioural data*

3.1.1.1. Response rates. On average, participants responded in 67.4 ± 2.4 % of all trials correctly (signalled and non-signalled). After excluding 'too slow' responses (7.3 ± 0.9 %, RT > 1200 ms), the response rates were 15.4 ± 1.8 % of signalled errors, 5.6 ± 0.7 % of non-signalled errors, 77.2 ± 2.3 % of signalled correct responses, and 1.8 ± 0.2 % of non-signalled correct responses.

As seven of the 30 participants did not show a sufficient number of non-signalled correct responses (< 6 trials), for the following analyses, this trial type had to be excluded.

3.1.1.2. Certainty rating. The pooled rating values (i.e., from 1 "high certainty" to 4 "low certainty", irrespective of response type) differed significantly for Response Types, $F(1.37, 39.68) = 35.24$, $p < .001$, $\varepsilon = .68$, $\eta^2 = .32$. The post-hoc tests showed highest certainty for signalled correct responses (1.18 ± 0.06) followed by signalled errors (1.44 ± 0.08) and non- signalled errors (1.85 ± 0.08, for all comparisons $p < .007$).

3.1.1.3. Response time. The ANOVA yielded a significant effect for Response Type on the individual RTs, $F(1.67, 48.42) = 11.54$, $p < .001$, $\varepsilon = .83$, $\eta^2 = .09$. Post-hoc tests revealed that RT for signalled correct responses (748.4 ± 9.8 ms) was significantly shorter than the RT for both, non-signalled errors (784.3 ± 14.1 ms, $p = .002$), and signalled errors (793.5 ± 10.9 ms, $p < .001$). Non-signalled errors and signalled errors did not differ significantly, $p = .623$.

3.1.1.4. Response force. The peak RF was significantly affected by the Response Type, $F(1.04, 30.14) = 15.13$, $p< .001$, $\varepsilon = .52$, $\eta^2 = .11$. Non-signalled errors were, on average, less forceful (119.3 ± 13.3 cN) compared to signalled errors (221.5 ± 33.0 cN, $p = .005$) and signalled correct responses (291.0 ± 55.5 cN, $p< .001$). Signalled errors and signalled correct responses did not differ significantly, $p = .078$.

3.1.1.5. Behavioural adaptation. The RT difference ($RT_{pre\text{-}post}$) indicated more slowing after signalled errors (10.4 ± 6.1 ms) compared to signalled correct responses (−7.8 ± 2.6 ms), $t(29) = 2.78$, $p = .009$, $d = 0.51$. The $RT_{pre\text{-}post}$ difference for non-signalled errors showed also a slowing (13.1 ± 11.4 ms), which however did not differ significantly from signalled correct responses, $t(29) = 1.72$, $p = .096$, $d = 0.31$.

3.1.2. *Event-related potentials*

The CSD-transformed ERP waveforms and topographical maps are depicted in Fig. 2. The performed one-way ANOVA did not identify a

Table 1

Means and (±) standard error of means for all assessed behavioural and electrophysiological variables separately for each Response Type (signalled and non-signalled errors, and signalled correct responses), and Response Speed (fast, and slow responses; for details see Exploratory Analyses).

	All trials included			Fast responses			Slow responses		
	Signalled Errors	Non-signalled Errors	Signalled Correct	Signalled Errors	Non-signalled Errors	Signalled Correct	Signalled Errors	Non-signalled Errors	Signalled Correct
Response Rate [%][a,c]	15.4 ± 1.8	5.6 ± 0.7	77.2 ± 2.3	7.6 ± 0.9	2.8 ± 0.4	38.4 ± 1.2	7.8 ± 0.9	2.9 ± 0.4	38.7 ± 1.2
Response Time [ms]	793.5 ± 10.9	784.3 ± 14.1	748.4 ± 9.8	704.8 ± 10.3	683.9 ± 13.3	665.8 ± 8.5	886.4 ± 12.6	878.5 ± 13.6	856.1 ± 10.2
Peak Response Force [cN]	221.5 ± 33.0	119.3 ± 13.3	291.0 ± 55.5	219.0 ± 37.2	106.4 ± 8.8	263.0 ± 51.4	223.8 ± 29.3	131.9 ± 19.0	318.7 ± 59.6
Certainty Rating [1–4][b]	1.44 ± 0.08	1.85 ± 0.08	1.18 ± 0.06	1.43 ± 0.08	1.76 ± 0.09	1.14 ± 0.05	1.46 ± 0.09	1.92 ± 0.09	1.22 ± 0.06
$N_{e/c}$ Amplitude [μV/cm^2]	−0.76 ± 0.19	−0.99 ± 0.27	−0.79 ± 0.22	−0.82 ± 0.17	−0.93 ± 0.18	−0.55 ± 0.13	−0.72 ± 0.15	−0.80 ± 0.16	−0.71 ± 0.19
$P_{e/c}$ Amplitude [μV/cm^2]	0.98 ± 0.22	0.38 ± 0.22	0.53 ± 0.18	1.15 ± 0.23	0.47 ± 0.26	0.64 ± 0.19	0.91 ± 0.22	0.40 ± 0.25	0.52 ± 0.18

Notes. [a] relative to all responses within response time limit, [b] certainty values ranging from 1 = very certain to 4 = very uncertain (pooled values, i.e., irrespective of Response Type), [c] note that non-signalled correct were excluded (1.8 ± 0.2 %).

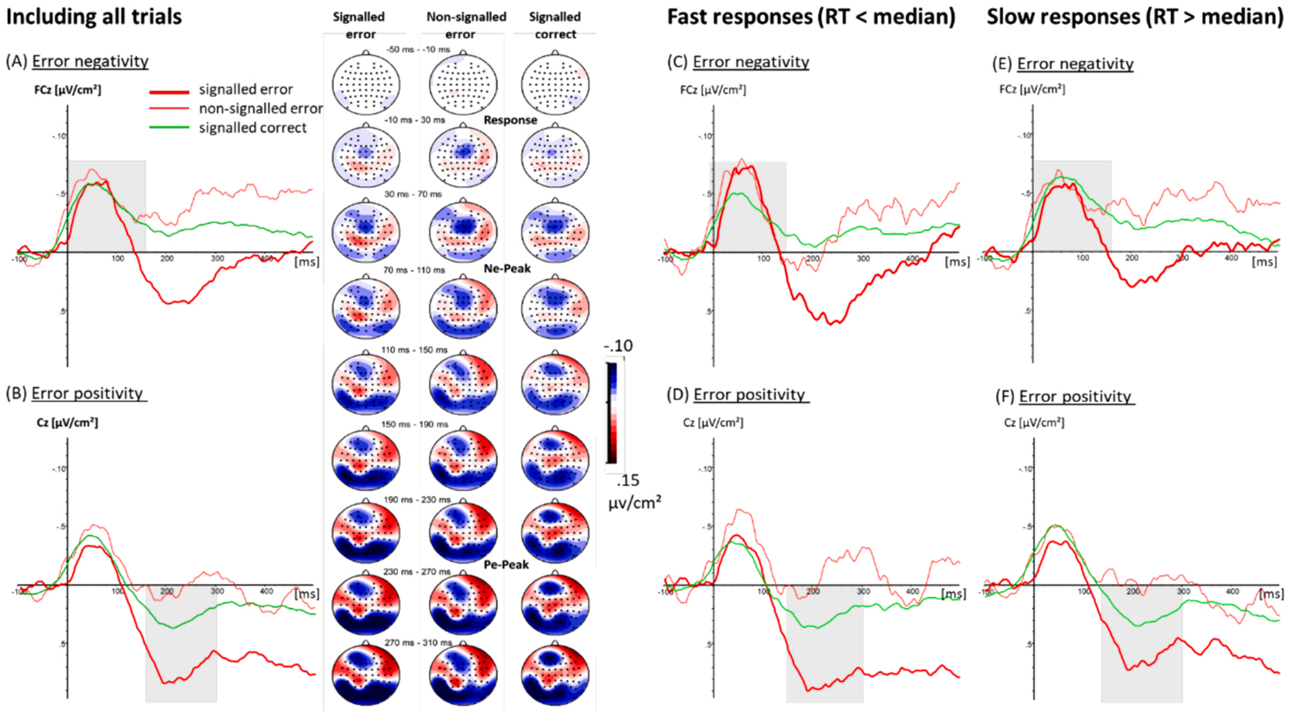

Fig. 2. Event-related potentials (A–F) and topographical maps separately for signalled errors, non-signalled errors and correct responses after current source density transformation for all responses (A, B), as well as for fast responses (C, D) and slow responses (E, F) after a median split. The grey shaded area indicates the search interval for the error negativity peak at FCz (A, C, E) and the error positivity peak at Cz (B, D, F).

Response Type effect on $N_{e/c}$ amplitude, $F(1.54, 44.66) = 1.66$, $p = .205$, $\varepsilon = .77$, $\eta^2 = .01$ (see Table 1). However, for $P_{e/c}$ amplitude a Response Type effect was revealed, $F(2, 58) = 8.54$, $p < .001$, $\varepsilon = .87$, $\eta^2 = .05$. Signalled errors (0.98 ± 0.22 μV/cm^2) showed a significantly larger amplitude compared to the other Response Types (non-signalled errors: 0.38 ± 0.22 μV/cm^2, $p < .001$; signalled correct responses: 0.53 ± 0.18 μV/cm^2, $p = .012$), which did not differ, $p = .567$.

3.2. *Summary and preliminary discussion of the main research questions*

The set of planned analyses showed that the participants were able to perform the task surprisingly well given the high complexity of the task; this was indicated by a large number of correct responses overall, a high error detection rate, post-error slowing and the participants' high level of confidence in their responses. In contrast to the literature investigating errors in tasks with two response-alternatives, errors were not faster but slower compared to correct responses in our task. Interestingly, Maier et al. (2010) could also not replicate shorter RTs for errors in their 8-response condition either It is possible that slower errors, on average, might be the result of different error sources due to the complexity of the response selection or the weakness in fast retrieval of S-R mapping, even though it was presented on the screen for the entire trial duration. Due to the eight stimuli and eight different responses, the probability of correct fast guesses was lower (i.e. 12.5 %; present task and Maier et al., 2010) compared to previously used two choice tasks (i.e. 50 %; e.g., Falkenstein et al., 1990). Nevertheless, we replicated that errors were less forceful (Bode & Stahl, 2014); the authors assumed that a smaller peak RF of errors could be an indicator of early error inhibition, starting already before response onset.

Whereas the $P_{e/c}$ amplitudes showed the expected and well-documented effect (i.e., a P_e occurred only in signalled error trials; e.g. Nieuwenhuis et al., 2001), the $N_{e/c}$ peak amplitudes did not vary between the three response types. This result was against our expectations (i.e., larger N_e compared to N_c; for review, see Gehring et al., 2012). Unfortunately, our study does not allow us to differentiate whether N_c is increased (e.g., due to higher response conflict or uncertainty), or N_e (signalled and non-signalled) is decreased (e.g., related to weaker error detection). To learn more about the underlying mechanisms and to elucidate the unexpected N_e finding, we performed several exploratory analyses to identify qualitatively different responses.

3.3. Exploratory analyses

3.3.1. Response subtypes

The present paradigm allowed us to differentiate several response subtypes, which we briefly consider next. Firstly, position errors are reflected in a response to the stimulus position instead of the stimulus identity (14.5 ± 1.8 % of all error trials; 74.5 ± 3.9 % of the position errors were signalled). Signalled errors and non-signalled errors did not differ in the percentage of position errors, $t(29) = 1.66$, $p = .107$, $d = .30$ (see also, Table 2). Interestingly, although the task was designed to evoke a large number of position errors due to the spatial-identity incompatibility similar to a Simon task (e.g. Simon, Sly, & Vilapakkam, 1981), neighbour errors occurred more often: These are errors that were committed by the direct neighbouring fingers to the correct finger (e.g. the middle finger would have been correct, but the ring finger was used). 63.5 ± 2.6 % of all errors were neighbour errors (67.4 ± 2.5 % were signalled). Considering signalled errors and non-signalled errors separately, more neighbour errors occurred in non-signalled errors (73.0 ± 2.3 %) compared to signalled errors (59.8 ± 2.9 %), $t(29) = 5.62$, $p < .001$, $d = 1.03$. Third, our sensitive keys allowed identifying *multiple responses*, which are trials where the participants pressed more than the first registered key (6.6 ± 0.6 % of all trials).

The probability of multiple responses differed significantly between Response Types, $F(1.59, 46.10) = 266.80$, $p < .001$, $\varepsilon = 0.79$, $\eta^2 = .76$. Post-hoc tests showed that the lowest probability of multiple responses was in signalled correct trials (11.2 ± 3.5 %), followed by signalled error trials (29.4 ± 3.0 %) and non-signalled error trials (80.2 ± 2.7 %, for all comparisons $p < .001$).

For multiple responses in error trials, it is important to note that the second responses were most often correct responses (error signalled: 71.1 ± 3.6 %; error non-signalled: 94.5 ± 1.4 %), indicating that participants attempted to correct their first, incorrect response. The correction response followed on average 123.0 ± 6.4 ms after the first response. As mentioned above, Bode and Stahl (2014) suggested that error correction would be marked by a lower RF in error trials compared to correct trials. We tested this assumption by comparing the RF on error trials where the error was immediately followed by a correction with error trials that were not followed by a correction. For signalled errors, the difference in RF between corrected (199.7 ± 34.0 cN) and non-corrected errors (792.7 ± 11.6 cN) was significant, $t(29) = 17.13$, $p < .001$, $d = 3.13$. For non-signalled errors, the difference in RF between corrected (98.6 ± 12.6 cN) and non-corrected errors (758.1 ± 23.8 cN) was also significant, $t(29) = 22.04$, $p < .001$, $d = 4.02$, supporting the previously established hypothesis.

3.3.2. Fast vs. slow responses

Although the stimulus-response assignment was visible during stimulus presentation, memorising the mapping would have been beneficial for a successful task fulfilment in time. In a post-experimental inquiry, participants reported specific response difficulties and to have used different strategies in response selection. For example, they reported that they had problems differentiating the used fingers (e.g. neighbouring fingers). In other trials, they just pressed any key after a while to be in time, even though they were sure that their response would be incorrect. In this case, a signalled error has a different meaning compared to error trials where participants had a clear representation of the correct response. Thus, we assumed that a slow signalled error could likely represent a memory error more (for a similar discussion, see Coleman, Watson, & Strayer, 2018). On the other hand, sometimes participants clearly realised the use of the incorrect key, being aware which response would actually be the correct one; in this case, the error, which is also signalled, would not result from memory failure but rather from premature response activation due to a lack of focussed attention. Thus, a *fast* signalled error might reflect an impulsive error, considering that short RT is a well-known indicator for impulsive responses (e.g. Leue and Beauducel, 2008). Therefore, an RT-based distinction may help to shed further light on different response subtypes. For the present data, fast and slow responses were defined by individual median splits for each of the three response types (correct signalled, error signalled, error non-signalled). This approach resulted in an equally distributed trial number for fast and slow responses (see Table 1), which was important to have a sufficient number of fast and slow error trials.

3.4. Behavioural data

3.4.1. Response subtypes

A two-by-two ANOVA for position errors did not obtain a Response Type effect, $F(1, 29) = 2.53$, $p = .122$, $\varepsilon = 1.0$, $\eta^2 = .01$, but a significant Speed effect, $F(1, 29) = 12.12$, $p = .002$, $\eta^2 = .06$ (fast: 14.6 ± 1.7 %, slow: 8.9 ± 1.2 %) and a significant interaction was shown, $F(1, 29) = 5.57$, $p = .025$, $\varepsilon = 1.0$, $\eta^2 = .02$. Position errors were more often in fast (17.3 ± 2.5 %) compared to slow (8.5 ± 1.5 %) signalled errors, $p < .001$, whereas no effect was shown for non-signalled errors (fast: 11.9 ± 2.1 %; slow: 9.3 ± 2.0 %, $p = .592$).

A two-by-two ANOVA for neighbour errors showed a significant Response Type effect, $F(1, 29) = 40.38$, $p < .001$, $\varepsilon = 1.0$, $\eta^2 = .17$ (signalled: 59.9 ± 2.2 %, non-signalled: 74.5 ± 2.0 %) but no effect of Speed, $F(1, 29) = 2.55$, $p = .121$, $\eta^2 = .01$. The significant interaction, $F(1, 29) = 9.35$, $p = .005$, $\varepsilon = 1.0$, $\eta^2 = .03$, indicated no significant difference between fast (60.9 ± 3.2 %) and slow (58.8 ± 3.2 %) signalled errors in neighbour errors, $p = .871$, but less neighbour errors were shown in fast non-signalled errors (70.1 ± 2.9 %) compared to slow non-signalled errors (78.9 ± 2.6 %, $p = .012$). Considering the error types

Table 2

Means and (±) standard error of means for all assessed behavioural and electrophysiological variables separately for each Response Type and Response Speed.

	All trials included			Fast responses			Slow responses		
	Signalled Errors	Non-signalled Errors	Signalled Correct	Signalled Errors	Non-signalled Errors	Signalled Correct	Signalled Errors	Non-signalled Errors	Signalled Correct
Position errors [%]	12.9 ± 1.8[a]	10.5 ± 1.7[b]	n.a.	17.3 ± 2.5[d]	11.9 ± 2.1[e]	n.a.	8.5 ± 1.5[g]	9.3 ± 2.0[h]	n.a.
Neighbour errors [%]	59.8 ± 2.9[a]	73.0 ± 2.3[b]	n.a.	60.9 ± 3.2[d]	70.1 ± 2.9[e]	n.a.	58.8 ± 3.2[g]	78.9 ± 2.6[h]	n.a.
Multiple response trials [%]	29.4 ± 3.0[a]	80.2 ± 2.7[b]	11.2 ± 3.5[c]	30.4 ± 3.4[d]	81.9 ± 2.8[e]	10.6 ± 3.5[f]	28.4 ± 3.1[g]	78.4 ± 3.3[h]	11.8 ± 3.4[i]

Notes. relative to *all* [a]signalled errors, [b]non-signalled errors and [c]signalled correct responses; relative to [d]fast or [g]slow signalled errors, [e]fast or [h]slow non-signalled errors and [d]fast or [i]slow signalled correct response; n.a. not applicable.

from a different perspective, 78.0 ± 3.7 % of the fast position errors and 73.0 ± 5.5 % of the slow position errors were signalled as errors. 69.0 ± 2.4 % of the fast neighbour errors and 65.9 ± 2.8 of the slow neighbour errors were signalled as errors.

The two-by-three ANOVA for multiple responses identified only a significant effect of Response Type, $F(1.58, 45.85) = 261.42, p < .001, \varepsilon = 0.79, \eta^2 = .73$ (mirroring the effect reported above), however, there was no further significant effect (Speed: $F(1, 29) = 0.92, p = .346, \eta^2 < .01$; Interaction: $F(2, 58) = 1.15, p = .324, \varepsilon = .86, \eta^2 < .01$).

3.4.2. *Response time*

The ANOVA for RT showed again a Response Type effect, $F(1.60, 46.53) = 8.12, p = .002, \varepsilon = 0.80, \eta^2 = .05$ (mirroring the effect in the planned tests), and a Speed effect, $F(1, 29) = 1305.25, p < .001, \eta^2 = .70$ (fast: 684.8 ± 6.5 ms, slow: 873.7 ± 7.1 ms), whereas no interaction was shown, $F(1.46, 42.32) = 0.99, p = .356, \varepsilon = 0.73, \eta^2 < .01$.

3.4.3. *Response force*

The ANOVA for peak RF showed a Response Type effect, $F(1.04, 30.14) = 15.17, p < .001, \varepsilon = 0.52, \eta^2 = .10$ (mirroring the effect in the planned tests), a Speed effect, $F(1, 29) = 21.83, p < .001, \eta^2 = .01$ (fast: 196.1 ± 22.3 cN, slow: 224.8 ± 24.2 cN), and an interaction, $F(1.19, 34.52) = 4.25, p = .040, \varepsilon = 0.60, \eta^2 < .01$. The post-hoc tests showed that the interaction can be explained by less forceful responses for signalled fast correct responses (263.0 ± 51.4 cN) compared to signalled slow correct responses (318.7 ± 59.6 cN, $p < .001$), whereas the two error types did not differ between fast and slow responses (see Table 1, both $p > .267$).

3.4.4. *Certainty rating*

The certainty rating values differed significantly for Response Types, $F(1.39, 40.26) = 36.51, p < .001, \varepsilon = .69, \eta^2 = .28$ (mirroring the effect in the planned tests). We also found a Speed effect, $F(1, 29) = 8.51, p = .007, \eta^2 = .01$. On average, the ratings indicated a higher certainty for fast responses (1.45 ± 0.05) compared to slow responses (1.53 ± 0.06; note, 1 = very certain), but no significant interaction, $F(1.26, 36.52) = 1.10, p = .317, \varepsilon = .63, \eta^2 < .01$.

3.5. *Event-related potentials*

After an RT-based median-split for the CSD-transformed ERPs, the pattern of results clearly differed between fast and slow responses (see Fig. 2, and Table 1). A significant effect of Response Type was shown for the $N_{e/c}$ amplitude, $F(2, 58) = 4.09, p = .022, \varepsilon = 0.89, \eta^2 = .01$ (mirroring the effect in the planned tests, see above). No Speed effect on $N_{e/c}$ amplitude was observed, $F(1, 29) = 0.06, p = .803, \eta^2 < .01$. The Response Type-by-Speed interaction was significant for $N_{e/c}$ amplitude, $F(2, 58) = 3.72, p = .030, \varepsilon = 0.97, \eta^2 = .01$. Post-hoc tests revealed that only in the fast signalled-correct trials the N_c amplitude (-0.55 ± 0.13 $\mu V/cm^2$) was smaller compared to N_e amplitudes in fast signalled errors (-0.82 ± 0.17 $\mu V/cm^2$, marginally significant, $p = .086$), and in fast non-signalled errors (-0.93 ± 0.18 $\mu V/cm^2$, $p < .004$), whereas for the slow response types no significant difference in $N_{e/c}$ amplitudes was observed (all $ps > .965$).

The Response Type had a significant effect on $P_{e/c}$ amplitude, $F(1.65, 47.74) = 5.83, p = .008, \varepsilon = 0.82, \eta^2 = .04$ (mirroring the effect in the planned tests, see above). The $P_{e/c}$ amplitude was also significantly affected by Speed, $F(1, 29) = 6.03, p = .020, \eta^2 < .01$, with smaller amplitudes for slow responses compared to fast ones (fast: 0.75 ± 0.14 $\mu V/cm^2$; slow: 0.61 ± 0.13 $\mu V/cm^2$). The Response Type by Speed interaction was not significant for $P_{e/c}$ amplitude, $F(2, 58) = 0.94, p = .397, \varepsilon = 0.84, \eta^2 < .01$.

In order to test whether the time spent on the task had an impact on our findings, we tested error rate, RT and RF block-wise, and $N_{e/c}$ and $P_{e/c}$ in the first and second half of the experiment. Only the first block was slower compared to the other 10 blocks. However, there was no further difference in RT, error rate, RF or the two components (see supplement).

4. Discussion

The aim of the present study was to investigate error processing and error detection in a complex task setting; to this end, we introduced a novel task—the 8ART—with an assignment of eight stimuli to eight responses. Although the task was relatively complex, overall performance was quite good and participants were able to detect errors successfully. As already discussed above, we could neither replicate faster errors compared to correct responses (e.g., Rabbitt, 1966), nor the typical $N_{e/c}$ effect (Falkenstein et al., 1991), but rather found no variation in the $N_{e/c}$ amplitude between the conditions (but see Maier et al., 2010). Nevertheless, we could replicate the well-known effect of a high P_e amplitude on signalled errors, but small $P_{e/c}$ amplitudes on non-signalled errors and correct responses (Nieuwenhuis et al., 2001), as well as smaller RF for error trials (e.g. Bode & Stahl, 2014). We performed several exploratory analyses to explain the unexpected finding, as outlined in the following.

4.1. *Behavioural findings and implications*

Interestingly, as revealed by the additional analyses of the different response subtypes, the errors were not predominantly the result of the incongruent position (position errors) as it is typical for Simon task due to spatial-identity incompatibility (Eriksen & Eriksen, 1974; Simon et al., 1981; de Simone, 2020); instead, the most important error-inducing source seemed to be the fingers' neighbourhood. In error trials, the sensitive keys registered also a number of multiple responses indicating that participants tended to immediately correct their errors. Especially in non-signalled error trials, many second responses were the actual correct response, which explains why the participants signalled them as correct. The unexpected finding that errors were slower than correct responses might be the result of a large number of memory errors.

Similar to previous findings (Bode & Stahl, 2014), erroneous responses were less forceful. We tested the authors' assumption that this result could be an indicator of an early error correction mechanism which might have started already before response onset. Interestingly, the second response occurred already about 125 ms after the first response onset. And indeed, our additional RF analyses of error trials with and without error correction showed lower peak RF for errors with a correction compared to errors without a correction, which further supported the assumption.

4.2. *Electrophysiological findings and implications*

The mixed findings for the two components (successful replication of known $P_{e/c}$ effects, but no successful replication of the very well-known $N_{e/c}$ effects) and the unexpected RT results led to the idea to separately investigate fast and slow responses. It turned out that RT was an important moderating variable. After performing an RT-based median split, we found the expected larger N_e amplitudes for signalled and non-signalled errors compared to N_c in correct trials, but only on trials with a short RT. For trials with a long RT, by contrast, no $N_{e/c}$ effect was observed.

4.3. *Different error sources within signalled errors*

The impact of RT on the results as well as the post-experimental inquiry addressing participants' response tendencies led us to assume that there were at least two types of errors: *impulsive* errors (indicated by short RT) and *memory* errors (indicated by long RT). Interestingly, Van Driel, Ridderinkhof, and Cohen (2012); see also (Novikov et al., 2017) provided evidence for these different error types and the resulting error-processing dynamics even in less complex tasks. Our assumption was that impulsive errors were accidentally incorrect responses based on the fast (insufficient) processing of (irrelevant) stimulus characteristics (e.g., stimulus position). Thus, this error is conceptually similar to errors occurring in a Simon task (e.g. Van Driel et al., 2012). In support of this assumption, our additional analyses revealed twice as many position errors in fast compared to slow errors. Moreover, in fast trials, the $N_{e/c}$ showed the well-known effect (larger N_e for signalled and non-signalled errors compared to correct responses), suggesting therefore that early error processing was possible.

On the other hand, slow errors might have been the result of working memory weakness, that is, temporarily weak access to the correct stimulus response mapping (e.g., Van Driel et al., 2012). Several studies demonstrated that working memory weakness actually led to smaller N_e amplitudes (Coleman et al., 2018; Maier & Steinhauser, 2017). Our participants reported that sometimes they were confused and could not remember the correctly assigned response. They then tended to press after a while any key to hit the RT deadline, while being aware that this response was most likely incorrect. Empirically, this would be identified as slow signalled errors. Although these post-experimental reports were purely anecdotal, and may have been biased, the idea was supported by the lower confidence scores after slow errors compared to fast responses, and the ERP results.

Assuming that the missing $N_{e/c}$ difference in slow trials was the result of a missing representation of the actual correct response due to a memory weakness, it was surprising that the $P_{e/c}$ still showed the well-known effect in slow trials with a signalled error. This means that the P_e-related process and successful error detection may have relied on other error indication information (e.g., identifying the memory weakness) than the N_e-related process. This is in line with Di Gregorio, Maier, and Steinhauser (2018) findings suggesting that the Ne reflected the detection of inconsistencies between the given response and the actual correct response, whereas the P_e seemed to reflect a more general knowledge that a response is incorrect without the representation of the actual correct response.

Interestingly, the P_e amplitude was significantly smaller for slow errors compared to fast errors. In a perceptually challenging task with an explicit variation in speed and accuracy instruction, Steinhauser and Yeung (2010, 2011) reported an effect of the manipulated speed-accuracy trade-off on P_e with a larger P_e amplitude in the speed condition compared to the accuracy condition. The authors interpreted this larger P_e amplitude as a sign of more error evidence accumulation in a shorter time period. Thus, if in our slow error trials there was no early evidence from the N_e-related process, there might have been a separate process identifying an ongoing memory weakness; however, this separate process might have provided less evidence, hence, less certainty and smaller P_e amplitudes in slow error trials (Boldt and Yeung, 2015). For future research it would be interesting to explicitly vary the speed-accuracy trade-off and to use mathematical modelling such as diffusion modelling for the error signalling response to systematically investigate the pre- and post-response processing (for review, see Heitz, 2014).

4.4. *Limitations and conclusion*

Although the results gave a first impression of the different error sources, one could assume further combinations of processes underlying the different trial types. A systematic variation of the processes is therefore required to validate this idea. A limitation of the new paradigm is that it still did not provoke a large number of false alarms (i.e., non-signalled correct), which would be a very interesting condition to learn more about unsuccessful error detection. However, this is in line with many studies (for a discussion, see Wessel, 2012). We also learned that the eight-point rating for the response evaluation was not very helpful as the participants mainly used the extreme values (Charles & Yeung, 2019). Thus, future studies could reduce the number of rating points without losing too much information. Of course, due to the exploratory character of the study, we discussed several post-hoc explanations. Furthermore, the number of dependent variables was high, which is in the nature of the task. In the future, the findings have to be replicated and specific hypotheses have to be tested to differentiate memory error, response conflict, and impulsivity systematically and with a new (and larger) sample. Finally, a cautionary note for the interpretation and comparison of the data with other studies based on the applied analysis method: we presented CSD transformed ERPs, because we found strongly overlapping brain activity in several brain areas in the classic ERPs. In brief, we found the well-known accuracy effect in $N_{e/c}$ amplitude (supplement Fig. S1), but we assume that the effect partially resulted from other processes (e.g. stimulus-looked P3), which were also sensitive to the response outcomes (for the ERP analyses and a detailed methodological discussion see supplement). This is in line with EEG/MRI studies showing several frontal and partial sources for the classic $N_{e/c}$ (Buzzell et al., 2017).

However, the first step to introduce a new paradigm that allows investigating error processing in complex decisions was successful. The 8ART induced a sufficient number of errors and we postulated different error sources (e.g., memory weakness and insufficient stimulus processing related to impulsiveness). Interestingly, fast (impulsive) errors left neural traces in terms of N_e variations of error processing quite early, whereas slow (memory) errors showed first neural traces on a later neural indicator of error processing, the P_e. In a next step, this might help to systematically investigate variations of 8ART to chronometrically disentangle processing of error sources and error monitoring.

Author note

This research was supported by the German Research Foundation (STA 1035/7-1). We are grateful to Wilfried Follmann and Ulrike Thesing for helpful comments during the development of the task as well as Cirsten Bauer and Thomas Meyer for data collection.

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.biopsycho.2020.107969.

References

Beaulieu, C., Bourassa, M.-È., Brisson, B., Jolicoeur, P., & de Beaumont, L. (2014). Electrophysiological correlates of motor sequence learning. *BMC Neuroscience, 15*, 102. https://doi.org/10.1186/1471-2202-15-102

Bode, S., & Stahl, J. (2014). Predicting errors from patterns of event-related potentials preceding an overt response. *Biological Psychology, 103*, 357–369. https://doi.org/10.1016/j.biopsycho.2014.10.002

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *The Journal of neuroscience: the official journal of the Society for Neuroscience, 35*(8), 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Brion, M., Dormal, V., Lannoy, S., Mertens, S., de Timary, P., & Maurage, P. (2018). Imbalance between cognitive systems in alcohol-dependence and Korsakoff syndrome: An exploration using the Alcohol Flanker Task. *Journal of Clinical and Experimental Neuropsychology, 40*(8), 820–831. https://doi.org/10.1080/13803395.2018.1438371

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science (New York, N.Y.), 307*(5712), 1118–1121. https://doi.org/10.1126/science.1105783

Buzzell, G. A., Richards, J. E., White, L. K., Barker, T. V., Pine, D. S., & Fox, N. A. (2017). Development of the error-monitoring system from ages 9-35: Unique insight provided by MRI-constrained source localization of EEG. *NeuroImage, 157*, 13–26.

Charles, L., & Yeung, N. (2019). Dynamic sources of evidence supporting confidence judgments and error detection. *Journal of Experimental Psychology. Human Perception and Performance, 45*(1), 39–52. https://doi.org/10.1037/xhp0000583

Coleman, J. R., Watson, J. M., & Strayer, D. L. (2018). Working memory capacity and task goals modulate error-related ERPs. *Psychophysiology, 55*(3). https://doi.org/10.1111/psyp.12805

de Simone, L. (2020). Grounding magnitudes. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2013.00410. Advance online publication.

Di Gregorio, F., Maier, M. E., & Steinhauser, M. (2018). Errors can elicit an error positivity in the absence of an error negativity: Evidence for independent systems of human error monitoring. *NeuroImage, 172*, 427–436. https://doi.org/10.1016/j.neuroimage.2018.01.081

Dutilh, G., van Ravenzwaaij, D., Nieuwenhuis, S., van der Maas Han, L. J., Forstmann, B. U., & Wagenmakers, E. J. (2012). How to measure post-error slowing: A confound and a simple solution. *Journal of Mathematical Psychology*. https://doi.org/10.1016/j.jmp.2012.04.001. Advance online publication.

Elson, M., & Bente, G. (2009). *CORTEX - computer-aided registration tool for experiments. [Computer software]*. Retrieved from. Cologne, Germany: University of Cologne http://cortex.uni-koeln.de/.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149. https://doi.org/10.3758/BF03203267

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodel divided attention on late ERP components: Ii. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology, 78*(6), 447–455.

W.J. Gehring , B. Goss , M. G. Coles , D.E. Meyer , E. Donchin . A neural system for error detection and compensation Psychological Science 1993; 4: (6) 385-390 . doi: 10.1111/j.1467-9280.1993.tb00586.x.

Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (2012). The error-related negativity (ERN/Ne). In S. J. Luck, & E. S. Kappenman (Eds.), *Oxford library of psychology. The Oxford handbook of event-related potential components* (pp. 231–291). New York, NY, US: Oxford University Press.

Gibbons, H., Fritzsche, A.-S., Bienert, S., Armbrecht, A.-S., & Stahl, J. (2011). Percept-based and object-based error processing: An experimental dissociation of error-related negativity and error positivity. *Clinical Neurophysiology, 122*(2), 299–310. https://doi.org/10.1016/j.clinph.2010.06.031

Gratton, G., Coles, M. G. H. [Michael G. H.], & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Hajcak, G., & Simons, R. F. (2002). Error-related brain activity in obsessive–compulsive undergraduates. *Psychiatry Research, 110*, 63–72. https://doi.org/10.1016/S0165-1781(02)00034-3

Hajcak, G., McDonald, N., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology, 40*(6), 895–903. https://doi.org/10.1111/1469-8986.00107

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience, 8*, 150. https://doi.org/10.3389/fnins.2014.00150

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. https://doi.org/10.1037//0033-295X.109.4.679

Holroyd, C. B., Coles, M. G. H., & Nieuwenhuis, S. (2002). Medial prefrontal cortex and error potentials. *Science (New York, N.Y.), 296*(5573), 1610–1611. https://doi.org/10.1126/science.296.5573.1610

Jasper, H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*(2), 370–375. https://doi.org/10.1016/0013-4694(58)90053-1

Jaśkowski, P., van der Lubbe, R. H. J., Wauschkuhn, B., Wascher, E., & Verleger, R. (2000). The influence of time pressure and cue validity on response force in an S1-S2 paradigm. *Acta Psychologica, 105*(1), 89–105. https://doi.org/10.1016/S0001-6918(00)00046-9

Kaczkurkin, A. N. (2013). The effect of manipulating task difficulty on error-related negativity in individuals with obsessive-compulsive symptoms. *Biological Psychology, 93*(1), 122–131. https://doi.org/10.1016/j.biopsycho.2013.01.001

Kantowitz, B. H. (1973). Response force as an indicant of conflict in double stimulation. *Journal of Experimental Psychology, 100*(2), 302–309. https://doi.org/10.1037/h0035780. General.

Ko, Y.-T., Alsford, T., & Miller, J. (2012). Inhibitory effects on response force in the stop-signal paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 38*(2), 465–477. https://doi.org/10.1037/a0027034

Leue, A., & Beauducel, A. (2008). A meta-analysis of reinforcement sensitivity theory: On performance parameters in reinforcement tasks. Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology. *Inc, 12*(4), 353–369. https://doi.org/10.1177/1088868308316891

Maier, M. E., & Steinhauser, M. (2017). Working memory load impairs the evaluation of behavioral errors in the medial frontal cortex. *Psychophysiology, 54*(10), 1472–1482. https://doi.org/10.1111/psyp.12899

Maier, M. E., Steinhauser, M., & Hübner, R. (2010). Effects of response-set size on error-related brain activity. *Experimental Brain Research, 202*(3), 571–581. https://doi.org/10.1007/s00221-010-2160-3

Mattes S., Ulrich R., (1997). Response force is sensitive to the temporal uncertainty of response stimuli. Perception & Psychophysics. (59), 1089–1097. 10.3758/BF03205523.

Niessen, E., Fink, G. R., Hoffmann, H. E. M., Weiss, P. H., & Stahl, J. (2017). Error detection across the adult lifespan: Electrophysiological evidence for age-related deficits. *NeuroImage, 152*, 517–529. https://doi.org/10.1016/j.neuroimage.2017.03.015

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J. [Jos], Band, G. P. H., & 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. https://doi.org/10.1017/S0048577201001111

Novikov, N. A., Nurislamova, Y. M., Zhozhikashvili, N. A., Kalenkovich, E. E., Lapina, A. A., & Chernyshev, B. V. (2017). Slow and fast responses: Two mechanisms of trial outcome processing revealed by EEG oscillations. *Frontiers in Human Neuroscience, 11*, 218. https://doi.org/10.3389/fnhum.2017.00218

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 1*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Overbeek, T. J. M., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing. *Journal of Psychophysiology, 19*(4), 319–329. https://doi.org/10.1027/0269-8803.19.4.319

Perrin, F., Pernier, J., Bertrand, O., & Echalli, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72*, 184–187.

Rabbitt, P. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology, 2*(71), 264–272.

Ruiz, M. H., Jabusch, H.-C., & Altenmüller, E. (2009). Detecting wrong notes in advance: Neuronal correlates of error monitoring in pianists. *Cerebral Cortex (New York, N.Y. : 1991), 19*(11), 2625–2639. https://doi.org/10.1093/cercor/bhp021

Rüsseler, J., Münte, T. F., & Wiswede, D. (2018). On the influence of informational content and key-response effect mapping on implicit learning and error monitoring in the serial reaction time (SRT) task. *Experimental Brain Research, 236*(1), 259–273. https://doi.org/10.1007/s00221-017-5124-z

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing' world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology Human Perception and Performance, 26*(1), 141–151. https://doi.org/10.1037//0096-1523.26.1.141

Shimizu, H. (2002). Measuring keyboard response delays by comparing keyboard and joystick inputs. *Behavior Research Methods, Instruments, & Computers, 34*(2), 250–256. https://doi.org/10.3758/BF03195452

Simon, J. R., Sly, P. E., & Vilapakkam, S. (1981). Effect of compatibility of SR mapping on reactions toward the stimulus source. *Acta Psychologica, 47*, 63–81.

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *The Journal of Applied Psychology, 51*(3), 300–304. https://doi.org/10.1037/h0020586

Stahl, J. (2010). Error detection and the use of internal and external error indicators: An investigation of the first-indicator hypothesis. *International Journal of Psychophysiology, 77*(1), 43–52. https://doi.org/10.1016/j.ijpsycho.2010.04.005

Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. *The Journal of Neuroscience, 30*(46), 15643–15653. https://doi.org/10.1523/JNEUROSCI.1899-10.2010

Steinhauser, M., & Yeung, N. (2012). Error awareness as evidence accumulation: Effects of speed-accuracy trade-off on error signaling. *Frontiers in Human Neuroscience, 6*, 240. https://doi.org/10.3389/fnhum.2012.00240

Van der Helden, J., Boksem, M. A. S., & Blom, J. H. G. (2010). The importance of failure: Feedback-related negativity predicts motor learning efficiency. *Cerebral Cortex (New York, N.Y. : 1991), 20*(7), 1596–1603. https://doi.org/10.1093/cercor/bhp224

Van Driel, J., Ridderinkhof, K. R., & Cohen, M. X. (2012). Not all errors are alike: Theta and alpha EEG dynamics relate to differences in error-processing dynamics. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 32*(47), 16795–16806. https://doi.org/10.1523/JNEUROSCI.0802-12.2012

Weinberg, A., & Hajcak, G. (2011). Longer term test-retest reliability of error-related brain activity. *Psychophysiology, 48*(10), 1420–1425. https://doi.org/10.1111/j.1469-8986.2011.01206.x

Wessel, J. R. (2012). Error awareness and the error-related negativity: Evaluating the first decade of evidence. *Frontiers in Human Neuroscience, 6*, 88. https://doi.org/10.3389/fnhum.2012.00088

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. https://doi.org/10.1037/0033-295X.111.4.931