Early correlates of error-related brain activity predict subjective timing of error awareness

Published Version:
Early correlates of error-related brain activity predict subjective timing of error awareness / Di Gregorio F.; Maier M.E.; Steinhauser M.. - In: PSYCHOPHYSIOLOGY. - ISSN 0048-5772. - STAMPA. - 59:7(2022), pp. e14020.1-e14020.11. [10.1111/psyp.14020]

Availability:
This version is available at: https://hdl.handle.net/11585/949898 since: 2023-11-28

Published:
DOI: http://doi.org/10.1111/psyp.14020

(Article begins on next page)

Running head: Timing of error awareness

Early correlates of error-related brain activity predict subjective timing of error awareness

Francesco Di Gregorio[1,2], Martin E. Maier[2] & Marco Steinhauser[2]

[1] UO Medicina Riabilitativa e Neuroriabilitazione, Azienda Unità Sanitaria Locale, 40133, Bologna, Italy

[2] Department of Psychology, Catholic University of Eichstätt-Ingolstadt, 85072, Eichstätt Germany

Correspondence should be addressed to:

Francesco Di Gregorio

UO Medicina Riabilitativa e Neuroriabilitazione, Azienda Unità Sanitaria Locale,

Ospedale Maggiore Carlo Alberto Pizzardi, Largo Nigrisoli, 2,

40133 Bologna, Italy

E-Mail: francesco.digregorio@ausl.bologna.it

Abstract

Humans are remarkably reliable in detecting errors in their behavior. Whereas error awareness has been assumed to emerge not until 200 - 400 ms after an error, so-called early error sensations refer to the subjective feeling of having detected an error even before the erroneous response was executed. Here, we collected EEG to track how early error sensations are reflected in neural correlates of performance monitoring. Participants first had to perform a task, and then had to indicate whether an error has occurred and whether this error was detected before or after response execution. EEG results showed that early error sensations were associated with an earlier peak of the error-related negativity (Ne/ERN), a component of error-related brain activity that occurs briefly after the error response. This demonstrates that early error-related activity influences metacognitive judgments on the time course of error awareness, and thus contributes to error awareness.

Keywords:

Early error sensations, error awareness, error-related negativity, error positivity, metacognition.

Introduction

Fast and accurate error detection is a crucial ability of the performance monitoring system. When performing well-learned choice tasks under time pressure, participants immediately recognize almost all of their errors (Rabbitt, 1968, 2002). In these experiments, participants often report to become aware of their errors even slightly before the erroneous response was actually executed. This phenomenon called early error sensations (Di Gregorio et al., 2020) can be observed also in everyday life. For example, when writing an email in a hurry, we sometimes have the feeling that an error is about to occur even before an actual typo is made. Early error sensations stand in stark contrast to the predominant view that error awareness emerges not until 200 to 400 ms after error responses (Nieuwenhuis et al., 2001; Steinhauser & Yeung, 2010), and suggest that error awareness could be influenced by early activity involved in error processing. In the present study, we investigated whether the occurrence of early error sensations is predicted by the strength and timing of early error-related brain activity that emerges at the time of the response. Such a finding would demonstrate that metacognitive judgments have access to early correlates of error monitoring, and thus, that early error-related brain activity contributes to the emergence of error awareness.

Performance monitoring mechanisms in the brain have been extensively investigated using event-related potentials (Gehring et al., 2018; Ullsperger et al., 2014). These studies revealed that errors in human behavior elicit a cascade of brain activity, which starts already at the moment of response execution. The error-related negativity (Ne/ERN; Falkenstein et al., 1991; Gehring et al., 1993) is a fronto-central negativity peaking around 50 ms after errors, and was proposed to reflect the detection of a mismatch (Falkenstein et al., 2000), post-response conflict (Yeung et al., 2004), or prediction error (Holroyd & Coles, 2002). The Ne/ERN is followed by the error positivity (Pe), which is a broad posterior positivity occurring about 200 to 400 ms after an error (Falkenstein et al., 1991; Overbeek et al., 2005), possibly reflecting the accumulation of evidence for an error that underlies the emergence of conscious error awareness (Steinhauser & Yeung, 2010; Ullsperger et al., 2014).

A common view on error detection is that the Pe is the earliest correlate of error awareness. For instance, when participants are asked to signal whether their response was correct or incorrect, the Pe rather than the Ne/ERN is larger for signaled compared to unsignaled errors (Endrass et al., 2007; Nieuwenhuis et al., 2001). Moreover, the Pe is gradedly modulated by subjective confidence about error commission (Boldt & Yeung, 2015). However, this view has been challenged by studies showing that error-related brain activity earlier than the Pe could reflect the emergence of error awareness. Indeed, a link between Ne/ERN and error awareness has been reported in several studies (Scheffers & Coles, 2000; Wessel et al., 2011), and Ne/ERN latencies have been found to vary with indirect measures of error awareness such as self-corrections (Falkenstein et al., 1997; Fiehler et al., 2005). While these results could imply that also the Ne/ERN contributes to the emergence of error awareness, recent studies argued that the link between Ne/ERN and error awareness is correlative rather than causal (Di Gregorio et al., 2018).

A further objection against the idea of the Pe as the earliest correlate of error awareness is the observation of early error sensations. In a recent behavioral study, it was investigated how frequently participants report the feeling of having detected an error already before response execution (Di Gregorio et al., 2020). To this end, a series of experiments was conducted in which different primary tasks (in which errors could occur) and secondary tasks (in which judgments on the primary task responses had to be given) were applied. While a flanker task and a visual discrimination task were used as primary tasks, the secondary tasks required participants either to categorize their responses as correct, early error or late error, or to wager on the occurrence of an early error sensation. Across experiments, participants consistently reported that the majority of errors were associated with early error sensations with a range between 57% and 70% of all errors, and further results demonstrated that these early error sensations were not due to instructions or expectation biases. This demonstrates that early error sensations are a robust phenomenon in choice tasks, which raises the possibility that error awareness could emerge considerably earlier than at the time of the Pe.

The present study aimed to reveal early signatures of error awareness by investigating the neural correlates of early error sensations. As in a previous study (Di Gregorio et al., 2020), participants had to classify errors according to whether they already knew that an error was about to occur before the execution of the response (early errors) or not (late errors). We compared error-related brain activity in these early and late errors to reveal which components affect these metacognitive judgments and thus the subjective time course of error awareness. Our analysis particularly focused on the Ne/ERN, a component of error-related brain activity that emerges around the time of the response. We hypothesized that, if the Ne/ERN is related to the emergence of early error sensation and thus error awareness, the Ne/ERN should occur earlier for early errors than for late errors. In addition, we also analyzed the Pe to reveal whether early error sensations are also reflected in this established neural correlate of error awareness. As an earlier onset of the Ne/ERN could imply that evidence accumulation starts earlier while, at the same time, more evidence for an error is provided, we expected that early error sensations are not only associated with a shorter latency of the Pe but also with an increased Pe amplitude.

Method

Participants

32 participants (26 female) between 19 and 31 years of age (M = 22.8, SE = 0.65) participated in the experiment. All participants had normal or corrected to normal vision, were recruited from the student population at the <<University removed for double-blind review>> and received course credit or 8 Euro per hour for participation. The study was approved by the ethical committee of the <<University removed for double-blind review>> and all participants provided informed consent.

Apparatus

A PC running presentation software (Neurobehavioral Systems, Albany, CA) controlled stimulus presentation and response registration. Stimuli were presented on a 21-inch color monitor (60 Hz refresh rate) at a viewing distance of 70 cm.

Stimuli and procedure

The experiment consisted of a primary flanker task and a secondary task, in which participants classified their responses in the flanker task. Stimuli of the primary flanker task were strings of five horizontal arrowheads in Arial font, subtending a visual angle of 4.1° horizontally and 1.4° vertically. The central arrowhead was designated as the target and the lateral arrowheads were designated as the flankers. Participants had to identify the direction of the target. In 50% of trials, flankers had the same direction as the target (*congruent condition*; e.g. <<<<<). In the other 50% of trials, flankers had the opposite direction (*incongruent condition*; e.g. <<><<). Each trial started with the presentation of a fixation cross for 350 ms. Then, the stimulus array was presented for 200 ms followed by a black screen until a response was given. Participants had to identify the direction of the target by pressing the “K” key for left or the “L” key for right on a standard keyboard with the index or the middle finger of one hand (*primary task response*). After the response, another black screen was presented for 1000 ms. Then, a question mark appeared in the screen center to prompt participants to classify their primary task response. To this end, participants had to execute one of four responses with the index finger, middle finger, ring finger, or little finger of the other hand (*secondary task response*). They indicated whether: a) they thought they had responded correctly (correct response), b) they had committed an error accompanied by early error sensations (early error), c) they had committed an error without early error sensations (late error) or d) they had committed an error, but they do not know when they detected the error (I-don’t-know error). The latter category was included to prevent that categorization is biased by guesses (Di Gregorio et al., 2020). Response mappings were counterbalanced so that half of the participants responded to the primary task with the right hand and to the secondary task with the left hand (primary task responses: “K” and “L”; secondary task responses: “A”, “S”, “D” and “F”) and the other half switched hands (primary task responses: “A” and “S”; secondary task responses: “H”, “J”, “K” and “L”).

The experiment consisted of eight test blocks with 64 trials per block. Each block contained 16 instances of each of the four possible stimuli in randomized order. Prior to the test blocks, participants performed two practice blocks (32 trials each) without a secondary response to practice the primary task and two further practice blocks in which the secondary task response was introduced. Before the administration of the two secondary task practice blocks, participants were instructed about early error sensations (Di Gregorio et al., 2020). We told participants that sometimes, the sensation could arise that they are about to commit an error already before the incorrect key was pressed, and that they should indicate whether this was the case or not on error trials. Prior to each of these practice blocks, participants were instructed to respond faster to the primary task whenever the error rate in the preceding block was below 25%.

EEG Data acquisition

The electroencephalogram (EEG) was recorded during the task using a BIOSEMI Active-Two system (BioSemi, Amsterdam, The Netherlands) with 64 Ag-AgCl electrodes from channels Fp1, AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C1, C3, C5, T7, TP7, CP5, CP3, CP1, P1, P3, P5, P7, P9, PO7, PO3, O1, Iz, Oz, POz, Pz, CPz, Fpz, Fp2, AF8, AF4, AFz, Fz, F2, F4, F6, F8, FT8, FC6, FC4, FC2, FCz, Cz, C2, C4, C6, T8, TP8, CP6, CP4, CP2, P2, P4, P6, P8, P10, PO8, PO4, O2 as well as the left and right mastoid. The CMS (Common Mode Sense) and DRL (Driven Right Leg) electrodes were used as reference and ground electrodes. Vertical and horizontal electroculogram (EOG) were recorded from electrodes above and below the right eye and on the outer canthi of both eyes. EEG and EOG data were continuously recorded at a sampling rate of 1024 Hz. All electrodes were offline re-referenced to the linked mastoids, re-sampled to 512 Hz and filtered with a 0.5 - 25 Hz band-pass filter.

Data analysis

Trials were classified according to stimulus congruency (congruent and incongruent), primary task response (correct, error) and secondary task response (correct, early error, late error and I don't know error). For all reported analyses, effects of variables with more than two levels were tested by analyses of variance

(ANOVAs) with repeated measurement. To compensate for violations of sphericity, Greenhouse–Geisser corrections were applied whenever appropriate (Greenhouse & Geisser, 1959), and corrected *p*-values but uncorrected degrees of freedom are reported. Differences between conditions were tested by planned comparisons using two-tailed t-tests for dependent samples and Cohen's d (*dz*) are reported. Moreover, 95% confidence intervals for the differences (CI) and scaled-information Bayes Factor (scale value = 0.707; Rouder et al., 2009) for testing the alternative hypotheses against the null hypothesis (BF_{10}) are reported.

Response time (RT) in the primary task was defined as the time interval between the onset of the stimulus and the subsequent key press. To control for outliers, trials were excluded whenever the RT of the primary task response was 3 standard deviations above or below the condition mean. 1.71% (*SE* = 0.4%) of all trials were excluded in this way. All frequency data were arcsine-transformed before statistical analyses (Winer, 1971).

ERP Data.

ERP data were analyzed using custom routines in MatLab R2013b (Mathworks, Natic, MA) and EEGLAB v13.0.1. Epochs from 200 ms before and 600 ms after the response were extracted from the continuous EEG. Because the Ne/ERN typically emerges slightly before the response, the average voltage in a time window from 150 ms to 50 ms before the response was used as baseline (Di Gregorio et al., 2016). Epochs contaminated with artifacts were identified using two methods from the EEGLAB toolbox (Delorme et al., 2007). An epoch was excluded (1) whenever the voltage exceeded 300 μV in order to remove epochs with large peaks, and (2) whenever an epoch deviated more than five standard deviations from the mean of the joint probability distribution to remove trials with improbable data (pop_eegthresh and pop_jointprob functions, respectively, in EEGLAB). The mean percentage of trials excluded in this way was 8.79% (*SE* = 1.45%). After artifact rejection, the average number of congruent error trials was 7.7 (*SE* = 1.4). Therefore, congruent trials in all conditions were excluded from EEG analyses because these trial numbers do not meet the requirements for reliable measurement of error-related brain activity (Olvet & Hajcak, 2009). For the incongruent trials, the condition with the smallest number of trials ('I-don't-know' error) had an average of 8 trials (*SE* = 1.8), but 19 participants had less then 6 trials in this condition. Thus, also incongruent 'I-don't-know' errors were excluded form the EEG analyses. The remaining error conditions considered for the EEG analyses had an average trial number of 46.6 (*SE* = 6.7) for incongruent early errors and of 30.5 (*SE* = 6.5) for incongruent late errors. Nevertheless, 5 participants with less than 6 error trials in one of the remaining conditions were excluded from the ERP analyses (Olvet & Hajcak, 2009). An independent component analysis (ICA) (Bell & Sejnowski, 1989) was performed to correct EOG artifacts. This was done using a multistep correlational template-matching process as implemented in CORRMAP v1.02 (Campos Viola et al., 2009). Topographies of ICs labeled as artifacts by the CORRMAP procedure were visually inspected and then removed from the data using inverse matrix multiplication. Epochs were then averaged separately for each participant and for the considered conditions.

We used two methods to quantify Ne/ERN and Pe amplitudes, which both rely on the error minus correct difference wave. First, the maximum difference was employed because it also allows to quantify both, the amplitude and the latency of a component, which is particularly important in the present study. Second, the mean difference in a specific time window was calculated because it is the most common measure, although it cannot be used to extract latencies. For the Ne/ERN, both approaches were applied in a time window between 0 ms and 200 ms relative to the response at electrode Fz (Endrass et al., 2007; West & Travers, 2008). For the Pe, we considered a time window between 200 ms and 400 ms relative to the response at electrode Pz (Overbeek et al., 2005). As a result, we obtained three measures for each component: 1) The maximum amplitude difference between correct and error trials, 2) the latency of the maximum amplitude difference, and 3) the mean difference between correct and error trials.

In addition, we analyzed stimulus-locked epochs ranging between -200 and 500 ms relative to the stimulus applying a baseline between -200 and 0 ms. We restricted this analysis to the P1 and N1 as later stimulus-locked components overlapped with response-locked error-related components. Whereas the P1 was quantified as the positive peak in a time window between 0 and 100 ms, the N1 was quantified as the negative peak between 100 ms and 200 ms (Krigolson & Holroyd, 2007). Both components were analyzed at electrode Oz where these components have maximal amplitudes (Barnes et al., 2014). These analyses were meant to reveal whether early and late errors differed with respect to early visual processing and thus with respect to the error source.

Results

Behavioral Data.

In a first step, we analyzed performance in the primary task. The mean error rate in the primary task was smaller for congruent trials (M = 3.23%, SE = 0.62%) compared to incongruent trials (M = 15.1%, *SE* = 1.3%), $t(31) = 11.93$, $p < .001$, *dz* =

2.2, *95% CI* = [10.1, 14.2], BF_{10} = 48215. Furthermore, a congruency effect with faster RTs for congruent correct trials (M = 376 ms, SE = 11 ms) than incongruent correct trials was revealed (M = 478 ms, SE = 19 ms), $t(31) = 10.1$, $p < .001$, dz = 1.8, *95% CI* = [80.8, 121], BF_{10} = 18850.

Please insert table 1 here

In a second step, we considered the frequencies of secondary task responses for each primary task response and stimulus condition (see Table 1). As observed in previous studies (e.g., Steinhauser et al., 2008), error detection (independently of error type) was very reliable. 92.5% (SE = 2.9%) of objective errors were categorized as either early error, late error, or I-don't-know error, and this rate was comparable for congruent stimuli (M = 93.2%, SE = 3.1%) and incongruent stimuli (M = 91.9%, SE = 2.6%), $t(31) = 0.79$, $p = .435$, dz = 0.14, *95% CI* = [-2.01, 4.55], BF_{10} = 3.05. Only 1.8% (SE = 0.41%) of correct responses were categorized as errors. This false alarm rate was higher for the incongruent (M = 2.71%, SE = 0.56%) than for the congruent condition (M = 0.89%, SE = 0.26%), $t(31) = 3.69$, $p = .001$, dz = 0.66, *95% CI* = [1.01, 3.11], BF_{10} = 54.65.

A considerable number of errors was categorized as early errors, that is, errors accompanied by early error sensations. Figure 1 shows the relative frequencies of objective errors categorized as early, late and I-don't-know errors among all detected errors. Importantly, the proportion of I-don't-know errors was 15.3% (SE = 4.3%), and was comparable for congruent and incongruent stimuli, $t(31)$ = 0.23, $p = .818$, dz = 0.04, *95% CI* = [-3.73 4.68], BF_{10} = 2.98. The frequencies of early and late errors were subjected to an ANOVA with the variables congruency (congruent, incongruent) and error type (early errors, late errors). The ANOVA showed a significant main effect of error type, $F(1,31) = 14.1$, $p = .001$, $\eta_p^2 = .31$, indicating that early errors (M = 57%; SE = 5.4%) were more frequent than late errors (M = 27.7%; SE = 4.5%). No significant main effect of congruency, $F(1,31) = 0.54$, $p = .818$, $\eta_p^2 = .01$ was obtained, but the interaction between error type and congruency was marginally significant, $F(1, 31) = 3.69$, $p = .064$ $\eta_p^2 = .11$, suggesting that the difference between early and late errors tended to be larger for congruent than for incongruent trials (Fig. 1A).

Furthermore, we compared RTs for correct responses, early and late errors (Fig. 1B). Only incongruent trials were considered for this analysis because 20 participants had no error trial in at least one error type condition of the congruent trials. Moreover, in the considered incongruent conditions, 6 participants with less than 5 trials in one condition were excluded. The results of a one-way ANOVA contrasting the three trial types showed a significant effect, $F(2, 75) = 14.72$, $p < .001$, $\eta_p^2 = 0.73$, with larger RTs for the corrects than for the two error types, *all ts* > 10.1, *all ps* < .001, *all dzs* > 1.8 *all* BF_{10} > 18850. Crucially, however, RTs did not significantly differ between early (M = 354 ms; SE = 12 ms) and late (M = 359 ms; SE = 15 ms) errors, $t(25) = 0.61$, $p = .551$, $dz = 0.12$, *95% CI* = [-25.7, 14.06], $BF_{10} = 3.29$ (Fig 1B). Finally, we compared error signaling RTs for correct responses correctly signaled, early and late errors. The results of a one-way ANOVA contrasting the three trial types in the congruent condition showed a significant effect, $F(2, 52) = 10.18$, $p = .001$, $\eta_p^2 = .281$, with slower RTs for detected errors (mean early and Late errors = 637 ms; SE = 88.94 ms) compared to correct responses (M = 244 ms; SE = 21.12 ms), *all ts* > 3.73, *all ps* < .001, *all dzs* > .718, *all* BF_{10} > 52.44. Moreover, analyses show a tendency towards faster RTs for early (M = 514 ms; SE = 67.61) compared to late (M

= 760 ms; SE = 140 ms) errors, *t*(26) = 1.95, *p* = .062, *dz* = .375, 95% *CI* = [-13.66, 505], BF_{10} = 1.44. However, these results should be treated with caution as participants were not instructed to signal their errors as fast as possible.

Please insert Figure 1 here

ERP Data.

Ne/ERN. Waveforms from electrode Fz (Fig. 2A) revealed more negative amplitudes on error trials compared to correct trials shortly after the response. Scalp topographies reflecting the maximum amplitude difference between correct and error trials show a typical fronto-central distribution in both error types. These Ne/ERNs were reliable and significant for early errors (maximum error minus correct difference amplitude: -5.85 µV, SE = 0.76 µV), *t*(26) = 7.78, $p < .001$, *dz* = 1.49, *95% CI* = [-7.41, -4.31], BF_{10} = 39416, as well as for late errors (-7.78 µV, SE = 1.26 µV), *t*(26) = 6.25, $p < .001$, *dz* = 1.2, *95% CI* = [-10.33, -5.22], BF_{10} = 16558. However, a comparison between the two error types revealed no significant difference for the amplitudes, neither for the maximum difference amplitude, *t*(26) = 1.76, *p* = .096, *dz* = 0.33, *95% CI* = [-0.36, 4.21], BF_{10} = 0.93, nor for the mean difference amplitude (early errors = -3.60 µV, SE = 0.79 µV; late errors = -4.227 µV, SE = 0.62 µV), *t*(26) = 0.836, *p* = .411, *dz* = 0.16, *95% CI* = [-0.92, 2.18], BF_{10} = 2.73[1]. Crucial, however,

[1] As the wide time window of 0 to 200 ms could blunt the effect around 20 to 100 ms when using a mean difference measure, we additionally analyzed mean amplitudes for correct, early and late errors in a time window of 20 to 100 ms after the response. A one-way ANOVA on the variable response type (correct, early and late errors) showed a significant response type effect, *F*(2,52) = 20.23, $p < .001$, np^2 = .438 with reliable Ne/ERNs for both error types (early errors = -0.70 µV, SE = 0.77 µV, *t*(26) = 5.756, $p < .001$, *dz* = 1.11, *95% CI* = [-4.54, -2.15], BF_{10} = 5456; late errors = -1.59 µV, SE = 0.56 µV, *t*(26) = 5.33, $p < .001$, *dz* = 1.02, *95% CI* = [-5.87, -2.60], *BF* = 2102) compared to correct responses (2.65 µV, SE = 0.72 µV). As with the maximum difference measure, no significant

was the question whether the Ne/ERN for both error types differed in their latency. This becomes obvious from Figure 2A (right side) in which the difference waveforms (error minus correct) for early and late errors are compared. The Ne/ERN maximum difference for late errors (78 ms, SE = 9 ms) occurs about 23 ms later than the maximum difference for early errors (55 ms, SE = 5 ms), an observation that is corroborated by statistical testing, $t(26) = 2.85$, $p = .008$, $dz = 0.55$, *95% CI* = [6.64, 40.69], $BF = 7.73$[2]. Additionally, we compared peak latencies in the raw signal of the different response types (correct, early and late errors) in a one-way ANOVA. Although peak latency in the raw signal was shorter for early errors (43 ms, SE = 4 ms) than for late errors (52 ms, SE = 8 ms) and correct responses (corrects = 57 ms; SE = 9 ms), the ANOVA did not reveal a significant effect of response type, $F(2,52) = 1.172$, $p = .318$, $\eta_p^2 = .043$. This shows that the latency effect of early errors is reliable only for the difference waveforms. Although raw signal and difference waveforms are both valid methods to quantify Ne/ERN amplitude and latency, the two measures can differ. In particular, while raw measures identify the absolute psychophysiological response after an error, a difference waveform between correct and error trials takes into consideration responses after correct trials and outputs a deviation when a change in the ongoing response is detected (Holroyd & Coles, 2002). In other words, difference waveforms can evidence the dissimilarity between error processing and correct response processing and thus indicate the magnitude of

---

difference was found between the early and late errors, $t(26) = 1.244$, $p = .224$, $dz = .238$, *95% CI* = [-0.58, 2.36], $BF = 1.88$.

[2] Results were confirmed also after application of current source density (CSD) transformation (as applied in previous studies; Kelly & O'Connell, 2013). Importantly, the maximum difference in the error minus correct waveform occurred significantly later for late errors (100 ms, SE = 8 ms) than for early (73 ms, SE = 8 ms), $t(26) = 2.856$, $p = .008$, $dz = .548$, *95% CI* = [7.73, 47.45], $BF_{10} = 4.228$.

error detection.

To provide a better illustration of the time course of the Ne/ERN, we also included the stimulus-locked waveforms from electrode Fz in Figure 3A. Here, the Ne/ERN can be seen in the error waveforms between about 300 and 500 ms. Because RTs of early and late errors were not significantly different, the shift in time between the two waveforms in this time period can be interpreted as the differential evolution of the Ne/ERN for early and late errors.

Pe. In addition, we also compared early and late errors with respect to the Pe, a component that has more frequently been associated with error awareness. Waveforms from electrode Pz (Fig. 2B) revealed more positive amplitudes on error trials than on correct trials starting at around 200 ms after the response. Scalp topographies representing this difference show a posterior distribution for both error types, which is typical for the Pe. The Pe was reliable and significant for early errors (error minus correct: 8.62 µV, SE = 1.11 µV), $t(26) = 7.76$, $p < .001$, $dz = 1.49$, *95% CI* = [6.33, 10.9], $BF_{10} = 37909$ as well as for late errors (7.11 µV, SE = 1.32 µV), $t(26) = 5.36$, $p < .001$, $dz = 1.03$, *95% CI* = [4.39, 9.83], $BF_{10} = 2251$. Crucially, a comparison between the two error types revealed a significant difference in the Pe time window as larger Pe amplitudes for early errors were found for the mean difference between correct trials and errors, $t(26) = 2.42$, $p = .023$, $dz = 0.47$, *95% CI* = [0.22, 2.79], $BF_{10} = 3.24$. For the maximum difference between correct trials and errors, no significant difference was found neither for the amplitude (early errors: 12.2 µV, SE = 1.05 µV; late errors: 12.1 µV, SE = 1.44 µV), $t(26) = 0.08$, $p = .932$, $dz = 0.18$, *95% CI* = [-1.71, 1.87], $BF_{10} = 3.79$ nor for the latencies (early errors: 317 ms, SE = 12 ms; late errors: 301 ms, SE = 13 ms), $t(26) = 1.48$, $p = .151$, $dz = 0.284$, *95% CI* = [-6.05, 37.16], BF_{10} = 1.38[3].

Taken together, the analyses of response-locked ERPs revealed typical Ne/ERN and Pe components. Most importantly, the Pe amplitude and the Ne/ERN latency were both modulated by the timing of error awareness. Early errors showed earlier Ne/ERN latencies and larger Pe amplitudes compared to late errors.

Please insert Figure 2 here

P1 and N1. In addition, we analyzed early stimulus-locked correlates of visual attention to investigate whether early and late error differ with respect to the source of an error (see Fig. 3A). However, no differences in amplitudes or latencies were found for the P1 (early errors P1 amplitude: 2.02 µV, SE = 0.43 µV; late errors P1 amplitude: 1.59 µV, SE = 0.68 µV; $t(26) = 0.72$, $p = .477$, $dz = 0.09$, *95% CI* = [-0.81,1.66], BF_{10} = 2.95; early errors P1 latency: 54.3 ms, SE = 8.07 ms; late errors P1 latency: 48.4 ms, SE = 7.62 ms; $t(26) = 0.76$, $p = .451$, $dz = 0.08$, *95% CI* = [-9.88, 21.59], BF_{10} = 2.88) or N1 (early errors N1 amplitude: -7.07 µV, SE = 0.97 µV; late errors N1 amplitude: -7.44 µV, SE = 1.18 µV; $t(26) = 0.52$, $p = .604$, $dz = 0.12$, *95% CI* = [-1.07, 1.81], BF_{10} = 3.34; early errors N1 latency: 151 ms, SE = 3.91 ms; late errors N1 latency: 156 ms, SE = 4.74 ms; $t(26) = 0.79$, $p = .431$, $dz = 0.08$), *95% CI* = [-14.98, 6.59], BF_{10} = 2.82.

Please insert Figure 3 here

[3] Note that after CSD transformation, no differences were found in the Pe mean difference amplitude between the two error types, $t(26) = 0.845$, $p = .406$, $dz = .163$, *95% CI* = [-0.688, 0.287], BF_{10} = 2.71

Discussion

Early error sensation is the subjective experience of becoming aware of an error even before the erroneous response was actually executed (Di Gregorio et al., 2020). The present data indicate that early error sensations are a frequent and robust phenomenon also in the current study. Replicating a previous work (Di Gregorio et al., 2020), behavioral results showed that participants consistently reported early error sensations on error trials in a similar proportion for congruent and incongruent conditions. Moreover, as in the previous study, no RT difference between early and late errors was obtained. Thus, whether or not participants report an early error sensation did not depend on task-related features like congruency or RT. This is interesting as previous studies speculated that response conflict could influence conscious error detection (Yeung et al., 2004). However, the similar RTs for both error types and the absence of a robust congruency effect here and in a previous study (Di Gregorio et al., 2020) on the frequency of early errors suggests that pre-response conflict (i.e., conflict induced by the congruent or incongruent stimulus) did not differ between early and late errors. This speaks against the idea that pre-response conflict is directly involved in the generation of early error sensations.

Our ERP data revealed several signatures of early error sensations. First of all, larger Pe amplitudes emerged for early errors compared to late errors. This again demonstrates that the Pe is linked to error awareness and metacognition (Boldt & Yeung, 2015; Charles et al., 2013; Yeung & Summerfield, 2012). It has been hypothesized that the Pe reflects the accumulated evidence for having made an error (Steinhauser & Yeung, 2010; Ullsperger et al., 2014), based on input from sensory, cognitive, and autonomous systems (Ullsperger et al., 2010; Wessel et al., 2011).

The mechanisms of error awareness are thus comparable to a decision process involving lower-level evidence accumulation (i.e., collecting evidence from sensory, cognitive and autonomous systems) and a higher-level decision (i.e., a metacognitive judgment operating on this evidence) (Dehaene et al., 2014) . From this perspective, the larger Pe for early errors could indicate that errors accompanied by early error sensations are associated with more evidence, possibly because the accumulation of evidence for an error started earlier for these errors. The absence of an effect on Pe latency could reflect that the earlier onset of evidence accumulation is counteracted by a longer latency of the accumulation process itself (due to the accumulation of more evidence). However, this interpretation is tentative and contradicts the idea of a bounded evidence accumulation process in which evidence is accumulated until a constant criterion is reached (Desender et al., 2021; Steinhauser & Yeung, 2010). Indeed, the idea of a bounded accumulation process receives support from the observation that no effect on Pe amplitude is obtained when analyzing peak amplitudes (rather than average amplitudes).

Crucially, our data demonstrate that also error-related brain activity earlier than the Pe is sensitive to the subjective timing of error detection. The Ne/ERN for early errors shows a robustly shorter latency than the Ne/ERN for late errors, suggesting that early error sensations are accompanied by an earlier peak of the Ne/ERN. Note that the time course of Ne/ERN shows a double peak in the early error waveform (see Fig. 3A). This is relatively common in conflict paradigms like the flanker task (Danielmeier et al., 2009; Kirschner et al., 2020). More importantly, the observation of earlier Ne/ERN peaks for early errors provides support for the idea that already early error-related brain activity at around the time of the response contributes to the emergence of error awareness, or, in other words, that metacognitive content is sensitive to this very early activity. Early and late errors differed only in Ne/ERN latency but not its amplitude. This suggests that errors leading to an early error sensation are not associated with a stronger error signal such as stronger post-response conflict (Yeung et al., 2004) or prediction error (Holroyd & Coles, 2002). Rather, they differ in the time point at which this error signal appears possibly reflecting random variation in the build-up of the error signal (e.g., post-response conflict), suggesting a correspondence between the timing of objective error processing and the timing of subjective error awareness. Such a correspondence in timing was found only for the Ne/ERN but not for the Pe, which further points towards the Ne/ERN as a neural correlate of the temporal characteristics of error awareness.

Early error sensations as a metacognitive illusion?

Whereas our data demonstrate an association between Ne/ERN latency and early error sensations, this does not necessarily imply that error awareness emerges at the time point of the Ne/ERN. Such an explanation would neglect the numerous failures to establish a robust relationship between Ne/ERN amplitude and error awareness (e.g., Nieuwenhuis et al., 2001; Steinhauser & Yeung, 2010). An alternative explanation is that error awareness emerges at the time point of the Pe, whereas an early error sensation is a metacognitive illusion that reflects the backdating of the subjective time point of error detection to temporally align error awareness with the emergence of the objective error signal (Di Gregorio et al., 2020). Such a backward referral process has previously been proposed in the domain of visual awareness as a means to synchronize the subjective time point of visual awareness with the onset of the objective stimulus to create a coherent perception in the stream of consciousness (Libet et al., 1979, 1983). In line with these ideas, error awareness might be subjectively backdated to the time point of the earliest neural evidence for an error, which is the Ne/ERN. This would explain why the occurrence of early error sensations is facilitated by an early occurring Ne/ERN, as demonstrated by the present results.

The temporal alignment of actions (i.e., an error response) and their effects (i.e., the feeling of being incorrect) could serve to evoke a sense of agency, i.e., the feeling of being in control of one's actions and action outcomes (Haggard & Tsakiris, 2009). Indeed, judgments on the causality of actions and their sensory effects have been shown to correlate with the perceived temporal contiguity between both (Haering & Kiesel, 2016). From this perspective, the phenomenon of early error sensations could be closely related to the phenomenon of intentional binding (Haggard, Aschersleben, et al., 2002; Haggard, Clark, et al., 2002), which is considered to be an implicit measure of agency. Intentional binding refers to the observation that the interval between an action and a subsequent stimulus is underestimated if the action is perceived to have caused this stimulus. Both intentional binding and early error sensations could reflect a backward referral mechanism that serves to temporally align actions and their consequences. In the case of early error sensations, the temporal course of early error-related brain activity (i.e. Ne/ERN latency) is the crucial information for backdating. Indeed, early error-related brain activity can be considered an internal information influencing conscious error detection (Ullsperger et al. 2014) and metacognitive error evaluation (Yeung & Summerfield, 2012). However, one has to note that while backdating (or recalibration) of stimulus onset has been discussed as one source of intentional binding (Moore & Haggard, 2008), other results favored the idea of a change in time perception (Wenke & Haggard, 2009).

Conclusion

The present study demonstrated that the time course of the Ne/ERN is predictive of whether an error is perceived as being detected before the occurrence of a response, a phenomenon we called early error sensation. This shows that characteristics of the Ne/ERN can influence error awareness, but also that metacognitive content is sensitive to this early error-related brain activity. We interpret this finding as reflecting the subjective backdating of error awareness to the time point of the Ne/ERN, which resembles similar phenomena in perceptual awareness (Libet et al., 1979) and the emergence of agency (Haggard, Clark, et al., 2002).

References

Barnes, J. J. M., O'Connell, R. G., Nandam, L. S., Dean, A. J., & Bellgrove, M. a. (2014). Monoaminergic modulation of behavioural and electrophysiological indices of error processing. *Psychopharmacology*, *231*, 379–392. https://doi.org/10.1007/s00213-013-3246-y

Bell, A. J., & Sejnowski, T. J. (1989). An information-maximisation approach to blind separation and blind deconvolution. *Neural computation*, *6*(February 1995), 1004–1034.

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

Campos Viola, F., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., & Debener, S. (2009). Semi-automatic identification of independent components representing EEG artifact. *Clinical Neurophysiology*, *120*, 868–877. https://doi.org/10.1016/j.clinph.2009.01.015

Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *NeuroImage*, *73*, 80–94. https://doi.org/10.1016/j.neuroimage.2013.01.054

Danielmeier, C., Wessel, J. R., Steinhauser, M., & Ullsperger, M. (2009). Modulation of the error-related negativity by response conflict. *Psychophysiology*, *46*, 1288–1298. https://doi.org/10.1111/j.1469-8986.2009.00860.x

Dehaene, S., Charles, L., King, J. R., & Marti, S. (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology*, *25*(1947), 76–84. https://doi.org/10.1016/j.conb.2013.12.005

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, *34*, 1443–1449. https://doi.org/10.1016/j.neuroimage.2006.11.004

Desender, K., Ridderinkhof, K. R., & Murphy, P. R. (2021). Understanding neural signals of post-decisional performance monitoring: An integrative review. *eLife*, *10*, e67556. https://doi.org/10.7554/eLife.67556

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*(January), 427–436. https://doi.org/10.1016/j.neuroimage.2018.01.081

Di Gregorio, F., Maier, M. E., & Steinhauser, M. (2020). Are errors detected before they occur? Early error sensations revealed by metacognitive judgments on the timing of error awareness. *Consciousness and Cognition*, *77*(November 2019), 102857. https://doi.org/10.1016/j.concog.2019.102857

Di Gregorio, F., Steinhauser, M., & Maier, M. E. (2016). Error-related brain activity and error awareness in an error classification paradigm. *NeuroImage*, *139*, 202–210. https://doi.org/10.1016/j.neuroimage.2016.05.074

Endrass, T., Reuter, B., & Kathmann, N. (2007). ERP correlates of conscious error recognition: Aware and unaware errors in an antisaccade task. *European Journal of Neuroscience*, *26*(July), 1714–1720. https://doi.org/10.1111/j.1460-9568.2007.05785.x

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

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological psychology*, *51*(2–3), 87–107.

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1997). [Event-related potential components related to errors]. *Zeitschrift Fur Experimentelle Psychologie: Organ Der Deutschen Gesellschaft Fur Psychologie*, *44*(1), 117–138.

Fiehler, K., Ullsperger, M., & von Cramon, D. Y. (2005). Electrophysiological correlates of error correction. *Psychophysiology*, *42*(1), 72–82. https://doi.org/10.1111/j.1469-8986.2005.00265.x

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–390.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (2018). The Error-Related Negativity. *Perspectives on Psychological Science*, *13*(2), 200–204. https://doi.org/10.1177/1745691617715310

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112. https://doi.org/10.1007/BF02289823

Haering, C., & Kiesel, A. (2016). Time perception and the experience of agency. *Psychological Research*, *80*(2), 286–297. https://doi.org/10.1007/s00426-015-0654-0

Haggard, P., Aschersleben, G., Gehrke, J., & Prinz, W. (2002). Action, binding, and awareness. In *Common mechanisms in perception and action: Attention and Performance* (XIX, Vol. 266, pagg. 266–287).

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382–385. https://doi.org/10.1038/nn827

Haggard, P., & Tsakiris, M. (2009). The experience of agency: Feelings, judgments, and responsibility. *Current Directions in Psychological Science*, *18*(4), 242–246. https://doi.org/10.1111/j.1467-8721.2009.01644.x

Holroyd, C. B., & Coles, M. G. H. 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

Kelly, S. P., & O'Connell, R. G. (2013). Internal and External Influences on the Rate of Sensory Evidence Accumulation in the Human Brain. *Journal of Neuroscience*, *33*(50), 19434–19441. https://doi.org/10.1523/JNEUROSCI.3355-13.2013

Kirschner, H., Humann, J., Derrfuss, J., Danielmeier, C., & Ullsperger, M. (2020). Neural and behavioral traces of error awareness. *Cognitive, Affective and Behavioral Neuroscience*, *September*. https://doi.org/10.3758/s13415-020-00838-w

Krigolson, O. E., & Holroyd, C. B. (2007). Hierarchical error processing: Different errors, different systems. *Brain Research*, *1155*, 70–80. https://doi.org/10.1016/j.brainres.2007.04.024

Libet, B., Gleason, A. C., Wright, E. W., & Pearl, D. K. (1983). Time of Conscious Intention To Act in Relation To Onset of Cerebral Activity (Readiness-Potential). *Brain*, *106*(3), 623–642. https://doi.org/10.1093/brain/106.3.623

Libet, B., Wright, E. W., Feinstein, B., & Pearl, D. K. (1979). Subjective referral of the timing for a conscious sensory experience: A functional role for the somatosensory specific projection system in-man. *Brain*, *102*(1), 193–224. https://doi.org/10.1093/brain/102.1.193

Moore, J., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and Cognition*, *17*(1), 136–144. https://doi.org/10.1016/j.concog.2006.12.004

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*, 752–760. https://doi.org/10.1111/1469-8986.3850752

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

Overbeek, T. J. M., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *Journal of*

Psychophysiology, *19*, 319–329. https://doi.org/10.1027/0269-8803.19.4.319

Rabbitt, P. (1968). Three kinds of error-signalling responses in a serial choice task. *The Quarterly Journal of Experimental Psychology*, *20*(2), 179–188. https://doi.org/10.1080/14640746808400146

Rabbitt, P. (2002). Consciousness is slower than you think. *The Quarterly journal of experimental psychology. A, Human experimental psychology*, *55*(4), 1081–1092. https://doi.org/10.1080/02724980244000080

Ruiz, M. H., Jabusch, H.-C., & Altenmüller, E. (2009). Detecting Wrong Notes in Advance: Neuronal Correlates of Error Monitoring in Pianists. *Cerebral Cortex*, *19*(11), 2625–2639. https://doi.org/10.1093/cercor/bhp021

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

Steinhauser, M., Maier, M., & Hübner, R. (2008). Modeling behavioral measures of error detection in choice tasks: Response monitoring versus conflict monitoring. *Journal of experimental psychology. Human perception and performance*, *34*(1), 158–176. https://doi.org/10.1037/0096-1523.34.1.158

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

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in cognitive sciences*, *18*(5), 259–267. https://doi.org/10.1016/j.tics.2014.02.009

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure and Function*, *214*(5), 629–643. https://doi.org/10.1007/s00429-010-0261-1

Wenke, D., & Haggard, P. (2009). How voluntary actions modulate time perception. *Experimental Brain Research*, *196*(3), 311–318. https://doi.org/10.1007/s00221-009-1848-8

Wessel, J. R., Danielmeier, C., & Ullsperger, M. (2011). Error awareness revisited: Accumulation of multimodal evidence from central and autonomic nervous systems. *Journal of cognitive neuroscience*, *23*(10), 3021–3036. https://doi.org/10.1162/jocn.2011.21635

West, R., & Travers, S. (2008). Tracking the temporal dynamics of updating cognitive control: An examination of error processing. *Cerebral Cortex*, *18*(May), 1112–1124. https://doi.org/10.1093/cercor/bhm142

Winer, B. J. (1971). Use of analysis of variance to estimate reliability of measurements. In B. J. Winer (A c. Di), *Statistical Principles in Experimental Design* (2nd ed., pagg. 283–295). McGraw-Hill.

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

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1310–1321. https://doi.org/10.1098/rstb.2011.0416

Author contribution.

All authors contributed to the study design and study concept. Testing and data collection were performed by FDG and MEM. FDG performed the data analysis and interpretation under the supervision of MS and MEM. FDG drafted the manuscript, and MS and MEM provided critical revisions. All authors approved the final version of the manuscript for submission.

Author note.

Authors declare no competing interests.

Table 1: Relative frequencies (in %) of secondary task responses for each stimulus condition and primary task response.

		Primary Task			
		Congruent		Incongruent	
		Correct	Error	Correct	Error
Secondary	Correct	99.1 (0.3)	6.81 (3.1)	97.3 (0.6)	8.1 (2.6)
Task	Early Error	0.4 (0.1)	58.6 (4.9)	0.7 (0.2)	52.6 (4.7)
	Late Error	0.2 (0.1)	23.0 (3.8)	0.4 (0.2)	28.2 (4.3)
	Don’t know	0.3 (0.1)	11.6 (2.4)	1.6 (0.4)	11.1 (2.6)

Notes: Brackets contain standard errors of the mean.

Figure Captions

Figure 1: Behavioral data. A. Relative frequencies of errors classified as early, late and I-don’t-know errors, separately for congruent and incongruent condition. **B.** Response times for correct responses, early errors and late errors in the incongruent condition. ms = milliseconds. Error bars represent standard errors of the mean.

Figure 2: Response-locked ERP data. A Grand average response-locked ERP waveforms at Fz and waveforms of the difference between errors and correct trials. Topographies represent the negative peak in the error minus correct difference waveforms in each error condition in the Ne/ERN time window (shaded gray bar between 0-200 ms). Bar graphs represent the differences (errors minus corrects) in Ne/ERN negative peak amplitude and latency. **B** Grand average response-locked ERP waveforms at Pz and waveforms of the difference between errors and correct trials. Topographies represent the mean of the error minus correct difference waveforms in the Pe time window in each error condition (shaded gray bar between 200-400 ms). Bar graphs represent the difference (errors minus corrects) in the Pe amplitude. The horizontal black arrow represents the time shift of the maximal Ne/ERN difference between error types. R = response; CR = correct response; μv = microvolt; ms = milliseconds. Error bars represent the standard error of the mean.

Figure 3. stimulus-locked ERP data. A. Grand average stimulus-locked ERPs waveforms at electrode Fz for Stimulus-locked Ne/ERN analyses (shaded gray bar between 250-450 ms). and **B** waveforms at electrode Oz for P1 (shaded gray bar

between 0-100 ms) and N1 analyses (shaded gray bar between 100-200 ms). Topographies show the peaks of the error minus correct difference waveforms in each component and condition. CR = correct responses; μv = microvolt, ms = milliseconds, S = stimulus.