Neurobiology of Aging 108 (2021) 34–46

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging.org

Neural correlates of metacognition across the adult lifespan

Helen Overhoff[a,b,c,*], Yiu Hong Ko[a,b,c], Daniel Feuerriegel[b], Gereon R. Fink[a,d], Jutta Stahl[c], Peter H. Weiss[a,d], Stefan Bode[b], Eva Niessen[c]

[a] Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Center Jülich, Jülich, Germany
[b] Melbourne School of Psychological Sciences, University of Melbourne, Parkville, Victoria, Australia
[c] Department of Individual Differences and Psychological Assessment, University of Cologne, Cologne, Germany
[d] Department of Neurology, Faculty of Medicine, University Hospital Cologne, University of Cologne, Cologne, Germany

ARTICLE INFO

Article history:
Received 17 March 2021
Revised 14 July 2021
Accepted 1 August 2021
Available online 8 August 2021

Keywords:
Confidence
Error detection
Aging
Error(-related) negativity (N_e, ERN)
Error positivity (P_e)
Behavioral adaptation

ABSTRACT

Metacognitive accuracy describes the degree of overlap between the subjective perception of one's decision accuracy (i.e. confidence) and objectively observed performance. With older age, the need for accurate metacognitive evaluation increases; however, error detection rates typically decrease. We investigated the effect of ageing on metacognitive accuracy using event-related potentials (ERPs) reflecting error detection and confidence: the error/correct negativity ($N_{e/c}$) and the error/correct positivity ($P_{e/c}$). Sixty-five healthy adults (20 to 76 years) completed a complex Flanker task and provided confidence ratings. We found that metacognitive accuracy declined with age beyond the expected decline in task performance, while the adaptive adjustment of behaviour was well preserved. P_e amplitudes following errors varied by confidence rating, but they did not mirror the reduction in metacognitive accuracy. N_e amplitudes decreased with age for low confidence errors. The results suggest that age-related difficulties in metacognitive evaluation could be related to an impaired integration of decision accuracy and confidence information processing. Ultimately, training the metacognitive evaluation of fundamental decisions in older adults might constitute a promising endeavour.

1. Introduction

We are continuously monitoring and controlling our behaviour in order to achieve goals and avoid errors. The internal evaluation of our behaviour and our decisions, also referred to as *metacognition*, is crucial in everyday life, because it guides our present and future behaviour (Desender et al., 2019b; Rabbitt, 1966). Metacognition comprises both the detection of committed errors and a feeling of confidence that accompanies a decision (Fleming and Frith, 2014; Shekhar and Rahnev, 2020). When we feel less confident about a decision, we might try to adjust it, seek more information, or recruit additional cognitive processes to optimise performance (Desender et al., 2019a, 2019b). As ageing is usually associated with declining cognitive functions and higher rates of decision errors in daily activities, decisions and corresponding motor actions need to be adjusted more often (Hertzog, 2015; Ruitenberg et al., 2014). This might be achieved, for example, by increasing efforts for an efficient metacognitive evaluation of one's behaviour.

In general, metacognitive judgements are highly predictive of actual task performance, yet there is strong evidence that metacognition constitutes a dissociable process from the execution of the initial task (Galvin et al., 2003; Song et al., 2011). The degree to which subjective perceptions and objectively observed performance overlap, that is, the *accuracy* of metacognitive judgements, varies across individuals and task demands (Fleming & Dolan, 2012; Hertzog & Hultsch, 2000; Rahnev et al., 2020). Metacognitive accuracy has been addressed in two separate but arguably related fields of research: studies on error detection, focussing on the recognition of errors, and studies on decision confidence, investigating processes related to beliefs regarding the likelihood of having made a correct choice. In most cases, low confidence implies a higher probability of having committed an error. It has been suggested that error detection and confidence judgements might even share similar underlying computations, whereby error detection arises from low confidence that a correct decision has been made (Boldt and Yeung, 2015; Yeung and Cohen, 2006; Yeung and Summerfield, 2014).

* Corresponding author at: Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Center Jülich, Leo-Brandt-Str. 5, 52425 Juelich, NRW Germany

E-mail address: h.overhoff@fz-juelich.de (H. Overhoff).

0197-4580/$ – see front matter
https://doi.org/10.1016/j.neurobiolaging.2021.08.001

1.1. Neural correlates of metacognition

Neural correlates of metacognition have been studied by measuring event-related potentials (ERPs) of the human scalp electroencephalogram (EEG). The error negativity (N_e) is a negative deflection peaking around 100 ms after an overt behavioural response at fronto-central electrodes and typically has larger amplitudes for errors than correct responses (N_c for correct responses; i.e. correct negativity; Falkenstein et al., 1991; Falkenstein et al., 2000; Vidal et al., 2003). The component is classically associated with conflict monitoring, assuming that it tracks conflict between the given response and continuously accumulated post-decision evidence favouring the correct response (Falkenstein et al., 1991; Yeung et al., 2004). Moreover, it has been shown that the N_e amplitude scales with confidence, that is, it decreases from perceived errors to uncertain responses (guesses) to trials where the participant is confident about its correctness (Boldt and Yeung, 2015; Scheffers and Coles, 2000). The more posterior error positivity (P_e; P_c for correct responses, i.e. correct positivity) with a maximum amplitude around 250 ms after a response, is considerably larger for detected compared to undetected errors and has therefore been associated with explicit error awareness (Endrass et al., 2012a; Nieuwenhuis et al., 2001). Notably, the P_e has also been found to increase in amplitude with decreasing confidence in perceptual decisions (Boldt and Yeung, 2015; Rausch et al., 2019).

Concerning the mechanisms underlying these two components, Di Gregorio et al. (2018) designed a sophisticated task to provide evidence that the P_e, but not the N_e, was present when it was evident for participants that an error had been made, but they did not know the correct answer. These findings suggest that the P_e does not require a representation of the correct response to emerge, but instead accumulates post-decisional error evidence from widely distributed neural sources (Di Gregorio et al., 2018; Murphy et al., 2015; Steinhauser and Yeung, 2010; Yeung and Summerfield, 2014). Thus, while both classical components of error processing, N_e and P_e, have been shown to vary with reported confidence, the P_e appears to be more specifically associated with conscious metacognitive processes (Boldt and Yeung, 2015; Nieuwenhuis et al., 2001; Scheffers and Coles, 2000).

1.2. Metacognition and ageing

Metacognitive abilities in older age have been shown to vary across cognitive domains (Fitzgerald et al., 2017; Hertzog & Hultsch, 2000). For instance, while older adults tend to underestimate the prevalence of their decision errors in everyday life, metacognitive judgements of certain memory aspects (e.g., memory encoding) seem to be well preserved (Castel et al., 2016; Harty et al., 2013; Mecacci and Righi, 2006). Previous studies on decision making and metacognition yielded relatively consistent findings of a significant decline in error detection rate with higher age across multiple tasks (Harty et al., 2013; Rabbitt, 1990), even when task performance was comparable (Harty et al., 2017; Niessen et al., 2017; Wessel et al., 2018). In a large sample of healthy adults, Palmer et al. (2014) investigated decision confidence using a measure of metacognitive accuracy that takes task performance into account (Maniscalco & Lau, 2012). The authors found that age was not correlated with metacognitive abilities in a memory task, but that it was negatively correlated with metacognitive abilities in a perceptual discrimination task.

Effects of ageing on the neural correlates of metacognition have primarily been investigated in the field of error detection. Here, both the difference between N_e and N_c (Endrass et al., 2012b; Falkenstein et al., 2001; Schreiber et al., 2011), and the $P_{e/c}$ amplitude (Clawson et al., 2017; Harty et al., 2017; Niessen et al., 2017) were smaller in older adults, while the decrease in P_e, in particular, was linked to a lower error detection rate. Notably, the processing of the stimulus can also affect subsequent response-related processes, and variations with age in two ERPs (namely the N2 and the P300; Groom & Cragg, 2015; Polich, 2007) have been documented (Korsch et al., 2016; Larson et al., 2016; Lucci et al., 2013; Niessen et al., 2017). With the decline in behavioural performance reported above, this suggests an impaired error evidence accumulation process in older age, possibly due to limited cognitive resources (Harty et al., 2017; Niessen et al., 2017). Surprisingly, neither $N_{e/c}$ nor $P_{e/c}$ have been investigated using confidence ratings to assess age-related variations of metacognitive abilities. Some evidence from neuroimaging studies point to age-related structural differences in the neural basis of metacognition (Chua et al., 2009; Hoerold et al., 2013; Sim et al., 2020). However, a conclusive account that explains individual differences in metacognitive accuracy is still missing, for which the use of ERPs with high temporal resolution might be well-suited to provide valuable insights (Dully et al., 2018; Fleming and Dolan, 2012; Yeung and Summerfield, 2014).

1.3. The current study

This study aimed to investigate task performance and metacognition in older adults with a novel perceptual task to determine how generalizable the findings of decreased metacognitive accuracy in older age are (Palmer et al., 2014). For this, we used a colour-flanker task, in which participants had to identify the colour of a target stimulus that was flanked by two squares of the same or a different colour. We assessed decision accuracy, measured confidence using a four-point rating scale, and examined the impact of metacognitive accuracy on adaptations of subsequent behaviour (Desender et al., 2019a; Fleming et al., 2012; Ruitenberg et al., 2014). Furthermore, we investigated whether the amplitudes of $N_{e/c}$ and $P_{e/c}$, which are described as neural correlates of metacognition, track changes in decision confidence across the lifespan.

We hypothesised that metacognitive accuracy in our decision task would decrease with age (Niessen et al., 2017; Palmer et al., 2014). Independent of confidence, we expected an error-specific attenuation of ERP component amplitudes in older adults, which should result in a smaller difference between the neural responses related to errors and correct decisions (Endrass et al., 2012b; Larson et al., 2016). Independent of age, reported confidence was expected to show a positive association with the $N_{e/c}$ and a negative association with the $P_{e/c}$ amplitude (Boldt and Yeung, 2015; Scheffers and Coles, 2000). Based on findings from error detection studies showing an age-related decrease in the P_e amplitude of detected, but not undetected errors (Harty et al., 2017; Niessen et al., 2017), as well as reports linking the P_e to confidence (Boldt and Yeung, 2015), we expected a specific decrease in P_e amplitude for low confidence errors with increasing age.

2. Methods

2.1. Participants

We recruited 82 healthy adults with a broad age range from 20 to 81 years (49.8 ± 1.9 years [all results are indicated as mean ± standard error of the mean; *SEM*]; 35 female, 47 male). We aimed for an approximately uniform distribution of age and thus tested at least 10 participants per decade. Inclusion criteria were right-handedness according to the Edinburgh Handedness Inventory (EDI; Oldfield, 1971), fluency in German, (corrected-to-) normal visual acuity, no colour-blindness and no history of neurological or psychiatric diseases. Any signs of cognitive impairment

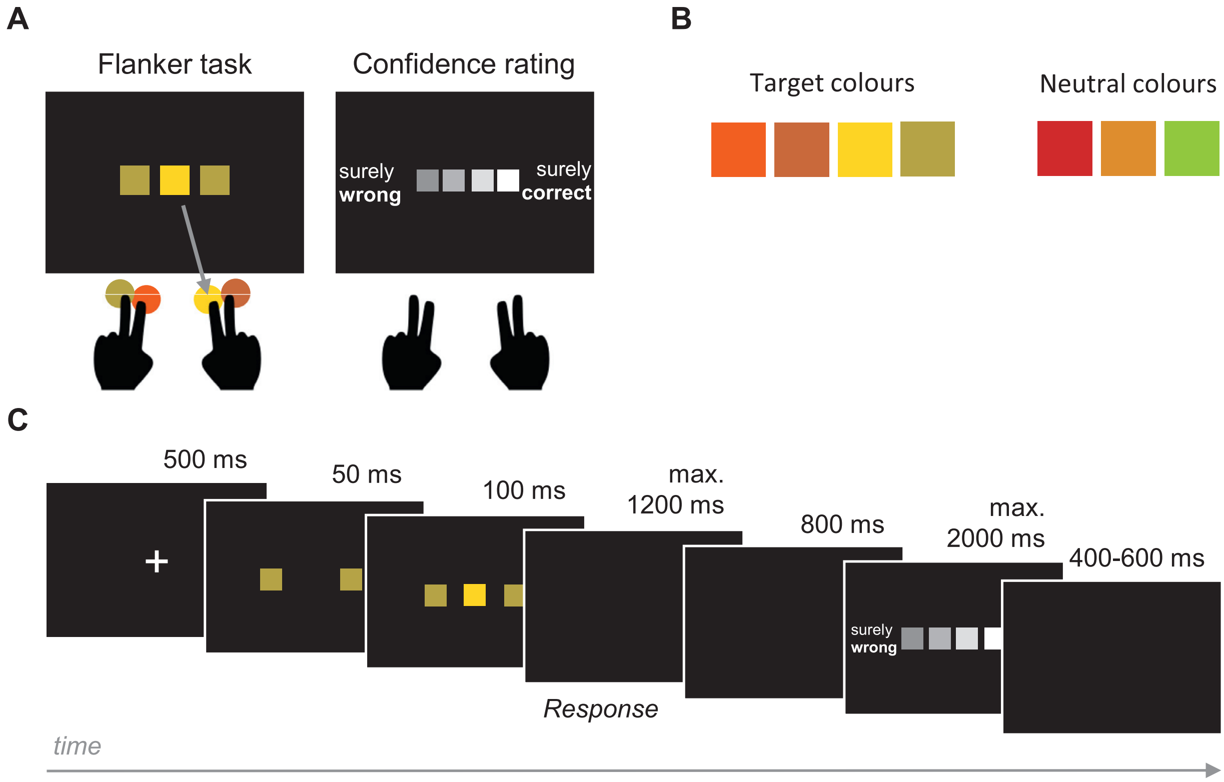

Fig. 1. (A) The left panel shows an example of a trial in the flanker task, where one central target and two flankers were presented, and the participant had to press the finger that was assigned to the respective target colour (illustrated by the grey arrow). The confidence rating (right panel) consisted of four squares, and the ends of the scale were labelled with the German words for 'surely wrong' on the left and 'surely correct' on the right side. The fingers were mapped onto the four squares according to their spatial location. (B) Colours used in the flanker task. Flanker stimuli could consist of target or neutral colours, whereas target stimuli could only consist of one of the four target colours. (C) Sequence of one trial (here, incongruent). Each trial started with a fixation cross, followed by the presentation of the flankers, to which the target was added shortly after. Then, the screen turned black until a response was registered (maximum 1,200 ms), followed by another blank screen. If a response had been given, the rating scale appeared until a rating was given (maximum 2,000 ms). If no response had been given within the designated time window, the German words for 'too slow' were shown instead. The trial ended with another blank screen for a random intertrial interval. For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.

(Mini-Mental-State Examination score lower than 24; MMSE; Folstein et al., 1975) or depression (Beck's Depression Inventory score higher than 17; BDI; Hautzinger, 1991) led to the exclusion of participants (one participant was excluded). Additionally, we excluded four participants who had more than one third of invalid trials (e.g., responses were too slow to fall into the pre-defined response window for analysis, or they showed recording artefacts). Another four participants were excluded because of an error rate (ER) higher than the chance level of 75%. Finally, eight participants were excluded because of combinations of very low accuracy, a high number of invalid trials, the selective use of single response keys, and errors in the colour discrimination task (described below), which suggested a lack of understanding of the task or the use of heuristic response strategies instead of trial-by-trial decisions. After exclusions, the final sample consisted of 65 healthy adults (45.5 ± 2.0 years; 20 to 76 years; 26 female, 39 male).

The study was approved by the ethics committee of the German Psychological Society (DGPs) and conformed to the declaration of Helsinki. All participants gave written informed consent before participating in the experiment.

2.2. Experimental paradigm

The main experimental task consisted of a modified version of the Eriksen flanker task using coloured squares as stimuli and four response options (Eriksen & Eriksen, 1974; Maier & Steinhauser, 2017; Fig. 1A). Participants were asked to respond as fast and accurately as possible to a centrally presented target by pressing a button with one of their index or middle fingers, mapped onto four designated target colours. In each trial, the target consisted of one of these target colours, and the flankers, located right and left to the target, consisted either of the same color as the target (congruent condition), of another target colour (incongruent condition), or of one of three additional neutral colours, which were not mapped to any response (neutral condition [Maier et al., 2008]; Fig. 1B). Both the incongruent and the neutral condition were used to induce conflict as they provided information distinct from the target. We chose this version of the classical flanker paradigm in order to increase task difficulty and thereby maximise the number of errors without tapping into other cognitive processes that might be affected by ageing (e.g., spatial, lexical, or semantic cognition). The colour-finger mapping was fixed over the course of the experiment for each participant and counterbalanced across participants.

Each trial started with a fixation cross for 500 ms. Then, flankers were presented for 50 ms before the target was added to the display for another 100 ms. Showing the (task-irrelevant) flankers before the target was expected to increase the induced conflict (Mattler, 2003). We used a response deadline of 1200 ms because this timing provided a good balance between a desirable number of errors and feasibility for all participants. If no response was registered before this deadline, the trial was terminated and the feedback 'zu langsam' (German for 'too slow') was presented on the screen. If a response was given, a confidence rating scale appeared after a black screen of 800 ms. The delay was introduced to avoid that EEG activity related to the first response overlapped with the confidence assessment. Participants were asked to indicate their confidence in their decision on a four-point rating scale from 'surely wrong' to 'surely correct' using the same keys as for the initial response. The maximum time for the confidence judgment was 2,000 ms. Trials were separated by a jittered intertrial interval of 400 to 600 ms. The sequence of an experimental trial is depicted in Fig. 1C.

2.3. Procedures

Prior to testing, participants were asked to provide demographic details and complete the handedness questionnaire. Afterwards, they completed a brief colour discrimination task (without EEG) to ensure that all participants were able to correctly discriminate the seven different colours used in the experimental paradigm (see Figure 1B). The discrimination task was followed by the EEG preparation and the main task. The neuropsychological tests were administered after the experiment. In addition, we assessed sustained attention span and processing speed using the d2-test (Brickenkamp, 2002), which have been shown to be positively associated with error processing abilities (Larson et al., 2011).

All stimuli in both tasks were presented on a black screen (LCD monitor, 60 Hz) in an electrically shielded and noise-insulated chamber with dimmed illumination, using Presentation software (Neurobehavioural Systems, version 14.5) for the colour discrimination task and uVariotest software (version 1.978) for the main task. A chin rest ensured a viewing distance of 70 cm to the screen and minimised movements. To record participants' responses, we used custom-made force-sensitive keys with a sampling rate of 1024 Hz (see Stahl et al., 2020).

The experiment started with a practice block of 18 trials in which participants received feedback about the accuracy of their response, which could be repeated if the participant considered it necessary. After that, two additional blocks with 72 trials without feedback and confidence assessments served as training blocks, allowing the participants to memorise the colour-finger mapping and to get accustomed to the response keys. Afterwards, another practice block introduced the confidence rating to ensure that participants understood and correctly applied the rating scale. The following main experiment consisted of five blocks with 72 trials each. Participants were allowed to take self-timed breaks after each block. The entire session lasted approximately three hours.

2.4. Electroencephalography recording and preprocessing

The EEG was recorded using 61 active electrodes (Acticap, Brain Products) aligned according to the international 10-20 system (Jasper, 1958). The electrodes were online referenced against the posterior Iz electrode close to the inion. Horizontal eye movements were measured using two electrodes at the outer canthi of the eyes (horizontal electrooculogram [EOG]), and another electrode underneath the left eye measured vertical movements (vertical EOG). The EEG signal was recorded continuously at a sampling rate of 500 Hz using a digital BrainAmp DC amplifier (Brain Products). Data were filtered between 0.1 Hz and 70 Hz, and a notch filter of 50 Hz was applied to remove line noise.

EEG data were preprocessed following a standardised pipeline using the MATLAB-based toolboxes EEGLAB and ERPLAB (Delorme and Makeig, 2004; Lopez-Calderon and Luck, 2014). The signal was segmented from -150 to 2,000 ms relative to target stimulus presentation (note that the flankers were presented at -50 ms). Epochs were visually inspected for artefacts and noisy electrodes. Epochs with artefacts were removed and identified noisy channels were interpolated using spherical spline interpolation. To identify and remove eyeblinks, we ran an Independent Component Analysis (ICA) using the infomax algorithm implemented in EEGLAB and afterwards baseline-corrected the epochs using the period of -150 ms to -50 ms to avoid influences of early perceptual processes related to the flanker presentation. Next, data were locked to the response, epoched from -150 ms to 800 ms relative to response onset and baseline-corrected using the 100 ms before the response. The additional analysis of conflict-related stimulus-locked ERPs can be found in the supplementary material S4. Remaining artefacts exceeding $\pm$ 150 μV were removed (Niessen et al., 2017), and a current source density (CSD) analysis was conducted using the CSD toolbox (Kayser and Tenke, 2006) allowing for better spatial isolation of ERP components and for obtaining a reference-independent measure (Perrin et al., 1989).

2.5. Behavioural data analysis

Trials with invalid responses (i.e. responses that were too slow) or recording artefacts, as well as responses faster than 200 ms were excluded from further analysis. The error rate (ER) was calculated as the proportion of errors relative to valid responses. Response time (RT) was defined as the time between stimulus onset and the initial crossing of the force threshold (40 cN) by any of the response keys.

To inspect how the confidence scale was used across participants, raw distributions of confidence ratings within all incorrect and correct responses were extracted. We computed Friedman ANOVAs for the proportion of each of each rating level for errors and correct responses with the factor confidence (4 levels). This analysis revealed that only a limited number of trials was available for the two middle confidence rating levels ('maybe wrong', 'maybe correct'), and we therefore collapsed those to create one category for all further analyses representing 'unsure' responses, i.e. confidence ratings expressing uncertainty.

For the analysis of metacognitive accuracy, we computed the Phi (Φ) correlation coefficient, which is a simple trial-wise correlation between task accuracy and reported confidence. It describes the extent to which the distributions of confidence ratings for correct and incorrect trials differ, while still depending on primary task performance and individual biases in confidence judgements (Fleming and Lau, 2014; Kornell et al., 2007; Nelson, 1984). Phi was calculated by correlating accuracy, coded as 0 (error) and 1 (correct response), and confidence (that the given response was correct), coded as 1 ('surely wrong'), 2 ('unsure'), and 3 ('surely correct'), for each participant. This provided us with one measure of metacognitive ability per participant that comprises both the accuracy and the confidence rating of each trial (e.g., Phi = 1 means that correct trials were successfully identified as such without uncertainty; while a Phi = 0 means that all errors were rated as 'surely correct', or all correct trials were rated as 'surely incorrect').

To assess the impact of accuracy and confidence on trial n on adaptations of behavioural responses, we computed a measure of response caution by multiplying the accuracy and RT on trial n+1 (Desender et al., 2019a). Response caution captures the trade-off between speed and accuracy in a decision, with higher values indicating a more cautious response strategy that is characterised by slower, and at the same time, more accurate responses. For this analysis, only pairs of two consecutive valid trials were included. Response caution was computed separately relative to a) initial trial accuracy (error, correct), and b) initial trial confidence ('surely wrong', 'unsure', 'surely correct').

At the group level, age-related effects on the d2-test score, the error rate, and Phi were computed using linear regressions.

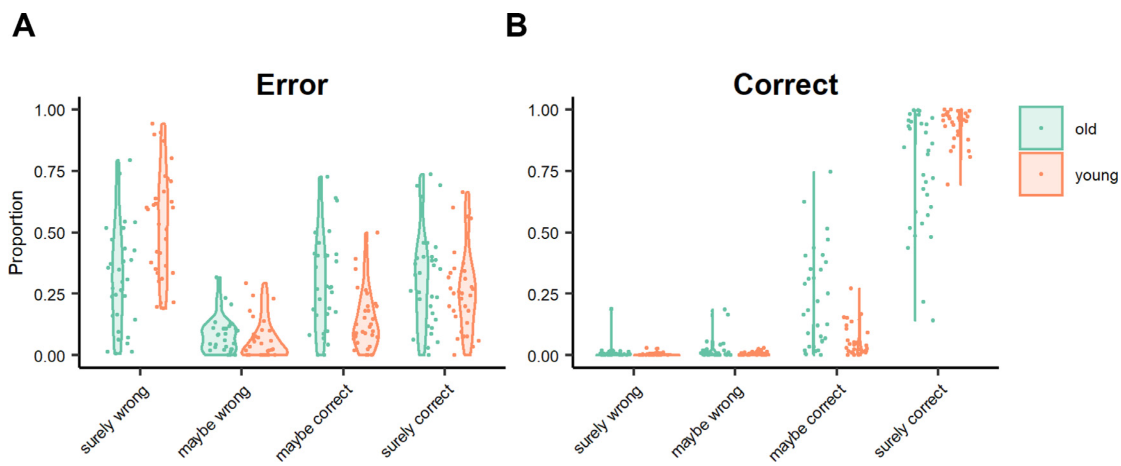

Fig. 2. Distributions of confidence ratings for errors (A) and correct responses (B). Errors were most often rated as 'surely wrong', and correct responses as 'surely correct'. Dots represent the individual proportions of the particular confidence response amongst all errors or correct responses, respectively. A median split by age (*Mdn* = 46) was conducted for illustration purposes. Older adults are shown in green, younger adults in orange. With increasing age, participants used the 'surely correct/wrong' ratings less often, and the middle of the confidence scale more often.

To rule out that metacognitive accuracy was confounded by age-related impairments in task performance or attention and processing speed, we performed additional multiple linear regressions to predict Phi by age, adding the factors of error rate or d2-test score, respectively.

For the analysis of performance and confidence at the trial level, data were analysed using linear and generalised linear mixed effects models. We always used the between-subject factor age as a predictor. The within-subject factor of interest was either accuracy (error, correct) or (pooled) confidence (3 levels). We fitted random intercepts for participants and, if possible, random slopes by participant for the within-subject factor of interest. For the outcome variables of RT, confidence and response caution, we fitted linear mixed models, for which *F* statistics are reported and degrees of freedom were estimated by Satterthwaite's approximation, and for accuracy we fitted generalised linear mixed models, for which X^2 statistics are reported. Model structures and coefficients are reported in the supplementary material S1.

Significant effects of confidence were followed up by pairwise comparisons across rating levels using paired-samples *t*-tests for linear mixed models and *Z*-tests for generalised linear mixed models. Significant interactions were followed up by (generalised) linear mixed regressions, separately for each level of a given within-subject factor to assess potential effects of age. We decided on these follow-up tests because our main interest was in the differential relations between accuracy, confidence, and behaviour across the lifespan rather than between the levels. Post-hoc test results were compared against Holm corrected significance thresholds to account for multiple comparisons.

Analyses were run in MATLAB R2019a and R (version 4.0.5; R Core Team, 2021) using the lme4 package (version 1.1; Bates et al., 2015).

2.6. Electroencephalographic data analysis

One participant had to be removed from electroencephalographic analyses, because noisy EEG data led to the exclusion of more than half of the trials. Data were response-locked and analysed at the single trial level. We first computed the grand-average for all participants, separately for errors and correct responses. The latency of the grand-average peak amplitude served as the time point around which individual mean amplitudes were extracted from the signal (± 50 ms). This was done to obtain meaningful time windows for statistical analyses, because data of single trials is too noisy to identify a meaningful peak (Clayson et al., 2013). On each trial, the $N_{e/c}$ local peak amplitudes were extracted from the response-locked data from the interval 0 to 150 ms following the response at FCz, and the $P_{e/c}$ local peak amplitudes were extracted from the interval 150 to 350 ms at Cz. This was based on visual inspection of the local maxima of the grand-average scalp topographies as well as previous literature (Falkenstein et al., 2000; Siswandari et al., 2019).

For statistical analyses of ERP amplitudes, we fitted the same linear mixed effects regression models as for the behavioural data. They included fixed effects of age and the within-subject factor accuracy (error, correct) for all trials combined (see supplementary material S3 for the analysis with the within-subject factor confidence for all trials) or confidence (3 levels) for separate analysis of errors and correct responses, a random intercept for each participant, and a random slope of the within subject factor by participant, if possible. The models were fitted to the CSD-transformed single trial mean ERP amplitudes of the $N_{e/c}$ and $P_{e/c}$. Model structures and coefficients are reported in the supplementary material S2.

3. Results

For brevity, only significant effects in the mixed effects regression analyses and relevant follow-up tests are reported in this section. For results of all tests as well as Bayesian analyses of relevant null effects, please refer to the supplementary material S1, S2 and S6.

3.1. Behavioural results

3.1.1. Attention

The average score for sustained attention and processing speed as assessed by the d2-test was 178.5 ± 5.6 (*M* ± *SEM*) and showed the typical decline for older adults, as reflected in a significant prediction of the test scores by age [$F(1,63) = 27.819$, $p < 0.001$; $\beta = -1.536$, $SE = 0.291$].

3.1.2. Distribution of confidence ratings

In a first step, we were interested in how the confidence ratings were distributed across the four confidence levels across the lifespan (Fig. 2). For this, we ran two Friedman ANOVAs for dependent

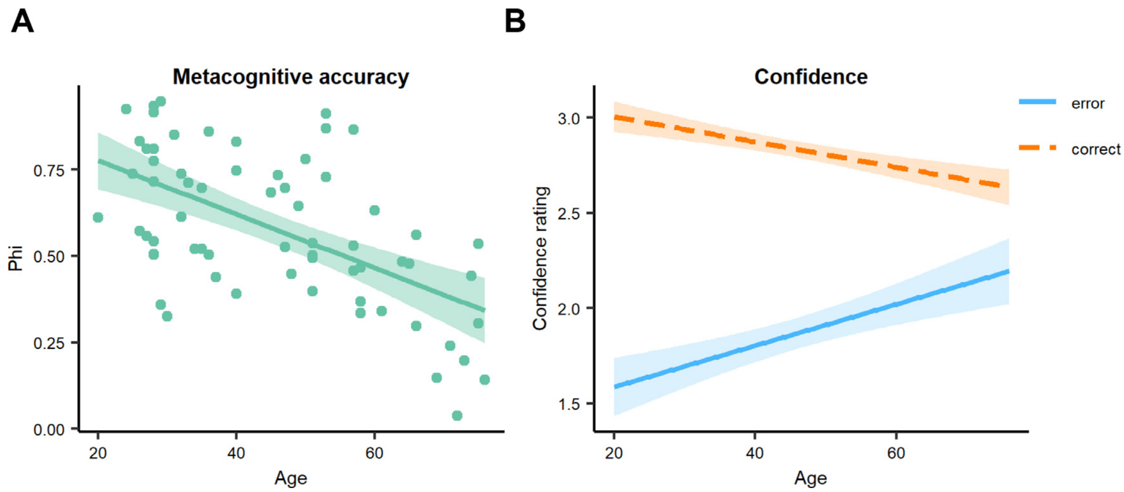

Fig. 3. Metacognition across the lifespan. (A) Metacognitive accuracy (Phi) decreased with age. Dots represent means of individual participants. (B) Confidence ratings for errors and correct trials were significantly predicted by age (in years). With increasing age, confidence was reduced for correct responses and increased for errors.

measures for the proportion for each rating category, separately for errors and correct responses.

The ANOVA for errors showed that the proportion differed between confidence levels [$X^2(3) = 78.029$, $p < 0.001$; Fig. 2A]. On average, most errors were rated as 'surely wrong' (42.8 %) and least errors as 'maybe wrong' (7.3%). Follow-up linear regressions on age-related differences for each rating category showed that the proportion of 'maybe correct' ratings was increased with higher age [$F(1,63) = 15.973$, $p < 0.001$; $\beta = 0.005$, $SE = 0.001$], whereas the ratio of 'surely wrong' ratings was decreased [$F(1,63) = 26.276$, $p < 0.001$; $\beta = -0.008$, $SE = 0.002$].

For correct responses, the ANOVA also revealed a main effect of confidence [$X^2(3) = 167.472$, $p < 0.001$]. Correct responses were most often rated as 'surely correct' (84.2 %) and least often as 'surely wrong' (0.7 %). Again, linear regression analyses on age-related differences showed that the proportion of 'maybe correct' ratings was increased with higher age [$F(1,63) = 24.653$, $p < 0.001$; $\beta = 0.006$, $SE = 0.001$], and the proportion of 'surely correct' ratings was decreased with age [$F(1,63) = 24.815$, $p < 0.001$; $\beta = -0.006$, $SE = 0.001$; Fig. 2B].

As mentioned above, to ensure a sufficient number of trials for each level of confidence for each participant, we combined 'maybe wrong' and 'maybe correct' ratings into one category representing 'unsure' responses. Thus, for all following behavioural analyses including the factor confidence, the reported analyses use three confidence levels.

3.1.3. Error rate (ER)

The average error rate was 15.6 ± 1.6%, and the mixed effects regression model testing for effects of confidence and age on error rate showed that the error rate significantly increased with higher age [$X^2(1) = 4.704$, $p = 0.030$]. The analysis further showed an effect of confidence on error rate [$X^2(2) = 2200.020$, $p < 0.001$]. The error rate decreased across confidence levels from 94.0 ± 0.7% on trials rated as 'surely wrong' to 67.3 ± 0.8% on trials rated as 'unsure' and 6.6 ± 0.2% on trials rated as 'surely correct'. Pairwise comparisons indicated that all comparisons were statistically significant (all $p < 0.001$). Thus, on average, participants' confidence reflected their performance well (which further supports the notion that the current study's confidence scale was a meaningful assessment tool). Furthermore, the regression analysis revealed a significant interaction between confidence and age [$X^2(2) = 168.125$, $p < 0.001$]. In subsequent mixed effects regression analyses for each level of confidence, error rates only significantly increased with higher age for the 'surely correct' confidence level [$X^2(1) = 37.664$, $p < 0.001$].

3.1.4. Response time (RT)

A mixed effects regression model predicting RT and testing for the effects of accuracy and age showed a significant effect of accuracy [$F(1,61.6) = 5.572$, $p = 0.021$] with on average slower errors (752.3 ± 3.8 ms) than correct responses (716.5 ± 1.3 ms). Moreover, the model revealed the expected slowing with age [$F(1,62.9) = 17.358$, $p < 0.001$], which did not significantly differ between errors and correct responses.

The mixed effects regression with the within-subject factor confidence similarly revealed an age-related slowing [$F(1,63.7) = 13.305$, $p < 0.001$; Fig. 4A]. Moreover, the analysis revealed an effect of confidence [$F(2,61.8) = 27.291$, $p < 0.001$] and a significant interaction between confidence and age [$F(2,56.6) = 5.187$, $p = 0.009$]. Pairwise comparisons indicated that all pairs were statistically significantly different (all $p < 0.010$), with trials associated with the 'unsure' confidence level (815.6 ± 6.7 ms) being considerably slower than trials rated as 'surely correct' (702.1 ± 1.3 ms) or 'surely wrong' (736.6 ± 6.7 ms). Furthermore, trials were significantly slower with older age for the extreme ratings ['surely wrong': $F(1,68.4) = 13.592$, $p < 0.001$; 'surely correct': $F(1.62.4) = 18.358$, $p < 0.001$], but not for 'unsure' ratings.

In short, RT was associated with confidence, such that high certainty (i.e. 'surely correct/wrong') was associated with the fastest responses, and this confidence-related modulation of RT decreased with higher age.

3.1.5. Confidence

A linear mixed effects regression model predicting confidence (coded from 1 to 3) across all trials revealed a significant effect of accuracy [i.e. error vs. correct trials; $F(1,63.4) = 162.928$, $p < 0.001$] and a significant interaction between accuracy and age [$F(1,62.4) = 37.361$, $p < 0.001$], but no significant effect of age. The average confidence rating was lower for errors (1.991 ± 0.014) compared to correct responses (2.867 ± 0.003). Follow-up regression analyses predicting confidence as a function of age for errors and correct responses separately revealed that confidence increased with age for errors [$F(1,60.1) = 17.977$, $p < 0.001$], while for correct responses it decreased [$F(1,62.1) = 23.816$, $p < 0.001$; Fig. 3B].

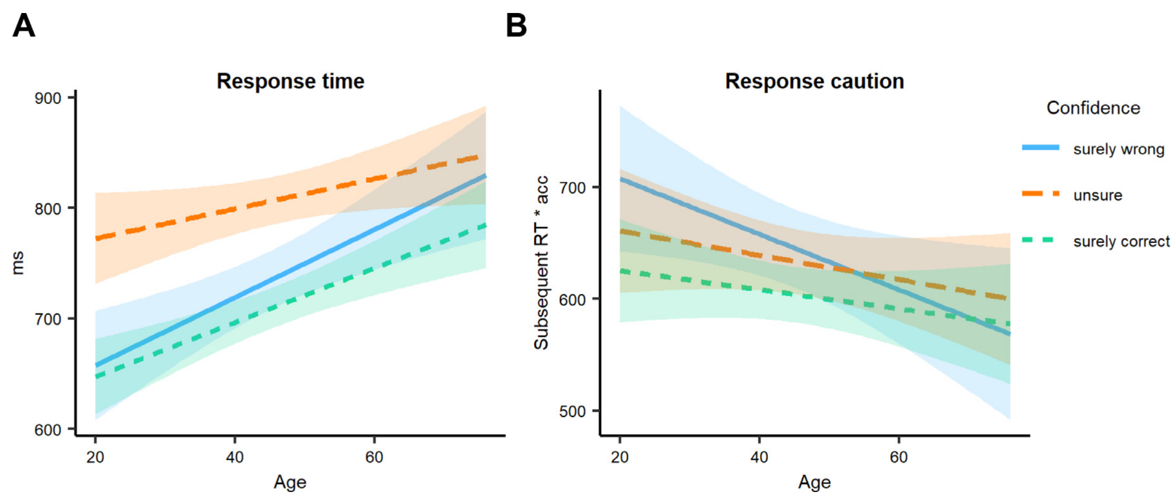

Fig. 4. Modulation of response time (RT; A) and response caution (B) by confidence and age (in years). (A) Trials rated as 'unsure' showed slower RTs than trials associated with any of the 'surely' rating categories, and this difference was smaller with increasing age. (B) Adaptation of response caution depending on previous trial confidence rating. Response caution was computed as the product of the accuracy and RT of subsequent trials. Across the lifespan, participants responded less cautiously after higher confidence ratings.

3.1.6. Metacognitive accuracy (Phi)

Phi had a mean of 0.579 ± 0.027 across the entire sample and was significantly predicted by age at the group level [F(1,63) = 32.206, p < 0.001; β = -0.008, SE = 0.001], indicating a decrease of metacognitive accuracy with age (Fig. 3A). Moreover, a multiple linear regression including the additional factor of error rate did not show a significant interaction with age (p = 0.535), suggesting that the association between metacognitive accuracy and age was not affected by decreased task performance in older adults. Similarly, a multiple linear regression including the additional factor of d2-test scores (which provide a task-independent measure of attention) suggested that the decrease in Phi with age was also independent of an age-related reduction in attentional capacity (interaction: p = 0.091).

3.1.7. Behavioural adaptation

To investigate the effect of accuracy and confidence in a given trial on the behaviour in the following trial, we computed response caution as the product of accuracy (coded as 0 and 1) and RT in the subsequent trial. The mixed effects regression with the within-subject factor accuracy (referring to the previous trial) revealed a significant effect of accuracy [F(1,55.9) = 12.366, p < 0.001] and an interaction between accuracy and age [F(1,43.9) = 6.709, p = 0.013], but no significant effect of age. Follow-up regression analyses for the subsets of errors or correct responses showed a nominal decrease in response caution with age for errors, but neither this nor the effect of age for correct responses was significant. Thus, these findings indicate that participants were on average more cautious after errors than after correct responses, and this effect did not significantly vary across age.

Next, we examined whether the response caution in the subsequent trial could also be predicted by the confidence rating in the preceding trial. As shown above, confidence and accuracy are strongly related; however, a significant modulation by confidence could also indicate that this internal confidence signal drives behavioural adaptations. The mixed effects regression on response caution with the within-subject factor confidence (referring to the previous trial) indeed revealed an effect of confidence [F(2,54.9) = 7.306, p = 0.002], but again, no effect of age and also no significant interaction (Fig. 4B). Pairwise comparisons between the confidence levels showed that the response caution after trials rated as 'surely correct' was significantly lower compared to trials rated as 'unsure' or as 'surely wrong'.

To summarise the effects of ageing on behaviour, we found the expected age-related general increase in error rates and response times, accompanied by a decrease in metacognitive ability, which was mainly reflected in reduced use of confidence ratings at the extreme ends of the scale but more indications of being unsure. Response caution, on the other hand, was not affected by ageing. Caution increased after errors compared to correct responses, and was notably specifically modulated by previous trial confidence. With higher confidence, the response caution in the subsequent trial decreased.

3.2. Electrophysiological results

3.2.1. $N_{e/c}$ amplitudes

The mean amplitude of the $N_{e/c}$ was significantly larger for errors compared to correct responses, as reflected in an effect of accuracy in the mixed effects regression predicting the $N_{e/c}$ as a function of accuracy and age [F(1,38.6) = 9.054, p = 0.005; Fig. 5A]. There was no main effect of age, but a significant interaction [F(1,31.8) = 5.472, p = 0.026] as the amplitude of the N_e [F(1,55.4) = 5.030, p = 0.029] but not the N_c was smaller with higher age.

For the analysis of confidence, we fitted separate linear mixed effects models to the N_e amplitude for errors and to the N_c amplitude for correct responses, with confidence as the within-subject factor and age as the between-subject factor. The regression analysis for errors showed effects of age [F(1,57.4) = 4.068, p = 0.048], confidence [F(2,2706.4) = 4.007, p = 0.018], and a significant interaction between age and confidence [F(2,2731.5) = 3.662, p = 0.026; Fig. 6A]. Pairwise comparisons between the confidence levels indicated a significant difference between errors rated as 'surely wrong' and 'surely correct', and follow-up mixed effects regressions showed that specifically the N_e amplitudes of low confidence errors (i.e. rated as 'surely wrong') was decreased with older age [F(1,58.1) = 9.735, p = 0.003].

The regression analysis for correct responses with the within-subject factor confidence yielded no significant effects (Fig. 6B).

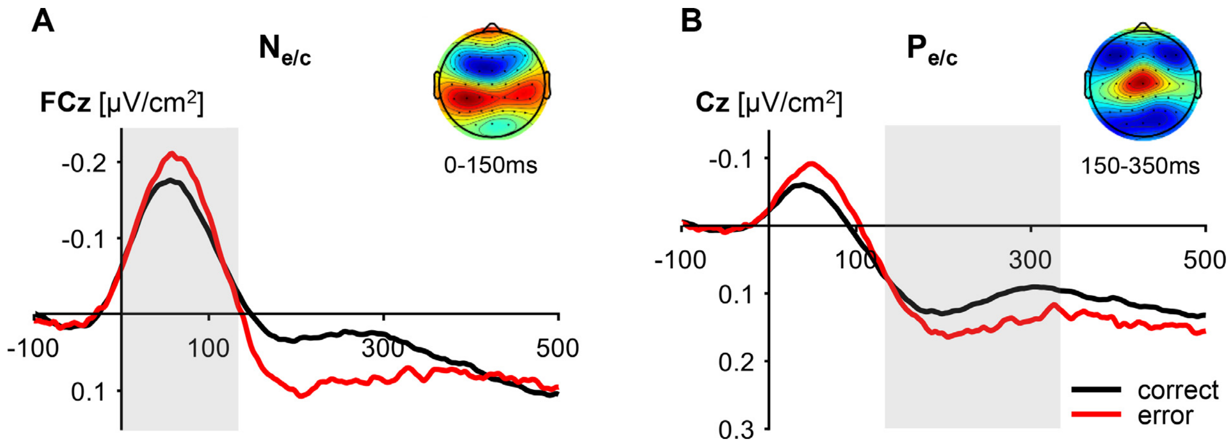

Fig. 5. Response-locked event-related potentials for errors and correct responses and topographical maps of errors after current source density transformation. (A) $N_{e/c}$ is computed at electrode FCz and (B) $P_{e/c}$ at electrode Cz. Errors are shown in red, correct trials in black. Scalp topographies depict the mean activity for all error trials averaged across the time windows for the N_e (0-150 ms) and the P_e (150-350 ms). Grey squares indicate time windows for the identification of peak amplitudes, which served to compute the adaptive mean amplitudes.

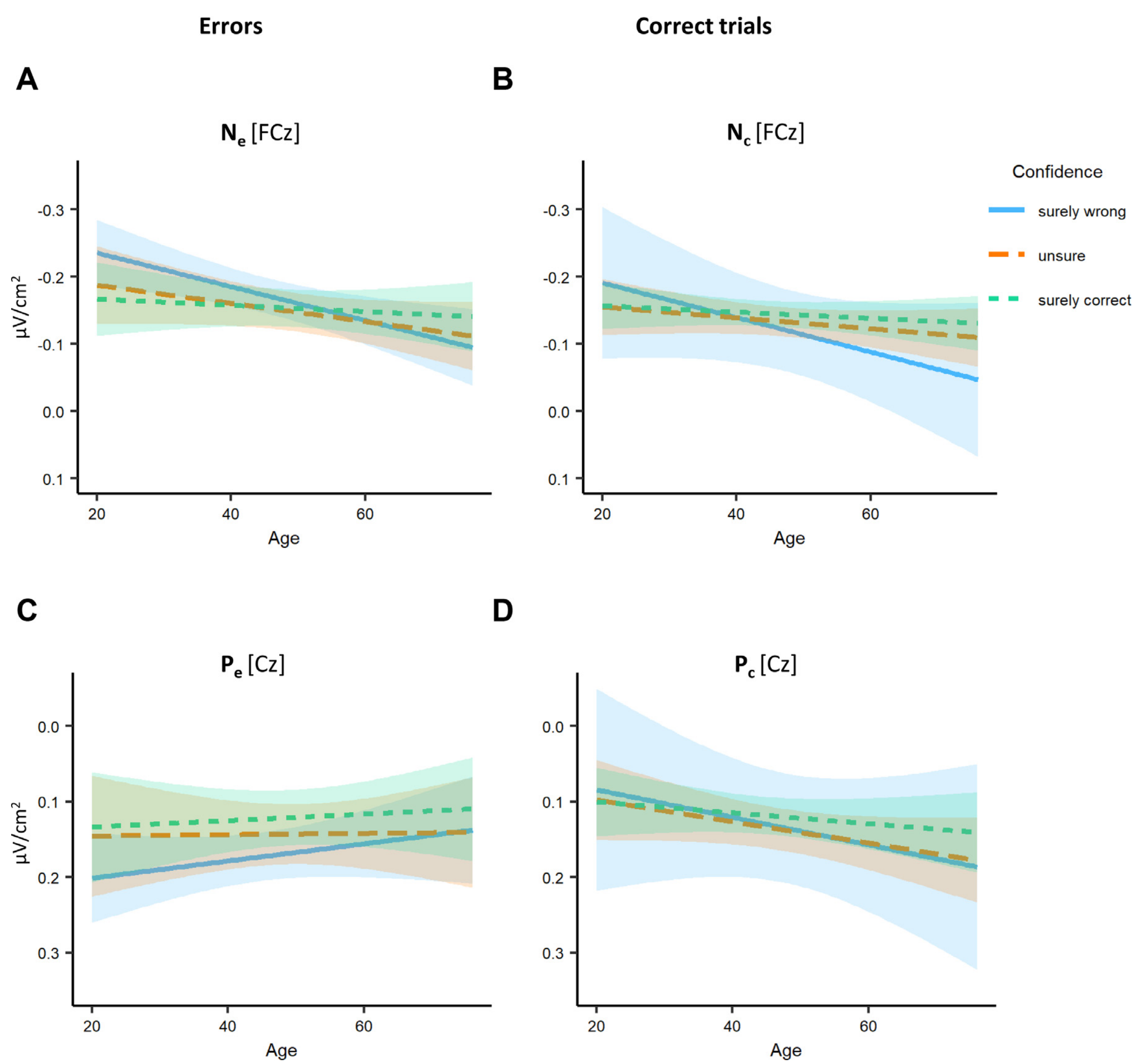

Fig. 6. Regression of response-locked event-related potentials on age (in years) by confidence, separately for errors and correct responses after current source density transformation. Errors are shown in the left panel, correct trials in the right panel. The N_e (A) and N_c (B) are shown at electrode FCz, and P_e (C) and P_c (D) are shown at electrode Cz. For errors, the amplitudes increased with lower confidence, while for correct responses, they were not modulated by confidence. Age predicted a decrease in N_e amplitude of 'surely wrong' errors.

3.2.2. $P_{e/c}$ amplitudes

The mixed effects regression on the $P_{e/c}$ amplitude with the within-subject factor accuracy revealed a significant effect of accuracy with larger amplitudes for errors compared to correct responses [$F(1,55.3) = 10.378$, $p = 0.002$; Fig. 5B]. There was no effect of age, but a significant interaction between accuracy and age [$F(1,49.2) = 6.443$, $p = 0.014$]. However, in follow-up regression analyses, no significant associations were found for errors or correct responses.

Next, responses were again split by their accuracy, and separate linear mixed effects models were fitted to the P_e and P_c amplitudes, respectively, with the within-subject factor confidence. Neither the analysis for errors nor the analysis for correct responses yielded any significant effects on the P_e and P_c amplitudes (Fig. 6C and D).

However, due to previous evidence suggesting a strong relation between $P_{e/c}$ amplitude and error detection or confidence ratings (Boldt and Yeung, 2015; Nieuwenhuis et al., 2001), we were specifically interested in the modulation of the P_e by confidence. To replicate previous findings, we fitted an additional, exploratory mixed effects model to the P_e amplitudes, including only the factor of confidence. The analysis revealed a significant, albeit small difference in P_e amplitude between errors rated as ‘surely wrong' and errors rated as ‘surely correct' [$F(2,2723.6) = 6.627$, $p = 0.001$], as confirmed in follow-up multiple comparisons between confidence levels ($t = 3.617$, $p < 0.001$). This exploratory analysis implies that the P_e was modulated by confidence when assessed independent of age.

4. Discussion

We conducted a complex four-choice flanker task with adult participants covering an age range from 20 to 76 years, allowing us to investigate confidence and metacognitive accuracy as well as neural indices thereof across the lifespan. We found that error rates and response times (RT) increased with age. Metacognitive accuracy, quantified as Phi, gradually decreased across the lifespan and was characterised by a differential use of confidence ratings. In contrast, we did not find differences between younger and older adults in the ability to adapt behaviour in accordance with reported confidence. As expected, the $N_{e/c}$ and $P_{e/c}$ amplitudes declined with higher confidence in having made a correct response, which was specifically observed for trials with response errors. While the N_e amplitude was smaller with older age whenever participants were sure they made an error, the variation in the P_e amplitude with reported confidence was surprisingly not affected by ageing. In the following, we will first discuss potential processes underlying age-related differences in metacognitive accuracy and their relation to task performance and confidence, before comparing the pattern we observed at the behavioural level to the patterns we observed in the ERPs. Finally, we argue that older adults' preserved ability to adapt their behaviour to their perceived confidence could be related to the $P_{e/c}$ amplitude.

4.1. Differential use of confidence scale as a marker of age-related metacognitive decline

In the present study, metacognitive accuracy (Phi) was reduced with increasing age. This is consistent with the findings of Palmer et al. (2014) who used a metacognitive efficiency measure, which further considered the individual performance in their perceptual discrimination task. As this measure was not directly applicable in our four-choice flanker task, we confirmed (by calculating multiple linear regressions taking into account the error rate and the d2-test score) that the observed decline in metacognitive accuracy was not merely a reflection of general age-related performance or attention deficits (d2-test; see also Larson & Clayson, 2011). Our results, therefore, show that Palmer et al.'s (2014) findings also hold for a more complex, speeded decision task, which was not based on stimulus ambiguity.

The question remains as to how the age-related differences in confidence emerge. Given the nature of Phi, a smaller value could either indicate more undetected errors or correct responses rated as being incorrect, or a generally higher uncertainty (i.e. rating all correct responses as ‘maybe correct' will result in a lower Phi value than rating the same number of correct responses as ‘surely correct'). Indeed, we observed that older adults used the extreme ends of the confidence scale considerably less often than younger adults.

For errors, this pattern resulted in a higher mean confidence with age. This disproportional rise in reported confidence has similarly been shown in error detection studies, indicated by a lower error detection rate in older adults (Harty et al., 2017,2013; Niessen et al., 2017). For correct decisions, we observed a *lower* mean confidence due to the tendency of the older adults to use the middle of the confidence scale, whereas previous studies rather reported an *over*-confidence in older age (Dodson et al., 2007; Hansson et al., 2008; Ross et al., 2012).

Interestingly, participants in our study responded slowest in case of uncertainty, i.e. ‘unsure' ratings. In contrast, studies on decision confidence typically report increasing RT with decreasing confidence (Kiani et al., 2014; Rahnev et al., 2020; Weidemann and Kahana, 2016). Most of these studies specifically measured confidence in having made a correct decision (i.e. the lowest confidence indicates guessing, while in our study it indicates high certainty in being incorrect), and typical paradigms in these studies are two-choice signal detection tasks in which the degree of sensory evidence, for instance, perceptual discriminability is manipulated (Kiani et al., 2014; Moran et al., 2015; Rollwage et al., 2020). In our task, we ensured (using a designated colour discrimination test) that all stimuli were perceptually discriminable without time pressure, and our data showed no signs of age-related differences in stimulus processing (even though it remains possible that slight impairments in colour perception, or other untested factors such as attention, working memory, etc., might have contributed to the age-related slowing we observed; see supplementary material S4). Instead, potential sources for errors could be, for instance, stimulus conflict caused by the flankers and the similarity of the stimulus colours, or difficulties in remembering the stimulus response mapping. Using a comparable paradigm, Stahl et al. (2020) found slow errors to be associated with lower confidence than fast, impulsive errors and inferred that those error types should predominantly be caused by weak stimulus-response representations (i.e. due to weak memory traces).

As such conclusions could not be drawn from classical error processing studies requiring only a binary error detection rating, our findings provide an important link between those and decision confidence studies. In a typical error processing paradigm that posed higher demands on the older adults (as indicated, for instance, by higher error rates), our results could be interpreted as their impaired metacognitive evaluation (assessed via confidence ratings) being partly related to more frequent memory-related errors, which appear to be more challenging to assess consciously (Maier and Steinhauser, 2017; Stahl et al., 2020).

4.2. Neural correlate of confidence is stable across age

The $P_{e/c}$ is an established marker of metacognition, reflecting variations in subjective error awareness and decision confidence (Boldt and Yeung, 2015; Nieuwenhuis et al., 2001). In the present study, the $P_{e/c}$ showed the well-known accuracy effect of larger amplitudes for errors than correct responses. Moreover, we could replicate prior findings of the P_e increasing with decreasing confidence, - for the first time - for a very broad age range (Boldt and Yeung, 2015; Rausch et al., 2019). This also replicates findings from error detection studies showing increased P_e amplitudes for detected compared to undetected errors (Endrass et al., 2012a; Nieuwenhuis et al., 2001).

The main interest of our study was to investigate the modulation of the $P_{e/c}$ by metacognition in the context of healthy ageing. Remarkably, the $P_{e/c}$ amplitude did not show an overall reduction with age, nor a differential modulation by confidence across the lifespan, suggesting that the accumulation of error evidence was well preserved in older age. This is contrary to the error detection literature (Harty et al., 2017; Niessen et al., 2017). Since these studies did not assess confidence on multiple levels, participants did not have the chance to express uncertainty. Assuming more 'unsure' cases with older age, their observed age-related decrease in P_e amplitude for detected errors might thus be confounded, as higher uncertainty was generally associated with reduced P_e amplitudes (Boldt and Yeung, 2015). Following this logic, it is also possible to explain the lack of a significant age-related modulation of the $P_{e/c}$ amplitude in the present study: If older adults' internal threshold for rating an error as 'surely wrong' was generally raised, the errors that *were* rated as 'surely wrong' should be trials with particularly high P_e amplitudes, as they were absolutely sure of having committed an error. As a result, a putative age-related decrease in the P_e amplitude of low confidence errors could be masked in our data, because the same reported rating levels might reflect a different sense of confidence for younger and older adults. Thus, the current pattern of results suggests that the P_e amplitude does *not* serve as a direct index of metacognitive accuracy across participants, but rather reflects the degree of confidence, irrespective of objective performance (Di Gregorio et al., 2018; Larson and Clayson, 2011; Pouget et al., 2016; Stahl et al., 2020).

4.3. Impaired neural processing of conflict modulates metacognitive decline

The marked behavioural decline in older adults' metacognitive accuracy was not mirrored in age-related variations of the $P_{e/c}$ amplitude, but rather in a differential modulation of the N_e across the lifespan. The modelling results revealed that the N_e amplitude was also affected by the interaction between confidence and age. With older age, the N_e declined for all errors in which high conflict was perceived. In other words, only the N_e amplitude of errors which were rated as 'surely wrong' varied in amplitude across the lifespan. As the $N_{e/c}$ is sensitive to conflict between the given and the actual correct response, older adults seemed to having had difficulties internally representing the correct response in highly conflicting situation (Yeung et al., 2004). Notably, this effect was error-specific, that is, we cannot draw conclusions about internal processes for correct responses, as the N_c amplitude did not show a relation to confidence that could have varied with age.

We suggest that the reduced N_e amplitude of low confidence errors with higher age could be related to the observed decrease in metacognitive accuracy in our flanker task. If older adults did not perceive high conflict due to difficulties in forming an accurate internal representation of the correct response, this information was necessarily missing for the metacognitive evaluation. Thus, the impaired neural integration of conflict detection and confidence could have led to the observed behavioural difficulties matching confidence ratings and objective accuracy.

4.4. Adults of all ages base future behaviour on subjective confidence

Ultimately, proper metacognitive evaluation should improve behaviour. Interestingly, response caution was not only enhanced after errors, but we also found evidence that it was modulated by the reported confidence in the preceding trial. Given that the participants did not receive any external feedback about the accuracy of their response (as it is often the case in real-life decisions), it seems plausible that they used their best available estimate, i.e. the subjective sense of confidence, to regulate subsequent behaviour (Desender et al., 2019a). Specifically, low confidence (reflecting a belief that an error had been committed) or uncertainty about a decision were associated with higher response caution in the subsequent trial. Possibly, participants sought more evidence before committing to their next decision, leading to slower but more accurate responses (Desender et al., 2019a, 2019b).

Translating our findings to error detection studies, the increase in response caution with lower previous trial confidence converges with findings of error detection studies reporting increased slowing (i.e. a sign of behavioural adaptation) after detected compared to undetected errors (Nieuwenhuis et al., 2001; Stahl et al., 2020; Wessel et al., 2018; for a review on post-error adjustments see Danielmeier & Ullsperger, 2011).

Notably, response caution was similarly affected by accuracy and confidence across the lifespan. Thus, while metacognitive accuracy was reduced in older age, a neural correlate of error confidence magnitude, the P_e amplitude, and the behavioural adaptations relative to the reported confidence were consistent across the lifespan. This suggests that it is the perceived confidence that shapes future behaviour, irrespective of metacognitive accuracy: Despite their failure in matching confidence to task performance, older adults seem to be equally able to use internal states of confidence to change future behaviour adaptively.

4.5. Limitations and implications

One limitation of the present study is the number of participants retained for the analyses. When designing the experiment, we tried to find an optimal balance between task difficulty, feasibility for all ages, and gaining many trials while ensuring that especially older adults were not exhausted at the end of the experiment. However, the combination of a substantial number of response alternatives, time pressure, and discriminability of stimuli was demanding, leading to an undesirably large number of participants to be excluded from the analyses (17 of the initial 82 participants).

A second shortcoming is the confined number of trials available for analysis after defining conditions of interest. Due to an unforeseen highly skewed use of the confidence scale, it was impossible to apply a factorial design while retaining four distinct confidence levels. In particular for correct trials, the variance in confidence ratings was low, which is a common problem in metacognition research (for a review, see Wessel, 2012). However, the application of linear mixed effects modelling provided us with a powerful tool that can account for varying trial numbers across participants and importantly, the multi-level structure of our data.

Nevertheless, our findings provide important insights into ageing effects on metacognition, integrating approaches from error detection and decision confidence research. In contrast to the metacognitive evaluation itself, the effect of confidence on subsequently adapting response caution was well preserved in older adults. Thus, training the metacognitive evaluation of fundamental decisions in older adults might constitute a promising endeavour (and has been shown to work for mathematical problem solving [Pennequin et al., 2010]).

5. Conclusion

The study of error detection and confidence in the context of healthy ageing have advanced largely in parallel. Our study demonstrates that confidence shapes our behavioural and neural processing of decisions and should be considered to investigate age-related effects on error processing and metacognitive abilities. Interestingly, the N_e, but not the P_e amplitude was differentially modulated by confidence across the lifespan, suggesting that the decreasing accuracy of metacognitive judgements with older age might be related to impaired integration of neural correlates of conflict detection and decision confidence.

Disclosure statement

The authors declare no conflict of interest.

Verification for the manuscript

"Neural correlates of metacognition across the adult lifespan"

1 All authors declare that they have no potential conflicts of interest to disclose.
2 This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – Project-ID 431549029 – SFB 1451; GRF and PHW), and the Australian Research Council (Discovery Project Grant; DP160103353; SB).
3 The data and analyses contained in this manuscript have not been submitted elsewhere and will not be submitted elsewhere while under consideration at *Neurobiology of Aging*.
4 The study was approved by the ethics committee of the German Psychological Society (DGPs) and conformed to the Declaration of Helsinki.
5 All authors have reviewed the contents of the manuscript being submitted, approved its contents and validated the accuracy of the data.

CRediT authorship contribution statement

Helen Overhoff: Conceptualization, Methodology, Software, Investigation, Data curation, Formal analysis, Writing – original draft, Visualization, Writing – review & editing. **Yiu Hong Ko:** Conceptualization, Writing – review & editing. **Daniel Feuerriegel:** Methodology, Writing – review & editing. **Gereon R. Fink:** Resources, Writing – review & editing. **Jutta Stahl:** Conceptualization, Methodology, Software, Writing – review & editing, Supervision. **Peter H. Weiss:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision. **Stefan Bode:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision. **Eva Niessen:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Supervision.

Acknowledgements

We thank all colleagues from the Institute of Neuroscience and Medicine (INM-3), Cognitive Neuroscience, and the Decision Neuroscience Lab at the Melbourne School of Psychological Sciences for valuable discussions and their support.

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation – Project-ID 431549029 – SFB 1451; GRF and PHW), and the Australian Research Council (Discovery Project Grant; DP160103353; SB).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neurobiolaging.2021.08.001.

References

Bates, D., Mächler, M., Boker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using {lme4}. J. Stat. Softw. 67, 1–48. doi:10.18637/jss.v067.i01.
Boldt, A., Yeung, N., 2015. Shared neural markers of decision confidence and error detection. J. Neurosci. 35, 3478–3484. doi:10.1523/JNEUROSCI.0797-14.2015.
Brickenkamp, R., 2002. The d2 Test of attention, 1st ed.9th ed. Verlag für Psychologie Hogrefe, Goettingen.
Castel, A.D., Middlebrooks, C.D., McGillivray, S., 2016. Monitoring Memory in Old Age: Impaired, Spared, and Aware. In: Dunlosky, J., Tauber, S.K. (Eds.), The Oxford Handbook of Metamemory. Oxford University Press, pp. 519–534. doi:10.1093/oxfordhb/9780199336746.013.3.
Chua, E., Schacter, D.L., Sperling, R., 2009. Neural basis for recognition confidence. Psychol. Aging 24, 139–153. doi:10.1037/a0014029.Neural.
Clawson, A., Clayson, P.E., Keith, C.M., Catron, C., Larson, M.J., 2017. Conflict and performance monitoring throughout the lifespan: An event-related potential (ERP) and temporospatial component analysis. Biol. Psychol. 124, 87–99. doi:10.1016/j.biopsycho.2017.01.012.
Clayson, P.E., Baldwin, S.A., Larson, M.J., 2013. How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. Psychophysiology 50, 174–186. doi:10.1111/psyp.12001.
Danielmeier, C., Ullsperger, M., 2011. Post-error adjustments. Front. Psychol. 2, 1–10.
Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. 134, 9–21.
Desender, K., Boldt, A., Verguts, T., Donner, T.H., 2019a. Confidence predicts speed-accuracy tradeoff for subsequent decisions. Elife doi:10.1101/466730.
Desender, K., Murphy, P.R., Boldt, A., Verguts, T., Yeung, N., 2019b. A Postdecisional Neural Marker of Confidence Predicts Information-Seeking in Decision-Making. J. Neurosci. 39, 3309–3319. doi:10.1101/433276.
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. doi:10.1016/j.neuroimage.2018.01.081.
Dodson, C.S., Bawa, S., Krueger, L.E., 2007. Aging, metamemory, and high-confidence errors: A misrecollection account. Psychol. Aging 22, 122–133. doi:10.1037/0882-7974.22.1.122.
Dully, J., McGovern, D.P., O'Connell, R.G., 2018. The impact of natural aging on computational and neural indices of perceptual decision making: A review. Behav. Brain Res. 355, 48–55. doi:10.1016/j.bbr.2018.02.001.
Endrass, T., Klawohn, J., Preuss, J., Kathmann, N., 2012a. Temporospatial dissociation of Pe subcomponents for perceived and unperceived errors. Front. Hum. Neurosci. 6, 1–10. doi:10.3389/fnhum.2012.00178.
Endrass, T., Schreiber, M., Kathmann, N., 2012b. Speeding up older adults: Age-effects on error processing in speed and accuracy conditions. Biol. Psychol. 89, 426–432. doi:10.1016/j.biopsycho.2011.12.005.
Eriksen, B.A., Eriksen, C.W., 1974. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept. Psychophys. 16, 143–149.
Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L., 1991. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalogr. Clin. Neurophysiol. 78, 447–455. doi:10.1016/0013-4694(91)90062-9.
Falkenstein, M., Hoormann, J., Christ, S., Hohnsbein, J., 2000. ERP components on reaction errors and their functional significance: A tutorial. Biol. Psychol. 51, 87–107. doi:10.1016/S0301-0511(99)00031-9.
Falkenstein, M., Hoormann, J., Hohnsbein, J., 2001. Changes of error-related ERPs with age. Exp. Brain Res 138, 258–262. doi:10.1007/s002210100712.
Fitzgerald, L.M., Arvaneh, M., Dockree, P.M., 2017. Domain-specific and domain-general processes underlying metacognitive judgments. Conscious. Cogn. 49, 264–277. doi:10.1016/j.concog.2017.01.011.
Fleming, S.M., Dolan, R.J., 2012. The neural basis of metacognitive ability. royalsocietypublishing.org 367, 1338–1349. doi:10.1098/rstb.2011.0417.
Fleming, S.M., Dolan, R.J., Frith, C.D., 2012. Metacognition: Computation, biology and function. Philos. Trans. R. Soc. B Biol. Sci. 367, 1280–1286. doi:10.1098/rstb.2012.0021.
Fleming, S.M., Frith, C.D., 2014. The Cognitive Neuroscience of Metacognition, The Cognitive Neuroscience of Metacognition. Springer doi:10.1007/978-3-642-45190-4.
Fleming, S.M., Lau, H.C., 2014. How to measure metacognition. Front. Hum. Neurosci. 8, 1–9. doi:10.3389/fnhum.2014.00443.

Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198. doi:10.1016/0022-3956(75)90026-6.
Galvin, S.J., Podd, J.V., Drga, V., Whitmore, J., 2003. Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. Psychon. Bull. Rev. 10, 843–876. doi:10.3758/BF03196546.
Groom, M.J., Cragg, L., 2015. Differential modulation of the N2 and P3 event-related potentials by response conflict and inhibition. Brain Cogn 97, 1–9. doi:10.1016/j.bandc.2015.04.004.
Hansson, P., Rönnlund, M., Juslin, P., Nilsson, L.G., 2008. Adult Age Differences in the Realism of Confidence Judgments: Overconfidence, Format Dependence, and Cognitive Predictors. Psychol. Aging 23, 531–544. doi:10.1037/a0012782.
Harty, S., Murphy, P.R., Robertson, I.H., O'Connell, R.G., 2017. Parsing the neural signatures of reduced error detection in older age. Neuroimage 161, 43–55. doi:10.1016/j.neuroimage.2017.08.032.
Harty, S., O'Connell, R.G., Hester, R., Robertson, I.H., 2013. Older adults have diminished awareness of errors in the laboratory and daily life. Psychol. Aging 28, 1032–1041. doi:10.1037/a0033567.
Hautzinger, M., 1991. The Beck Depression Inventory in clinical practice. Nervenarzt 62, 689–696.
Hertzog, C., 2015. Aging and Metacognitive Control. The Oxford Handbook of Metamemory doi:10.1093/oxfordhb/9780199336746.013.31.
Hertzog, C., Hultsch, D., 2000. Metacognition in adulthood and old age. In: Craik, F.I.M., Salthouse, T.A. (Eds.), The Handbook of Aging and Cognition. Lawrence Erlbaum Associates Publishers, pp. 417–466.
Hoerold, D., Pender, N., Robertson, IH, 2013. mende. Neuropsychologia 51, 385–391.
Jasper, H., 1958. The ten twenty electrode system of the international federation. Electroencephalogr. Clin. Neurophysiol. 10, 371–375.
Kayser, J., Tenke, C.E., 2006. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: II. Adequacy of low-density estimates. Clin. Neurophysiol. 117, 369–380. doi:10.1016/j.clinph.2005.08.033.
Kiani, R., Corthell, L., Shadlen, M.N., 2014. Choice certainty is informed by both evidence and decision time. Neuron 84, 1329–1342. doi:10.1016/j.neuron.2014.12.015.
Kornell, N., Son, L.K., Terrace, H.S., 2007. Transfer of metacognitive skills and hint seeking in monkeys: Research article. Psychol. Sci. 18, 64–71. doi:10.1111/j.1467-9280.2007.01850.x.
Korsch, M., Frühholz, S., Herrmann, M., 2016. Conflict-specific aging effects mainly manifest in early information processing stages-an ERP study with different conflict types. Front. Aging Neurosci. 8, 1–12. doi:10.3389/fnagi.2016.00053.
Larson, M.J., Clayson, P.E., 2011. The relationship between cognitive performance and electrophysiological indices of performance monitoring. Cogn. Affect. Behav. Neurosci. 11, 159–171. doi:10.3758/s13415-010-0018-6.
Larson, M.J., Clayson, P.E., Keith, C.M., Hunt, I.J., Hedges, D.W., Nielsen, B.L., Call, V.R.A., 2016. Cognitive control adjustments in healthy older and younger adults: Conflict adaptation, the error-related negativity (ERN), and evidence of generalized decline with age. Biol. Psychol. 115, 50–63. doi:10.1016/j.biopsycho.2016.01.008.
Larson, M.J., South, M., Clayson, P.E., 2011. Sex differences in error-related performance monitoring. Neuroreport 22, 44–48. doi:10.1097/WNR.0b013e3283427403.
Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: An open-source toolbox for the analysis of event-related potentials. Front. Hum. Neurosci. 8, 1–14. doi:10.3389/fnhum.2014.00213.
Lucci, G., Berchicci, M., Spinelli, D., Taddei, F., Di Russo, F., 2013. The Effects of Aging on Conflict Detection. PLoS One 8. doi:10.1371/journal.pone.0056566.
Maier, M.E., Steinhauser, M., 2017. Working memory load impairs the evaluation of behavioral errors in the medial frontal cortex. Psychophysiology 54, 1472–1482. doi:10.1111/psyp.12899.
Maier, M.E., Steinhauser, M., Hübner, R., 2008. Is the Error-related Negativity Amplitude Related to Error Detectability? Evidence from Effects of Different Error Types. J. Cogn. Neurosci. 20, 2263–2273.
Maniscalco, B., Lau, H., 2012. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious. Cogn. 21, 422–430. doi:10.1016/j.concog.2011.09.021.
Mattler, U., 2003. Delayed flanker effects on lateralized readiness potentials. Exp. Brain Res 151, 272–288. doi:10.1007/s00221-003-1486-5.
Mecacci, L., Righi, S., 2006. Cognitive failures, metacognitive beliefs and aging. Pers. Individ. Dif. 40, 1453–1459. doi:10.1016/j.paid.2005.11.022.
Moran, R., Teodorescu, A.R., Usher, M., 2015. Post choice information integration as a causal determinant of confidence: Novel data and a computational account. Cogn. Psychol. 78, 99–147. doi:10.1016/j.cogpsych.2015.01.002.
Murphy, P.R., Robertson, I.H., Harty, S.S., O'Connell, R.G., O'Connell, R.G., 2015. Neural evidence accumulation persists after choice to inform metacognitive judgments. Elife 4, 1–23. doi:10.7554/eLife.11946.
Nelson, T., 1984. A comparison of current measures of the accuracy of feeling-of-knowing predictions. Psychol. Bull. 95, 109–133.
Niessen, E., Fink, G.R., Hoffmann, H.M., Weiss, P.H., Stahl, J., 2017. Error detection across the adult lifespan: Electrophysiological evidence for age-related deficits. Neuroimage 152, 517–529. doi:10.1016/j.neuroimage.2017.03.015.
Nieuwenhuis, S., Richard Ridderinkhof, K., Blom, J., 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, 752–760. doi:10.1017/S0048577201001111.
Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9, 97–113. doi:10.1016/0028-3932(71)90067-4.
Palmer, E.C., David, A.S., Fleming, S.M., 2014. Effects of age on metacognitive efficiency. Conscious. Cogn. 28, 151–160. doi:10.1016/j.concog.2014.06.007.
Pennequin, V., Sorel, O., Mainguy, M., 2010. Metacognition, executive functions and aging: The effect of training in the use of metacognitive skills to solve mathematical word problems. J. Adult Dev. 17, 168–176. doi:10.1007/s10804-010-9098-3.
Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72, 184–187. doi:10.1016/0013-4694(89)90180-6.
Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148. doi:10.1016/j.clinph.2007.04.019.
Pouget, A., Drugowitsch, J., Kepecs, A., 2016. Confidence and certainty: distinct probabilistic quantities for different goals. nature.com doi:10.1038/nn.4240.
Rabbitt, P., 1990. Age, IQ and awareness, and recall of errors. Ergonomics 33, 1291–1305. doi:10.1080/00140139008925333.
Rabbitt, P.M.A., 1966. Errors and error correction in choice-response tasks. Exp. Psychol. 71, 264–272. doi:10.1037/h0022853.
Rahnev, D., Desender, K., Lee, A.L.F., Adler, W.T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L.Y., Balcı, F., Bang, J.W., Bègue, I., Birney, D.P., Brady, T.F., Calder-Travis, J., Chetverikov, A., Clark, T.K., Davranche, K., Denison, R.N., Dildine, T.C., Double, K.S., Duyan, Y.A., Faivre, N., Fallow, K., Filevich, E., Gajdos, T., Gallagher, R.M., de Gardelle, V., Gherman, S., Haddara, N., Hainguerlot, M., Hsu, T.Y., Hu, X., Iturrate, I., Jaquiery, M., Kantner, J., Koculak, M., Konishi, M., Koß, C., Kvam, P.D., Kwok, S.C., Lebreton, M., Lempert, K.M., Ming Lo, C., Luo, L., Maniscalco, B., Martin, A., Massoni, S., Matthews, J., Mazancieux, A., Merfeld, D.M., O'Hora, D., Palser, E.R., Paulewicz, B., Pereira, M., Peters, C., Philiastides, M.G., Pfuhl, G., Prieto, F., Rausch, M., Recht, S., Reyes, G., Rouault, M., Sackur, J., Sadeghi, S., Samaha, J., Seow, T.X.F., Shekhar, M., Sherman, M.T., Siedlecka, M., Skóra, Z., Song, C., Soto, D., Sun, S., van Boxtel, J.J.A., Wang, S., Weidemann, C.T., Weindel, G., Wierzchoń, M., Xu, X., Ye, Q., Yeon, J., Zou, F., Zylberberg, A., 2020. The Confidence Database. Nat. Hum. Behav. 4, 317–325. doi:10.1038/s41562-019-0813-1.
Rausch, M., Zehetleitner, M., Steinhauser, M., Maier, M.E., 2019. Cognitive modelling reveals distinct electrophysiological markers of decision confidence and error monitoring. bioRxiv, 860379 doi:10.1101/860379.
Rollwage, M., Loosen, A., Hauser, T.U., Moran, R., Dolan, R.J., Fleming, S.M., 2020. Confidence drives a neural confirmation bias. Nat. Commun. 11. doi:10.1038/s41467-020-16278-6.
Ross, L.A., Dodson, J., Edwards, J.D., Ackerman, M.L., Ball, K., 2012. Self-rated Driving and Driving Safety in Older Adults. Accid. Anal. Prev. 48, 523–527. doi:10.1038/jid.2014.371.
Ruitenberg, M.F.L., Abrahamse, E.L., De Kleine, E., Verwey, W.B., 2014. Post-error slowing in sequential action: An aging study. Front. Psychol. 5, 1–8. doi:10.3389/fpsyg.2014.00119.
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. J. Exp. Psychol. Hum. Percept. Perform. 26, 141–151. doi:10.1037/0096-1523.26.1.141.
Schreiber, M., Pietschmann, M., Kathmann, N., Endrass, T., 2011. ERP correlates of performance monitoring in elderly. Brain Cogn 76, 131–139. doi:10.1016/j.bandc.2011.02.003.
Shekhar, M., Rahnev, D., 2020. The Nature of Metacognitive Inefficiency in Perceptual Decision Making. Psychol. Rev. 1–83. doi:10.1037/rev0000249.
Sim, J., Brown, F., O'Connell, R., Hester, R., 2020. Impaired error awareness in healthy older adults: an age group comparison study. Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2020.08.001.
Siswandari, Y., Bode, S., Stahl, J., 2019. Performance monitoring beyond choice tasks: The time course of force execution monitoring investigated by event-related potentials and multivariate pattern analysis. Neuroimage 197, 544–556. doi:10.1016/j.neuroimage.2019.05.006.
Song, C., Kanai, R., Fleming, S.M., Weil, R.S., Schwarzkopf, D.S., Rees, G., 2011. Relating inter-individual differences in metacognitive performance on different perceptual tasks. Conscious. Cogn. 20, 1787–1792. doi:10.1016/j.concog.2010.12.011.
Stahl, J., Mattes, A., Hundrieser, M., Kummer, K., Mück, M., Niessen, E., Porth, E., Siswandari, Y., Wolters, P., Dummel, S., 2020. Neural correlates of error detection during complex response selection: Introduction of a novel eight-alternative response task. Biol. Psychol. 156. doi:10.1016/j.biopsycho.2020.107969.
Steinhauser, M., Yeung, N., 2010. Decision processes in human performance monitoring. J. Neurosci. 30, 15643–15653. doi:10.1523/JNEUROSCI.1899-10.2010.
Team, R.C., 2021. No Title.
Vidal, F., Burle, B., Bonnet, M., Grapperon, J., Hasbroucq, T., 2003. Error negativity on correct trials: a reexamination of available data. Biol. Psychol. 64, 265–282. doi:10.1016/S0301-0511(03)00097-8.
Weidemann, C.T., Kahana, M.J., 2016. Assessing recognition memory using confidence ratings and response times. R. Soc. Open Sci. 3, 150670. doi:10.1098/rsos.150670.
Wessel, J.R., 2012. Error awareness and the error-related negativity: evaluating the first decade of evidence. Front. Hum. Neurosci. 6, 1–16. doi:10.3389/fnhum.2012.00088.
Wessel, J.R., Dolan, K.A., Hollingworth, A., 2018. A blunted phasic autonomic response to errors indexes age-related deficits in error awareness. Neurobiol. Aging 71, 13–20. doi:10.1016/j.neurobiolaging.2018.06.019.

Yeung, N., Botvinick, M.M., Cohen, J.D., 2004. The neural basis of error detection: conflict monitoring and the error-related negativity. Psychol. Rev. 111, 931–959. doi:10.1037/0033-295X.111.4.931.

Yeung, N., Cohen, J.D., 2006. The Impact of Cognitive Deficits on Conflict MonitoringPredictable Dissociations Between the Error-Related Negativity and N2. Psychol. Sci. 17, 164–171.

Yeung, N., Summerfield, C., 2014. Shared mechanisms for confidence judgements and error detection in human decision making. In: The Cognitive Neuroscience of Metacognition. Springer-Verlag, Berlin Heidelberg, pp. 147–167. doi:10.1007/978-3-642-45190-4_7.