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 Centre Jülich, Leo-Brandt-Str. 5, 52425 Jülich, Germany

[b]Melbourne School of Psychological Sciences, University of Melbourne, Parkville Campus, Parkville 3010, Victoria, Australia

[c]Department of Individual Differences and Psychological Assessment, University of Cologne, Pohligstr. 1, 50969 Cologne, Germany

[d]Department of Neurology, Faculty of Medicine, University Hospital Cologne, University of Cologne, Kerpener Str. 62, 50937 Cologne, Germany

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

Email: h.overhoff@fz-juelich.de

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 perceptual 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. $P_{e/c}$ amplitudes varied by confidence rating, but they did not mirror the reduction in metacognitive accuracy. $N_{e/c}$ amplitudes decreased with age except for high confidence correct responses. 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.

Keywords: confidence, error detection, aging, error(-related) negativity (N_e, ERN), error positivity (P_e), behavioral adaptation

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., 2018; 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

66 confidence that a correct decision has been made (Boldt and Yeung, 2015; Yeung and Cohen, 2006;
67 Yeung and Summerfield, 2014).

68 **1.1 Neural correlates of metacognition**

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

85 Concerning the mechanisms underlying these two components, Di Gregorio et al. (2018)
86 designed a sophisticated task to provide evidence that the P_e, but not the N_e, was present when it
87 was evident for participants that an error had been made, but they did not know the correct answer.
88 These findings suggest that the P_e does not require a representation of the correct response to
89 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 Pe 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, Schreiber, et al., 2012; 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) was 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 $\pm$ 1.9 years [all results are indicated as mean $\pm$ 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 and no history of neurological or psychiatric diseases. Any signs of cognitive impairment (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; Figure 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 colour 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]; see Figure 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)

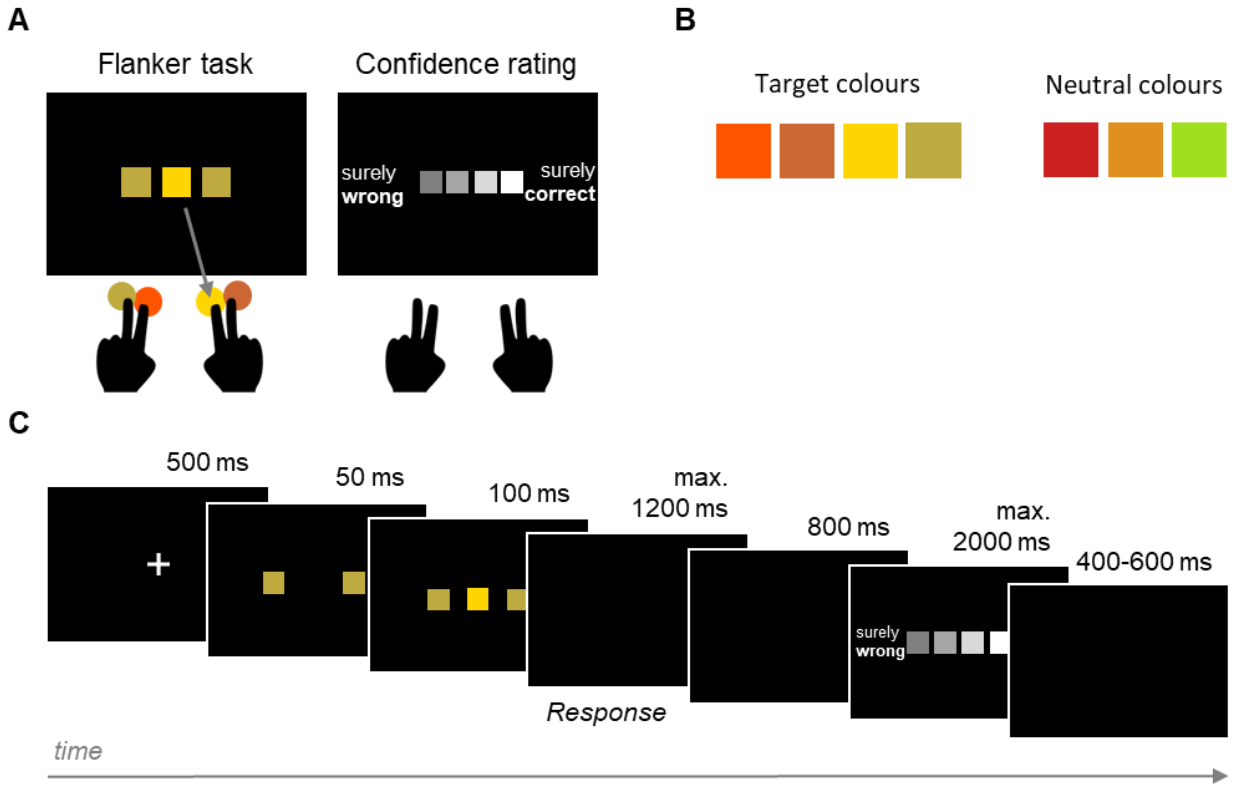


Figure 1. (A) The left panel shows an example of a trial in the flanker task, where one central target and two flankers were presented, and the participant had to press the finger that was assigned to the respective target colour (illustrated by the grey arrow). The confidence rating (right panel) consisted 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.

flankers before the target was expected to increase the induced conflict (Mattler, 2003). We used a response deadline of 1,200 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 Figure 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 S3. 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. For pre-defined conditions of interest (see below), individual median RTs were computed, and the means across participants were entered into group-level statistical analyses.

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 percentage 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 (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 on trial *n*+1 by multiplying the individual percentage correct and median RT (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 to pairs of consecutive valid trials were included. Response caution was computed separately relative to a) initial trial accuracy (error, correct), and b) each confidence category ('surely wrong', 'unsure', 'surely correct') of the initial trial.

Age-related effects on the d2-test score, the error rate, and Phi were computed using Pearson correlations between these behavioural variables and age. To rule out that metacognitive accuracy was confounded by task performance or attention and processing speed, we computed partial correlations between Phi and age, controlling for the individual error rate and d2-test score, respectively.

For the analysis of performance and confidence across the lifespan, we performed a series of one-way repeated measures ANCOVAs with age as the between-subject covariate of interest. The dependent variables were error rate, RT, and response caution, and either accuracy (error, correct), or (pooled) confidence (3 levels) served as the within-subject factor.

Significant main effects of accuracy or confidence were followed up by planned paired-samples *t*-tests comparing the dependent variable between all levels. Significant interactions with age were followed up by Pearson correlations, separately for each level of a given within-subject factor. We used 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 tests were computed using Bonferroni corrected *p*-values. If the assumption of sphericity was violated, degrees of freedom were corrected according to Greenhouse-Geisser, and adjusted *p*-values are reported. Effect sizes are reported as partial eta squared (η^2_p), Cohen's d, or Pearson's correlation coefficient r, respectively. Analyses were run in MATLAB R2019a (The Mathworks, Inc.) and IBM SPSS Statistics for Windows (Version 25.0).

2.6 Electroencephalographic data analysis

Data were response-locked and averaged for each participant, separately for the following categories: a) errors and correct responses (across confidence levels), b) errors ('low confidence' and 'higher confidence'), and c) correct ('high confidence' and 'lower confidence'). Error trials rated as 'surely wrong' were termed 'low confidence'; error trials which received any higher

confidence rating were termed 'higher confidence' (and vice versa for correct trials). The reason for aggregating the trials in this way rather than using three levels as for the behavioural analyses was that errors were most often associated with a 'surely wrong' and correct trials with a 'surely correct' rating, causing trial numbers for the other three rating categories to be too low to be included as separate categories of interest (note that because of data cleaning, typically less trials are available for the analysis of ERP data).

The $N_{e/c}$ local peak amplitude was 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 amplitude was extracted from the interval 150 to 350 ms at Cz based on conventions and visual inspection of the local maxima of the grand-average scalp topographies (Falkenstein et al., 2000; Siswandari et al., 2019).

Only data sets with a minimum of six trials in each condition of the respective comparison were used for statistical analyses (Pontifex et al., 2010). To maximise the number of data-sets available for each analysis of interest, three subgroups of participants were used: Subgroup 1 included 63 participants and was used to examine the effect of accuracy and age on the ERP amplitudes for pooled trials. In addition, we ran focussed control analyses on two smaller subgroups of participants who had a sufficient number of trials to investigate modulations by confidence across the lifespan separately for errors (Subgroup 2, $N = 44$) and correct responses (Subgroup 3, $N = 51$). For the latter analyses, we ensured that participants' age was still well-distributed across all decades within the subgroups, and that age similarly affected the main behavioural parameters as in the full sample (see supplementary material S1).

For statistical analyses of ERP peak amplitudes, we also employed one-way repeated measures ANCOVAs with age as the covariate of interest. For Subgroup 1, the dependent variables were the CSD-transformed N_e and N_c as well as P_e and P_c, amplitudes with accuracy (error, correct) as the within-subject factor (see supplementary material S2 for the analysis with the within-subject

341 factor confidence). For Subgroup 2 (including only errors), the dependent variables were N_e and
342 P_e, and the within-subject factor was confidence ('surely wrong', 'higher confidence'), and lastly,
343 for Subgroup 3 (including only correct responses), the dependent variables were N_c and P_c, and the
344 within-subject factor was again confidence; however, with different levels ('surely correct', 'lower
345 confidence').
346
347 **3. Results**
348 Only significant effects in the ANCOVAs and relevant follow-up tests are reported in this
349 section. For results of all tests, please refer to supplementary Tables S1-S4. Note that all computed
350 ANCOVAs included age as the covariate of interest.
351 **3.1 Behavioural results**
352 ***3.1.1 Attention***
353 The average score for sustained attention and processing speed as assesses by the d2-test
354 was 178.5 $\pm$ 5.6 ($M \pm SEM$) and showed the typical decline for older adults, as shown by a
355 significant correlation between test scores and age [$r(63) = -.553$, $p < .001$].
356 ***3.1.2 Distribution of confidence ratings***
357 In a first step, we were interested in how the confidence ratings were distributed across the
358 four confidence levels across the lifespan (Figure 2). For this, we ran two Friedman ANOVAs for
359 dependent measures for the percentages for each rating category, separately for errors and correct
360 responses.
361 The ANOVA for errors showed that the percentages differed between confidence levels
362 [$X^2(3) = 78.029$, $p < .001$]. On average, most errors were rated as 'surely wrong' (42.8 %) and the
363 smallest proportion of errors as 'maybe wrong' (7.3%). Interestingly, follow-up correlation
364 analyses between age and percentage for each rating category showed that the proportion of 'maybe

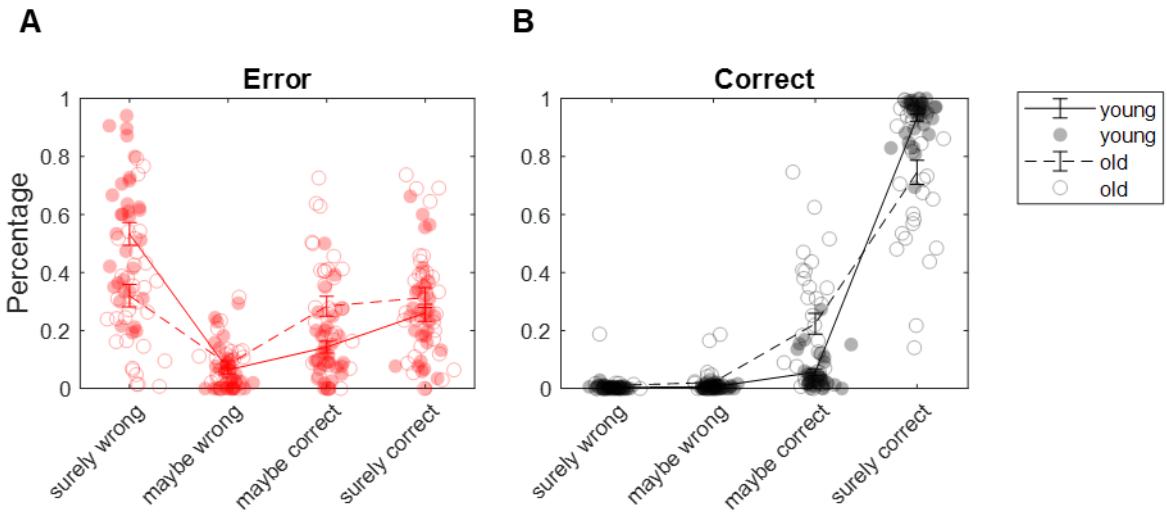


Figure 2. Distribution of confidence ratings for errors (red; A) and correct responses (black; B). Errors were most often rated as 'surely wrong', and correct responses as 'surely correct'. Dots represent the individual proportion of the particular confidence response amongst all errors or correct responses, respectively. A median split by age (*Mdn* = 46) was conducted for illustration purposes. Filled dots and solid lines represent younger adults, empty circles and dashed lines represent older adults. With increasing age, participants used the 'surely correct/wrong' ratings less, but more the middle of the confidence scale.

correct' ratings (21.0%) was increased with higher age [$r(63) = .450$, $p < .001$], whereas the ratio of 'surely wrong' ratings was decreased [$r(63) = -.543$, $p < .001$; all other correlations *ns*; Figure 2A].

For correct responses (Figure 2B), the ANOVA also revealed a main effect of confidence [$X^2(3) = 167.472$, $p < .001$]. Correct responses were most often rated as 'surely correct' (84.2 %) and least often as 'surely wrong' (0.7 %). Again, correlation analyses between age and percentage within each confidence category showed that the proportion of 'maybe correct' ratings (13.8%)

was increased with higher age [$r(63) = .530$, $p < .001$], and was decreased for 'surely correct' ratings [$r(63) = -.532$, $p < .001$; all other correlations *ns*].

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 significantly increased with higher age [$r(63) = .594$, $p < .001$]. The ANCOVA for error rate with the within-subject factor confidence revealed a main effect of confidence [$F(2,124) = 97.426$, $p < .001$, $\eta^2_p = .611$]. The error rate decreased across confidence levels from 92.1 ± 1.4% on trials rated as 'surely wrong' to 34.7 ± 2.6% on trials rated as 'unsure' and 8.1 ± 1.4% on trials rated as 'surely correct'. This showed that, on average, participants' confidence reflected their performance reasonably well (which further supports the notion that the current study's confidence scale was a meaningful assessment tool). Furthermore, the ANCOVA revealed a significant interaction between confidence and age [$F(2,124) = 5.264$, $p = .009$, $\eta^2_p = .078$]. In subsequent correlation analyses between error rate and age for each level of confidence, a higher error rate with older age was only found for the 'surely correct' confidence level [$r(63) = .568$, $p < .001$; both other correlations *ns*].

3.1.4 Response time (RT)

An ANCOVA for mean RT with the within-subject factor accuracy showed the expected slowing with age [$F(1,63) = 18.164$, $p < .001$, $\eta^2_p = .224$; correlation between age and RT: $r(63) = .534$, $p < .001$] and a significant main effect of accuracy [$F(1,63) = 4.188$, $p = .045$, $\eta^2_p = .062$]. A follow-up *t*-test revealed that errors (733.4 ± 13.9 ms) were on average slower than correct

responses [709.1 ± 11.5 ms; $t(64) = 2.937$, $p = .005$, $d = 0.364$]. The interaction between accuracy and age was not significant.

The age-related slowing was similarly present as a main effect of age in the ANCOVA with the within-subject factor confidence [$F(1,57) = 19.159$, $p < .001$, $\eta^2_p = .252$; Figure 4A]. Moreover, the ANCOVA revealed a main effect of confidence [$F(2,114) = 13.132$, $p < .001$, $\eta^2_p = .187$] and a significant interaction between confidence and age [$F(2,114) = 3.793$, $p = .030$, $\eta^2_p = .062$]. Planned follow-up *t*-tests showed that trials associated with the 'unsure' confidence level (805.4 ± 15.0 ms) were significantly slower than trials rated as 'surely correct' [700.0 ± 11.7 ms; $t(58) = 9.198$, $p < .001$, $d = 1.198$] or 'surely wrong' [725.2 ± 16.9 ms; $t(58) = -5.236$, $p < .001$, $d = 0.682$]. Furthermore, trials with the extreme ratings ['surely wrong': $r(57) = .502$, $p < .001$; 'surely correct': $r(57) = .567$, $p < .001$], but not with 'unsure' ratings, were significantly slower with older age.

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

3.1.5 Confidence

A repeated measures ANCOVA examining mean confidence ratings (coded from 1 to 3, i.e., 'surely wrong', 'unsure', and 'surely correct') across all trials revealed a main effect of accuracy [i.e., error vs. correct trials; $F(1,63) = 164.008$, $p < .001$, $\eta^2_p = .722$] and a significant interaction between accuracy and the covariate age [$F(1,63) = 37.433$, $p < .001$, $\eta^2_p = .373$]. The average confidence rating was lower for errors (1.861 ± 0.047) compared to correct responses (2.836 ± 0.026), as confirmed in a follow-up *t*-test [$t(64) = -16.774$, $p < .001$, $d = 2.081$]. Separate follow-up correlation analyses between confidence and age for each accuracy level separately (error, correct) showed that the mean confidence for errors was increased with higher age [$r(63) =$

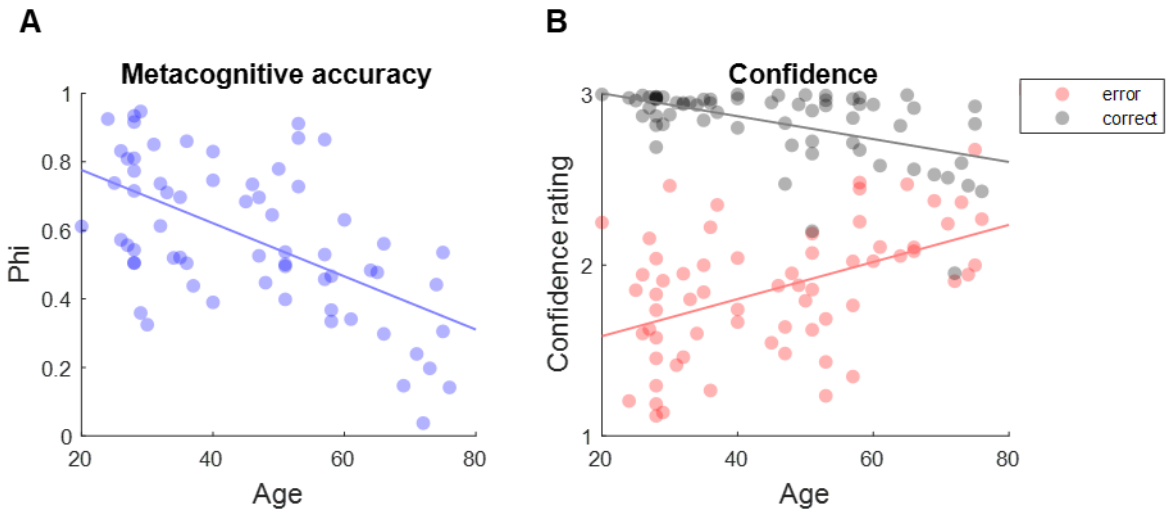


Figure 3. Metacognition across the lifespan. (A) Metacognitive accuracy (Phi) decreased with age. (B) Mean confidence ratings of errors correlated significantly positively with age, while mean confidence ratings of correct trials correlated significantly negatively. Dots represent means of individual participants.

.471, $p < .001$], while the mean confidence was decreased with higher age for correct responses [$r(63) = -.523$, $p < .001$; Figure 3B].

3.1.6 Metacognitive accuracy (Phi)

Phi had a mean of 0.579 ± 0.027 across the entire sample and correlated negatively with age, indicating a decrease of metacognitive accuracy with age [$r(63) = -.582$, $p < .001$; Figure 3A]. Moreover, we computed a partial correlation between Phi and age while controlling for error rate, which still showed a decrease of Phi with age [$r(60) = -.343$, $p = .005$], suggesting that changes in metacognitive accuracy with age were not due to individual differences in task performance. Similarly, a partial correlation controlling for the individual 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 [$r(60) = -.487$, $p < .001$].

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 individual percentage correct and median RT in the subsequent trial. The ANCOVA with the within-subject factor accuracy (referring to the previous trial) revealed a main effect of accuracy [$F(1,63) = 6.929$, $p = .011$, $\eta^2_p = .099$] but no main effect of age, nor a significant interaction. The main effect was confirmed in a follow- up *t*-test between errors and correct responses [$t(64) = 2.950$, $p = .004$, $d = 0.357$]. These findings indicate that participants were on average more cautious after errors than after correct responses, and this effect was independent of age.

Next, we examined whether the response caution in the subsequent trial could also be modulated by confidence. As shown above, confidence and accuracy are strongly correlated; however, a significant modulation by confidence could also indicate that this internal confidence signal drives behavioural adaptations (Figure 4B). The ANCOVA with the within-subject factor confidence (referring to the previous trial) revealed a trend for a main effect of confidence [$F(2,110) = 2.897$, $p = .059$, $\eta^2_p = .050$], but again, no main effect of age, nor a significant interaction. Follow-up *t*-tests between the confidence levels showed that the response caution after trials rated as 'surely correct' was significantly lower compared to trials rated as 'unsure' [$t(56) = 3.448$, $p = .001$, $d = 0.457$] or as 'surely wrong' [$t(56) = 3.066$, $p = .003$, $d = 0.406$].

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 notably, tended to be

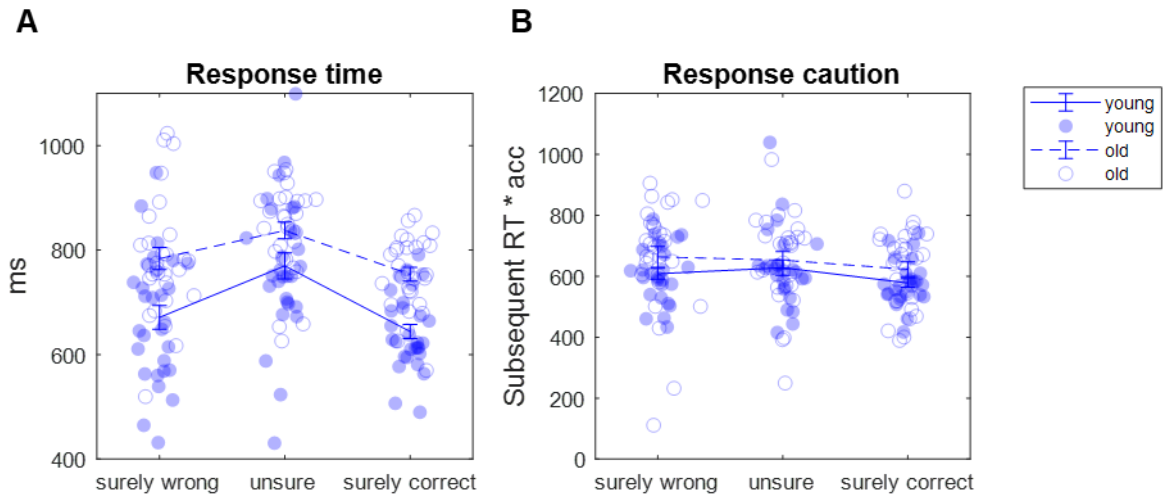


Figure 4. Modulation of response time (RT; A) and response caution (B) by confidence. Participants were split into a younger and an older group relative to the median (*Mdn* = 46) for illustration purposes. Filled dots and solid lines represent younger adults, empty circles and dashed lines represent older adults. (A) Trials rated as 'unsure' showed slower RTs than trials associated with a 'surely' rating, 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 individual average accuracy and RT of subsequent trials.

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 peak amplitude of the $N_{e/c}$ was significantly larger for errors compared to correct responses, as reflected in a main effect of accuracy in the ANCOVA with the within-subject factor accuracy and the covariate of interest age [$F(1,61) = 15.209$, $p < .001$, $\eta^2_p = .200$; $t(62) = -4.544$, $p < .001$, $d = 0.572$; Figure 5A & B]. A main effect of age revealed smaller amplitudes with older age [$F(1,61) = 5.999$, $p = .029$, $\eta^2_p = .076$], and an interaction between accuracy and age showed

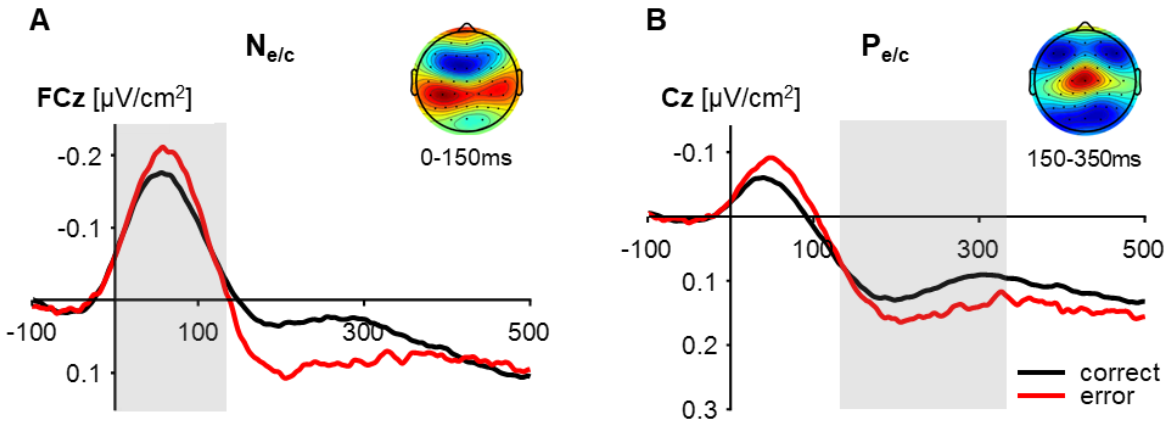


Figure 5. Response-locked event-related potentials for errors and correct responses and topographical maps of errors (N = 63; A, B) after current source density transformation. (A) $N_{e/c}$ is computed at electrode FCz and (B) $P_{e/c}$ at electrode Cz. Errors are shown in red, correct trials in black. 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 analysis of peak amplitudes for the respective components.

that the age effect differed between errors and correct responses [$F(1,61) = 5.999$, $p = .017$, $\eta^2_p = .090$]. Follow-up correlation analyses between the N_e and N_c amplitude and age, separately for errors and correct responses, confirmed that the N_e [$r(61) = .326$, $p = .009$], but not the N_c amplitude was reduced with higher age.

For the analysis of confidence, we ran separate ANCOVAs for errors and correct responses on N_e and N_c amplitude, respectively. Confidence served as the within-subject factor, but was collapsed into two levels (error: 'surely wrong', 'higher confidence'; correct: 'surely correct', 'lower confidence'; see Methods). The ANCOVA for errors showed a main effect of age [$F(1,42) = 10.787$, $p = .002$, $\eta^2_p = .204$], but the main effect of confidence failed to reach significance [$F(1,42) = 3.461$, $p = .070$, $\eta^2_p = .076$]. No significant interaction was found (Figure 6A). Follow-

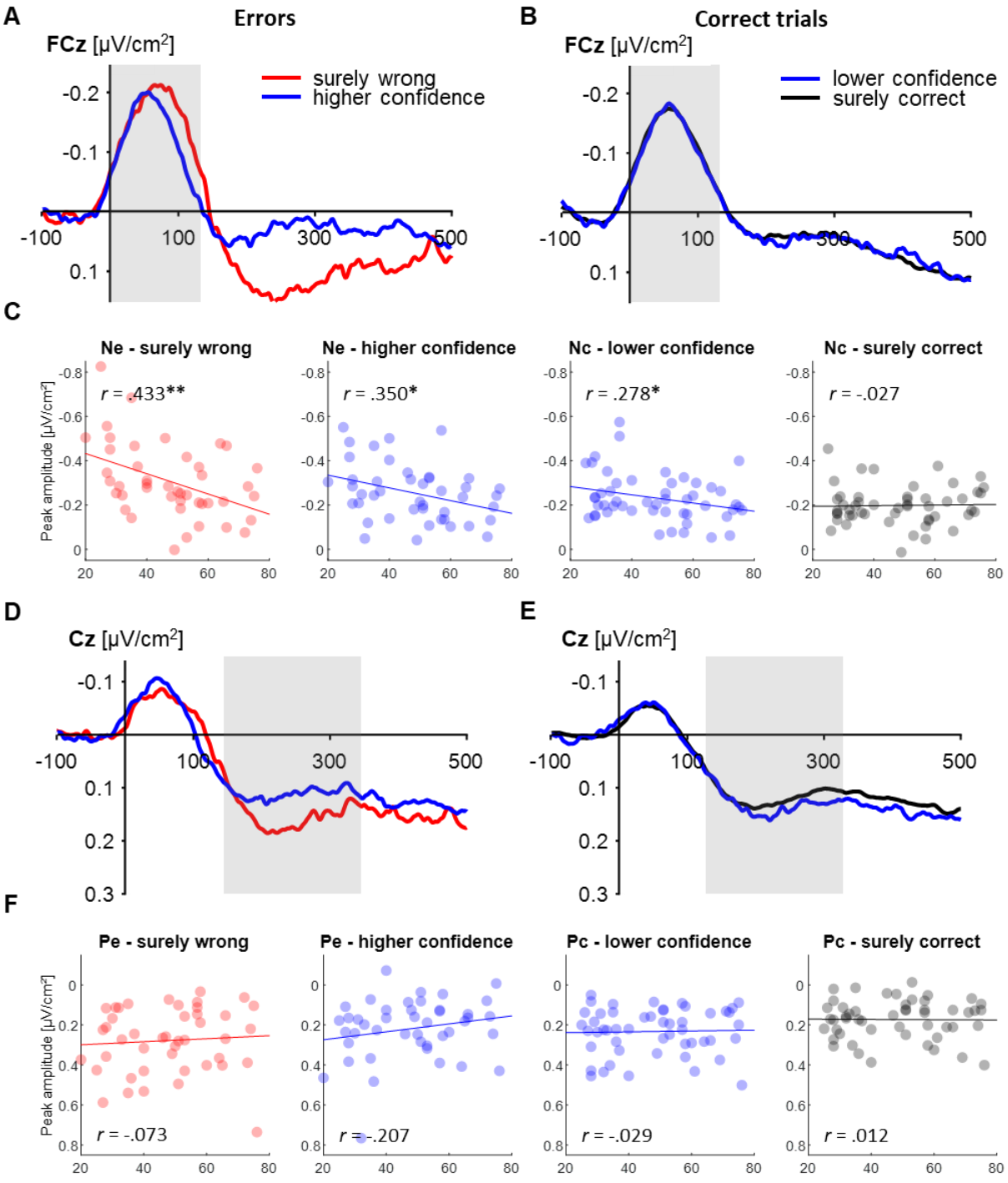
up *t*-tests showed that N_e amplitudes were indeed larger for errors of low confidence compared to higher confidence [$t(43) = -2.309$, $p = .026$, $d = 0.348$].

The ANCOVA for correct responses with the within-subject factor confidence yielded a significant effect of confidence [$F(1,49) = 12.624$, $p = .001$, $\eta^2_p = .205$]. The N_c was not modulated by age, but we found a significant interaction between confidence and age [$F(1,49) = 7.746$, $p = .008$, $\eta^2_p = .136$; Figure 6B]. The main effect was driven by smaller N_c amplitudes after correct responses rated as 'surely correct' compared to those with lower confidence, as reflected in a significant follow-up *t*-test [$t(50) = -2.699$, $p = .009$, $d = 0.438$]. Follow-up correlation analyses pointed to an age-related decrease in N_c amplitude on correct trials with lower confidence ratings [$r(49) = .278$, $p = .048$], but not when they were rated as 'surely correct' (Figure 6C).

3.2.2 $P_{e/c}$ amplitudes

The ANCOVA for the $P_{e/c}$ with the within-subject factor accuracy revealed a significant main effect for accuracy with larger amplitudes for errors compared to correct responses [$F(1,61) = 23.886$, $p < .001$, $\eta^2_p = .281$; $t(62) = 4.507$, $p < .001$,$d = 0.568$; Figure 5C & D]. There was no main effect for age, but a significant interaction between confidence and age [$F(1,61) = 11.836$, $p = .001$, $\eta^2_p = .163$]. Follow-up correlation analyses for errors and correct responses found no significant correlation between the amplitude and age for correct responses; however, the age-related decrease in amplitude for errors marginally missed the significance threshold [$r(61) = -.237$, $p = .061$].

Next, responses were again split by their accuracy, and separate ANCOVAs were conducted with the within-subject factor confidence. The ANCOVA for errors did not yield any significant effects on the P_e amplitude (Figure 6D). However, due to previous evidence of a strong relation between $P_{e/c}$ amplitude and error detection or confidence ratings (Boldt and Yeung, 2015; Nieuwenhuis et al., 2001), we performed an exploratory *t*-test comparing the P_e amplitudes



528

529 *Figure 6.* Response-locked event-related potentials, separately for errors and correct responses
530 after current source density transformation. Errors are shown in the left panel, correct trials in the
531 right panel. For errors (N = 44), data are binned into trials rated as 'surely wrong' (red) and trials
532 rated with any higher confidence (blue). The N_e (A) and N_c (B) are shown at electrode FCz and did

not differ between confidence levels, and P_e (D) and P_c (E) are shown at electrode Cz and were both increased for the trials associated with lower confidence, respectively. (C) and (F) illustrate the correlations between age and the ERP components, separately for errors, correct trials, and confidence levels. The amplitudes of the N_e (both confidence levels) and the N_c (only 'lower confidence') decreased with age, while there was no significant correlation between age and any of the $P_{e/c}$ amplitudes. Dots represent individual average peak amplitudes.

*p < .05. **p < .01. ***p < .001

between the two confidence levels. Indeed, we found a significant difference between 'surely wrong' and 'higher confidence' errors [$t(43) = 2.157$, $p = .037$, $d = 0.325$], suggesting that the P_e was only modulated by confidence, if assessed independent of age.

The ANCOVA for correct trials, in contrast, revealed a main effect of confidence [$F(1,49) = 5.065$, $p = .029$, $\eta^2_p = .094$], and follow-up t-tests showed that the peak amplitude was on average higher for correct responses with lower confidence compared to those rated as 'surely correct' [$t(50) = 5.838$, $p < .001$, $d = 0.818$; Figure 6E]. Again, there was no effect of age on P_c amplitude, neither as a main effect, nor as an interaction.

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 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. While the $N_{e/c}$ amplitude was smaller with older age whenever participants were not entirely certain about their response accuracy, the variation in the $P_{e/c}$ 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 partial correlations 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 'misclassifications', or a general response tendency towards the middle (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, in contrast to previous studies that reported an *over*-confidence in older age (Dodson et al., 2007; Hansson et al., 2008; Ross et al., 2012), we observed *lower* mean confidence due to the tendency of the older adults to use the middle of the confidence scale more often. These findings emphasise the actual difficulties of older adults in evaluating their performance and establishing confidence. In our study, the (relatively) high task difficulty might be a reason for these difficulties. Stahl et al. (2020) showed that their slow errors, which were associated with a weak stimulus-response representation (i.e., due to a weak memory), were associated with lower confidence than fast, impulsive responses. Thus, assuming that the present task posed higher demands on the older adults (as indicated, for instance, by higher error rates), their impaired metacognitive evaluation might partly be 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 $P_{e/c}$ increasing with decreasing confidence, - for the first time - for a very broad age range (Boldt and Yeung, 2015; Rausch et al., 2019). The modulation of the P_e of errors by confidence 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 $P_{e/c}$ 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 $N_{e/c}$ across the lifespan. The control analyses revealed that the $N_{e/c}$ amplitude was also affected by the interaction between confidence and age. With older age, the $N_{e/c}$ declined for all trial types in which

some degree of conflict was perceived. In other words, only the N_c of correct trials that were rated as 'surely correct' showed no amplitude variations 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 any conflicting situation (Yeung et al., 2004).

We suggest that the reduced $N_{e/c}$ amplitude of perceived errors with higher age could be related to the observed decrease in metacognitive accuracy in our flanker task. If older adults had difficulties 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 some (limited) 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, medium and low confidence about a choice was 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 confidence magnitude, the $P_{e/c}$ amplitude, and the behavioural adaptations relative to the reported confidence were consistent across the lifespan. This suggests that it is not metacognitive accuracy per se, but rather the perceived confidence that shapes future behaviour: 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 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).

A second shortcoming 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.

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

677 was well preserved in older adults. Thus, training the metacognitive evaluation of fundamental
678 decisions in older adults might constitute a promising endeavour (and has been shown to work for
679 mathematical problem solving [Pennequin et al., 2010]).

680 **5. Conclusion**

681 The study of error detection and confidence in the context of healthy ageing have advanced
682 largely in parallel. Our study demonstrates that confidence shapes our behavioural and neural
683 processing of decisions and should be considered to investigate age-related effects on error
684 processing and metacognitive abilities. Interestingly, the $N_{e/c}$, but not the $P_{e/c}$ amplitude was
685 differentially modulated by confidence across the lifespan, suggesting that the decreasing accuracy
686 of metacognitive judgements with older age might be related to impaired integration of neural
687 correlates of conflict detection and decision confidence.

688

689 **Acknowledgements**

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

693

694 **Disclosure statement**

695 The authors declare no conflict of interest.

696

References

Boldt, A., Yeung, N., 2015. Shared neural markers of decision confidence and error detection. J. Neurosci. 35, 3478–3484. https://doi.org/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. https://doi.org/10.1093/oxfordhb/9780199336746.013.3

Chua, E., Schacter, D.L., Sperling, R., 2009. Neural basis for recognition confidence. Psychol. Aging 24, 139–153. https://doi.org/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. https://doi.org/10.1016/j.biopsycho.2017.01.012

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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/S0301-0511(99)00031-9

Falkenstein, M., Hoormann, J., Hohnsbein, J., 2001. Changes of error-related ERPs with age.

746 Exp. Brain Res. 138, 258–262. https://doi.org/10.1007/s002210100712
747 Fitzgerald, L.M., Arvaneh, M., Dockree, P.M., 2017. Domain-specific and domain-general
748 processes underlying metacognitive judgments. Conscious. Cogn. 49, 264–277.
749 https://doi.org/10.1016/j.concog.2017.01.011
750 Fleming, S.M., Dolan, R.J., 2012. The neural basis of metacognitive ability.
751 royalsocietypublishing.org 367, 1338–1349. https://doi.org/10.1098/rstb.2011.0417
752 Fleming, S.M., Dolan, R.J., Frith, C.D., 2012. Metacognition: Computation, biology and
753 function. Philos. Trans. R. Soc. B Biol. Sci. 367, 1280–1286.
754 https://doi.org/10.1098/rstb.2012.0021
755 Fleming, S.M., Frith, C.D., 2014. The Cognitive Neuroscience of Metacognition, The Cognitive
756 Neuroscience of Metacognition. Springer. https://doi.org/10.1007/978-3-642-45190-4
757 Fleming, S.M., Lau, H.C., 2014. How to measure metacognition. Front. Hum. Neurosci. 8, 1–9.
758 https://doi.org/10.3389/fnhum.2014.00443
759 Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. “Mini-mental state”. A practical method for
760 grading the cognitive state of patients for the clinician. J. Psychiatr. Res. 12, 189–198.
761 https://doi.org/10.1016/0022-3956(75)90026-6
762 Galvin, S.J., Podd, J. V., Drga, V., Whitmore, J., 2003. Type 2 tasks in the theory of signal
763 detectability: Discrimination between correct and incorrect decisions. Psychon. Bull. Rev.
764 10, 843–876. https://doi.org/10.3758/BF03196546
765 Groom, M.J., Cragg, L., 2015. Differential modulation of the N2 and P3 event-related potentials
766 by response conflict and inhibition. Brain Cogn. 97, 1–9.
767 https://doi.org/10.1016/j.bandc.2015.04.004
768 Hansson, P., Rönnlund, M., Juslin, P., Nilsson, L.G., 2008. Adult Age Differences in the Realism
769 of Confidence Judgments: Overconfidence, Format Dependence, and Cognitive Predictors.

770 Psychol. Aging 23, 531–544. https://doi.org/10.1037/a0012782
771 Harty, S., Murphy, P.R., Robertson, I.H., O'Connell, R.G., 2017. Parsing the neural signatures of
772 reduced error detection in older age. Neuroimage 161, 43–55.
773 https://doi.org/10.1016/j.neuroimage.2017.08.032
774 Harty, S., O'Connell, R.G., Hester, R., Robertson, I.H., 2013. Older adults have diminished
775 awareness of errors in the laboratory and daily life. Psychol. Aging 28, 1032–1041.
776 https://doi.org/10.1037/a0033567
777 Hautzinger, M., 1991. The Beck Depression Inventory in clinical practice. Nervenarzt 62, 689–
778 96.
779 Hertzog, C., 2015. Aging and Metacognitive Control, The Oxford Handbook of Metamemory.
780 https://doi.org/10.1093/oxfordhb/9780199336746.013.31
781 Hertzog, C., Hultsch, D., 2000. Metacognition in adulthood and old age., in: Craik, F.I.M.,
782 Salthouse, T.A. (Eds.), The Handbook of Aging and Cognition. Lawrence Erlbaum
783 Associates Publishers, pp. 417–466.
784 Hoerold, D., Pender, N., Robertson IH, 2013. mende. Neuropsychologia 51, 385–391.
785 Jasper, H., 1958. The ten twenty electrode system of the international federation.
786 Electroencephalogr. Clin. Neurophysiol. 10, 371–375.
787 Kayser, J., Tenke, C.E., 2006. Principal components analysis of Laplacian waveforms as a
788 generic method for identifying ERP generator patterns: II. Adequacy of low-density
789 estimates. Clin. Neurophysiol. 117, 369–380. https://doi.org/10.1016/j.clinph.2005.08.033
790 Kornell, N., Son, L.K., Terrace, H.S., 2007. Transfer of metacognitive skills and hint seeking in
791 monkeys: Research article. Psychol. Sci. 18, 64–71. https://doi.org/10.1111/j.1467-
792 9280.2007.01850.x
793 Korsch, M., Frühholz, S., Herrmann, M., 2016. Conflict-specific aging effects mainly manifest in

794 early information processing stages-an ERP study with different conflict types. Front. Aging
795 Neurosci. 8, 1–12. https://doi.org/10.3389/fnagi.2016.00053
796 Larson, M.J., Clayson, P.E., 2011. The relationship between cognitive performance and
797 electrophysiological indices of performance monitoring. Cogn. Affect. Behav. Neurosci. 11,
798 159–171. https://doi.org/10.3758/s13415-010-0018-6
799 Larson, M.J., Clayson, P.E., Keith, C.M., Hunt, I.J., Hedges, D.W., Nielsen, B.L., Call,
800 V.R.A.V.R.A., 2016. Cognitive control adjustments in healthy older and younger adults:
801 Conflict adaptation, the error-related negativity (ERN), and evidence of generalized decline
802 with age. Biol. Psychol. 115, 50–63. https://doi.org/10.1016/j.biopsycho.2016.01.008
803 Larson, M.J., South, M., Clayson, P.E., 2011. Sex differences in error-related performance
804 monitoring. Neuroreport 22, 44–48. https://doi.org/10.1097/WNR.0b013e3283427403
805 Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: An open-source toolbox for the analysis of
806 event-related potentials. Front. Hum. Neurosci. 8, 1–14.
807 https://doi.org/10.3389/fnhum.2014.00213
808 Lucci, G., Berchicci, M., Spinelli, D., Taddei, F., Di Russo, F., 2013. The Effects of Aging on
809 Conflict Detection. PLoS One 8. https://doi.org/10.1371/journal.pone.0056566
810 Maier, M.E., Steinhauser, M., 2017. Working memory load impairs the evaluation of behavioral
811 errors in the medial frontal cortex. Psychophysiology 54, 1472–1482.
812 https://doi.org/10.1111/psyp.12899
813 Maier, M.E., Steinhauser, M., Hübner, R., 2008. Is the Error-related Negativity Amplitude
814 Related to Error Detectability? Evidence from Effects of Different Error Types. J. Cogn.
815 Neurosci. 20, 2263–2273.
816 Maniscalco, B., Lau, H., 2012. A signal detection theoretic approach for estimating
817 metacognitive sensitivity from confidence ratings. Conscious. Cogn. 21, 422–430.

https://doi.org/10.1016/j.concog.2011.09.021

Mattler, U., 2003. Delayed flanker effects on lateralized readiness potentials. Exp. Brain Res. 151, 272–288. https://doi.org/10.1007/s00221-003-1486-5

Mecacci, L., Righi, S., 2006. Cognitive failures, metacognitive beliefs and aging. Pers. Individ. Dif. 40, 1453–1459. https://doi.org/10.1016/j.paid.2005.11.022

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. https://doi.org/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.E.M.M., Weiss, P.H., Stahl, J., 2017. Error detection across the adult lifespan: Electrophysiological evidence for age-related deficits. Neuroimage 152, 517–529. https://doi.org/10.1016/j.neuroimage.2017.03.015

Nieuwenhuis, S., 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. https://doi.org/10.1017/S0048577201001111

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9, 97–113. https://doi.org/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. https://doi.org/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. https://doi.org/10.1007/s10804-010-9098-3

842 Perrin, F., Pernier, J., Bertrand, O., Echallier, J.F., 1989. Spherical splines for scalp potential and
843 current density mapping. Electroencephalogr. Clin. Neurophysiol. 72, 184–187.
844 https://doi.org/10.1016/0013-4694(89)90180-6
845 Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118,
846 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019
847 Pontifex, M.B., Scudder, M.R., Brown, M.L., O'Leary, K.C., Wu, C.T., Themanson, J.R.,
848 Hillman, C.H., 2010. On the number of trials necessary for stabilization of error-related
849 brain activity across the life span. Psychophysiology 47, 767–773.
850 https://doi.org/10.1111/j.1469-8986.2010.00974.x
851 Pouget, A., Drugowitsch, J., Kepecs, A., 2016. Confidence and certainty: distinct probabilistic
852 quantities for different goals. nature.com. https://doi.org/10.1038/nn.4240
853 Rabbitt, P., 1990. Age, IQ and awareness, and recall of errors. Ergonomics 33, 1291–1305.
854 https://doi.org/10.1080/00140139008925333
855 Rabbitt, P.M.A., 1966. Errors and error correction in choice-response tasks. Exp. Psychol. 71,
856 264–272. https://doi.org/10.1037/h0022853
857 Rahnev, D., Desender, K., Lee, A.L.F., Adler, W.T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova,
858 P., Atlas, L.Y., Balcı, F., Bang, J.W., Bègue, I., Birney, D.P., Brady, T.F., Calder-Travis, J.,
859 Chetverikov, A., Clark, T.K., Davranche, K., Denison, R.N., Dildine, T.C., Double, K.S.,
860 Duyan, Y.A., Faivre, N., Fallow, K., Filevich, E., Gajdos, T., Gallagher, R.M., de Gardelle,
861 V., Gherman, S., Haddara, N., Hainguerlot, M., Hsu, T.Y., Hu, X., Iturrate, I., Jaquiery, M.,
862 Kantner, J., Koculak, M., Konishi, M., Koß, C., Kvam, P.D., Kwok, S.C., Lebreton, M.,
863 Lempert, K.M., Ming Lo, C., Luo, L., Maniscalco, B., Martin, A., Massoni, S., Matthews, J.,
864 Mazancieux, A., Merfeld, D.M., O'Hora, D., Palser, E.R., Paulewicz, B., Pereira, M., Peters,
865 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. https://doi.org/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. https://doi.org/10.1101/860379

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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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.

https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/j.biopsycho.2020.107969

Steinhauser, M., Yeung, N., 2010. Decision processes in human performance monitoring. J. Neurosci. 30, 15643–15653. https://doi.org/10.1523/JNEUROSCI.1899-10.2010

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. https://doi.org/10.1016/S0301-0511(03)00097-8

Wessel, J.R., 2012. Error awareness and the error-related negativity: evaluating the first decade of evidence. Front. Hum. Neurosci. 6, 1–16. https://doi.org/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. https://doi.org/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.

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

915 Yeung, N., Cohen, J.D., 2006. The Impact of Cognitive Deficits on Conflict

916 MonitoringPredictable Dissociations Between the Error-Related Negativity and N2.

917 Psychol. Sci. 17, 164–171.

918 Yeung, N., Summerfield, C., 2014. Shared mechanisms for confidence judgements and error

919 detection in human decision making, in: The Cognitive Neuroscience of Metacognition.

920 Springer-Verlag Berlin Heidelberg, pp. 147–167. https://doi.org/10.1007/978-3-642-45190-

921 4_7

922