NeuroImage 218 (2020) 116963

Contents lists available at ScienceDirect

NeuroImage

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

Cognitive modelling reveals distinct electrophysiological markers of decision confidence and error monitoring

Manuel Rausch [a,*], Michael Zehetleitner [a], Marco Steinhauser [b], Martin E. Maier [b]

[a] *Katholische Universität Eichstätt-Ingolstadt, Fakultät für Psychologie und Pädagogik, Professur für Allgemeine Psychologie II, Eichstätt, Germany*
[b] *Katholische Universität Eichstätt-Ingolstadt, Fakultät für Psychologie und Pädagogik, Lehrstuhl für Allgemeine Psychologie, Eichstätt, Germany*

ARTICLE INFO

Keywords:
Metacognition
Confidence
Event-related potentials
Cognitive modelling
ERN/Ne
Pe

ABSTRACT

Is confidence in perceptual decisions generated by the same brain processes as decision itself, or does confidence require metacognitive processes following up on the decision? In a masked orientation task with varying stimulus-onset-asynchrony, we used EEG and cognitive modelling to trace the timing of the neural correlates of confidence. Confidence reported by human observers increased with stimulus-onset-asynchrony in correct and to a lesser degree in incorrect trials, a pattern incompatible with established models of confidence. Electrophysiological activity was associated with confidence in two different time periods, namely 350–500 ms after stimulus onset and 250–350 ms after the response. Cognitive modelling revealed that only the activity following on the stimulus exhibited the same statistical regularities as confidence, while the statistical pattern of the activity following the response was incompatible with confidence. It is argued that electrophysiological markers of confidence and error awareness are at least in parts distinct.

1. Introduction

Decision confidence is a ubiquitous feature of human decision making: Whenever we make a choice, the decision is accompanied by a greater or smaller degree of confidence that the choice is correct. Confidence can be defined as an evaluation of one's decision making, resulting in a degree of certainty that the decision is correct (Pouget et al., 2016). How does the brain give rise to confidence? Two conflicting views have been proposed: According to one view, confidence may be generated directly by the very same brain processes that are involved in decision formation (Kepecs et al., 2008; Kiani and Shadlen, 2009; Rolls et al., 2010; Vickers, 1970). According to the second view, confidence is generated by a separate, metacognitive process that gives rise to both confidence and error awareness (Boldt and Yeung, 2015; Charles and Yeung, 2018). A common mechanism underlying error monitoring and decision confidence may be on-going accumulation of sensory evidence after the decision, allowing observers to reverse their belief about the stimulus (Pleskac and Busemeyer, 2010; Resulaj et al., 2009; Steinhauser et al., 2008; van den Berg et al., 2016).

The aim of the present study was to test if there are neural correlates of confidence in a perceptual decision already before the time of the behavioural response consistent with a common origin of confidence and choice formation, or if these correlates do not emerge until the time of neural markers of error awareness following the response. For this purpose, the present study used cognitive modelling and electroencephalography to trace the timing of the neural correlates of confidence in perceptual decisions.

1.1. *Event-related potential correlates of confidence*

The present study examines three event-related potential (ERP) components that were previously proposed as correlates of confidence: the P3 (Hillyard et al., 1971), the error-related negativity ERN (Scheffers and Coles, 2000), and the error-related positivity Pe (Boldt and Yeung, 2015). The P3 is an ERP component recorded over central and parietal electrodes peaking 300–500 ms after the presentation of a task-relevant stimulus. It is a natural candidate for a shared electrophysiological correlate of confidence and the decision because the parietal P3 was suggested as a marker of accumulated evidence in perceptual decision making tasks (O'Connell et al., 2012; Philiastides et al., 2014; Twomey et al., 2015). Previous studies showed that P3 amplitudes are correlated with confidence judgments (Eimer and Mazza, 2005; Hillyard et al., 1971). In addition, the P3 showed statistical properties expected from a Bayesian model of decision confidence in a vibrotactile forced-choice

[*] Corresponding author. Katholische Universität Eichstätt-Ingolstadt. Psychologie II, Ostenstraße 25, 85072, Eichstätt, Germany.
E-mail address: manuel.rausch@ku.de (M. Rausch).

https://doi.org/10.1016/j.neuroimage.2020.116963
Received 13 January 2020; Received in revised form 5 May 2020; Accepted 14 May 2020
Available online 24 May 2020

task (Herding et al., 2019). However, a marker of accumulated evidence is by far not the only interpretation of the P3: According to a classical theory, the P3 reflects updating of working memory in response to task-relevant events (Donchin and Coles, 1988). Other theories include the global broadcast of visual contents within a neural global workspace (Sergent et al., 2005), the mobilization for action following motivationally significant stimuli (Nieuwenhuis et al., 2011), or a monitoring process if the decision is correctly transformed into an action (Verleger et al., 2005).

ERN and Pe are established makers of error processing: If one shared neurocognitive mechanism gives rise to both confidence and error monitoring, confidence should be associated with ERN and Pe. The ERN is an ERP component with frontocentral topography at the same time of shortly after incorrect responses (Falkenstein et al., 1991; Gehring et al., 1993). An equivalent yet smaller negativity referred to as CRN was observed after correct responses (Vidal et al., 2003). Previous studies suggested that the ERN was associated with participants' confidence judgments in a flanker task (Scheffers and Coles, 2000). However, the ERN failed to predict graded confidence judgments on a trial-to-trial basis in a visual discrimination task with briefly flashed stimuli (Boldt and Yeung, 2015). Finally, the ERN can be dissociated from decision confidence by the relation with subjective visibility: In a masked number discrimination task, the ERN varied in an all-or-nothing way and was only present if there was a conscious percept of the stimulus, while confidence varied continuously and did not depend on a conscious percept of the stimulus (Charles et al., 2014, 2013).

The Pe is a parietally focused positive deflection 200–500 ms after incorrect responses. The Pe is similar to the parietal P3 in terms of topography and latency although the Pe is locked to the response, and P3 to the stimulus (Overbeek et al., 2005). The Pe is a marker of conscious awareness of having committed an error (Nieuwenhuis et al., 2001) and can be dissociated from the ERN: In a study where participants responded to a masked target stimulus surrounded by visible flanker stimuli, erroneous responding to the flanker elicited only a Pe, but not an ERN (Di Gregorio et al., 2018). The Pe can be explained by the strength of accumulated evidence of having made an error (Steinhauser and Yeung, 2012, 2010; Ullsperger et al., 2010; Wessel et al., 2011). Moreover, in a visual discrimination task, the Pe was associated with both confidence in correct responses as well as the subjective belief of having made an error in a gradual way (Boldt and Yeung, 2015). However, the timings of ERN and Pe are not immediately plausible for correlates of decision confidence. As it seems that confidence is experienced already at a point in time when no response has yet been made, correlates of confidence may naïvely be expected before the response, at the same time as the decision or shortly afterwards. And yet, ERN and Pe do not occur until after the response.

1.2. Statistical properties of decision confidence

How can hypothesized neural correlates of confidence be tested? If specific neural activity is a correlate of confidence, it must be associated with the same statistical regularities as confidence judgments (Kepecs et al., 2008; Sanders et al., 2016): By implication, if the statistical regularities of a specific ERP component are incompatible with those of confidence, that component is not a correlate of confidence. In the present study, we tracked the statistical regularities of confidence by fitting a series of cognitive models to the behavioural data. The model that fitted the behaviour best was used to predict the neuronal data. Previous studies used the so-called folded X-pattern as a statistical marker of confidence (Braun et al., 2018; Fetsch et al., 2014; Herding et al., 2019; Lak et al., 2017; Urai et al., 2017). The folded X-pattern is characterised by an increase of confidence with stimulus strength in correct trials and a decrease of confidence with stimulus strength in incorrect trials and was derived from Bayesian decision theory (Hangya et al., 2016; Sanders et al., 2016), but also follows from signal detection theory (Kepecs et al., 2008) or postdecisional accumulation models (Moran et al., 2015). However, the folded X-pattern can be misleading about confidence because Bayesian decision theory is compatible with other statistical patterns, too (Adler and Ma, 2018; Rausch and Zehetleitner, 2019a). In addition, in some tasks, confidence empirically increased with stimulus strength in correct trials and to a lesser degree in incorrect trials (Kiani et al., 2014; Rausch et al., 2018; Stolyarova et al., 2019; van den Berg et al., 2016), a pattern we refer to as double increase pattern. The double increase pattern can be reproduced by a smaller number of mathematical models, including the weighted evidence and visibility (WEV) model (Rausch et al., 2018), the heuristic detection model (Maniscalco et al., 2016; Peters et al., 2017), and some Bayesian models (Adler and Ma, 2018; Rausch and Zehetleitner, 2019a). For these reasons, it is not legitimate to assume a specific statistical pattern a priori. However, irrespective of whether confidence follows the folded-X or double increase pattern in a specific task, a neural correlate of confidence should always show the same pattern as the one observed with confidence judgments. In addition, a cognitive model fitted to confidence judgments should also be able to accurately predict the neural correlate of confidence.

1.3. Rationale of the present study

To assess the timing of the neural correlates of confidence in perceptual decisions, human observers performed a masked orientation discrimination task (see Fig. 1) while EEG was recorded. After each single response, observers reported their confidence on a scale with the categories, “not at all”, “a little”, “nearly sure”, and “completely sure”. In case observers were aware of an incorrect response, observers were instructed to respond, “not at all”. We used a task where confidence followed the double increase pattern in previous studies (Rausch et al., 2018), because the double increase pattern can be explained by a smaller number of cognitive models. The strength of stimulation was manipulated by varying the stimulus-onset-asynchrony (SOA), i.e. the time between onset of the stimulus and the mask. Bayes factors were used for statistical inference, allowing us to quantify both the evidence for an effect as well as evidence against an effect (Rouder et al., 2009).

To trace the statistical regularities underlying confidence, we fitted nine previously proposed models of decision confidence to confidence judgements, including.

- SDT (Green and Swets, 1966; Macmillan and Creelman, 2005; Wickens, 2002)
- SDT with noise superimposed on ratings (Maniscalco and Lau, 2016)
- SDT model with postdecisional accumulation (Barrett et al., 2013)
- the two-dimensional Bayesian model (Aitchison et al., 2015)
- the noisy decay model (Maniscalco and Lau, 2016)
- the detection heuristic model (Peters et al., 2017)
- the two high-threshold model (Kellen and Klauer, 2015)
- the two channel model (Rausch and Zehetleitner, 2017)
- the WEV-model (Rausch et al., 2018).

Because the objective of the present study was specifically confidence, we selected models from the literature that are fitted directly to confidence judgments, and not to reaction times. From the model parameters that fitted the behavioural data best, we generated a prediction about the ERP amplitudes.

The present study was designed to test the following predictions: With respect to confidence judgments, we expected that confidence increases as a function of the SOA both in correct as well as in incorrect trials, i.e. confidence is characterised by the double increase pattern. Regarding proposed ERP correlates of confidence, as correct responses are commonly associated with more positive activity at the time of the P3 (Koivisto and Revonsuo, 2010), we hypothesized that confidence is positively associated with EEG activity at the time of the P3. As errors are known to cause negative shifts at the time of the ERN, again a positive association was expected between confidence and activity at the time of

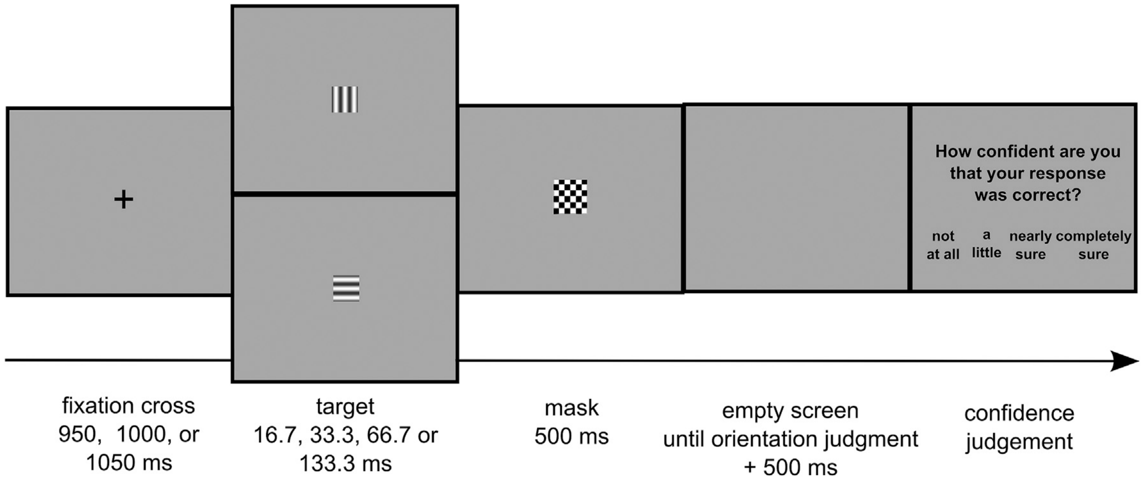

How confident are you that your response was correct?
not at all
a little
nearly sure
completely sure
fixation cross 950, 1000, or 1050 ms
target 16.7, 33.3, 66.7 or 133.3 ms
mask 500 ms
empty screen until orientation judgment + 500 ms
confidence judgement

Fig. 1. Sequence of events during the Experiment. The target stimulus was a sinusoidal grating, oriented horizontally or vertically. After 16.7, 33.3, 66.7, 133.3 ms, the target was replaced by a chequered mask presented for 500 ms. Afterwards, observers reported first the orientation of the target and then their degree of confidence in having made the correct orientation response. Observers were instructed that accuracy but not speed was critical for both responses.

the ERN (Scheffers and Coles, 2000). In contrast, as errors are known to cause positive shifts at the time of the Pe, we predicted a negative association between confidence and activity at the time of the Pe in line with previous research (Boldt and Yeung, 2015). Moreover, if P3, ERN, and Pe were indeed correlates of confidence, the statistical pattern as a function of SOA and choice accuracy should correspond to the statistical pattern observed in confidence judgments: This means that P3, ERN, and Pe should be characterised by the double increase pattern as well. Regarding cognitive models, we expected that the best fit to the behavioural data should be achieved by one of the models that is in principle able to accommodate the double increase pattern, i.e. the WEV-model, the heuristic detection model, or the noisy decay model. Finally, the models that provide an adequate fit to the behavioural data should also accurately predict the ERP correlates of confidence.

2. Material and methods

2.1. Participants

25 human participants (21 female, 4 male) took part in the experiment. The age of the participants ranged between 18 and 36 years ($Md =$ 22). All participants reported normal or corrected-to-normal vision, no history of neuropsychological or psychiatric disorders and not to be on psycho-active medication. All participants gave written informed consent and received either course credits or €8 per hour for participation. The experimental protocol was approved by the ethics committee of the Catholic University of Eichstätt-Ingolstadt.

2.2. Apparatus and stimuli

The experiment was performed a sound-attenuated and electrically shielded cabin. The stimuli were presented on an Iiyama MS103DT monitor with screen diagonal of 51 cm, set at a resolution of 1280 × 1024 px and refresh rate of 60 Hz. The viewing distance, not enforced by constraints, was approximately 60 cm. The experiment was conducted using PsychoPy v. 1.83.04 (Peirce, 2009, 2007) on a Fujitsu Celsius W530 desktop computer with Windows 8.1. The target stimulus was a square (size $3^\circ \times 3^\circ$), textured with a sinusoidal grating with one cycle per degree of visual angle (maximal luminance: 44 cd/m^2; minimal luminance: 14 cd/m^2). The mask consisted of a square ($4^\circ \times 4^\circ$) with a black (0 cd/m^2) and white (60 cd/m^2) chequered pattern consisting of 5 columns and rows. All stimuli were presented at fixation in front of a grey (29 cd/m^2) background. The orientation of the grating varied randomly between horizontal or vertical. Participants reported the orientation of the grating with their right hands by pressing the down key when the grating was vertical and the right key when the grating was horizontal. Likewise, participants reported their confidence in being correct with their left hands by pressing one, two, three, or four on the number keys in top row of the keyboard.

2.3. Experimental trial

Each trial began with the presentation of a fixation cross for a duration randomly chosen between 950, 1000, and 1050 ms, after which the target stimulus appeared. The duration of the fixation cross and thus the onset of the target stimulus was varied to minimize preparatory EEG activity before the onset of the target. Then the target stimulus was shown for a short period of time until it was replaced by the masking stimulus. There were four different possible SOAs, i.e. time periods between target onset and mask onset: 16.7, 33.3, 66.7 and 133.3 ms. The mask was presented for 500 ms. When the mask had disappeared, an empty screen was shown. Participants then indicated whether the target had been horizontal or vertical. The question “How confident are you about your response?” with the four response options “not at all”, “a little”, “nearly sure”, and “completely sure” was displayed 500 ms after the response to ensure that the confidence scale did not interfere with ERN and Pe. Participants then pressed a key to indicate their degree of confidence that their orientation response was correct. If participants had indicated the incorrect orientation of the target, the word *error* was displayed for 1000 ms before the trial ended.

2.4. Design and procedure

Participants were instructed to report the orientation of the grating as accurately as possible without time pressure and to guess the orientation of the target if they had no idea about the orientation at all. In addition, they were instructed that they should report their degree of confidence that their orientation response had been correct, they should report their confidence as accurately as possible and that if they were aware that they had made an error, they should rate their degree of confidence as “not at all”.

The experiment consisted of one training block and 24 experimental blocks of 40 trials each. Each SOA featured 10 times in each block in random order. The orientation of the target stimulus varied randomly across trials. After each block, the percentage of errors was displayed to provide participants with feedback about their accuracy. The whole experimental session took approximately 1.5 h.

2.5. EEG acquisition

The electroencephalogram (EEG) was recorded from 64 electrodes using a BIOSEMI Active-Two system (BioSemi, Amsterdam, Netherlands; Ag/AgCl electrodes, channels Fp1, AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C1, C3, C5, T7, TP7, CP5, CP3, CP1, P1, P3, P5, P7, P9, PO7, PO3, O1, Iz, Oz, POz, Pz, CPz, Fpz, Fp2, AF8, AF4, AFz, Fz, F2, F4, F6, F8, FT8, FC6, FC4, FC2, FCz, Cz, C2, C4, C6, T8, TP8, CP6, CP4, CP2, P2, P4, P6, P8, P10, PO8, PO4, O2 as well as the left and right mastoid, relative to common mode sense CMS active electrode and driven right leg DR passive electrode). Vertical and horizontal electrooculogram (EOG) was recorded from electrodes above and below the right eye and on the outer canthi of both eyes. All electrodes were off-line re-referenced to linked mastoids. EEG and EOG data were continuously recorded at a sampling rate of 512 Hz.

2.6. EEG analysis

The analysis of the EEG data was performed using MNE-Python v. 0.16.1 (Gramfort et al., 2014, 2013). First, the data was re-referenced to the linked mastoids. Next, the signal was band-pass filtered between 0.5 and 40 Hz by windowed finite impulse response filtering. The raw data was inspected visually to remove extreme noise events and artefact-contaminated electrodes. Then, we conducted an independent component analysis based on the fastica algorithm (Hyvärinen, 1999), identifying components representing blinks and/or horizontal eye movements and removing these artefacts before back-projection of the residual components.

The continuous EEG was segmented into two different kinds of epochs, one of which locked to the onset of the target stimulus, and one locked to the orientation response. Stimulus-locked epochs started 200 ms before stimulus onset, ended 600 ms after stimulus onset and were baseline-corrected to the 200 ms pre-stimulus interval. Response-locked epochs started 200 ms before the orientation response, ended 600 ms after the response, and were baseline-corrected to the time range between 150 and 50 ms before the response. The baseline for the response-locked time windows was chosen to avoid an overlap between baseline and ERN because sometimes ERN activity starts slightly prior to the completion of the motor response (Riesel et al., 2013). Epochs with amplitude changes greater than 100 μV were excluded from analysis, the same exclusion criterion as in a previous study of EEG correlates of confidence (Boldt and Yeung, 2015). Finally, ERP waveforms were obtained by averaging across epochs (but not for the validation of the prediction by the cognitive models, see below). EEG activity in specific time windows was quantified by calculating mean amplitudes because mean amplitudes are robust to different numbers of trials across conditions (Luck, 2014). The time windows were 350–500 ms poststimulus at electrode Pz for the P3, -40 – 60 ms after the orientation discrimination response at electrode FCz for the ERN, and 250–350 ms after the orientation discrimination response at electrode Pz for the Pe, the same time windows as in a previous study of EEG correlates of confidence (Boldt and Yeung, 2015). The time window of 350–500 after stimulus onset excluded the point in time when participants responded to the orientation of the stimulus (reaction time at the SOA of 133.3 ms: $M = 624.6$ ms, at the SOA of 16.7 ms: $M = 1001.0$ ms). As can be seen from Fig 5L, the Pe in the present study seemed to stretch over a more prolonged time window than 250–350 ms postresponse. For this reason, we repeated all analyses using a time window of 200–500 ms postresponse, which came to the same interpretation of the data. After artefact rejection, there were on average $M = 816.3$ stimulus-locked epochs and $M = 800.0$ response-locked epochs per participant. In the condition with the smallest number of trials (incorrect trials at the SOA of 133.3 ms), there were on average 10.1 trials. To create topographical maps, artefact-contaminated electrodes that were excluded in the beginning were interpolated using spherical splines (Perrin et al., 1989).

2.7. Model specification

Nine models were fitted to the combined distributions of orientation discrimination and confidence judgments, separately for each single participant.

i. SDT
ii. SDT with noise superimposed on ratings
iii. Noisy decay model
iv. WEV-model
v. Two channel model
vi. SDT model with postdecisional accumulation
vii. Detection heuristic model
viii. Two-dimensional Bayesian model
ix. Two high-threshold model

For all nine models, we assumed that the stimulation is comprised out of two experimental variables, the identity of the stimulus S_{id} and the strength of the stimulus S_s. Participants select a discrimination response $R_{id} \in \{0, 1\}$ about the identity of the stimulus $S_{id} \in \{0, 1\}$ and confidence judgment out of an ordered set of confidence categories $C \in \{1, 2, 3, 4\}$. Models (i)-(vi) were derived from SDT and assumed the same architecture for the choice about the identity of the stimulus. In contrast, models (vii)-(ix) were based on different decision architectures.

2.7.1. SDT derived models

Models (i) -(vi) assumed that a decision about the identity of the stimulus is made based on a comparison between a continuous decision variable for the discrimination judgment δ_{id} with the free criterion parameter θ_{id}. Participants responded $R = 0$, when $\delta_{id} < \theta_{id}$, and $R = 1$ if $\delta_{id} > \theta_{id}$.The decision variable for the discrimination judgment δ_{id} was modelled as a random sample from a Gaussian distribution $\mathcal{N}$:

$$\delta_{id} \tilde{} \mathcal{N}(\mu = (S_{id} - 1/2) \times S_s, \sigma = \sigma_{id}) \tag{1}$$

The stimulus strength S_s was a free parameter specific to each SOA. When $S_{id} = 0$, the distribution of $\delta_{id}$was shifted to the left by the distance of $S_s/2$. When $S_{id} = 1$, the distribution is shifted by the same distance to the right. Thus, S_s denotes the distance of the distributions generated by the two possible identities of the stimulus and is in this respect equivalent to the sensitivity parameter d' in standard SDT. Concerning the standard deviation σ_{id}, model fitting was repeated using two different assumptions about σ_{id} to ensure that the results were robust. For the first set of analyses, $\sigma_{id}$was fixed at 1 for both identities of the stimulus, thus the free parameter S_s fully accounted for each participant's d' at each SOA. For the second set of analyses, σ_{id} could vary depending on S_{id} : An additional free parameter λ characterised the relationship between the variability of δ_{id} associated the two possible identities of the stimulus.

$$\begin{aligned} \sigma_{id0} &= \lambda \\ \sigma_{id1} &= 1/\lambda \end{aligned} \tag{2}$$

A specific degree of confidence was determined by comparing the decision variable for confidence δ_c against a set of three criteria θ_c. Each criterion delineated between two adjacent categories of confidence, e.g. participants selected the category 2 if δ_c fell between θ_{c1} (which separated category 1 and 2) and θ_{c2} (which separated category 2 and 3). To be consistent with standard SDT, we fitted three different criteria for each of the two response options. The different models were characterised by different ways how δ_c was determined.

2.7.1.1. SDT rating model.

According to model (i), the decision variables for identification and confidence were identical:

$$\delta_c = \delta_{id} \tag{3}$$

2.7.1.2. Noisy SDT model.

According to model (ii), δ_c was sampled from a Gaussian distribution, with a mean equal to the decision variable δ_{id} and the standard deviation σ_c, which was an additional free parameter:

$$\delta_c \tilde{} \mathcal{N}(\mu = \delta_{id}, \sigma = \sigma_c) \tag{4}$$

2.7.1.3. Noisy decay model. According to model (iii), δ_c was also sampled from a Gaussian distribution with the standard deviation σ_c. Just as in the noisy SDT model, the mean of δ_c depended on δ_{id}. However, in contrast to the noisy SDT model, according to the noisy decay model, $\delta_{id}$was reduced by multiplication with a signal reduction parameter ρ_S. The signal reduction parameter ρ_S was a separate free parameter for each SOA and was bounded between 0 and 1.

$$\delta_c \tilde{} \mathcal{N}(\mu = \delta_{id} \times \rho_s, \sigma = \sigma_c) \tag{5}$$

2.7.1.4. WEV model. The conceptual idea underlying the WEV-model is that the observer combine evidence about the choice-relevant feature of the stimulus with strength of evidence about choice-irrelevant features to select one out of several confidence categories (Rausch et al., 2018; Rausch and Zehetleitner, 2019b). Evidence about choice-irrelevant features of the stimulus can improve confidence judgment because they allow the observer to estimate the reliability of the percept more precisely. A possible neural mechanism may involve posterior parietal cortex and ventral striatum, which were found to track sensory reliability independently of the choice (Bang and Fleming, 2018).

The express this idea in formal terms, the WEV model assumed that δ_c was again sampled from a Gaussian distribution with the standard deviation σ_c :

$$\delta_c \tilde{} \mathcal{N}(\mu = (1-w) \times \delta_{id} + w \times (2R_{id} - 1) \times (S_s - \overline{S_s}), \sigma = \sigma_c) \tag{6}$$

Formula (6) ensured that the centre of the distribution was shifted towards 0 when the strength of stimulation S_s was low and away from 0 when S_s was high. The parameter w captured the degree to which participants relied on sensory evidence about the identity or on identity-irrelevant evidence when they determined their degree of confidence. When $w = 0$, the model was identical to the noisy SDT model; when $w = 1$, δ_c only depended on the strength of stimulation S_s, but not on the decision variable for the identification judgment δ_{id}. The term $2R_{id} - 1$ ensured that strong stimuli tended to shift the location of the distribution in a way that high confidence was more likely, and likewise, weak stimuli tended to shift the location of the distribution in a way that the probability of low confidence increased. $\overline{S}_s$ denotes the mean of S_s across the five SOAs and was added to the formulae to increase stability during parameter fitting. The standard deviation σ_c quantifies the amount of unsystematic variability contributing to confidence judgments but not to identification judgments. The unsystematic variability may stem from different sources, including the uncertainty in the estimate of stimulus strength or the noise inherent to metacognitive processes.

2.7.1.5. Two-channel model. The two-channel represents the idea that confidence is based on sensory evidence independent from the sensory evidence used for the decision. Thus, for the two-channel model, δ_c was again sampled from a Gaussian distribution, but now δ_c was sampled independently from δ_{id}:

$$\delta_c \tilde{} \mathcal{N}(\mu = (S_{id} - 1/2) \times S_s \times a, \quad \sigma = 1) \tag{7}$$

The free parameter a expressed the fraction of signal available to the second channel relative to the signal available to the first channel.

2.7.1.6. SDT model with postdecisional evidence. According to model (vi), the,δ_c was again sampled from a Gaussian distribution:

$$\delta_c \tilde{} \mathcal{N}\left(\mu = \delta_{id} + (2S_{id} - 1) \times S_s \times b, \sigma = \sqrt{b}\right) \tag{8}$$

The free parameter b indicated the amount of postdecisional accumulation, and the term $2S_{id} - 1$ ensured that postdecisional accumulation tended to decrease δ_c when $S_{id} = 0$, and to increase δ_c when $S_{id} = 1$.

2.7.2. Non-SDT models

Model (vii)-(ix) assumed a different decision architecture for the identification judgment than models (i)-(vi).

2.7.2.1. Detection heuristic model. According to model (vii), there were two separate decision variables for the identification judgment, each belonging to one possible identity of the stimulus:

$$\begin{aligned} \delta_{id0} \tilde{} \mathcal{N}(\mu = (1 - S_{id}) \times S_s - b, \sigma = \sigma_{id}) \\ \delta_{id1} \tilde{} \mathcal{N}(\mu = S_{id} \times S_s + b, \sigma = \sigma_{id}) \end{aligned} \tag{9}$$

The parameter b reflected the a priori bias in favour of $R_{id} = 1$. Participants were assumed to respond $R_{id} = 0$, when $\delta_{id0} > \delta_{id1}$, and $R_{id} = 1$ if $\delta_{id0} < \delta_{id1}$. Confidence judgments were only based on the decision variable pertaining to the selected response: When $R_{id} = 0$, δ_{id0} was compared against a series of confidence criteria $\theta_{c0}$to select a specific degree of confidence; and when $R_{id} = 1$, the comparison was based on δ_{id1} as well as a second set of criteria θ_{c1}. The bias parameter b was not included in the original version of the model (Peters et al., 2017), but we included it here because there was strong evidence that the free bias parameter improved model fit of the detection heuristic model.

2.7.2.2. 2-D Bayesian model. According to model (viii), there were again two separate decision variables, δ_{id0} and δ_{id1}, referred to as 'sensory signals' by Aitchinson et al. (2015), each referring to one of the two possible identities of the stimulus:

$$\begin{aligned} \delta_{id0} \tilde{} \mathcal{N}(\mu = (1 - S_{id}) \times \delta t, \sigma = s) \\ \delta_{id1} \tilde{} \mathcal{N}(\mu = S_{id} \times \delta t, \sigma = s) \end{aligned} \tag{10}$$

Δt denotes the physical SOA in seconds and s is a free noise parameter. The model assumed that the observer's choices about the identity of the stimulus and about the visibility depended on the posterior probability of the identity of the stimulus given the decision variables $\mathrm{P}(\mathrm{S_{id}}|\delta_{id0}, \delta_{id1})$:

$$\mathrm{P}(\mathrm{S_{id}} = 1|\updelta_{\mathrm{id0}}, \updelta_{\mathrm{id1}}) = \frac{\sum_{\mathrm{t}} \mathrm{P}(\updelta_{\mathrm{id0}}|\Delta\mathrm{t} = \mathrm{t}, \mathrm{s}, \mathrm{S_{id}} = 1)\mathrm{P}(\updelta_{\mathrm{id1}}|\Delta\mathrm{t} = \mathrm{t}\ \ , \mathrm{s}, \quad \mathrm{S_{id}} = 1)}{\sum_{\mathrm{t,i}} \mathrm{P}(\updelta_{\mathrm{id0}}|\Delta\mathrm{t} = \mathrm{t}, \mathrm{s}, \mathrm{S_{id}} = \mathrm{i})\mathrm{P}(\updelta_{\mathrm{id1}}|\Delta\mathrm{t} = \mathrm{t}, \mathrm{s}, \quad \mathrm{S_{id}} = \mathrm{i})} \tag{11}$$

A specific identity and degree of visibility were chosen by comparing the posterior probability $\mathrm{P}(\mathrm{S_{id}} = 1|\delta_{id0}, \delta_{id1})$ against a set of criteria θ. It was assumed that the possible identities and degrees of visibility formed an ordered set of decision options. Each criterion delineated two adjacent decision options, e.g. participants chose to respond that the identity was 1 and visibility was 1 if $\mathrm{P}(\mathrm{S_{id}} = 1|\updelta_{\mathrm{id0}}, \updelta_{\mathrm{id1}})$ was smaller than the criterion associated with identity 1 and visibility 2, and at the same time $\mathrm{P}(\mathrm{S_{id}} = 1|\updelta_{\mathrm{id0}}, \updelta_{\mathrm{id1}})$ was greater than the criterion for identity 0 and visibility 1. Finally, it was assumed that observers did not always give the same response as they intended to. When a lapse occurred, identification and visibility responses were assumed to be random with equal probabilities. The lapse rate λ was an additional free parameter.

2.7.2.3. Two high thresholds model. Model (ix), the two high thresholds model, assumed that the decision variable for the identification judgment δ_{id} was not continuous, but categorical $\delta_{id} \in \{0, 0.5, 1\}$: Observer could either detect the identity of the stimulus and choose the response accordingly $R_{id} = 0$ *if* $\delta_{id} = 0$, and $R_{id} = 1$ *if* $\delta_{id} = 1$. Alternatively, observers could be in a state of uncertainty, $\delta_{id} = 0.5$, in which no information about the identity was available, and observers responded by random guessing. The probability to detect the identity of the stimulus depended on the five SOAs as well as on the identity of the stimulus, resulting in a total of ten detection parameters $p(\delta_{id} = S_{id}|S_s, S_{id})$. A guessing parameter g determined the probability with which observers responded $R_{id} = 1$ when they were in the state of uncertainty. A specific degree of confidence was sampled randomly depending on the three possible states of δ_{id} and the response R_{id}. As the response was fixed when observers detected the identity, there were four different sets of probabilities to determine confidence judgments $p(C = c|\delta_{id} = 0)$, $p(C = c|\delta_{id} = 1)$, $p(C = c|\delta_{id} = 0.5, R_{id} = 0)$, and $p(C = c|\delta_{id} = 0.5, R_{id} = 1)$. All $p(\delta_{id} = S_{id}|S_s, S_{id})$,$p(C = c|\delta_{id}, R_{id})$ and g were free parameters.

2.8. Model fitting

The nine models were fitted to the combined distributions of orientation discrimination and confidence judgments separately for each single participant. First, the frequency of each confidence category was counted for each orientation of the stimulus and each orientation response. Then, for each model, the set of parameters was determined that minimized the negative log-likelihood. For models (i)-(vii) and (ix), the likelihood was calculated analytically (see Supplementary Tables S1 and S2). Only for the 2-D Bayesian model, the likelihood was approximated by simulation. Minimization was performed using a general SIMPLEX minimization routine (Nelder and Mead, 1965). To quantify the goodness-of fit of the nine models, we calculated BIC (Schwarz, 1978) and AICc (Burnham and Anderson, 2002), a variant of the Akaike information criterion (Akaike, 1974) using the negative likelihood of each model fit with respect to each single participant and the trial number.

2.9. Predictions of ERP amplitudes

Predictions about mean ERP amplitudes in the time windows of P3, ERN, and Pe were generated from model fits using the following computational steps:

- First, the statistical models were used to calculate the probabilities of all four confidence categories depending on SOA and choice accuracy separately for each participant using the parameter sets obtained during model fitting of the behavioural data and the formulae in Supplementary Table S1.
- Then, an optimization procedure was used to obtain a transformation to convert each confidence category into an EEG amplitude separately for each participant (see below for details). As result, one specific value of EEG amplitude was assigned to each confidence category.
- The statistical models provided us with probabilities of each confidence category given SOA and accuracy.To obtain an estimate of mean ERP amplitude on the level of single trials, the expected ERP amplitude was calculated by averaging EEG amplitudes assigned to the four different confidence categories weighted by the probability of each confidence category as a function of SOA and accuracy.
- Finally, the correlations across trials between predicted and observed ERP amplitudes were assessed separately for each participant.

Concerning the transformation of confidence into EEG amplitudes, simplex minimization of sum-of-squares with respect to single-trial ERP amplitudes was used to determine the parameters of the transformation. There were to two separate runs of the analysis, one of which assumed a linear transformation and one a monotonous transformation. The linear transformation involved two free parameters, intercept and slope. The monotonous transformation involved four free parameters, one for each confidence categories, each parameter specifying the expected ERP amplitude. These four parameters were constraint by the optimization algorithm to ensure that the expected ERP amplitude was either monotonously increasing or decreasing with confidence.

2.10. Statistical analysis

All statistical tests were based on Bayes factors (Rouder et al., 2009), as implemented in the R package *BayesFactor* (Morey and Rouder, 2015). To test if an ERP component was related to confidence or SOA, we used a Bayesian linear mixed regression model with confidence or SOA as fixed effect and a random effect of participant on the intercept, using default mixture-of-variance priors and a scale parameter of r = 1/2 (Rouder and Morey, 2012). Conceptually, the prior represents the a priori belief that smaller regression slopes are more plausible than large slopes, while even very large slopes were not deemed impossible. Each Bayes factor represents a comparison between the full regression model and a regression model with only the random effect of participant. To compare fits between models of confidence, the Bayesian equivalent of a paired *t*-test was used, assuming a Cauchy distribution with a scale parameter of 1 as prior for the standardized effect size δ, a choice recommended as default (Rouder et al., 2009). The strength of statistical evidence was interpreted according to an established guideline (Lee and Wagenmakers, 2013). In addition, we constructed 95% HDI intervals of the regression slopes or mean differences by 10^6 samples from the posterior distribution using the same models and priors as for Bayes factors.

Concerning figures, error bars and ribbons were based on within-subject standard errors of mean corrected for the number of within-subject conditions (Morey, 2008).

2.11. Data and code availability

The computer programme for the experiment, the behavioural and EEG data, and all analysis scripts to reproduce all results reported in the present paper are freely available at the Open Science Framework website (https://osf.io/93weg).

3. Results

3.1. Behavioural results

Discrimination performance of the orientation ranged between chance at the shortest SOA ($M = 50.8\%$, $SD = 2.6$) and close-to-ceiling at the longest SOA ($M = 94.8\%$, $SD = 8.5$, see Fig. 2A). Confidence ranged between $M = 1.6$ ($SD = 0.6$) on a four-point scale at the shortest SOA and $M = 3.7$ ($SD = 0.4$) at the longest SOA. Fig. 2B shows that confidence was characterised by an increase with SOA in correct as well as in incorrect trials. The evidence for the increase with SOA was extremely strong for correct trials, 95% HDI [0.016 0.020] scale steps/ms, $BF_{10} = 1.3 \cdot 10^{24}$, and strong for incorrect trials, 95% HDI [0.002 0.007] scale points/ms, $BF_{10} = 23.7$. Supplementary Fig. S1 shows that at the shortest SOA, the two larger confidence categories represented only a small fraction of trials, while at the longest SOA, there was only small fraction of trials with the two smaller confidence categories.

3.2. ERP results

The effects of confidence were examined in correct trials during the time windows of the three candidate correlates of confidence: P3, ERN, and Pe. Consistent with our prediction, there was extremely strong evidence that EEG activity in the P3 time range (350–500 ms after onset of the target stimulus, recorded at the parietal electrode Pz) increased with confidence, 95% HDI [1.8 2.9] μV/scale step, $BF_{10} = 3.6 \cdot 10^{10}$ (see Fig. 3A). Fig. 4A shows that the association between ERPs and confidence in correct trials during the P3 time window had a centroparietal distribution over the scalp, consistent with known topographies of the P3 in difficult perceptual discrimination tasks (Koivisto and Revonsuo, 2010). The analyses if confidence judgments predict EEG activity at the time of the P3 separately for each SOA were not conclusive about an effect of confidence for three out of four SOAs, $1.01 < \mathrm{BF}_{10} < 2.53$, and there was moderate evidence against an effect at the SOA of 66.7 ms, $\mathrm{BF}_{10} = 0.30$.

Fig. 3B shows the effect of confidence in correct trials during the ERN time window (-40 – 60 ms after the orientation response, at the frontocentral electrode FCz). The evidence for an effect of confidence was not conclusive, 95% HDI [-0.6 0.1] μV/scale step, $BF_{10} = 0.45$. However, although a positive relation between confidence and ERN would have

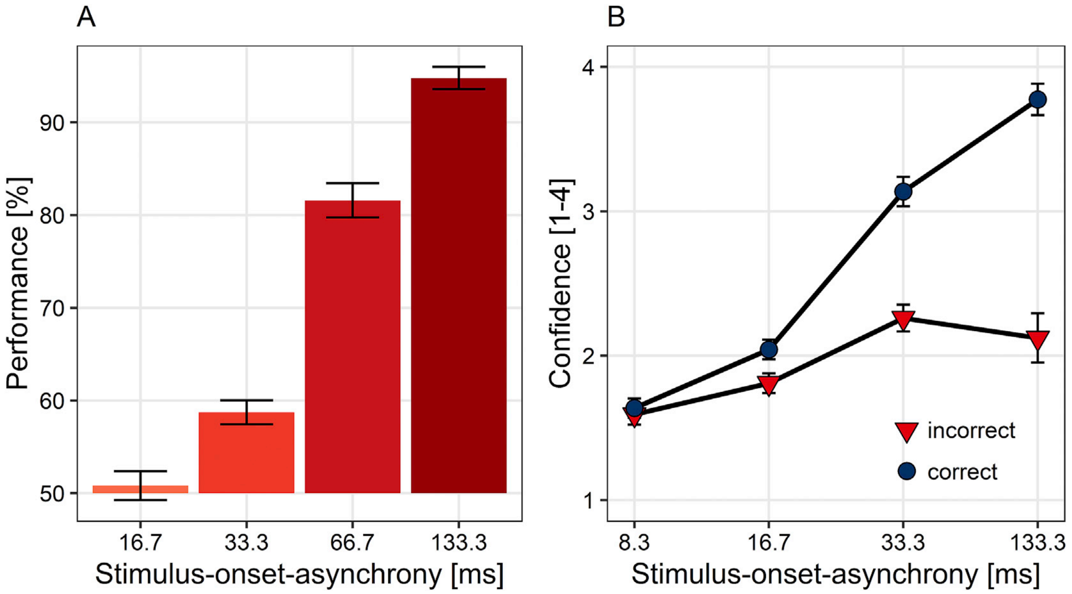


Fig. 2. (A) Accuracy in the orientation discrimination task depending on stimulus-onset-asynchrony. B): Decision confidence as a function of stimulus-onset-asynchrony in correct (blue symbols) and incorrect trials (red). Bars and symbols indicate observed means. Error bars indicate 1 within-subject *SEM*.

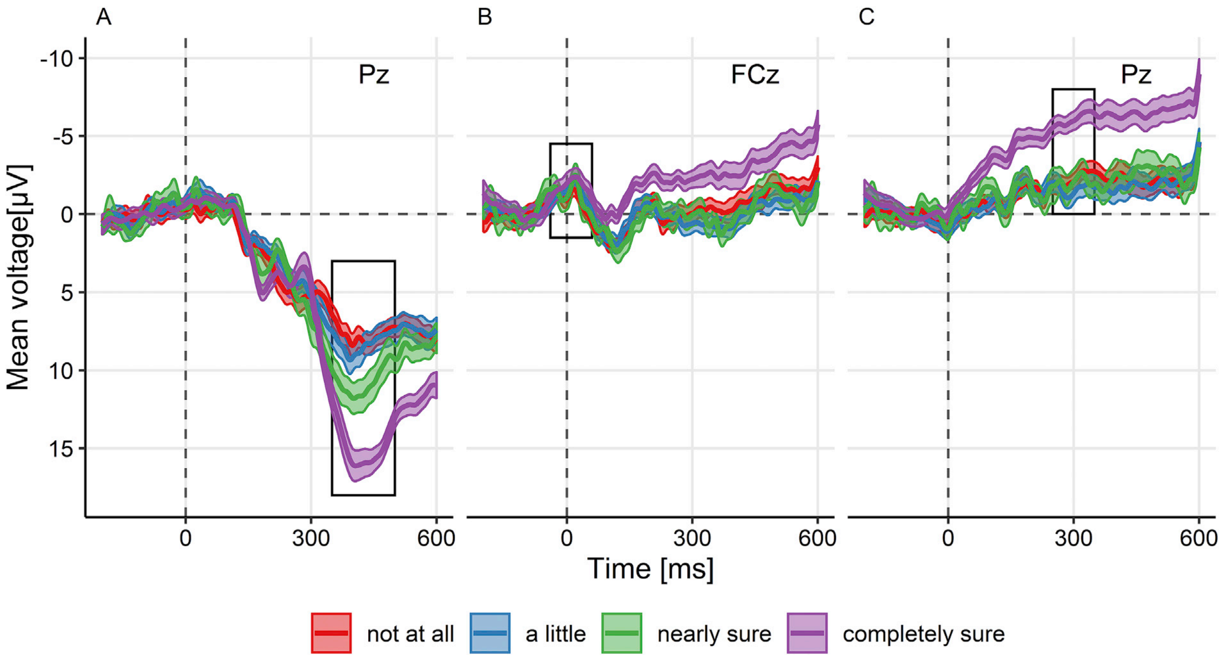


Fig. 3. Event-related potentials in correct trials as a function of confidence judgments. (A) Activity at the parietal electrode Pz locked to the onset of the target stimulus. The box highlights the range of the P3 time window. The ERP is locked to the onset of the target stimulus. (B) Activity at the frontocentral electrode FCz locked to the orientation response. The box highlights the time range of the ERN. (C) Activity at Pz locked to the orientation response. The box highlights the time range of the Pe. Colours indicate the degree of confidence reported by the observers. Ribbons indicate ±1 within-subject *SEM* around the mean.

been expected, the trend was in the opposite direction. As shown by Fig. 4B, no clear effect of confidence emerged anywhere over the scalp at the time of the ERN. Finally, Fig. 3C shows the effect of confidence in correct trials during the Pe time window (250–350 ms after the orientation response, at Pz). In line with our prediction, there was strong evidence that confidence was negatively associated with the ERP at the time of the Pe, 95% HDI [-1.8–0.5] μV/scale step, $BF_{10} = 60.6$. As can be seen from Fig. 4C, the association between ERPs and confidence in correct trials during the Pe time window had a posterior parietal distribution over the scalp, consistent with known topographies of the Pe (Boldt and Yeung, 2015).

Next, we tested if P3, ERN and Pe are markers of confidence by comparing the effects of SOA on confidence judgments and on ERPs, separately for correct and incorrect trials. Consistent with the pattern of confidence judgments, there was extremely strong evidence that the ERP in the P3 range increased with SOA in correct trials, 95% HDI [0.08 0.10] μV/ms, $BF_{10} = 1.8 \cdot 10^{25}$, as well as in incorrect trials, 95% HDI [0.04 0.08] μV/ms, $BF_{10} = 1.2 \cdot 10^{7}$ (see Fig. 5A–D, see also Figs. 7 and 8). In the ERN range, Fig. 5H shows that the strongest effect at the time of the ERN was a negative shift in incorrect trials at the longest SOA. The effect of SOA in incorrect trials was in the opposite direction as the pattern of confidence judgments: the evidence was extremely strong for a negative,

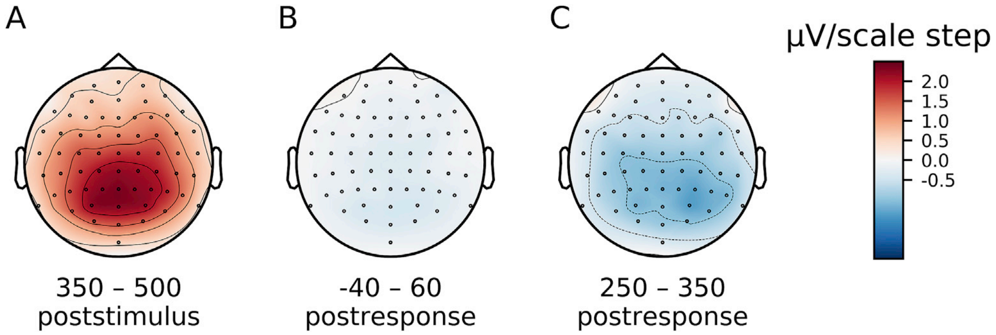


Fig. 4. Distribution of the association between ERPs and confidence in correct trials across the scalp. Maps are based on regression slopes with ERP amplitudes as a function of confidence (A) Time window 350–500 ms after target stimulus onset. (B) Time window between 40 ms before the orientation response and 60 ms after the response. (C) Time window 250–350 ms after the response.

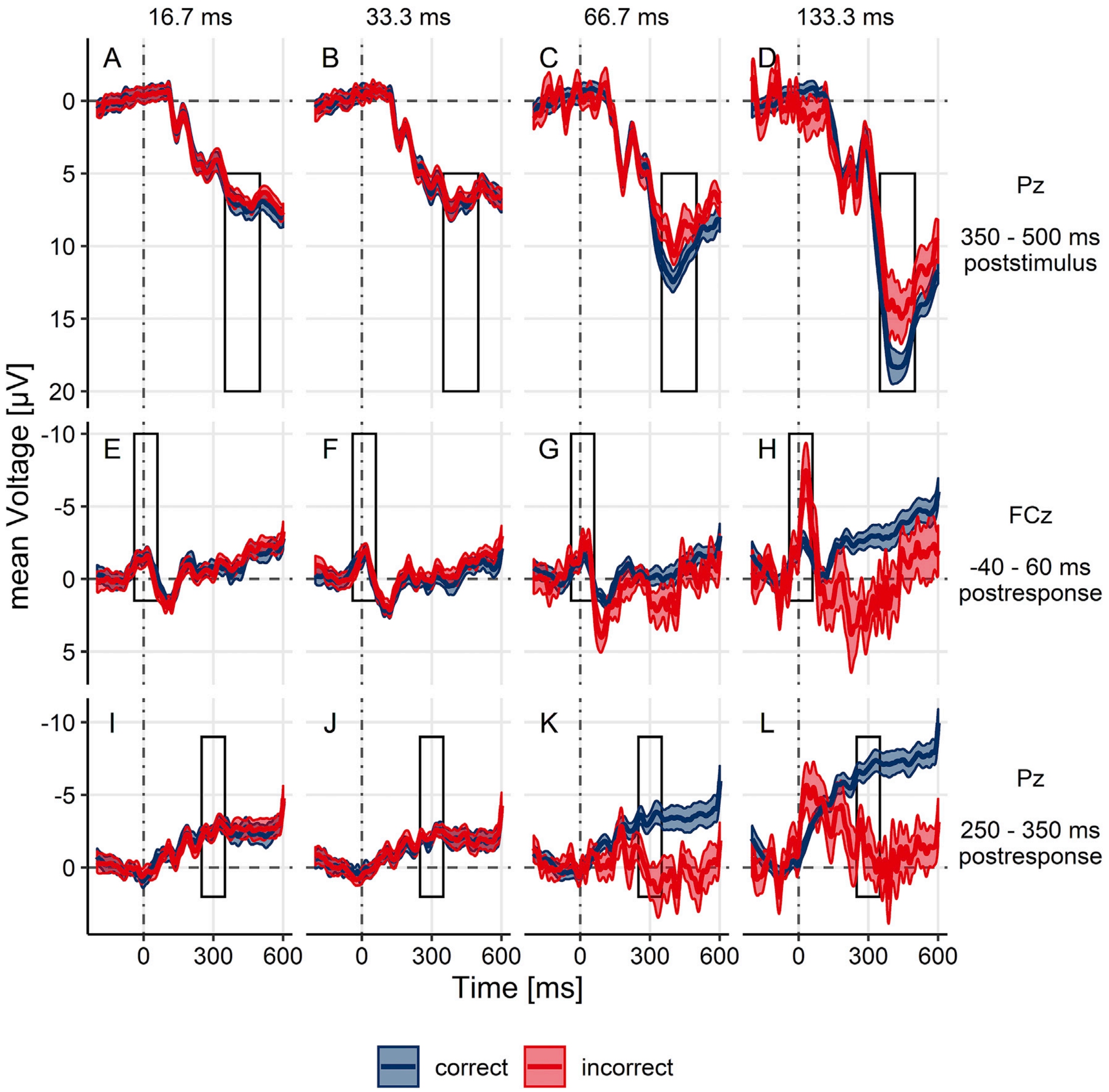


Fig. 5. Event-related potentials as a function of stimulus-onset-asynchrony (different columns) and accuracy of the orientation response (blue: correct responses; red: incorrect responses). (A–D) ERP activity locked to target stimulus onset, recorded at Pz. The box highlights the time window of the P3 (350–500 ms poststimulus). (E–H) ERP activity locked to the orientation response, recorded at FCz. The box highlights the time window of the ERN (-40 – 60 ms postresponse). (I–L) ERP activity locked to the orientation response, recorded at Pz. The box highlights the time window of the Pe (250–350 ms postresponse). Ribbons indicate ±1 within-subject *SEM* around the mean.

not positive shift, 95% HDI [-0.04–0.01] μV/ms, $BF_{10} = 164.5$. The evidence with respect to an effect on correct trials was not conclusive, 95% HDI [-0.01 0.00] μV/ms, $BF_{10} = 2.9$. Fig. 5I-L shows that the EEG activities in correct and incorrect trials at the time of the Pe seemed to diverge from each other with increasing SOA, i.e. the Pe seemed to follow the folded X-pattern. This is inconsistent with double increase pattern of confidence judgments, as confidence had increased with SOA in both correct and incorrect trials. As EEG activity at the time of the Pe is negatively associated with confidence, the pattern of confidence judgments implied a negative shift with SOA in both correct and incorrect trials. In accordance with the pattern of confidence judgments, there was extremely strong evidence for the negative shift with increasing SOA in correct trials, 95% HDI [-0.05–0.03] μV/ms, $BF_{10} = 7.6 \cdot 10^7$. However, although the pattern of confidence judgments implied a negative shift with increasing SOA in incorrect trials as well, there was moderate evidence against a relationship between SOA and ERPs in incorrect trials at the time of the Pe, 95% HDI [-0.01 0.03] μV/ms, $BF_{10} = 0.29$.

3.3. Cognitive modelling

3.3.1. Modelling confidence judgments

Fig. 6 shows confidence judgments as a function of SOA and choice accuracy compared to the model prediction based on parameter sets identified during fitting. The WEV-model, the noisy decay model, the detection heuristic model and the 2-D Bayesian model correctly predicted that confidence in incorrect trials increases with SOA, $BF_{10} \geq 92.4$ (Fig. 6, A, E, H, I). The SDT-model, the SDT-model with postdecisional evidence, the noisy SDT model, and the two-channel model produced a decreasing relationship instead (Fig. 6 B, C, D, G), $BF_{10} \geq 3.3 \times 10^{15}$. For the two high-threshold model, the relationship between SOA and predicted confidence in incorrect trials appeared to be flat, but the evidence was not conclusive, $BF_{10} = 0.39$.

Quantifying model fit using the Akaike information criterion (AIC$_c$) and the Bayes information criterion (BIC) showed that the best fit to the data was obtained by the WEV-model, followed by the noisy decay model (Fig. 7). Regarding AIC$_c$, the evidence if the WEV model performed better than the noisy decay model was not conclusive, $M_{\Delta AIC} = 8.9$, 95% HDI [-4.9, 21.6], $BF_{10} = 0.35$, but there was very strong evidence that the WEV-model performed better than the two-high-thresholds-model, $M_{\Delta AIC} = 99.4$, 95% HDI [42.7.146.7], $BF_{10} = 43.4$, and extreme evidence that the WEV-model performed better than each of the other models, $M_{\Delta AIC} \geq 34.2$, $BF_{10} \geq 101.3$. Regarding BIC, there was moderate evidence that the WEV-model performed better than the detection heuristic model, $M_{\Delta BIC} = 24.6$, 95% HDI [7.1, 39.4], $BF_{10} = 6.8$, strong evidence that the WEV-model was better than the noisy decay model, $M_{\Delta BIC} = 23.3$, 95% HDI [8.5, 35.8], $BF_{10} = 16.5$, and extreme evidence that the WEV model was better than each of the other five models, $M_{\Delta AIC} \geq 34.2$, $BF_{10} \geq 101.3$. These results were essentially the same when it was assumed that the variances of the decision variable differed between horizontal or vertical stimuli. Summary statistics of the fitted parameters of the WEV model can be found in the Supplementary Table S3.

3.3.2. Model classification analysis

To investigate if one of the other models could have been misclassified as WEV-model, a simulation was performed based on the second-best and the third-best performing model, i.e. the noisy decay model and the heuristic detection model. For each participant, we used the parameter sets determined based on the empirical data to simulate the same number of trials as in the real experiment. Then, both the known generative model and the WEV model were fitted to the simulated data of each participant and compared using AIC$_c$ and BIC. When model fits of the WEV model were compared with fits of the detection heuristic

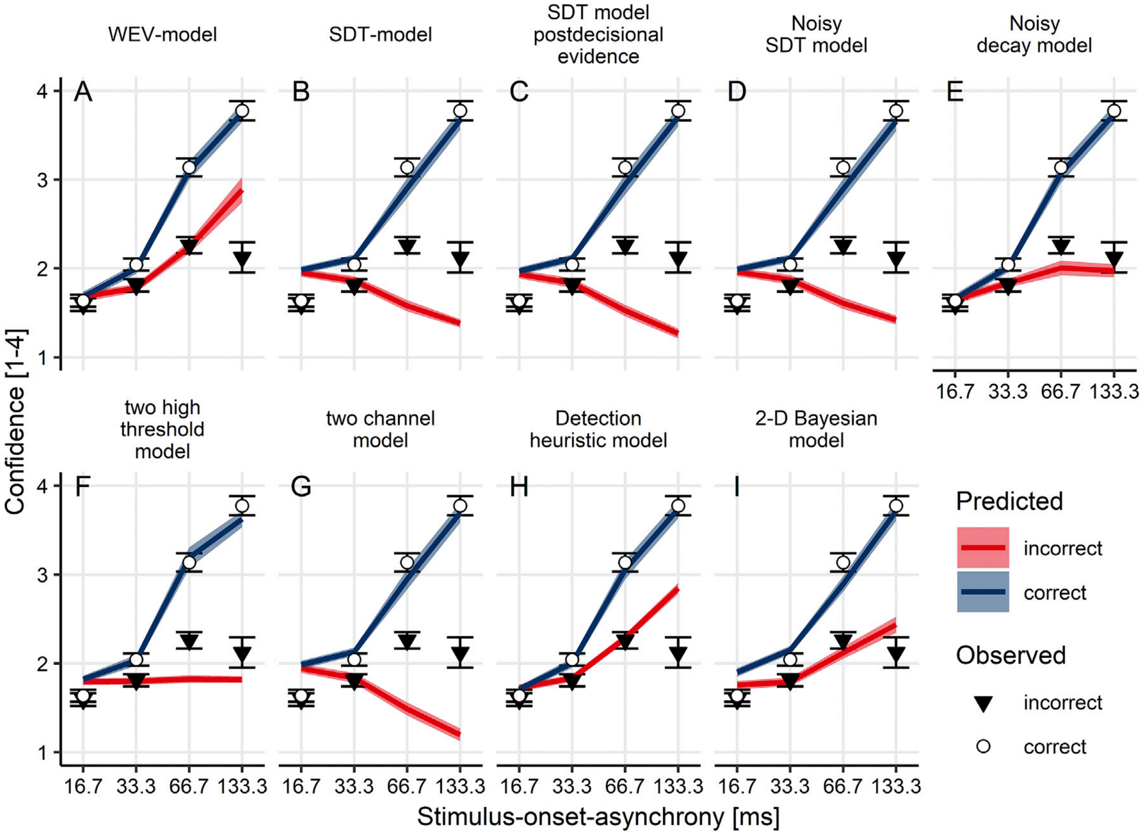


Fig. 6. Mean confidence judgments depending on stimulus-onset-asynchrony (x-Axis) and accuracy of the orientation response. Different panels show the prediction of the different models based on the sets of parameters identified during model fitting, assuming constant variances of the decision variable. Blue lines indicate the prediction for correct trials, red lines for incorrect trials. Ribbons indicate ±1 within-subject *SEMs* around the predicted mean confidence. Circles indicate observed confidence judgments in correct trials, and triangles in incorrect trials. Error bars = 1 within-subject *SEM*.

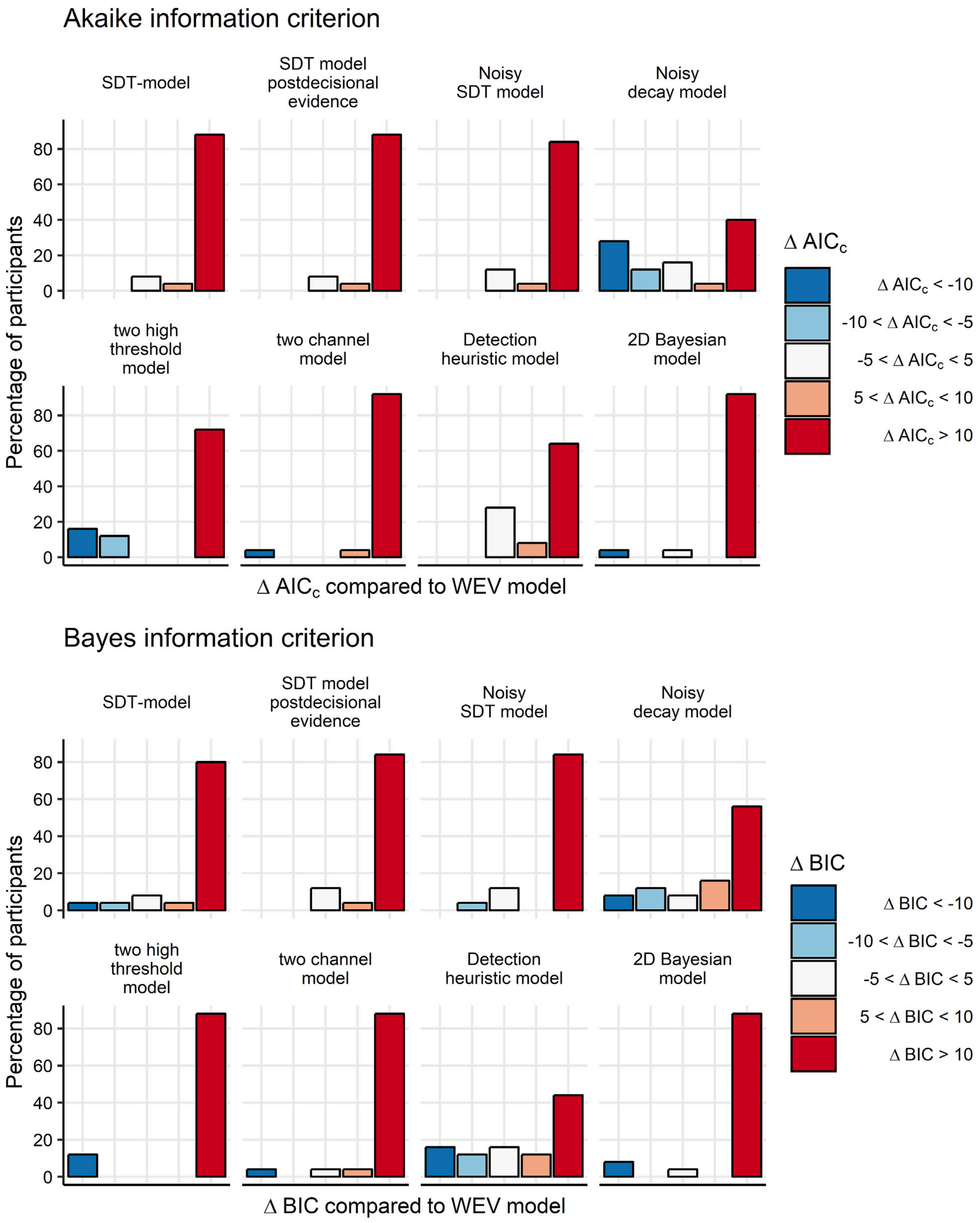


Fig. 7. Formal model comparisons. The different panels depict the frequency of AICc- and BIC differences when the WEV model was compared to each of the seven other models, assuming constant variances of the decision variable. AICc and BIC differences were assorted into categories based on established guidelines for interpretation.

model based on data that conforms to the detection heuristic model, AIC$_c$-differences indicated the correct model for 83.3% of the simulated data sets, while BIC-differences indicated the correct model for each single data set (see Supplementary Fig. S2). When model fits of the WEV model were compared with fits of the noisy decay model based on data generated according to the noisy decay model, AIC$_c$-differences indicated the correct model for 95.8% of the data sets, while BIC-differences indicated the correct model for 75.0% data sets. It should be noted that the present study compared AICc and BIC differences averaged across participants, which is why it is not necessary that model classification is

100%, it is merely required that model classification accuracy is markedly above 50%.

3.3.3. Predicting ERPs from model fits

Based on parameter sets of the WEV-model obtained by fitting the behavioural data, we determined expected ERP amplitudes at the time of P3, ERN, and Pe as a function of SOA and choice accuracy. First, a simple linear transformation was applied to confidence with parameters of the transformation determined based on the EEG data. Fig. 8A shows that the linear transformation of predicted confidence resulted in a reasonably accurate prediction regarding ERP amplitude in the P3 window. Consequently, there was a medium-sized correlation between predicted and observed single-trial amplitudes at the time of the P3, $M = 0.33$, 95% HDI [0.29 0.37]. In contrast, as can be seen from Fig. 8B, the predicted EEG in the ERN time window did not reproduce the large negative shift specifically in incorrect trials at the longest SOA. Therefore, the correlation between predicted and observed single-trial amplitudes at the time of the ERN was small, $M = 0.09$, 95% HDI [0.06 0.11]. Likewise, Fig. 8C shows that longer SOAs were associated with a positive shift in incorrect trials during the Pe time window, which was just opposite to the pattern observed with confidence judgments (cf. Fig. 2B) and therefore was not reproduced by the prediction. The correlation between predicted and observed single-trial amplitudes at the time of the Pe was also small $M = 0.16$, 95% HDI [0.12 0.20]. The same results were obtained when we repeated this analysis with the noisy decay model and the detection heuristic model (see Supplementary Fig. S3). Finally, an exploratory analysis was performed to assess when in quasi-continuous time the EEG activity was associated with predicted confidence according to the WEV-model. For this purpose, we used a series of multivariate regression analyses performed separately for 10 ms time windows with confidence predicted by the WEV-model as outcome variable and all sensors as predictors (see Supplementary Fig. S4). The analysis suggested that stimulus-locked EEG activity strongly predicted confidence according to the WEV-model with peaks around 150 ms, 250 ms, and 400 ms post-stimulus. The third peak coincided with the P3. For response-locked ERPs, only a small portion of the variance of confidence according to the WEV-model could be explained by EEG at the time of the response, and a moderate portion during a broad time window between 200 and 500 ms postresponse.

The relationship between confidence and ERP amplitudes of course does not need to be linear. For this reason, we fitted non-linear transformations to the data from each subject by assigning the voltage that minimized the prediction error with respect to ERP amplitude to each level of confidence. The only restriction of the transformation was that the relationship between confidence and ERP amplitudes was assumed to be monotonous. Nevertheless, the predictions based on these specifically adapted transformations were only consistent with amplitudes at the time of the P3, but not with ERN or Pe (see Supplementary Fig. S5).

4. Discussion

The present study was consistent with an EEG correlate of decision confidence 350–500 ms after onset of the stimulus, at the time of the P3 component: First, ERP amplitudes at the time of the P3 were associated with observers' confidence judgments, although the data were not conclusive if the correlation between confidence and EEG activity at the time of the P3 can be explained by the correlation between confidence and SOA. Second, the amplitude at the time of the P3 varied as a function of SOA and choice accuracy in the same way as confidence judgments did. Finally, P3 amplitude could be accurately predicted by the weighted evidence and visibility (WEV) model, which at the same time provided the best account of confidence judgments. In contrast, EEG activity at the time of the ERN component, an established marker of error detection, as well as at the time of the Pe, a marker of error awareness, did not follow the same statistical pattern as decision confidence as a function of SOA and accuracy, despite the fact that a correlation between amplitude and confidence was detected at the time of the Pe. Moreover, there were only weak correlations between the prediction derived from the WEV-model and ERP amplitude at the time of ERN and Pe.

4.1. Role of the P3

The P3 was the only one of the three previously proposed ERP correlates of decision confidence that was consistent with the present data. Although the present study was not conclusive about an effect of confidence on EEG activity over and above task difficulty, a previous study using only one difficulty level of a different perceptual task observed that confidence was associated with EEG activity at the time of the P3 (Boldt and Yeung, 2015). Nevertheless, it cannot be ruled out that the observed correlation between decision confidence and P3 is due to a common cause. One potential alternative explanation is that the P3 reflects subjective visibility (Lamy et al., 2008; Sergent et al., 2005; Tagliabue et al., 2019), which is correlated to decision confidence but may not necessarily reflect the same process (Jachs et al., 2015; Rausch and Zehetleitner, 2016; Zehetleitner and Rausch, 2013). A second possibility is that the P3 reflects processes involved in discrimination performance. Recent studies showed that stimulus properties can be experimentally varied to change confidence without changing discrimination performance (Koizumi et al., 2015; Odegaard et al., 2018; Samaha et al., 2016). Experimental manipulations of the stimulus that influence confidence but not accuracy seem promising for future studies to elucidate if the P3 reflects confidence or discrimination performance.

How could an association between P3 and decision confidence be reconciled with the various other roles of the P3 that have been proposed in the literature? An explanation may be given in terms of probabilistic models of perception, according to which observers take into account knowledge of the uncertainty associated with the observations (Ma, 2012). One possible interpretation is that the P3 directly reflects certainty within the decision process (Herding et al., 2019). In line with this interpretation, the P3 showed the statistical pattern referred to as folded-X pattern in a vibrotactile task (Herding et al., 2019), meaning both statistical patterns associated with decision confidence in different tasks, the folded X-pattern and the double increase pattern, have been detected in P3 amplitudes. Moreover, the P3 is related to the accumulation of sensory evidence within the decision process (O'Connell et al., 2012; Twomey et al., 2015). Finally, the P3 is suppressed in highly visible stimuli if observers are not required to make a perceptual decision (Pitts et al., 2014). These findings converge with a line of research suggesting that decision confidence may emerge directly from the decision process. For example, neurons in parietal cortex of rhesus monkeys represented both formation of the direction decision and the degree of certainty (Kiani and Shadlen, 2009). Likewise, human EEG correlates of decision formation and confidence coincided in time and in reconstructed sources in a face vs. car discrimination task (Gherman and Philiastides, 2015). In contrast, for at least one brain area implicated in decision making, the superior colliculus, it was shown that it reflects decision making, not decision confidence (Odegaard et al., 2018).

A second interpretation in terms of probabilistic models is that the P3 reflects sensory representations that include the reliability of the percept (Kopp et al., 2016). This second view is consistent with classical interpretations of the P3 as update of working memory in response to task-relevant events (Donchin and Coles, 1988) or global broadcast of information within a neural global workspace (Sergent et al., 2005). These updated or broadcast representations may encompass the reliability of the percept (Shea and Frith, 2019), which is why the P3 should be correlated with confidence judgments. In line with this interpretation, the WEV-model assumes that confidence is determined by the perceived strength or reliability of the percept based on evidence about choice-relevant and choice-irrelevant features. This means that the inferred computational principles underlying decision confidence include a representation of the reliability of the percept as well.

Finally, it should be noted that the observation that decision

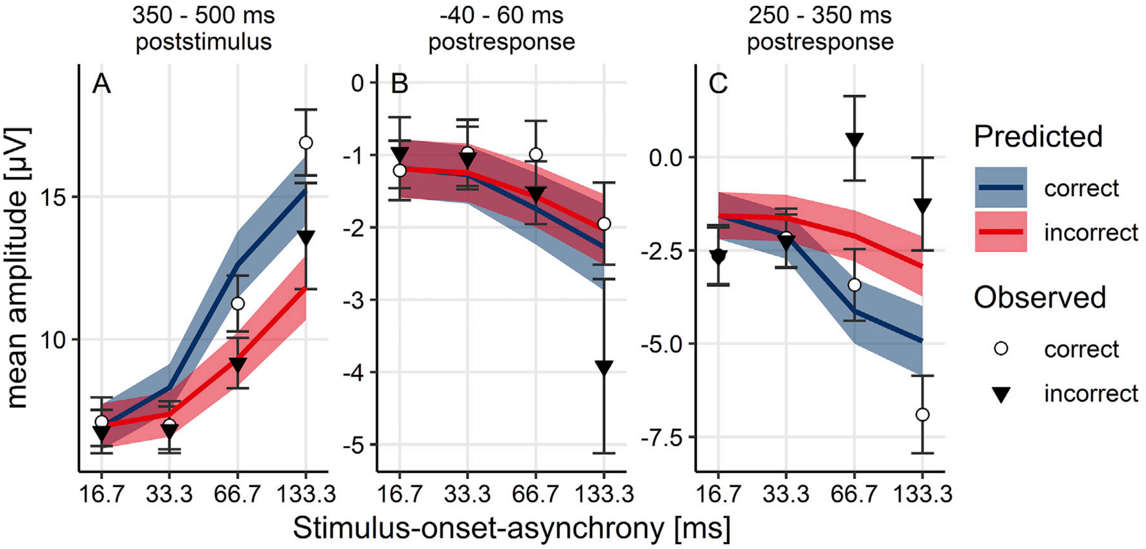


Fig. 8. Comparison between predicted and observed amplitudes (A) in the P3 time window, (B) in the ERN time window, and (C) in the Pe time window, depending on stimulus-onset-asynchrony (x-Axis) and accuracy of the orientation response (colours and symbols). Symbols: observed data. Lines: Prediction based on the parameters of the WEV-model fitted to confidence judgments as well as a linear transformation of confidence. Error bars = 1 within-subject *SEM*. Ribbons = 1 within-subject *SEM*.

confidence and P3 share their statistical patterns in the present study does not imply that decision confidence and P3 necessary share their statistical patterns across all possible experiments. The statistical patterns associated with confidence vary across different tasks (e.g. Kiani et al., 2014; Moran et al., 2015; Rausch et al., 2018; Sanders et al., 2016; van den Berg et al., 2016). If the P3 were indeed a neural marker of decision confidence, confidence and P3 should be associated with the same statistical patterns in all tasks. The present study and Herding et al. are not sufficient to make this conclusion. Future studies are necessary to test if the P3 is a general marker of decision confidence, or if the present results are specific to the present task.

4.2. Role of ERN

In the present study, EEG activity in the ERN time window can be interpreted as specifically error detection, but not as decision confidence. EEG activity at the time of the ERN does not reflect confidence because the effects of SOA were opposite to what was expected from observed confidence judgments. At least in the present study, the ERN may not be related to postdecisional sensory evidence, because sensory evidence in correct trials is expected to increase with SOA (Hangya et al., 2016), but at the time of the ERN, the only reliable effect was large negative shift specifically in incorrect trials at the longest SOA. The absence of an ERN at shorter SOAs is in line with a previous study showing that the elicitation of a ERN requires participants to know which response is the correct one (Di Gregorio et al., 2018). Likewise, in the present study, observers also did not know for sure which response had been correct at shorter SOAs because the mask impeded perception of the target. These findings are also consistent with a previous study showing that the ERN occur only when observers make erroneous responses to stimuli rated as “visible” (Charles et al., 2014, 2013). Although we did not measure conscious awareness in the present study, we can extrapolate from other studies using the same task that observers’ conscious percepts of the stimuli were degraded in shorter SOAs (Rausch and Zehetleitner, 2019b; Zehetleitner and Rausch, 2013); possibly, weakly conscious stimuli are not sufficient to trigger an ERN.

4.3. Role of Pe

A possible interpretation for the role of the Pe in the present study is as accumulation of postdecisional sensory evidence. At least in the present study, the Pe does not reflect decision confidence because their statistical patterns as functions of SOA and choice accuracy are not compatible. In addition, the Pe does not exclusively reflect error awareness, because EEG activity at the time of the Pe was correlated with confidence in correct trials. However, the pattern of the Pe as a function of SOA and choice accuracy matches the diverging pattern between correct and incorrect responses expected from postdecisional accumulation of sensory evidence (Moran et al., 2015). The contribution of postdecisional sensory evidence to confidence varies across tasks (Baranski and Petrusic, 1994). In the present paradigm, it may be relatively small, because the mask prevents ongoing accumulation of evidence from sensory memory. In line with this interpretation, cognitive modelling showed that the WEV model fitted confidence much better than the SDT model with postdecisional evidence. If the Pe reflects postdecisional accumulation of evidence, this explains effects at the time of the Pe seemed to be limited to high confidence trials why in the present study. The efficiency of the mask varies across trials, and presumably the mask had been relatively ineffective in trials when observers reported high degrees of confidence. Moreover, if the Pe represents postdecisional sensory evidence, it can be explained why a previous study detected an association between the Pe and all degrees of confidence (Boldt and Yeung, 2015). As stimuli in that study were not masked, postdecisional accumulation of sensory evidence may have been more effective than in the present study. Finally, the Pe may not only be sensitive to postdecisional sensory evidence, but may reflect also other sources of information, including response conflict, efference copy, proprioception, perception of action effects, and interoception (Ullsperger et al., 2010; Wessel et al., 2011).

4.4. Statistical signatures of confidence?

The present study demonstrates that statistical patterns of confidence can provide a strong test for identifying correlates of confidence, although it is crucial to validate statistical signatures of confidence empirically by behavioural measures of confidence. It has been argued that if confidence is determined objectively as the posterior probability of being correct, the pattern referred to as folded X-pattern is the statistical signature of confidence (Hangya et al., 2016; Sanders et al., 2016). Therefore, a substantial number of recent studies have searched for the folded X-pattern to empirically identify correlates of decision confidence (Braun et al., 2018; Fetsch et al., 2014; Herding et al., 2019; Lak et al., 2017; Sanders et al., 2016; Urai et al., 2017). However, it has been shown mathematically that the folded X-pattern is neither a necessary nor a sufficient condition for Bayesian confidence (Adler and Ma, 2018; Rausch and Zehetleitner, 2019a). The present study showed empirically that a second statistical pattern of confidence exists and can be used to identify correlates of confidence. Had we not measured decision confidence directly and relied on the purported folded-X signature, the Pe, not the P3, would have been falsely considered a correlate of confidence.

5. Conclusion

The present results are consistent with an EEG correlate of decision confidence over parietal electrodes 350–500 ms after onset of the stimulus. However, there is no single EEG correlate of decision confidence and error awareness: EEG components after the response, which have been established as markers of error detection or error awareness, were dissociated from decision confidence.

CRediT authorship contribution statement

Manuel Rausch: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing - original draft. **Michael Zehetleitner:** Conceptualization, Funding acquisition, Supervision, Writing - review & editing. **Marco Steinhauser:** Conceptualization, Methodology, Resources, Supervision, Writing - review & editing. **Martin E. Maier:** Conceptualization, Formal analysis, Methodology, Investigation, Writing - review & editing.

Acknowledgements

We are grateful to Sebastian Hellmann for help in programming the experiment, to Leonie Manzke for help with Python, and to Marina Nickl, Anna Bogner, and Johanna Dietrich for help with data collection.

This work was supported by the Deutsche Forschungsgemeinschaft (grant numbers ZE 887/8-1, RA2988/3-1). The publication of this article was supported by the Open Access Fund of the Catholic University Eichstätt-Ingolstadt. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116963.

References

Adler, W.T., Ma, W.J., 2018. Limitations of proposed signatures of Bayesian confidence. Neural Comput. 30, 3327–3354. https://doi.org/10.1162/neco_a_01141.

Aitchison, L., Bang, D., Bahrami, B., Latham, P.E., 2015. Doubly Bayesian analysis of confidence in perceptual decision-making. PLoS Comput. Biol. 11, e1004519 https://doi.org/10.1371/journal.pcbi.1004519.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. AC- 19, 716–723. https://doi.org/10.1007/978-1-4612-1694-0_16.

Bang, D., Fleming, S.M., 2018. Distinct encoding of decision confidence in human medial prefrontal cortex. Proc. Natl. Acad. Sci. Unit. States Am. 115, 6082–6087. https://doi.org/10.1073/pnas.1800795115.

Baranski, J.V., Petrusic, W.M., 1994. The calibration and resolution of confidence in perceptual judgments. Percept. Psychophys 55, 412–428. https://doi.org/10.3758/BF03205299.

Barrett, A.B., Dienes, Z., Seth, A.K., 2013. Measures of metacognition on signal-detection theoretic models. Psychol. Methods 18, 535–552. https://doi.org/10.1037/a0033268.

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.

Braun, A., Urai, A.E., Donner, T.H., 2018. Adaptive history biases result from confidence-weighted accumulation of past choices. J. Neurosci. 38, 2418–2429. https://doi.org/10.1523/JNEUROSCI.2189-17.2017.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed. Springer, New York.

Charles, L., King, J.-R., Dehaene, S., 2014. Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. J. Neurosci. 34, 1158–1170. https://doi.org/10.1523/JNEUROSCI.2465-13.2014.

Charles, L., Opstal, F. Van, Marti, S., Dehaene, S., 2013. Distinct brain mechanisms for conscious versus subliminal error detection. Neuroimage 73, 80–94.

Charles, L., Yeung, N., 2018. Dynamic sources of evidence supporting confidence judgments and error detection. J. Exp. Psychol. Hum. Percept. Perform. 45, 39–52.

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.

Donchin, E., Coles, M.G., 1988. Is the P300 component a manifestation of context updating? Behav. Brain Sci. 11, 357–374.

Eimer, M., Mazza, V., 2005. Electrophysiological correlates of change detection. Psychophysiology 42, 328–342.

Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L., 1991. Effects of crossmodal divided attention on ERP components: error processing in choice reaction tasks. Electroencephalogr. Clin. Neurophysiol. 78, 447–455. https://doi.org/10.1016/0013-4694(91)90061-8.

Fetsch, C.R., Kiani, R., Newsome, W.T., Shadlen, M.N., 2014. Effects of cortical microstimulation on confidence in a perceptual decision. Neuron 83, 797–804. https://doi.org/10.1016/j.neuron.2014.07.011.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E., 1993. A neural system for error detection and compensation. Psychol. Sci. 4, 385–390.

Gherman, S., Philiastides, M.G., 2015. Neural representations of confidence emerge from the process of decision formation during perceptual choices. Neuroimage 106, 134–143. https://doi.org/10.1016/j.neuroimage.2014.11.036.

Gramfort, A., Luessi, M., Larson, E., Engemann, D., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M., 2014. MNE software for processing MEG and EEG data. Neuroimage 86, 446–460. https://doi.org/10.1021/nl061786n.Core-Shell.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goy, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-Python. Front. Neurosci. 7, 1–13. https://doi.org/10.3389/fnins.2013.00267.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York.

Hangya, B., Sanders, J.I., Kepecs, A., 2016. A mathematical Framework for statistical decision confidence. Neural Comput. 28, 1840–1858. https://doi.org/10.1162/NECO_a_00864.

Herding, J., Ludwig, S., von Lautz, A., Spitzer, B., Blankenburg, F., 2019. Centro-parietal EEG potentials index subjective evidence and confidence during perceptual decision making. Neuroimage 201, 116011. https://doi.org/10.1016/j.neuroimage.2019.116011.

Hillyard, S.A., Squires, K.C., Bauer, J.W., Lindsay, P.H., 1971. Evoked potential correlates of auditory signal detection. Sci 172, 1357–1360.

Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Network. 10, 626–634. https://doi.org/10.1109/72.761722.

Jachs, B., Blanco, M.J., Grantham-Gill, S., Soto, D., 2015. On the independence of visual awareness and metacognition: a signal detection theoretic analysis. J. Exp. Psychol. Hum. Percept. Perform. 41, 269–276.

Kellen, D., Klauer, K.C., 2015. Signal detection and threshold modeling of confidence-rating ROCs: a critical test with minimal assumptions. Psychol. Rev. 122, 542–557.

Kepecs, A., Uchida, N., Zariwala, H., Mainen, Z.F., 2008. Neural correlates, computation and behavioural impact of decision confidence. Nat 455, 227–231. https://doi.org/10.1038/nature07200.

Kiani, R., Corthell, L., Shadlen, M.N., 2014. Choice certainty is informed by both evidence and decision time. Neuron 84, 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015.

Kiani, R., Shadlen, M.N., 2009. Representation of confidence associated with a decision by neurons in the parietal cortex. Sci 324, 759–764. https://doi.org/10.1126/science.1169405.

Koivisto, M., Revonsuo, A., 2010. Event-related brain potential correlates of visual awareness. Neurosci. Biobehav. Rev. 34, 922–934. https://doi.org/10.1016/j.neubiorev.2009.12.002.

Koizumi, A., Maniscalco, B., Lau, H., 2015. Does perceptual confidence facilitate cognitive control? Atten. Percept. Psychophys. 77, 1295–1306. https://doi.org/10.3758/s13414-015-0843-3.

Kopp, B., Seer, C., Lange, F., Kluytmans, A., Kolossa, A., Fingscheidt, T., Hoijtink, H., 2016. P300 amplitude variations, prior probabilities, and likelihoods: a Bayesian ERP study. Cognit. Affect Behav. Neurosci. 16, 911–928. https://doi.org/10.3758/s13415-016-0442-3.

Lak, A., Nomoto, K., Keramati, M., Sakagami, M., Kepecs, A., 2017. Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. Curr. Biol. 27, 821–832. https://doi.org/10.1016/j.cub.2017.02.026.

Lamy, D., Salti, M., Bar-haim, Y., 2008. Neural correlates of subjective awareness and unconscious Processing : an ERP study. J. Cognit. Neurosci. 21, 1435–1446.

Lee, M.D., Wagenmakers, E.-J., 2013. Bayesian Cognitive Modeling: A Practical Course. Cambridge University Press, Cambridge, UK.

Luck, S.J., 2014. Comparing conditions with different numbers of trials or different levels of noise. In: An Introduction to the Event-Related Potential Technique. MIT Press.

Ma, W.J., 2012. Organizing probabilistic models of perception. Trends Cognit. Sci. 16, 511–518. https://doi.org/10.1016/j.tics.2012.08.010.

Macmillan, N.A., Creelman, C.D., 2005. Detection Theory. A User's Guide. Lawrence Erlbaum Associates, Mahwah, NY.

Maniscalco, B., Lau, H., 2016. The signal processing architecture underlying subjective reports of sensory awareness. Neurosci. Conscious. 1, niw002. https://doi.org/10.1093/nc/niw002.

Maniscalco, B., Peters, M.A.K., Lau, H., 2016. Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. Atten. Percept. Psychophys. 78, 923–937. https://doi.org/10.3758/s13414-016-1059-x.

Moran, R., Teodorescu, A.R., Usher, M., 2015. Post choice information integration as a causal determinant of confidence: novel data and a computational account. Cognit. Psychol. 78, 99–147. https://doi.org/10.1016/j.cogpsych.2015.01.002.

Morey, R.D., 2008. Confidence intervals from normalized data: a correction to cousineau (2005). Tutor. Quant. Methods Psychol. 4, 61–64. https://doi.org/10.20982/tqmp.04.2.p061.

Morey, R.D., Rouder, J.N., 2015. BayesFactor: computation of Bayes factors for common designs. R package version 0 9, 10–11.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7, 308–313.

Nieuwenhuis, S., de Geus, E.J., Aston-Jones, G., 2011. The anatomical and functional relationship between the P3 and autonomic components of the orienting response. Psychophysiology 48, 162–175. https://doi.org/10.1111/j.1469-8986.2010.01057.x.

Nieuwenhuis, S., Ridderinkhof, R.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.

O'Connell, R.G., Dockree, P.M., Kelly, S.P., 2012. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nat. Neurosci. 15, 1729–1735. https://doi.org/10.1038/nn.3248.

Odegaard, B., Grimaldi, P., Cho, S.H., Peters, M.A.K., Lau, H., Basso, M.A., 2018. Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. Proc. Natl. Acad. Sci. Unit. States Am. 115, E1588–E1597. https://doi.org/10.1073/pnas.1711628115.

Overbeek, T.J.M., Nieuwenhuis, S., Ridderinkhof, K.R., 2005. Dissociable components of error processing: on the functional significance of the Pe vis-à-vis the ERN/Ne. J. Psychophysiol. 19, 319–329. https://doi.org/10.1027/0269-8803.19.4.319.

Peirce, J.W., 2009. Generating stimuli for neuroscience using PsychoPy. Front. Neuroinf. 2, 1–8. https://doi.org/10.3389/neuro.11.010.2008.

Peirce, J.W., 2007. PsychoPy: psychophysics software in Python. J. Neurosci. Methods 162, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017.

Perrin, F., Pernier, J., Bertrand, O., Echallier, J., 1989. Spherical splines for scalp potential and current density mapping. Electroencephalogr. Clin. Neurophysiol. 72, 184–187.

Peters, M.A.K., Thesen, T., Ko, Y.D., Maniscalco, B., Carlson, C., Davidson, M., Doyle, W., Kuzniecky, R., Devinsky, O., Halgren, E., Lau, H., 2017. Perceptual confidence neglects decision-incongruent evidence in the brain. Nat. Hum. Behav. 1 https://doi.org/10.1038/s41562-017-0139.

Philiastides, M.G., Heekeren, H.R., Sajda, P., 2014. Human scalp potentials reflect a mixture of decision- related signals during perceptual choices. J. Neurosci. 34, 16877–16889. https://doi.org/10.1523/JNEUROSCI.3012-14.2014.

Pitts, M.A., Padwal, J., Fennelly, D., Martínez, A., Hillyard, S.A., 2014. Gamma band activity and the P3 reflect post-perceptual processes , not visual awareness. Neuroimage 101, 337–350. https://doi.org/10.1016/j.neuroimage.2014.07.024.

Pleskac, T.J., Busemeyer, J.R., 2010. Two-stage dynamic signal Detection : a theory of choice , decision time, and confidence. Psychol. Rev. 117, 864–901. https://doi.org/10.1037/a0019737.

Pouget, A., Drugowitsch, J., Kepecs, A., 2016. Confidence and certainty: distinct probabilistic quantities for different goals. Nat. Neurosci. 19, 366–374. https://doi.org/10.1038/nn.4240.

Rausch, M., Hellmann, S., Zehetleitner, M., 2018. Confidence in masked orientation judgments is informed by both evidence and visibility. Atten. Percept. Psychophys. 80, 134–154. https://doi.org/10.3758/s13414-017-1431-5.

Rausch, M., Zehetleitner, M., 2019a. The folded X-pattern is not necessarily a statistical signature of decision confidence. PLoS Comput. Biol. 15, e1007456 https://doi.org/10.1371/journal.pcbi.1007456.

Rausch, M., Zehetleitner, M., 2019b. Modelling visibility judgments using models of decision confidence. PsyArXiv. https://doi.org/10.31219/osf.io/7dakz.

Rausch, M., Zehetleitner, M., 2017. Should metacognition be measured by logistic regression? Conscious. Cogn 49, 291–312. https://doi.org/10.1016/j.concog.2017.02.007.

Rausch, M., Zehetleitner, M., 2016. Visibility is not equivalent to confidence in a low contrast orientation discrimination task. Front. Psychol. 7, 591. https://doi.org/10.3389/fpsyg.2016.00591.

Resulaj, A., Kiani, R., Wolpert, D.M., Shadlen, M.N., 2009. Changes of mind in decision-making. Nature 461, 263–266. https://doi.org/10.1038/nature08275.

Riesel, A., Weinberg, A., Endrass, T., Meyer, A., Hajcak, G., 2013. The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks. Biol. Psychol. 93, 377–385. https://doi.org/10.1016/j.biopsycho.2013.04.007.

Rolls, E.T., Grabenhorst, F., Deco, G., 2010. Decision-making, errors, and confidence in the brain. J. Neurophysiol. 104, 2359–2374. https://doi.org/10.1152/jn.00571.2010.

Rouder, J.N., Morey, R.D., 2012. Default Bayes factors for model selection in regression. Multivariate Behav. Res. 47, 877–903.

Rouder, J.N., Speckman, P.L., Son, D., Morey, R.D., 2009. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon. Bull. Rev. 16, 225–237. https://doi.org/10.3758/PBR.16.2.225.

Samaha, J., Barrett, J.J., Sheldon, A.D., LaRocque, J.J., Postle, B.R., 2016. Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory. Front. Psychol. 7, 851. https://doi.org/10.3389/fpsyg.2016.00851.

Sanders, J.I., Hangya, B., Kepecs, A., 2016. Signatures of a statistical computation in the human sense of confidence. Neuron 90, 499–506. https://doi.org/10.1016/j.neuron.2016.03.025.

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.

Schwarz, G., 1978. Estimating the dimensions of a model. Ann. Stat. 6, 461–464. https://doi.org/10.1214/aos/1176348654.

Sergent, C., Baillet, S., Dehaene, S., 2005. Timing of the brain events underlying access to consciousness during the attentional blink. Nat. Neurosci. 8, 1391–1400. https://doi.org/10.1038/nn1549.

Shea, N., Frith, C.D., 2019. The global workspace needs metacognition. Trends Cognit. Sci. 23, 560–571. https://doi.org/10.1016/j.tics.2019.04.007.

Steinhauser, M., Maier, M., Hübner, R., 2008. Modeling behavioral measures of error detection in choice tasks: response monitoring versus conflict monitoring. J. Exp. Psychol. Hum. Percept. Perform. 34, 158–176.

Steinhauser, M., Yeung, N., 2012. Error awareness as evidence accumulation: effects of speed-accuracy trade-off on error signaling. Front. Hum. Neurosci. 6, 1–12. https://doi.org/10.3389/fnhum.2012.00240.

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.

Stolyarova, A., Rakhshan, M., Hart, E.E., O'Dell, T.J., Peters, M.A.K., Lau, H., Soltani, A., Izquierdo, A., 2019. Dissociable roles for anterior cingulate cortex and basolateral amygdala in decision confidence and learning under uncertainty. Nat. Commun. 10, 1–14. https://doi.org/10.1101/655860.

Tagliabue, C.F., Veniero, D., Benwell, C.S.Y., Cecere, R., Savazzi, S., Thut, G., 2019. Subjective perceptual experience tracks the neural signature of sensory evidence accumulation during decision formation. Sci. Rep. 9, 1–12. https://doi.org/10.1038/s41598-019-41024-4.

Twomey, D.M., Murphy, P.R., Kelly, S.P., O'Connell, R.G., 2015. The classic P300 encodes a build-to-threshold decision variable. Eur. J. Neurosci. 42, 1636–1643. https://doi.org/10.1111/ejn.12936.

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

Urai, A.E., Braun, A., Donner, T.H., 2017. Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. Nat. Commun. 8, 14637. https://doi.org/10.1038/ncomms14637.

van den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M.N., Wolpert, D.M., 2016. A common mechanism underlies changes of mind about decisions and confidence. Elife 5, e12192. https://doi.org/10.7554/eLife.12192.

Verleger, R., Jaśkowski, P., Wascher, E., 2005. Evidence for an integrative role of P3b in linking reaction to perception. J. Psychophysiol. 19, 165–181. https://doi.org/10.1027/0269-8803.19.3.165.

Vickers, D., 1970. Evidence for an accumulator model of psychophysical discrimination. Ergonomics 13, 37–58. https://doi.org/10.1080/00140137008931117.

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., Danielmeier, C., Ullsperger, M., 2011. Error awareness revisited: accumulation of multimodal evidence from central and autonomic nervous systems. J. Cognit. Neurosci. 23, 3021–3036. https://doi.org/10.1162/jocn.2011.21635.

Wickens, T.D., 2002. Elementary Signal Detection Theory. Oxford University Press, New York.

Zehetleitner, M., Rausch, M., 2013. Being confident without seeing: what subjective measures of visual consciousness are about. Atten. Percept. Psychophys. 75, 1406–1426. https://doi.org/10.3758/s13414-013-0505-2.

Cognitive modelling reveals distinct electrophysiological markers of decision confidence and error monitoring

Supplementary Material

Manuel Rausch[1*], Michael Zehetleitner[1], Marco Steinhauser[2], and Martin E. Maier[2]

Catholic University of Eichstätt-Ingolstadt

[1] Katholische Universität Eichstätt-Ingolstadt, Fakultät für Psychologie und Pädagogik, Professur für Allgemeine Psychologie II, Eichstätt, Germany.

[2] Katholische Universität Eichstätt-Ingolstadt, Fakultät für Psychologie und Pädagogik, Lehrstuhl für Allgemeine Psychologie, Eichstätt, Germany

*Correspondence: Manuel Rausch, Katholische Universität Eichstätt-Ingolstadt. Psychologie II, Ostenstraße 25, 85072 Eichstätt, Germany. E-Mail: manuel.rausch@ku.de.

Supplementary Table S1. Formulae for calculating the likelihood of the data given the parameters for all models assuming continuous decision variables depending on the identity of the stimulus S_{id}, the stimulus strength S_s, the identification judgment R_{id}, as well as the confidence judgment R_v. f indicates the Gaussian probability density function.

Model	R_{id}	C	$P(R_{id}, C \mid S_{id}, S_s)$
SDT	0	1	$\int_{\theta_{c01}}^{\theta_{id}} f(x \mid (S_{id} - \frac{1}{2}) \times S_s, \sigma_{id})\ dx$
	0	2	$\int_{\theta_{c02}}^{\theta_{c01}} f(x \mid (S_{id} - \frac{1}{2}) \times S_s, \sigma_{id})\ dx$
	0	5	$\int_{-\infty}^{\theta_{c04}} f(x \mid (S_{id} - \frac{1}{2}) \times S_s, \sigma_{id})\ dx$

	1	1	$\int_{\theta_{id}}^{\theta_{c11}} f(x\vert(S_{id} - \frac{1}{2}) \times S_s, \sigma_{id})\ dx$
	1	2	$\int_{\theta_{c11}}^{\theta_{c12}} f(x\vert(S_{id} - \frac{1}{2}) \times S_s, \sigma_{id})\ dx$
	1	5	$\int_{\theta_{c14}}^{\infty} f(x\vert\mu(S_{id} - \frac{1}{2}) \times S_s, \sigma_{id})\ dx$
Noisy SDT	0	1	$\int_{-\infty}^{\theta_{id}} f\left(x\middle\vert\left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) \times \left(\int_{\theta_{c01}}^{\infty} f(y\ \vert x, \sigma_c)dy\right) dx$
	0	2	$\int_{-\infty}^{\theta_{id}} f\left(x\middle\vert\left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) \times \left(\int_{\theta_{c02}}^{\theta_{c01}} f(y\ \vert x, \sigma_c)dy\right) dx$

	0	5	$\int_{-\infty}^{\theta_{id}} f\left(x\middle\vert\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{-\infty}^{\theta_{c04}} f(y\,\vert x,\sigma_c)dy\right)dx$
	1	1	$\int_{\theta_{id}}^{\infty} f\left(x\middle\vert\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{-\infty}^{\theta_{c11}} f(y\,\vert x,\sigma_c)dy\right)dx$
	1	2	$\int_{\theta_{id}}^{\infty} f\left(x\middle\vert\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{\theta_{c11}}^{\theta_{c12}} f(y\,\vert x,\sigma_c)dy\right)dx$
	1	5	$\int_{\theta_{id}}^{\infty} f\left(x\middle\vert\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{\theta_{c14}}^{\infty} f(y\,\vert x,\sigma_c)dy\right)dx$
Noisy decay model	0	1	$\int_{-\infty}^{\theta_{id}} f\left(x\middle\vert\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{\theta_{c01}}^{\infty} f(y\,\vert x\,\times\rho_S,\sigma_c)dy\right)dx$

	0	2	$\int_{-\infty}^{\theta_{id}} f\left(x \mid \left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) \times \left(\int_{\theta_{c02}}^{\theta_{c01}} f(y \mid x \times \rho_S, \sigma_c)dy\right) dx$
	0	5	$\int_{-\infty}^{\theta_{id}} f\left(x \mid \left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) \times \left(\int_{-\infty}^{\theta_{c04}} f(y \mid x \times \rho_S, \sigma_c)dy\right) dx$
	1	1	$\int_{\theta_{id}}^{\infty} f\left(x \mid \left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) \times \left(\int_{-\infty}^{\theta_{c11}} f(y \mid x \times \rho_S, \sigma_c)dy\right) dx$
	1	2	$\int_{\theta_{id}}^{\infty} f\left(x \mid \left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) \times \left(\int_{\theta_{c11}}^{\theta_{c12}} f(y \mid x \times \rho_S, \sigma_c)dy\right) dx$
	1	5	$\int_{\theta_{id}}^{\infty} f\left(x \mid \left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) \times \left(\int_{\theta_{c14}}^{\infty} f(y \mid x \times \rho_S, \sigma_c)dy\right) dx$

WEV-model	0	1	$\int_{-\infty}^{\theta_{id}} f\left(x\middle\|\mu\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{\theta_{c01}}^{\infty} f(y\,\|(1-w)\times x+w\times \mathrm{sgn}(x-\theta_{id})\times\ (S_s-\ \overline{S_s}),\sigma_c)dy\right)dx$
	0	2	$\int_{-\infty}^{\theta_{id}} f\left(x\middle\|\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{\theta_{c02}}^{\theta_{c01}} f(y\,\|(1-w)\times x+w\times \mathrm{sgn}(x-\theta_{id})\times\ (S_s-\ \overline{S_s}),\sigma_c)dy\right)dx$
	0	5	$\int_{-\infty}^{\theta_{id}} f\left(x\middle\|\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{-\infty}^{\theta_{c04}} f(y\,\|(1-w)\times x+w\times \mathrm{sgn}(x-\theta_{id})\times\ (S_s-\ \overline{S_s}),\sigma_c)dy\right)dx$
	1	1	$\int_{\theta_{id}}^{\infty} f\left(x\middle\|\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{-\infty}^{\theta_{c11}} f(y\,\|(1-w)\times x+w\times \mathrm{sgn}(x-\theta_{id})\times\ (S_s-\ \overline{S_s}),\sigma_c)dy\right)dx$
	1	2	$\int_{\theta_{id}}^{\infty} f\left(x\middle\|\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{\theta_{c11}}^{\theta_{c12}} f(y\,\|(1-w)\times x+w\times \mathrm{sgn}(x-\theta_{id})\times\ (S_s-\ \overline{S_s}),\sigma_c)dy\right)dx$

	1	5	$\int_{\theta_{id}}^{\infty} f\left(x\middle\| \left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) \times \left(\int_{\theta_{c14}}^{\infty} f(y \mid (1 - w) \times x + w \times \mathrm{sgn}(x - \theta_{id}) \times \ (S_s - \ \overline{S_s}), \sigma_c)dy \right) dx$
Two-channel model	0	1	$\int_{-\infty}^{\theta_{id}} f\left(x\middle\| \left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) dx \times \int_{\theta_{c01}}^{\infty} f(y \mid \left(S_{id} - \frac{1}{2}\right) \times S_s \ \times a, 1)dy$
	0	2	$\int_{-\infty}^{\theta_{id}} f\left(x\middle\| \left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) dx \times \int_{\theta_{c02}}^{\theta_{c01}} f(y \mid \left(S_{id} - \frac{1}{2}\right) \times S_s \ \times a, 1)dy$
	0	5	$\int_{-\infty}^{\theta_{id}} f\left(x\middle\| \left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) dx \times \int_{-\infty}^{\theta_{c04}} f(y \mid \left(S_{id} - \frac{1}{2}\right) \times S_s \ \times a, 1)dy$
	1	1	$\int_{\theta_{id}}^{\infty} f\left(x\middle\| \left(S_{id} - \frac{1}{2}\right) \times S_s, \sigma_{id}\right) dx \times \int_{-\infty}^{\theta_{c11}} f(y \mid \left(S_{id} - \frac{1}{2}\right) \times S_s \ \times a, 1)dy$

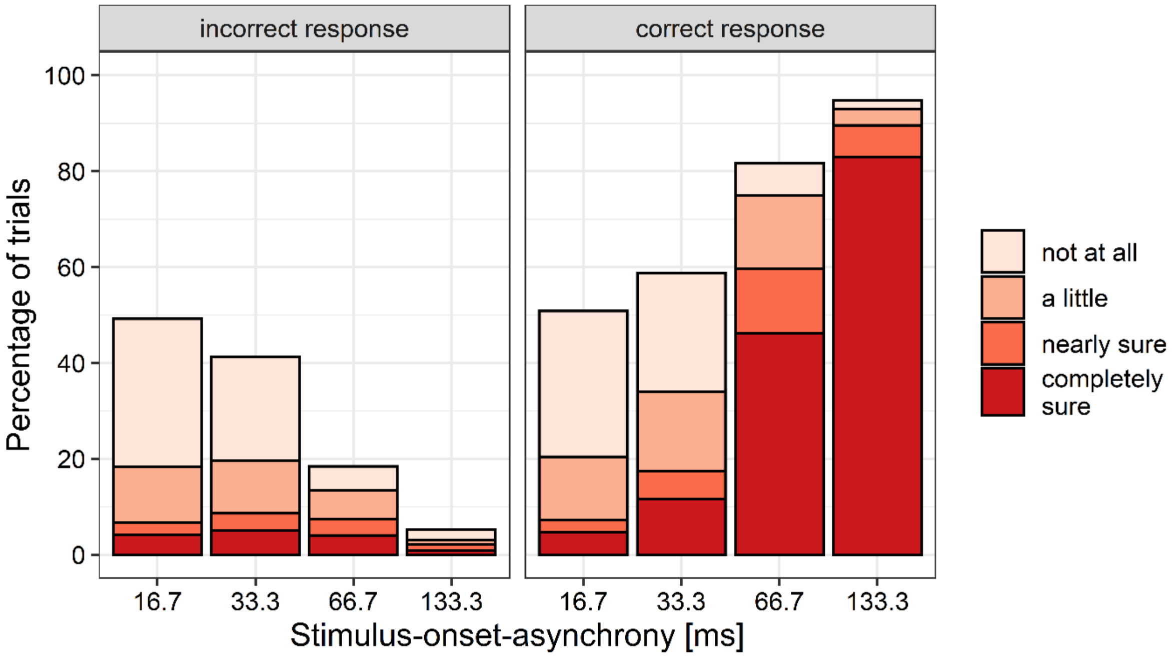
	1	2	$\int_{\theta_{id}}^{\infty} f\left(x \middle\| \left(S_{id}-\frac{1}{2}\right)\times S_s, \sigma_{id}\right) dx \times \int_{\theta_{c11}}^{\theta_{c12}} f(y \mid \left(S_{id}-\frac{1}{2}\right)\times S_s \ \times a, 1)dy$
	1	5	$\int_{\theta_{id}}^{\infty} f\left(x \middle\| \left(S_{id}-\frac{1}{2}\right)\times S_s, \sigma_{id}\right) dx \times \int_{\theta_{c14}}^{\infty} f(y \mid \left(S_{id}-\frac{1}{2}\right)\times S_s \ \times a, 1)dy$
SDT model with postdecisional evidence	0	1	$\int_{-\infty}^{\theta_{id}} f\left(x \middle\| \left(S_{id}-\frac{1}{2}\right)\times S_s, \sigma_{id}\right) \times \left(\int_{\theta_{c01}}^{\infty} f(y \mid x \ + \ (2\ S_{id} - \ 1)\times S_s \times b, \sqrt{b})dy\right) dx$
	0	2	$\int_{-\infty}^{\theta_{id}} f\left(x \middle\| \left(S_{id}-\frac{1}{2}\right)\times S_s, \sigma_{id}\right) \times \left(\int_{\theta_{c02}}^{\theta_{c01}} f(y \mid x \ + \ (2\ S_{id} - \ 1)\times S_s \times b, \sqrt{b})dy\right) dx$
	0	5	$\int_{-\infty}^{\theta_{id}} f\left(x \middle\| \left(S_{id}-\frac{1}{2}\right)\times S_s, \sigma_{id}\right) \times \left(\int_{-\infty}^{\theta_{c04}} f(y \mid x \ + \ (2\ S_{id} - \ 1)\times S_s \times b, \sqrt{b})dy\right) dx$

	1	1	$\int_{\theta_{id}}^{\infty} f\left(x\middle|\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{-\infty}^{\theta_{c11}} f(y\,\|x + (2\,S_{id}-1)\times S_s\times b,\sqrt{b})dy\right)dx$
	1	2	$\int_{\theta_{id}}^{\infty} f\left(x\middle|\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{\theta_{c11}}^{\theta_{c12}} f(y\,\|x + (2\,S_{id}-1)\times S_s\times b,\sqrt{b})dy\right)dx$
	1	5	$\int_{\theta_{id}}^{\infty} f\left(x\middle|\left(S_{id}-\frac{1}{2}\right)\times S_s,\sigma_{id}\right)\times\left(\int_{\theta_{c14}}^{\infty} f(y\,\|x + (2\,S_{id}-1)\times S_s\times b,\sqrt{b})dy\right)dx$
Detection heuristic model	0	1	$\int_{-\infty}^{\theta_{c01}} f(x\|(1-S_{id})\times S_s-b,\sigma_{id})\times\left(\int_{-\infty}^{x} f(y\|(S_{id}-1)\times S_s+b,\sigma_{id})\,dy\right)dx$
	0	2	$\int_{\theta_{c01}}^{\theta_{c02}} f(x\|\mu = (1-S_{id})\times S_s-b,\sigma_{id})\times\left(\int_{-\infty}^{x} f(y\|\mu = (S_{id}-1)\times S_s+b,\sigma_{id})\,dy\right)dx$

	0	5	$\int_{\theta_{c04}}^{\infty} f(x\|(1- \ S_{id}) \times S_s - b, \sigma_{id}) \times \left(\int_{-\infty}^{x} f(y\|(S_{id} - 1) \times S_s + b, \sigma_{id})\, dy \right) dx$
	1	1	$\int_{-\infty}^{\theta_{c11}} f(x\|(S_{id} - 1\ ) \times S_s + b, \sigma_{id}) \times \left(\int_{-\infty}^{x} f(y\|(1- \ S_{id}) \times S_s - b, \sigma_{id})\, dy \right) dx$
	1	2	$\int_{\theta_{c11}}^{\theta_{c12}} f(x\|(S_{id} - 1\ ) \times S_s + b, \sigma_{id}) \times \left(\int_{-\infty}^{x} f(y\|(1- \ S_{id}) \times S_s - b, \sigma_{id})\, dy \right) dx$
	1	5	$\int_{\theta_{c14}}^{\infty} f(x\|(S_{id} - 1\ ) \times S_s + b, \sigma_{id}) \times \left(\int_{-\infty}^{x} f(y\|(1- \ S_{id}) \times S_s - b, \sigma_{id})\, dy \right) dx$

Supplementary Table S2. Formulae for calculating the likelihood of the data the two high threshold model depending on the identity of the stimulus S_{id}, and the discrimination response R_{id}.

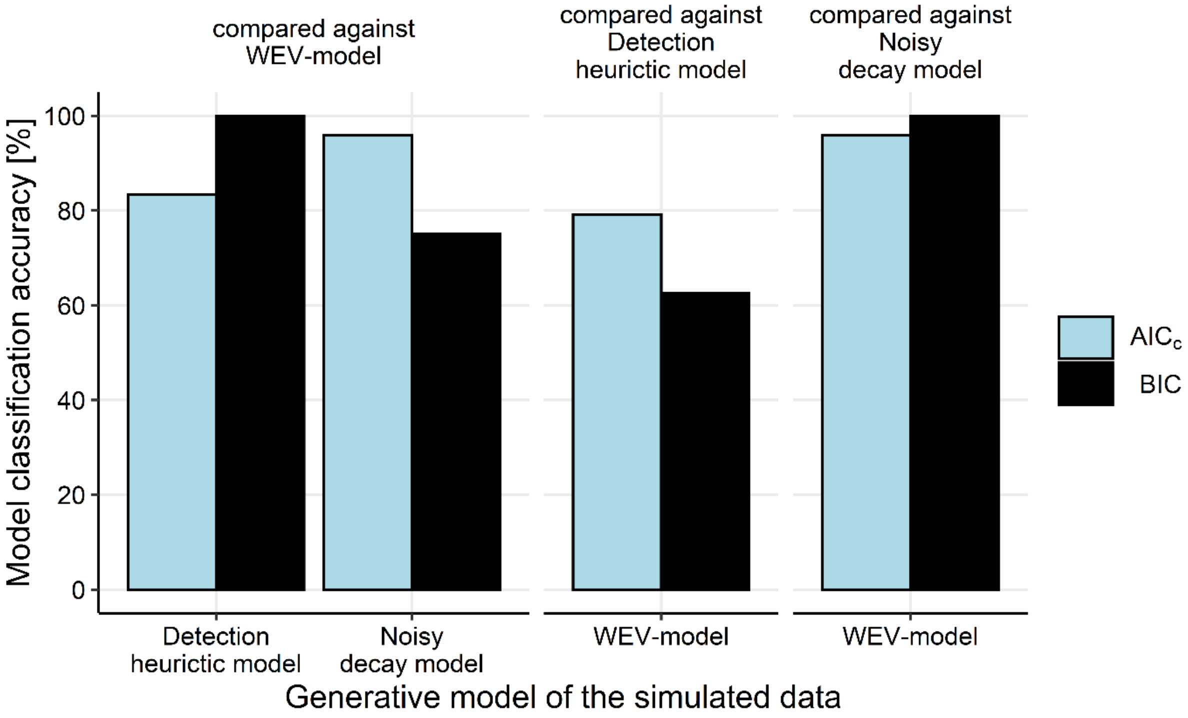
S_{id}	R_{id}	$P(R_{id}, C = c \mid S_{id}, S_s)$
0	0	$p(\delta_{id} = 0 \mid S_s, S_{id} = 0) \times p(C = c \mid \delta_{id} = 0) +$ $(1 - p(\delta_{id} = 0 \mid S_s, S_{id} = 0)) \times (1 - g) \times p(C = c \mid \delta_{id} = 0.5, R_{id} = 0)$
1	0	$(1 - p(\delta_{id} = 1 \mid S_s, S_{id} = 1)) \times (1 - g) \times p(C = c \mid \delta_{id} = 0.5, R_{id} = 1)$
0	1	$(1 - p(\delta_{id} = 0 \mid S_s, S_{id} = 0)) \times g \times p(C = c \mid \delta_{id} = 0.5, R_{id} = 0)$
1	1	$p(\delta_{id} = 1 \mid S_s, S_{id} = 1) \times p(C = c \mid \delta_{id} = 1) +$ $+(1 - p(\delta_{id} = 1 \mid S_s, S_{id} = 1)) \times g \times p(C = c \mid \delta_{id} = 0.5, R_{id} = 1)$



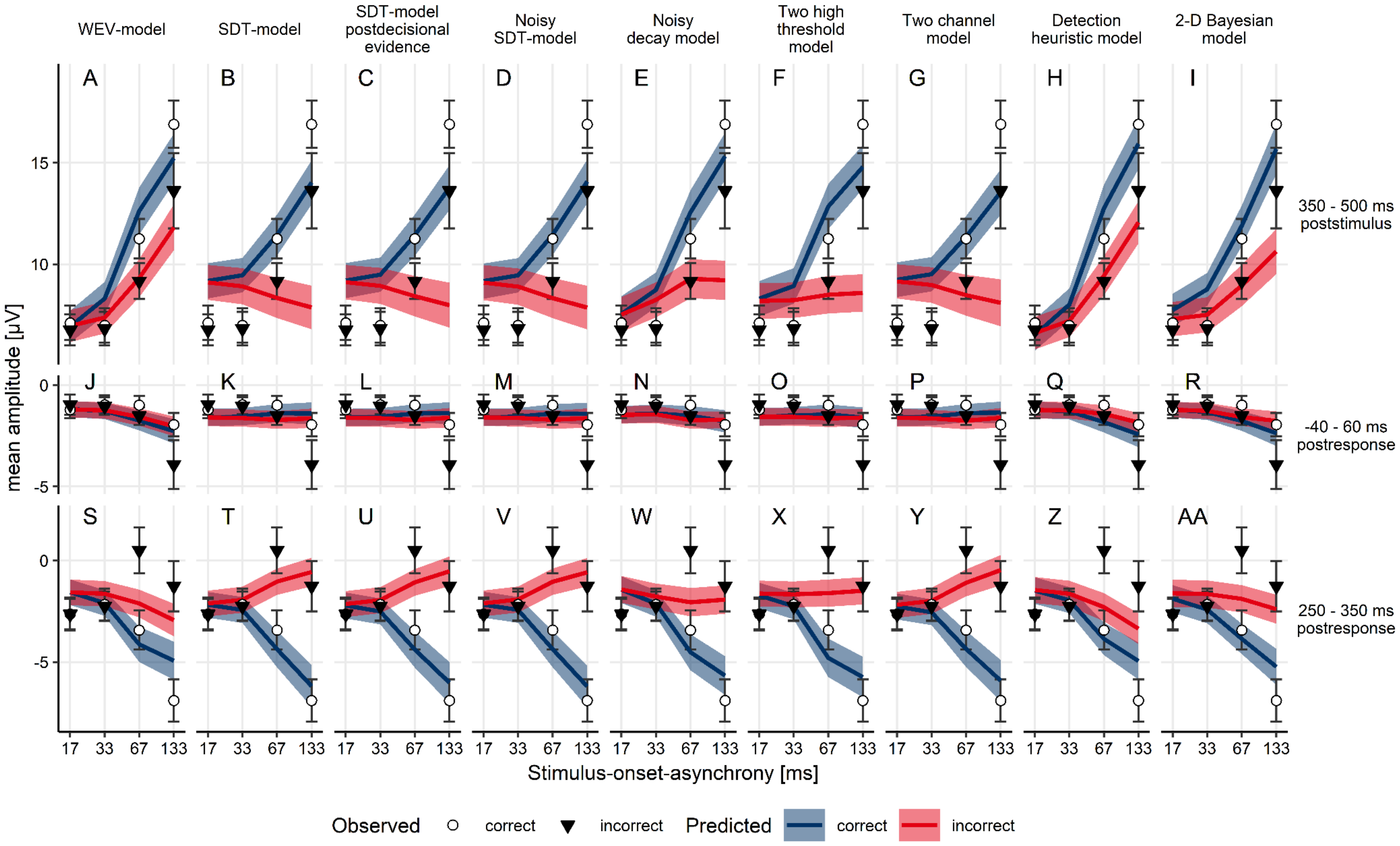
Supplementary Figure S1. Proportion of incorrect confidence judgments (left panel) and correct confidence judgments (right panels) as a function of SOA (x-axis). Stacked bars indicate the fraction of trials with a specific confidence rating.

Supplementary Table S3. Summary statistics of the parameters of the WEV-model obtained during model fitting

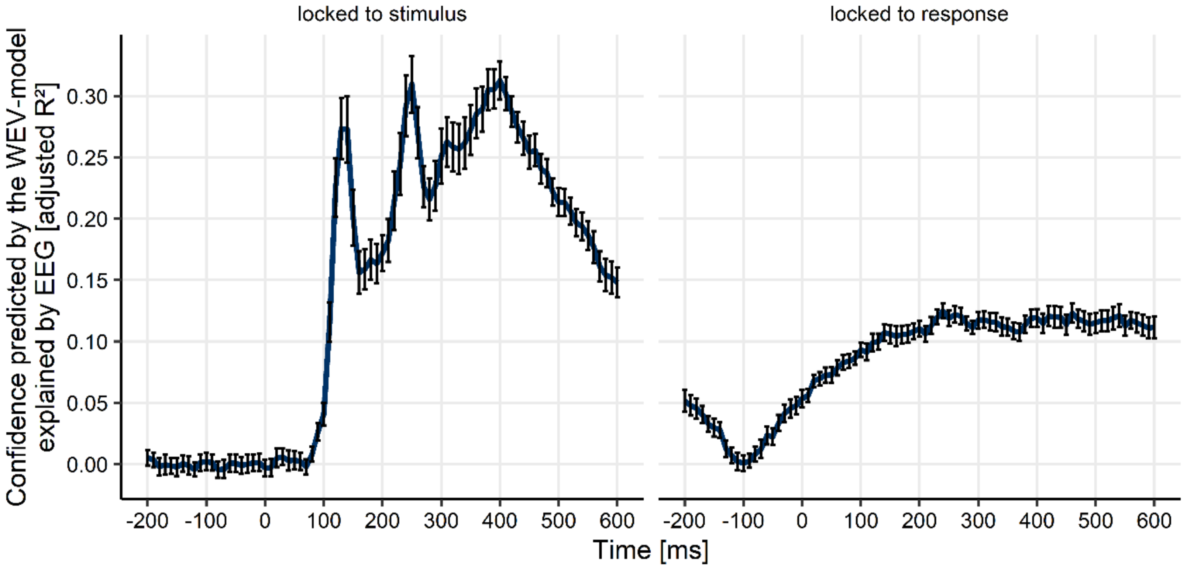
Parameter	Interpretation	*M*	*SD*	*Min*	*Max*
S_{S1}	stimulus strength at the SOA of 16.7 ms	-0.07	0.22	-0.62	0.39
S_{S2}	stimulus strength at the SOA of 33.3 ms	0.68	0.59	-0.07	2.11
S_{S3}	stimulus strength at the SOA of 66.7 ms	2.56	1.38	0.02	5.90
S_{S4}	stimulus strength at the SOA of 133.4 ms	4.05	1.49	1.03	7.21
ϑ_{id}	criterion with respect to the stimulus discrimination judgment	-0.22	0.36	-0.85	0.40
w	degree to which confidence relies on sensory evidence about the identity or on identity-irrelevant evidence	0.36	0.20	0.06	0.91
σ_c	noise superimposed on the confidence judgment	0.54	0.51	0.02	2.85
ϑ_{c01}	confidence criterion delineating C=1 and C=2, if R_{id} = 0	0.02	0.64	-0.85	1.87
ϑ_{c02}	confidence criterion delineating C=2 and C=3, if R_{id} = 0	-0.68	0.47	-1.40	0.51
ϑ_{c03}	confidence criterion delineating C=3 and C=4, if R_{id} = 0	-1.10	0.50	-2.34	0.41
ϑ_{c11}	confidence criterion delineating C=1 and C=2, if R_{id} = 1	-0.05	0.75	-2.54	1.08
ϑ_{c12}	confidence criterion delineating C=2 and C=3, if R_{id} = 1	0.72	0.33	0.18	1.42
ϑ_{c13}	confidence criterion delineating C=3 and C=4, if R_{id} = 01	1.17	0.49	0.28	2.37



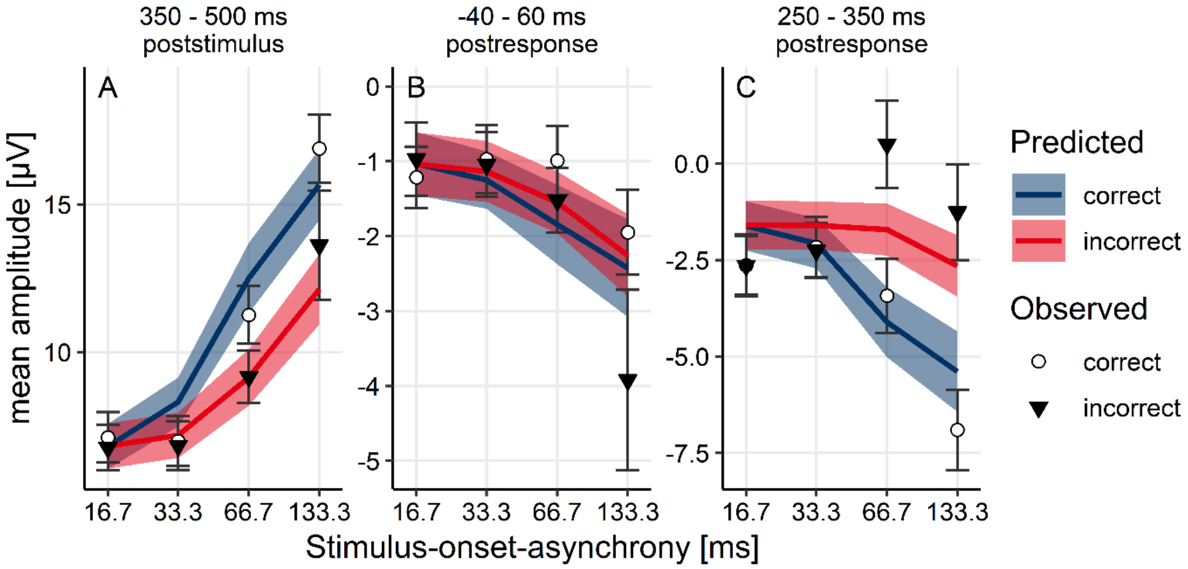
Supplementary Figure S2. Each bar represents model classification accuracy when data was simulated according to the model on the x-axis, and then the model fit of the generative model on the x-axis was compared against the model fit of an alternative model in different panels. Colours indicate different goodness-of-fit measures. We always fitted both the true generative model as well as an alternative model to the simulated data. Model classification was considered as correct if AIC$_c$ or BIC of the generative model was smaller than AIC$_c$ or BIC of the alternative model. For example, 100% model classification accuracy with BIC for the detection heuristic model as generative model compared against the WEV model means that we generated data sets according the detection heuristic model, fitted both the detection heuristic model and the WEV model to the simulated data, and for all simulated data sets, the BIC associated with the detection heuristic model was smaller than the BIC associated with the WEV model. The WEV model, the noisy decay model, and the detection heuristic model were selected for model classification analysis because these three models had performed best when we fitted the behavioural data.


WEV-model
SDT-model
SDT-model postdecisional evidence
Noisy SDT-model
Noisy decay model
Two high threshold model
Two channel model
Detection heuristic model
2-D Bayesian model
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
AA
15
10
0
-5
0
-5
350 - 500 ms poststimulus
-40 - 60 ms postresponse
250 - 350 ms postresponse
mean amplitude [µV]
17
33
67
133
Stimulus-onset-asynchrony [ms]
Observed
○ correct
▼ incorrect
Predicted
correct
incorrect

Supplementary Fig S3. Comparison between predicted and observed amplitudes (A-I) in the P3 time window, (J-R) in the ERN time window, and (S-AA) the Pe time window, depending on all the models used in the present study to fit the behavioural data (in columns), stimulus-onset-asynchrony (x-Axis) and accuracy of the orientation response (colours and symbols). Symbols: observed data. Lines: Prediction based on the parameters of the WEV-model fitted to confidence judgments as well as a linear transformation of confidence. Error bars = 1 within-subject *SEM*. Ribbons = 1 within-subject *SEM*. When the noisy decay model was used to predict ERP amplitudes, the correlation between predicted and observed amplitudes was stronger at the time of the P3 (see panel E), M = .33, 95% HDI = [.29 .36], compared to the correlations at the time of the ERN (see panel N), M = .09, 95% HDI = [.06 .11], and Pe (panel W), M = .16, 95% HDI = [.12 .20]. Likewise, when the detection heuristic model was used, the results were the same, P3 (panel H): M = .33, 95% HDI = [.29 .37], ERN (panel Q): M = .09, 95% HDI = [.06 .11], and Pe (panel Z): M = .17, 95% HDI = [.12 .20].



Supplementary Fig S4. Fraction of variance of confidence predicted by the WEV-model explained by EEG activity (y-axis) as a function of time locked to the onset of the stimulus (left panel) and locked to the response (right panel). Each dot represents the average adjusted R^2 of multivariate regression analyses with confidence predicted by the WEV-model as outcome variable and all EEG sensors as predictors. R^2 was adjusted for the number of predictors. Error bars = 1 within-subject *SEM*.



Supplementary Fig. S5. Comparison between predicted and observed amplitudes (A) in the P3 time window, (B) in the ERN time window, and (C) in the Pe time window, depending on stimulus-onset-asynchrony (x-Axis) and accuracy of the orientation response (colours and symbols). Symbols: observed data. Lines: Prediction based on the parameters of the WEV-model fitted to confidence judgments as well as a monotonous transformation of confidence. Error bars = 1 within-subject *SEM*. Ribbons = 1 within-subject *SE*