Metacognitive Decisions on Decision Accuracy: Confidence Judgment and Changes of Mind

Yiu Hong Ko

ORCID: 0000-0003-4436-0166

Doctor of Philosophy

December, 2022

Faculty of Medicine, Dentistry and Health Sciences

Melbourne School of Psychological Sciences

The University of Melbourne

Faculty of Human Sciences

University of Cologne

Submitted in total fulfilment for the jointly awarded degree of Doctor of Philosophy at the University of Melbourne and University of Cologne

Inauguraldissertation

zur

Erlangung des Doktorgrades

der Humanwissenschaftlichen Fakultät

der Universität zu Köln

nach der Promotionsordnung vom 18.12.2018

vorgelegt von

Yiu Hong, Ko

aus

Hong Kong

December / 2022

Abstract

Even in the absence of external feedback, humans are capable of subjectively estimating the accuracy of their own decisions, resulting in a sense of confidence that a decision is correct. While decision confidence has been proposed to be closely related to other metacognitive judgments, including error awareness (i.e., awareness that a decisions error has occurred) and changes of mind (i.e., reversal of previously made decisions), their relationships so far remain unclear.

The current project aimed to investigate how confidence could be related to metacognitive judgments from two perspectives. First, Studies 1 and 2 investigated how confidence and changes of mind were affected by changes in different stimulus properties, particularly absolute evidence strength. In a brightness judgment task, participants were presented with two flickering, grayscale squares and required to select the square that appeared brighter. After each trial, participants reported their subjective accuracy on a rating scale ranging from "surely incorrect" to "surely correct". Results showed that with stronger absolute evidence (i.e., increased overall luminance across both squares), confidence was increased and the proportion of changes of mind trials was reduced. These consistent changes support the hypothesis that higher confidence could contribute to less frequent changes of mind.

Second, Study 3 investigated the relationships between confidence and the event-related potential (ERP) components of the centro-parietal potential (CPP) and the error positivity (Pe), which have been respectively proposed to be indexes of pre- and post-decisional evidence accumulation processes. In the same brightness judgment task, it was found that the relationships between confidence and these two ERP components depended on decision accuracy: Confidence was positively related to CPP amplitudes in correct trials, but negatively related to Pe amplitudes in error trials. These findings

suggest that confidence in correct and error decisions involve different pre- and post-decisional processes.

Overall, the current findings suggest that (a) confidence could serve as a basis of changes of mind, and (b), confidence in correct and erroneous decisions was differentially related to pre- and post-decisional ERP indexes of evidence accumulation. Taken together, they suggest that confidence might emerge during decision formation and could, with the contribution from post-decisional processes, serve as a basis of changes of mind.

Abstract

Auch ohne (externes) Feedback sind Menschen in der Lage, die Genauigkeit ihrer eigenen Entscheidungen einzuschätzen, was zu einer Sicherheit führt, dass die jeweilige Entscheidung richtig ist (*confidence,* Entscheidungssicherheit). Es wurde zwar vermutet, dass die Entscheidungssicherheit in engem Zusammenhang mit anderen metakognitiven Konzepten steht, beispielweise dem Fehlerbewusstsein (d. h. dem Bewusstsein, dass ein Fehler aufgetreten ist) und der Meinungsänderung (d. h. der Änderung bzw. Umkehrung zuvor getroffener Entscheidungen), aber die Zusammenhänge zwischen diesen Konzepten sind bisher unklar.

Das aktuelle Projekt beabsichtigt, aus zwei Perspektiven zu untersuchen, wie die Entscheidungssicherheit mit metakognitiven Konzepten zusammenhängt. Zuerst wurde in den Studien 1 und 2 untersucht, wie Entscheidungssicherheit und Meinungsänderung durch Veränderungen verschiedener Reizeigenschaften, vor allem der absoluten Evidenzstärke dieser Reize, beeinflusst werden können. In einer Aufgabe sollten die Teilnehmenden die Helligkeit von zwei visuellen Reizen beurteilen. Dazu wurden den Teilnehmenden zwei flackernde Quadrate in verschiedenen Graustufen präsentiert, von denen sie das jeweils hellere Quadrat auswählen sollten. Nach jedem Versuch berichteten die Teilnehmenden die subjektive Genauigkeit ihrer Entscheidung mithilfe einer Bewertungsskala, die von "sicher falsch" bis "sicher richtig" reichte. Die Ergebnisse zeigten, dass mit stärkerer absoluter Evidenz (d. h. mit höherer Gesamt-Helligkeit beider Quadrate) die Entscheidungssicherheit zunahm und der Anteil der Durchgänge, bei denen die Meinung geändert wurde, abnahm. Diese konsistenten Veränderungen unterstützen die Hypothese, dass eine höhere Entscheidungssicherheit zu weniger häufigen Meinungsänderungen beitragen könnte.

Darüber hinaus untersuchte Studie 3 die Beziehungen zwischen der Entscheidungssicherheit und den ereigniskorrelierten Potenzialen (EKP), centro-parietales Potential (CPP) und Fehlerpositivität (Pe), die als Indizes für prä- und postdezisionale Evidenzakkumulationsprozesse gelten. Unter Verwendung des gleichen Paradigmas wie in Studien 1 und 2 konnte festgestellt werden, dass die Beziehungen zwischen der Entscheidungssicherheit und diesen beiden EKP-Komponenten von der Entscheidungsgenauigkeit abhängen: Entscheidungssicherheit stand im positiven Zusammenhang mit den CPP-Amplituden bei richtigen Versuchen (d.h. korrekten Entscheidungen), aber im negativen Zusammenhang mit Pe-Amplituden bei fehlerhaften Versuchen (d.h. Fehlern). Diese Ergebnisse deuten darauf hin, dass die Entscheidungssicherheit in korrekte und fehlerhafte Entscheidungen unterschiedliche prä- und postdezisionalen Prozesse beinhaltet.

Zusammengefasst deuten die aktuellen Befunde darauf hin, dass (a) die Entscheidungssicherheit als Grundlage für Meinungsänderungen dienen könnte und (b) Entscheidungssicherheit in richtige und falsche Entscheidungen in unterschiedlichem Maße mit prä- und postdezisionalen EKP-Indizes der Evidenzakkumulation verbunden war. Zusammengenommen weisen die vorgelegten Ergebnisse darauf hin, dass Entscheidungssicherheit während der Entscheidungsfindung entsteht und durch postdezisionale Prozesse als Grundlage für Meinungsänderung dienen könnte.

Declaration

This is to certify that

1) This thesis comprises only my original work toward the degree of Doctor of Philosophy, except where otherwise indicated in the preface,
2) Due acknowledgement has been made in the text to all other material used,
3) This thesis is fewer than 100 000 words in length, exclusive of tables, figures, references, footnotes, and appendices.

Ko Yiu Hong

……………………………
Yiu Hong Ko

Preface

For each study reported herein, experimental data was collected by the PhD candidate with assistance from members of the Decision Neuroscience Laboratory at The University of Melbourne. One chapter of this thesis is presented in article format: Chapter 2 presents an article published in *Cognition* on 25/04/2022. The empirical manuscript was drafted by the PhD candidate, and was then revised and edited with the assistance of the manuscript's named co-authors. All co-authors have agreed to the use of the manuscript in this thesis, and have provided signed copies of the co-author authorisation form. The work in this thesis was supported by an Australian Research Council Discovery Project Grant (DP160103353), a Jülich University of Melbourne Postgraduate Academy (JUMPA) scholarship, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 431549029 – SFB 1451.

This dissertation was accepted by the Faculty of Human Sciences of the University of Cologne in June, 2023. (Decision of the Doctoral Committee on 10 October 2010)

Acknowledgements

I would like to thank my supervisors, Stefan Bode, Jutta Stahl, Eva Niessen, and Peter Weiss-Blankenhorn for their guidance throughout the entire PhD project. I also thank all the members in the Decision Neuroscience Laboratory in Melbourne, the Individual Differences and Psychological Assessment Lab in Cologne, and INM-3 in the Juelich Research Center. Particularly I would like to thank Helen Overhoff, Will Turner, Daniel Feuerriegel for always helping me during my candidature.

Publications and Conference Abstracts During Candidature

Publications

Ko, Y. H., Feuerriegel, D., Turner, W., Overhoff, H., Niessen, E., Stahl, J., ... & Bode, S. (2022). Divergent effects of absolute evidence magnitude on decision accuracy and confidence in perceptual judgements. *Cognition*, *225*, 105125.

Overhoff, H., **Ko, Y. H.**, Feuerriegel, D., Fink, G. R., Stahl, J., Weiss, P. H., ... & Niessen, E. (2021). Neural correlates of metacognition across the adult lifespan. *Neurobiology of Aging*, *108*, 34-46.

Overhoff, H., **Ko, Y. H.**, Fink, G. R., Stahl, J., Weiss, P. H., Bode, S., & Niessen, E. (in press). The relationship between response dynamics and the formation of confidence varies across the lifespan. *Frontiers in Aging Neuroscience*.

Conference abstracts

Ko, Y. H., Turner, W,, Feuerriegel, D., Overhoff, H., Niessen, E., Stahl, J., Weiss, P. H., & Bode, S. (2019). The effects of absolute and relative evidence strength on decision confidence. *Australasian Cognitive Neuroscience Society*, Tasmania, Australia **(Oral presentation)**

Ko, Y. H., Turner, W,, Feuerriegel, D., Overhoff, H., Niessen, E., Stahl, J., Weiss, P. H., & Bode, S. (2019). The effects of absolute and relative evidence strength on decision confidence. *Students of Brain Research*, Melbourne, Australia **(Poster presentation)**

Ko, Y. H., Turner, W,, Feuerriegel, D., Overhoff, H., Niessen, E., Stahl, J., Weiss, P. H., & Bode, S. (2019). Event-related potential correlates of decision confidence show specific associations with objective and subjective accuracy. *Association of Scientific Studies of Consciousness*, Amsterdam, The Netherlands **(Poster presentation)**

Overhoff, H., **Ko, Y. H.**, Fink, G. R., Stahl, J., Weiss, P. H., Bode, S., & Niessen, E. (2022). The relationship between response dynamics and the formation of confidence varies across the lifespan. *Association of Scientific Studies of Consciousness*, Amsterdam, The Netherlands **(Poster presentation)**

Table of Contents

List of Figures

Appendix C: Effects of relative and absolute evidence on ERP measures

Appendix D: Relationships between ERP measures and certainty of being correct/incorrect regardless of objective accuracy

Appendix E: Supplementary material of Chapter 2

List of Tables

Mixed-effects models results of ERP data analysis: Accuracy, full-range confidence, and subjective accuracy

Mixed-effects models results of ERP data analysis: Certainty of being correct and certainty of being incorrect

Appendix B: Relationships between additional ERP measures and confidence

Mixed-effects models results of ERP data analysis: Subjective accuracy

Mixed-effects models results of ERP data analysis: Certainty of being correct and certainty of being incorrect

Appendix C: Effects of relative and absolute evidence on ERP measures

Mixed-effects models results of ERP data analysis: Stimulus-locked CPP and response locked CPP

Mixed-effects models results of ERP data analysis: Pe and late Pe

Appendix D: Relationships between ERP measures and certainty of being correct/incorrect regardless of objective accuracy

Mixed-effects models results of ERP data analysis

Appendix E: Supplementary material of Chapter 2

Experiment 1

Experiment 2

Chapter 1. Introduction

1.1. Background

Decisions are ubiquitous in our lives. For both perceptual and value-based decisions, the decision process is based on evaluation of choice alternatives, for example, comparing their physical qualities or economic utilities (Smith & Krajbich, 2021). The decision outcomes, however, are oftentimes not finalized after the evaluation process, as they themselves are also subject to further evaluation regarding their validity. When functioning properly, this secondary evaluation allows us to correct our erroneous decisions and improve decision making in the future.

This idea could be illustrated by the following everyday example. Imagine that you are walking in a dark and remote area and facing two paths. You want to choose the one that is more brightly lit, as it is more likely to lead to safer place. Therefore, you try to evaluate and compare the brightness of the two paths. Once the evaluation process is complete, one path is judged to be brighter and correspondingly you walk along the chosen path. However, this decision is subject to further evaluation. For example, you might doubt the correctness of your decision after a couple more seconds. As a result, you might decide that the previous decision was incorrect, and walk back to make the decision again.

The ability to evaluate the accuracy of our own decisions and make corresponding adjustment is an important aspect of decision making, and it has been studied in different areas of research that were recently argued to be under the broad research topic of metacognition. The current research project aimed to investigate how one key aspect of metacognitive decision, decision confidence, could be related to other metacognitive judgments. To provide a background of the project, this introduction chapter will (a) review the literature on the key constructs related to metacognition in decision making (Sections 1.2 & 1.3), (b) review the literature on relevant theories on metacognitive decisions (Section 1.4), (c) review research methods relevant to the current project (Sections 1.5 & 1.6), and (d) provide an overview of the current project (Section 1.7).

1.2. What is metacognition?

The term metacognition was coined by Flavell (1979) to refer to “knowledge and cognition about cognitive phenomena” in his study of cognitive development. It is an umbrella term that refers to a range of cognitive processes (e.g., knowledge, experience, regulation) operating on different cognitive contents (e.g., perception, memory, learning; Flavell, 1979; Schwartz & Díaz, 2014; Metcalfe & Schwatz, 2016; Norman et al., 2019). For example, it encompasses one’s knowledge on how attention and memory generally work, how well one knows about one’s ability to complete a cognitive task, as well as the use of this knowledge or experience to devise task strategies and achieve desired task performance. This broad concept of metacognition has been studied in different areas of research including developmental, educational, social, personality, and cognitive psychology, as well as clinical domains including clinical psychology, neuropsychology, and psychiatry (Chapman et al., 2020; Flavell, 1979; Norman et al., 2019).

Functionally, metacognition serves to modulate cognitive processes such that they are more flexible and less reliant on the external environment (Fernandez-Duque et al., 2000). These modulations include two main aspects: monitoring (e.g., detecting errors, confirming decisions, and generating internal learning signal) and control (e.g., revising decisions, adjusting caution, guiding future decisions, time and resource allocation, information seeking; Desender et al., 2018; Fleming & Daw, 2017; Peters, 2022; Pouget et al., 2016; van den Berg, Anandalingam et al., 2016; Yeung &

Summerfield, 2012). In a broader sense, these basic functions also regulate cognitive performance through adaptation at higher levels, including flexible goal-setting (e.g., deciding to complete or abandon tasks), developing cognitive/metacognitive strategies (e.g., deciding on how to improve performance), learning (e.g., deciding on how well one have learned / will learn), and also improving performance in collective decision making (e.g., through communication of beliefs; Flavell, 1979; Metcalfe, 2009; Peters, 2022).

1.3. Metacognitive decision constructs

As metacognition has been investigated from different perspectives, various constructs and terminologies have been developed in the literature. As the scope of the current project concerns only metacognition in decision-making, this section first introduces what metacognitive decision encompasses, then focused on three areas that are particularly relevant, namely, (1) performance and error monitoring, (2) confidence judgment, and (3) changes of mind. For each of these areas, their definitions and characteristics are given, and common research methods and findings are summarized. Their relevant theoretical frameworks are discussed in Section 1.4.

1.3.1. Metacognitive judgments / decisions

Within the area of cognitive psychology, researchers focus mainly on how cognition is regulated and coordinated (Fernandez-Duque et al., 2000); more specifically, how meta-level cognitive processes monitor and control object-level cognitive process (Fernandez-Duque et al., 2000; Fleming & Dolan, 2012; Koriat, 2007; Nelson & Narens, 1990). In general, these processes were commonly studied by prompting metacognitive judgments / decisions, which are decisions that require evaluation of cognitive processes / cognitive task outcomes (Fleming & Dolan, 2012; Fernandez-Duque et al., 2000; Yeung & Summerfield, 2012). Conventionally, decisions related to object-level and meta-level processes are often referred to as primary and secondary decisions, type 1 and type 2 decisions, or first-order and second-order decisions (de Gardelle & Mamassian, 2015; Peters, 2022; Yeung & Summerfield, 2012).

Depending on cognitive domains, different types of metacognitive decisions are elicited. For example, the memory domain often requires participants to report judgment of learning (predicting how well one can recall an item) or feeling of knowing (reporting how well one thinks an item could be recalled later after just failing to recall it). In the decision-making domain, the types of metacognitive decisions range from perceptual to knowledge- and value-based decisions often requiring confidence judgment (reporting how well one thinks a decision just made was correct), post-decision wagering (gambling with a wager amount that corresponds to how well one thinks a decision was correct), and opt-out decision (deciding not to make a decision when one is unsure whether a decision will be correct; Fleming & Dolan, 2012; Gherman & Philiastides 2015; Rahnev et al., 2020; Sanders et al., 2016). Within the scope of the current thesis, three types of metacognitive judgments, namely error monitoring, confidence judgement, and changes of mind, are discussed below. In this thesis, the term metacognitive decision refers to as a set of decisions that include these three types.

1.3.2. Performance and error monitoring

As part of cognitive control, performance monitoring is a broad set of processes that evaluate motivationally salient events and determine adaptation (Ullsperger, 2017). Specifically, it tracks and signals a variety of events ranging from response conflict to violation of outcome prediction, and leads to a range of adaptations, including orienting, motor adjustment, and learning. As studies on performance monitoring focuses primarily on how erroneous decisions are detected (which involve discrepancy between expected and actual states) in tasks that induce response conflicts (e.g., go/no-go tasks, Stroop tasks, flanker tasks), performance monitoring is sometimes defined more narrowly as error monitoring (Steinhauser & Yeung, 2010; Ullsperger, 2017).

Given this definition, error monitoring can also be defined as a function of metacognition (Fernandez-Duque et al., 2000; Fleming & Dolan, 2012; Yeung & Summerfield, 2012). Indeed, error monitoring studies typically measure behavior that could require metacognitive ability (e.g., error correction responses, post-error slowing [i.e., slowdown response in trials following error commission]), or explicitly require metacognitive decisions from participants, e.g., error-signaling responses (i.e., indicating whether a decision error has occurred; Rabbitt, 1968). Traditionally, researchers often study error monitoring processes through variables including error detection and correction rates/RT, error awareness ratings, task performance that follow errors (e.g., post-error slowing, post-error accuracy [i.e., increased or decreased accuracy in trials following error commission]), as well as response force and electromyogram (EMG) measures (for measures of partial errors and force profiles during motor execution), and electroencephalography (EEG) measures (for identifying neural correlates of error monitoring; discussed in Section 1.5.2; Scheffer et al., 1996; Wessel, 2017; Yeung & Summerfield, 2012).

An example of error detection study is the early work by Rabbitt (1966), who employed an error detection task that required participants to respond to sequences of light signals as quickly as possible, and correct errors when they occurred. In this study, response time (RT) of error responses, error-correcting responses, and post-error slowing were measured. Essentially, it was found that erroneous decisions could be quickly corrected without external feedback, and these decisions were also followed by slower responses, suggesting the presence of error monitoring processes.

Building on these early studies focusing on error detection and correction, research has been extended to error awareness (Hester et al., 2005; Ullsperger et al., 2010; Wessel et al., 2011). So far, consistent findings have shown that participants are able to process and correct errors they made, with or without error awareness. Accordingly, separate neural correlates of automatic error processing and awareness were postulated (e.g., Yeung & Summerfield, 2012), and different theoretical accounts have been suggested (Fleming & Daw, 2017; Wessel, 2017; discussed in Section).

1.3.3. Confidence judgment

Decision confidence is defined as the subjective belief that a decision is correct or appropriate (Luttrell, 2013; Meyniel et al., 2015; Pouget et al., 2016; Yeung & Summerfield, 2012). The study of decision confidence dates back to early psychophysical experiments, for example, studies by Peirce and Jastrow (1885) and Henmon (1911). These experiments were the first to include a subjective measure of confidence ratings, such that participants could report the extent to which they believed their decision was correct, for example, after making comparative judgment based on two lines of different length. These early experiments investigated how confidence was related to task difficulty, decision accuracy, and RT. Specifically, as later confirmed with different task paradigms, stimuli, and response modalities, confidence is usually positively related to accuracy, negatively related to task difficulty, and negatively related to RT of both primary and secondary decisions (Rahnev et al., 2020; Sanders et al., 2016).

Particularly, consistent findings have been reported in previous studies using different cognitive tasks (e.g., perceptual decision, knowledge-based decision, value-based decision, reasoning task) with different stimuli (e.g., visual, auditory, tactile), different confidence ratings with different measurement scales (e.g., continuous, ordinal, binary ratings), and different response modalities (e.g., eye saccade, hand movement, wagering; Double & Birney, 2019; Fairhurst et al., 2018; Folke et al., 2016; Grimaldi et al, 2015; Kunimoto, 2001; Moreno-Bote, 2010; Persaud et al., 2007; Peters, 2022; Sanders et al., 2016; van den Berg, Anandalingam et al., 2016; Zylbergerg et al., 2016).

Taken together, these findings suggest that humans are capable of reporting confidence estimates that track objective accuracy (Sanders et al., 2016). Building on this relationship, some studies also focused on how well confidence corresponds to objective accuracy (metacognitive accuracy, e.g., how often participants assign high confidence to correct decisions and low confidence to incorrect decisions) as well as the tendency to rate high and low confidence (metacognitive bias, e.g., how often participants assign high confidence to decisions regardless of objective accuracy) and investigated conditions in which metacognitive performance changed (Boldt et al., 2017; Maniscalco & Lau, 2012; Maniscalco et al., 2021; Sander et al., 2016). For example, it has been reported that metacognitive accuracy is usually higher when primary task difficulty is low (because primary task performance influences secondary task performance), and when the primary decision is made with emphasis on speed, but secondary decision is made with sufficient time (Baranski & Petrusic, 1994; Boldt et al., 2017; Desender et al., 2020, 2021; Moran et al., 2015; Pleskac & Busemyer, 2010; Yu et al., 2015). These findings served as the basis for better understanding the mechanism underlying confidence judgment (Moran et al., 2015; discussed in Section 1.4.).

Lastly, it should be noted that a similar term *certainty* has been used in the literature with two different definitions. First, it has been used to refer to choice-independent sensory variability (Pouget et al., 2016). Second, it has been used as a directional expression of confidence (Baranski & Petrusic, 1994; Moreira et al., 2018). For example, a confidence rating indicating that a decision has no chance of being correct could be equally framed as indicating absolute certainty of being incorrect. Throughout this thesis (particularly in Chapter 3), certainty refers to the second definition.

1.3.4. Changes of mind

Compared with error monitoring and confidence judgment, change-of-mind is a relatively recent research topic. Change-of-mind decisions are defined as decisions that reverse the outcome of previous decisions, with and without explicit response execution or additional processing of external stimuli (Resulaj et al., 2009; Stone et al., 2022). Similar to confidence studies, previous studies have mainly investigated changes of mind in perceptual, memory, and value-based decision tasks (Stone et al., 2022). For example, perceptual decision-making tasks such as dot motion tasks and luminance discrimination tasks were commonly used (Fleming et al., 2018; Resulaj et al., 2009; Turner et al., 2021; van den Berg, Anandalingam et al., 2016). In terms of response modalities, in addition to the typical button press that allows a measure of binary changes of decision (Turner et al., 2021), studies have also utilized devices that allow more detailed examination of motor response, for example, manipulandum and touchscreen that track movement trajectory, which provide measures of movement speed and bends, and allow more fine-grained changes of mind to be expressed, or the process of the change of mind decision to be tracked while it unfolds (Burk et al., 2014; Dotan et al., 2018; van den Berg, Anandalingam et al., 2016).

For example, in an early study by Resulaj et al. (2009), participants completed a dot motion task where they were required to judge the moving direction of random dot stimuli with different motion strength. Using a manipulandum, they responded by moving a handle to the position that indicate the choice (left or right). Critically, the stimuli presentation was terminated once the hand movement initiated. However, it was shown that even in the absence of continued stimulus presentation, the trajectory of hand movement sometimes changed from one direction to another. This study suggests that even in the absence of stimuli, processing of stimuli continued after response initiation and could reverse the initial decision (i.e., the processing pipeline hypothesis).

Changes of mind usually occur most often at an intermediate level of difficulty (Resulaj et al., 2009). When task difficulty is low, changes of mind rarely occur. As task difficulty increases, changes of mind occur more frequently because there are more error trials, which are more subjected to changes of mind. With extremely high difficulty, further processing after the decisional process is also unlikely to reverse the initial decision (Stone et al., 2022). As accuracy and RT both covary with difficulty, changes of mind also appear to show a consistent relationship: They are more likely to occur when accuracy is intermediate, and when RT is slow (Resulaj et al., 2009). Changes of mind are also largely corrective, as they usually correct erroneous decisions rather than spoiling correct decisions (Resulaj et al., 2009; Stone et al., 2022). This corrective nature is more obvious for decisions originally made with low-quality evidence (Stone et al., 2022).

1.3.5. Summary

Although originated as different research areas, error monitoring, confidence judgment, and changes of mind are closely related. They are all decisions concerning the subjective accuracy of certain primary decisions: Confidence judgment involves judging the correctness of the decision (i.e., certainty of being correct), error detection involves judging how likely a decision is incorrect (i.e., certainty of being incorrect), and changes of mind are subsequent decisions based on a subjective sense of accuracy (Moreira et al., 2018; Stone et al., 2020; van den Berg, Anandalingam et al., 2016). Their close relationships are further supported by their similar neural correlates and proposed underlying mechanisms (Boldt & Yeung, 2015; Charles & Yeung, 2019; Fleming & Daw, 2017; Rausch et al., 2020; van den Berg, Anandalingam et al., 2016; Yeung & Summerfield, 2012, 2014). This naturally leads to some tentative proposals that these three types of metacognitive decisions could be explained by a common mechanism (Desender et al., 2021; Fleming & Daw, 2017; van den Berg, Anandalingam et al., 2016; discussed in 1.3.6).

However, these metacognitive decisions are different in several aspects. Methodologically, researchers investigate these decisions with different task types and parameters such as difficulty, response deadlines, and measurement scales (Yeung & Summerfield, 2012). Conceptually, confidence is more flexible with respect to time references (can be measured prospectively and retrospectively) and assumed to emerge as early as decision formation (Di Gregorio et al., 2020; Dotan et al., 2018; Gherman & Philiastides, 2015, 2018; Lee et al., 2022). In contrast, changes of mind and error awareness are assumed to occur and are measured after the primary decision and might require late-arriving evidence in addition to evidence that support the primary decision (Stone et al. 2022; van den Berg, Anandalingam et al., 2016). Additionally, empirical support for an association between error awareness and confidence is still equivocal with some studies suggesting that error awareness and confidence are partially dissociable (Fitzgerald et al., 2017). Therefore, whether these three types of metacognitive decisions could be considered as sufficiently similar to be explained by the same underlying mechanism remains unclear.

1.4. Theories and models of metacognitive decisions

Based on the experimental findings discussed above, researchers have attempted to characterize about the mechanisms underlying different metacognitive decisions. Some theories were further developed into computational models that account for different sets of empirical observations. This section reviews two dominant types of models that were commonly used to account for metacognitive decisions in two-choice tasks: Signal detection theory (SDT), and sequential sampling models (Rahnev, 2020, 2021). While the application of SDT has been limited to confidence judgment, sequential sampling models have been applied to all three types of metacognitive decisions.

1.4.1. Signal Detection Theory (SDT)

As an early successful model of decision making, SDT was one of the first models utilized to account for confidence judgment and it has served as the basis for more recent models (Green & Swets, 1966; Macmillan & Creelman, 2004). Proposed by Green and Swets (1966), SDT was first applied to behavioural tasks in which responses correspond to different stimulus types, and accurate performance is defined as the extent to which behaviour follows the correspondence (e.g., signal detection tasks; Macmillan & Creelman, 2004).

In an example of a signal detection task, SDT assumes a decision space where the perceived stimulus strength is determined by a decision variable, the value of which comes from either of two overlapping normal probability distributions representing the mean and variation and of stimulus strength in the target-present and target-absent conditions. As the decision maker does not know the source of the decision variable value, a further criterion (or *response bias*, *c*) is assumed to separate the distributions, such that values above the criterion lead to a target-present response, while values below lead to a target-absent response. Therefore, the criterion partitions the distributions into four possible outcomes: hit, miss, false alarm, and correct rejection. The proportions of these outcomes allow the estimation of an unbiased measure of *sensitivity* (*d'*), as the mean difference between the two distributions in terms of their common standard deviation (representing how well the decision maker discriminate the two conditions), and a measure of response bias as the location of the criterion (representing how biased the decision maker to give a target-present response). This theory also provides graphical tools such as the Receiver Operating Characteristics (ROC) curve and its derived measures (e.g., area under curve [AUROC] as a non-parametric measure of sensitivity).

In addition to its application to primary decision performance, SDT can also account for confidence by assuming multiple additional type 2 criteria dividing the decision space. While the implementations of the type 2 SDT vary with different assumptions, the central idea is that more extreme states of the decision variable (lower than the lowest criterion or higher than the highest criterion) result in higher confidence, while more intermediate states result in lower confidence (Clarke et al., 1959; Galvin et al., 2003; Macmillan & Creelman, 2004; Maniscalco & Lau, 2012, 2014; Maniscalco et al., 2016; Peters & Lau, 2015). This type 2 SDT approach has provided adequate accounts for confidence judgment since early work and more recent variants have been developed (e.g., with additional assumptions on decision space [Maniscalco et al., 2016], and noise structures [Shekhar & Rahnev, 2019]). Additionally, this approach also provides a foundation for measures of metacognitive performance in recent research (Pleskac & Busemeyer, 2010; Maniscalco & Lau, 2012; Maniscalco et al., 2022; discussed in Section 1.4.1.2).

However, this traditional SDT approach is limited in two ways. First, this approach connected the primary decision and confidence judgment, but its application to error monitoring and changes of mind is still limited (although there has been attempts to apply SDT for error detection tasks, the applications did not consider primary decision performance; Charles et al., 2013; Steinhauser & Yeung, 2010). The main reason is that SDT assumes that primary decision and confidence are based on the same source of evidence, which is inconsistent with major theories suggesting that error detection and changes of mind were driven by an evidence source different from that of primary decision (Yeung et al., 2004; Stone et al., 2022; Ullsperger et al., 2010). Even in SDT variants where confidence involves different sources of evidence than the primary decision (e.g., sensory signal contaminated by metacognitive noise), these models still do not allow cases where confidence contradicts the primary decision (Shekhar & Rahnev, 2019). Second, SDT does not consider the temporal dynamics within decisions, and thus not accounting for the RTs of primary decision and metacognitive decisions (Rahnev, 2021).

1.4.2. Sequential sampling models

One type of model that overcomes the limitations of SDT are sequential sampling models, which typically assume noisy accumulation of evidence over time, and a decision is made when the accumulated evidence crosses a decision criterion, or threshold (Ratcliff, 1978). These models have been successful in accounting for primary decisions, and it is possible to extend their application to secondary decisions with additional auxiliary assumptions on how confidence is related to the accumulation processes (Lee et al., 2022). This section discusses two major categories of sequential sampling models: single-stage models, which suggest confidence is based on evidence accumulated during the decision process, and dual-stage models, which suggest confidence also involves evidence accumulated during post-decisional processes, and could be applied to changes of mind and error detection (Moran et al., 2015; Resulaj et al., 2009; Ullsperger et al., 2010).

It should be noted that past studies have also categorized models by other features, for example, when confidence processing occurs (i.e., locus of confidence, which does not necessarily align with whether confidence is based on pre- or post-decisional evidence, e.g., a model could assume that confidence processing occurs after decision but based on evidence accumulated during decision formation, Baranski & Petrusic, 1998; Moran et al., 2015; Yeung & Summerfield, 2012), whether confidence judgment is computational or heuristic-based (i.e., based on evidence from the stimulus or external information such as RT; Moran et al., 2015; Shea et al., 2014), whether a single decision variable or two separate variables support decision and confidence (i.e., first-order vs. second-order models, without specifying the temporal aspect; Fleming & Daw, 2017).

1.4.2.1 Single-stage models

An early example of single-stage models is the accumulator model by Vickers and Packer (1982). Central to their model is the balance-of-evidence hypothesis, which states that confidence is determined by the difference in accumulated evidence between the winning and losing accumulators contributing to the primary decision. This hypothesis was successful in accounting for confidence in different tasks and influenced later models that involved different accumulation processes and auxiliary assumptions (Charles & Yeung, 2019; Fleming & Daw, 2017; Kiani et al., 2014; Rahnev, 2021; Ratcliff & Starns, 2009, 2013; Sander et al., 2016). For example, Kiani et al. (2014) proposed an evidence accumulation model in which the balance of evidence together with decision time (and thus RT) were sufficient to account for confidence, which additionally assumes that RT could inform confidence as high confidence is often coupled with shorter RT. Additionally, single-stage models can also involve different accumulation processes and computation of confidence, for example, leaky and competing accumulators, or accumulation of confidence rather than evidence (Lee et al., 2022; Maniscalco et al., 2021; Usher & McClelland, 2001). However, as most models of this type generally assume confidence is a readout of decision variable, most struggle to explain empirical findings that confidence and accuracy were sometimes dissociable (unless an assumption of biased evidence use is incorporated; Maniscalco et al., 2016; Zylberberg et al., 2012), or account for error awareness and changes of mind (Desenders et al., 2020, 2021; Fleming & Daw, 2017; Rabbitt, 1966; Resulaj et al., 2009; Steinhauser & Yeung, 2010; Stone et al., 2022).

1.4.2.2 Dual-stage models

More recent models therefore incorporated a post-decisional evidence accumulation process, reflected by the state of a metacognitive or confidence variable. A simple form of this kind of models assumes a continuation of the pre-decisional evidence accumulation (e.g., due to late-arriving evidence in the processing pipeline) and additional criteria that define the amount of evidence required to reach certain confidence levels (Moran et al., 2015; Pleskac & Busemeyer, 2010; Resulaj et al., 2009; van den Berg, Anandalingam et al., 2016; Yu et al., 2015). Because this kind of models does not have the limitation that the same decision variable determines choice metacognitive decision, it allows cases where the metacognitive decision contradicts the primary decision, thus providing an explanation for changes of mind in addition to confidence (Resulaj et al., 2009; van den Berg, Anandalingam et al., 2016).

Similarly, post-decisional evidence accumulation has also been proposed to be the mechanism underlying error awareness (Steinhauser & Yeung, 2010; Ullsperger et al., 2010). In such case, however, it is assumed that this process accumulates error-specific evidence from multiple sources (e.g., post-response conflict) instead of sensory evidence from stimulus (Ullsperger et al., 2010). Similar suggestions that other evidence sources could contribute to metacognitive judgment have also been made for confidence judgment (e.g., information from response execution [Fleming & Daw, 2017], memory [Yu et al., 2015]). More recently, it was further suggested that error-specific evidence from multiple sources could be accumulated to give rise to both confidence and error awareness in a unified model (Desender et al., 2021). With the assumption that different sources of evidence contribute to the post-decisional accumulation process, it has been proposed that the evidence accumulation processes of the pre-decisional stage and post-decisional stage differ in terms of accumulation rates and reference frames: The pre-decisional stage accumulates sensory evidence with reference to the stimulus-response mapping, while the post-decisional stage accumulates error evidence with reference to decision accuracy (Desender et al., 2021).

1.4.3. Summary

This section reviewed the major theories on metacognitive decisions and the convergence from three areas of metacognitive decisions that a two-stage, evidence accumulation model could be applied in their respective areas. This forms the basis that different metacognitive decisions could be explained by a common mechanism, and the construct of confidence is common in these models. (Desender et al., 2021; van den Berg, Anandalingam et al., 2016). However, while such proposal is possible, it should be considered with caution as these metacognitive decisions could be qualitatively different at least in the following two aspects.

First, while sequential sampling models commonly assume that metacognitive decisions are largely determined by accumulated evidence, it is unclear what type of evidence is relevant, and whether different metacognitive decisions involve the same sources of evidence. Past studies have shown that metacognitive decisions are likely informed by a wide range of information sources (i.e., a multi-cue model; Bolt et al., 2017; Stone et al, 2022; Ullsperger et al., 2010), such as information from different stimulus properties (Bolt et al., 2017; Rausch et al., 2018; Zylbergerg et al. 2012), information related to motor execution (Fleming et al., 2015; Resulaj et al., 2009; Pereira et al., 2020; Turner et al., 2021), prior belief (Fleming & Daw, 2017), individual differences in self-confidence (Double & Birney, 2019), or even information confidence from irrelevant tasks (e.g., changes in motion range in a dot motion stimulus affects brightness judgment based on the same stimulus, i.e., uncertainty transfer effect; Spence et al., 2018). However, these sources of evidence might not contribute similarly to different types of metacognitive decisions. For example, some studies have shown that confidence could be explained solely by stimulus-based evidence, while changes of mind and error awareness were often explained by post-decisional evidence such as response conflict (Lee et al., 2022; Shekhar & Rahnev, 2022; Steinhauser et al., 2008; Steinhauser & Yeung, 2010; Stone et al., 2022; Ullsperger et al., 2010).

Second, the temporal aspect of evidence accumulation could differ across different types of metacognitive decisions (discussed in Section 1.3.5). On the one hand, confidence appears to be more closely related to pre-decisional process (Lee et al., 2022; Shekhar & Rahnev, 2022). On the other hand, changes of mind and error awareness might rely more on post-decisional processes (Steinhauser et al., 2008; Steinhauser & Yeung, 2010; Stone et al., 2022; Ullsperger et al., 2010). This potential difference in processing stages also reinforces the idea that different metacognitive decisions involve different sources of evidence, as previous studies have shown that contributions from different evidence source could change across processing stages (Stone et al, 2022).

Therefore, metacognitive decisions appear to be more complex processes and the proposals that they shared a common mechanism remains to be explored. Specifically, in the current project focusing on confidence judgment, sources of evidence and the temporal aspect of accumulation process were respectively investigated with behavioral experiments that manipulated stimulus properties and ERP measures. To provide a background of these two methodological approaches, the next section reviews common measures that are relevant to the current project.

1.5. Measures in metacognitive decision research

To understand metacognitive processes, studies most often required metacognitive decisions with different types of behavioral measures and obtain neural measures that correlate with these metacognitive decisions under different experimental conditions. This section first focuses on the most common behavioral measure of self-report, subjective ratings of accuracy and its derived measures of metacognitive performance. Then it moves on to the often used event-related potentials (ERP) measures based on EEG signals, which were found to be consistently related to metacognitive decisions.

1.5.1. Behavioral measures

1.5.1.1 Accuracy ratings

Although different methods have been used to estimate metacognitive states (e.g., pupil dilation [Urai et al., 2017] for confidence, movement tracking [Resulaj et al., 2009] for changes of mind), the majority of studies on metacognitive decisions employed self-report accuracy ratings (including confidence ratings, error awareness ratings, and change-of-mind responses). This section reviews commonly used measurements in the literature and provides a background for the ratings scale used in the current project.

In error monitoring studies, error awareness is most often measured after decisions with forced-choice ratings, which require a response to indicate whether or not (or the extent to which) the participant is aware of an error, or error-signaling responses, which requires pressing a button only when the participant is aware of an error (Nieuwenhuis et al., 2001; Wessel, 2012). Error-signaling responses are considered suboptimal as it could induce a bias not to report error awareness and could wrongly categorize trials with residual error awareness as unaware errors (Wessel, 2012). Within studies employing forced choice ratings, most studies used binary scales and only few used ordinal scales that measure the degree of error awareness (Wessel, 2012; Scheffers & Coles, 2000; Hewig et al., 2011).

On the other hand, confidence has been measured less consistent ways. Researchers have employed different variants of the confidence rating scale (Grimaldi et al., 2015; Rahnev et al., 2020). In terms of the timing when the ratings are given, studies have used retrospective ratings (confidence reported after choice), prospective confidence ratings (confidence prediction before choice), or simultaneous report of choice and confidence (Fleming & Dolan, 2012; Fleming et al., 2012; Peters, 2022; Siedlecka et al., 2016; Zylberberg et al., 2014). These options were motivated by different assumptions and could exert different effects. For example, retrospective ratings usually track accuracy better than prospective ratings (Fleming & Daw, 2017; Siedlecka et al., 2016). Simultaneous report was assumed to limit post-decisional processing underlying confidence judgment, but some argued that it could also lead to different processing and interpretation than that elicited by sequential report (as an experience judgment instead of a performance judgment; Desender et al., 2020; Fleming

& Lau, 2014; Galvin, 2003; Petrusic & Baranski, 2003; Samaha & Denison, 2022). In terms of the measurement scales, studies differ by the type of scale (continuous, ordinal with different number of points, and binary ratings scales) as well as scale labels (verbal labels [“surely incorrect” vs. “surely correct”] vs. percentage labels [0% to 100% correct]; Cheesman & Merikle, 1986; Dienes & Perner, 1999; Grimaldi et al., 2015; Baranski & Petrusic, 1994, 1998). Studies also assess confidence with regard to performance of different scales. While traditionally item-wise confidence judgment is most commonly used, confidence judgement can also be made with regard to performance within task intervals or overall performance over the whole task (Norman & Price, 2015; McWilliam et al., 2022).

More importantly, confidence rating scales also differ in terms of scale range. While the full-range scale ranges a from “surely incorrect (0% chance of being correct)” to “surely correct” (100% chance of being correct), others have used a half-range scale which ranges from “not confident” (50% chance of being correct) to “confident” (100% chance of being correct), based on the assumption that in two-choice tasks the objectively minimal accuracy should be 50% (Lichtenstein et al., 1982). Therefore, with a half-range scale, participants might report low confidence when they know their decision was wrong, but the full-range scale allows indicating that one has detected an error (Baranski & Petrusic, 1994). For this reason, more recent studies are now shifting to the use the full-range scale instead of the half-range scale as a hybrid measure of both confidence and error awareness, thus avoiding excluding error detection trials or simply recording them as low confidence trials as in previous studies (Bolt & Yeung, 2015; Fleming & Daw, 2017).

Such a measure could also be converted into a measure of changes of mind and thus serves as a hybrid scale that measures these three types of metacognitive decisions.

Particularly, assuming that these metacognitive decisions can be defined on a continuum of subjective probability of being correct, any confidence ratings indicating accuracy below chance could be considered as a potential change-of-mind trial (Charles et al., 2019; Fleming et al., 2018). By the same token, the graded confidence ratings could also be reverse coded as a measure of error awareness (e.g., confidence ratings indicating "surely incorrect" could be considered as indicating a strong sense of error awareness). While such conceptualization could be inconsistent with the theoretical differences between these types of decisions and could affect the interpretation of findings (Charles & Yeung, 2019; Maniscalco & Lau, 2014), this hybrid scale provides a flexible, simultaneous measure of the three types of metacognitive decisions.

1.5.1.2 Metacognitive accuracy/sensitivity

Measures of subjective accuracy allow researchers to further investigate how well metacognitive decisions match primary task performance, termed metacognitive accuracy/sensitivity (Fleming & Dolan, 2012; Fleming & Lau, 2014). Although humans are capable of making metacognitive decisions that reasonably correspond to objective accuracy, this correspondence appears to vary across experimental conditions and individuals (Rahnev et al., 2020; Rouault et al., 2018). To better understand this relationship, a number of measures of metacognitive accuracy have been developed. In confidence studies, these include calibration measures, correlation measures such phi and gamma (correlation between accuracy and binary/ordinal confidence ratings), and early SDT-derived measures (which measures how well high and low confidence ratings discriminate correct and erroneous decisions; Baranski & Petrusic, 1994; Clarke et al., 1959; Fleming & Lau, 2014; Galvin, 2003; Lichtenstein & Fischhoff,1977; Maniscalco & Lau, 2012). Similarly, error monitoring studies also examine metacognitive performance with error detection rates and SDT-derived measures

(Steinhauser & Yeung, 2010), and changes of mind studies measure the proportions of corrective changes and spoilt responses (Stone et al., 2022). However, a major limitation is that these measures are not independent of primary task performance (which could be confounding as primary and secondary decision performance are often positively correlated; Baranski & Petrusic, 1994; Evans & Azzopardi, 2007; Fleming & Lau, 2014).

Therefore, less biased SDT-derived measures have been developed, and one commonly used measure is meta-d’ (Mansicalco & Lau, 2012; Fleming, 2017). The basis of meta-d’ is the d’ measure in SDT. As discussed above in Section 1.4.1.1., d’ is a sensitivity measure that represents how well the representation of two stimuli are discriminable to an observer in the presence of noise, and this is usually estimated with primary task performance in the type 1 SDT model (Macmillan & Creelman, 2004). The measure of meta-d’ was developed based on the fact that the type 1 SDT model can also be determined accuracy ratings. Therefore, in reverse, empirical accuracy ratings could be expressed in terms of type 1 model parameters. In practice, meta-d’ was estimated by fitting the SDT model to accuracy ratings data. This estimated meta-d’ is therefore on the same scale of d’ and thus can be used to calculate a metacognitive measure relative to primary task performance (by calculating the ratio between the two [*meta-d’*/*d’*] or difference between the two [*meta-d’* – *d’*]), termed metacognitive efficiency (Fleming, 2017; Mansicalco & Lau, 2012; but recent studies have suggested that this measure is not completely independent of primary task performance [Guggenmos, 2021], speed/accuracy trade-off [Desender et al., 2022], and metacognitive bias [Shekhar & Rahnev, 2021; Xue et al., 2021]).

These less biased measures have been commonly used in studies on confidence judgment and error monitoring (Charles et al., 2013; Guggenmos, 2021). These studies showed that human participants usually show metacognitive inefficiency, meaning metacognitive performance worse than primary task performance, but in some cases metacognitive hyper-efficiency was observed (i.e., metacognitive performance being better than primary task performance, e.g., Charles et al., 2013). These discrepant findings may result from the use of other sources of evidence, error detection, and different criteria improved metacognitive performance (Fleming & Daw, 2017; Maniscalco & Lau, 2012).

1.5.1.3 Metacognitive bias

Additionally, when metacognitive accuracy/sensitivity is not perfect, a metacognitive bias can be measured (i.e., the tendency to over- or underestimate confidence; Baranski & Petrusic, 1994; Fleming & Lau, 2014; Lichtenstein & Fischhoff, 1977). Traditionally, in confidence studies this bias was termed over-/under-confidence bias (Lichtenstein & Fischhoff,1977; Baranski & Petrusic, 1994). A typical finding is that metacognitive bias depends on the task type and difficulty. Past studies showed a hard-easy effect, which describes the common phenomenon that individuals are usually overconfident in difficult, knowledge-based tasks, while underconfident or unbiased in easier, perceptual tasks (Gigerenzer et al., 1991).

While past studies used different measures of metacognitive bias (e.g., calibration measures), a metacognitive bias derived from SDT, meta-c, was more commonly utilized in recent studies (Baranski & Petrusic, 1994; Bolt & Yeung, 2017; Maniscalco & Lau, 2012, 2014). Based on the same model that generates meta-d', meta-c can be estimated similarly. This measure of metacognitive bias was found to be relevant particularly when metacognitive decisions were biased, e.g., when confidence appeared to be based on sources of evidence that were not informative about decision accuracy (Boldt et al., 2017; Samaha & Denison, 2022; Winter & Peters, 2022).

1.5.2. ERPs correlates of metacognitive decisions

Besides behavioral measures, previous studies have also utilized EEG recordings and identified ERP correlates of metacognitive decisions. Typically, EEG are recorded by electrodes on the scalp that record electrical activity of the human brain, which were then amplified and filtered; Luck, 2014). As the recordings are noisy in nature due to due to different sources of activity, recordings in response to a common event (e.g., stimulus onset and response) are averaged together to increase the signal-to-noise ratio. The amplitudes within a specific time window and electrode location is then extracted as a measure of ERP. While ERPs correlates of metacognitive decisions cannot serve as objective measures of metacognitive states (e.g., how confident one is), they are informative about the timing and stage of processing involved in such metacognitive decisions.

This section reviews three ERP components that are investigated in the current project: error-related negativity (ERN/Ne), CPP/P3 (Centro-parietal positivity), and error positivity (Pe). These components were candidates of metacognitive decision correlates because the ERN/Ne and the Pe have been the major ERP components studied in the error monitoring literature. Moreover, the Pe has been shown to be similar to the CPP/P3 in that they were both recently theorized to reflect decision variables, which are closely related to confidence (Gehring et al., 2012; Rausch et al., 2020).

1.5.2.1. Centro-parietal potential (CPP/P3)

The centro-parietal potential (CPP/P3; Sutton et al., 1965; also sometimes labelled as late positive potential when measured as a stimulus-locked component; Sun et al., 2017) is a positive component measured at midline, centro-parietal sites. Its typical measurement window varies from 300 to 800 ms relative to stimulus onset, or -300 to -100 ms relative to the time point a response is given (depending on task paradigms). and it is believed to originate from temporal and parietal regions (Polich, 2012). The CPP/P3 is considered to be indicative of different cognitive processes related to stimulus processing (Polich, 2012). For example, early studies found it to be larger in response to targets compared to non-targets in oddball tasks, and thus was assumed to reflect updating of stimulus representation in working memory (Donchin et al., 1978, 1981), particularly for task-relevant properties as its amplitudes were larger when attention resources are not reduced by another task (Kramer et al., 1985). Also, it was suggested to be related to memory encoding and its amplitudes are related to memory strength (Crites et al., 1998; Patterson et al., 1991). Its latency was found to be related to stimulus processing speed (Kutas et al., 1977; Magliero et al., 1984).

More recently, this component was considered a supramodal signal that reflects task-related evidence accumulation, as its amplitudes in a range of decision tasks show a characteristic build-up pattern up to the response, which is considered as accumulation-to-bound dynamics, and this was observed even when no motor response was required (Kelly & O'Connell, 2013; O’Connell et al., 2012; O’Connell & Kelly, 2021; Twomey et al., 2015). It also shows other characteristics that are related to evidence accumulation: Its build-up rates covary with RT, decision accuracy, as well as evidence strength (O’Connell & Kelly, 2021). It was therefore suggested that it closely corresponds to the decision variable in some computational models (Desenders et al., 2021; Gold & Shadlen, 2007; O’Connell & Kelly, 2021).

Some studies further linked this component to subjective experience including subjective visibility (Del Cul et al., 2007; Lamy et al., 2009; Sergent et al., 2005; Tagliabue et al., 2016, 2019) and decision confidence (Eimer & Mazza, 2005; Hillyard et al., 1971; Herding et al., 2019; Rausch et al., 2020; Sutton et al., 1982; Squires et al., 1975). For example, its amplitudes were positively related to visual awareness reported on the perceptual awareness scale in visual discrimination tasks (Tagliabue et al., 2016, 2019). Its amplitudes also positively correlated with confidence and are affected by task difficulty in a way similar to confidence changes (Eimer & Mazza, 2005; Herding et al., 2019; Rausch et al., 2020; Scheffer et al., 2000; Sun et al., 2017). However, it should be noted that this relationship between CPP/P3 amplitudes and confidence was mostly observed for correct trials, as some studies did not analysis error trials (e.g., Rausch et al., 2020), and the relationship was smaller or absent when error trials were analyzed (e.g., Eimer & Mazza, 2005; Herding et al., 2019; Hillyard et al., 1971). Additionally, this relationship was often based on stimulus-locked CPP amplitudes, which could be artefactual due to the negative relationship between RT and confidence (Feuerriegel et al., 2022). This is because low confidence trials are often coupled with slower responses, and the stimulus-locked CPP would thus be more temporally variable and smaller in amplitudes after averaging.

1.5.2.2. Error negativity (ERN/Ne)

The ERN/Ne is a negative deflection typically observed within 100 ms after an erroneous response in a wide range of speeded choice reaction time tasks with different stimulus modalities (Falkenstein et al., 1989, 1991; Gehring et al., 1990, 1993, 2012). It is typically measured at frontal sites and assumed to be generated at the anterior cingulate cortex (ACC; Nieuwenhuis et al. 2004; Hester et al., 2005; Scheffers et al., 1996; Gehring et al., 2012). It has been theorized that it reflects processes involving cognitive control (Gehring et al., 2012), including mismatch between the representations of correct response and executed response (Coles et al. 2001; Di Gregorio et al., 2018; Falkenstein et al., 1991; Gehring et al., 1993; Scheffers & Coles, 2000), mismatch between expected and actual stimuli (Bernstein et al., 1995; Schmidt & Gordon, 1977), the process of monitoring response conflict between multiple activated responses (Carter et al., 1998; van Veen & Carter, 2002; Yeung et al., 2004), reinforcement learning and violation of expectation (Holroyd et al., 2004; Holroyd & Coles, 2002; Alexander & Brown, 2010), and negative affective response to errors (Luu et al., 2003, 2004; Luu & Pederson, 2004; Vidal et al., 2000). Additionally, although the ERN/Ne was proposed to be error-specific, a component following correct response with similar scalp distribution and latency, named the correct-response negativity (CRN), has also been consistently reported (Ford, 1999; Gehring & Knight, 2000; Luu et al., 2000; Scheffers & Coles, 2000; Vidal et al., 2000).

Previous studies have reported several typical findings regarding the ERN/Ne. In terms of experimental conditions, ERN/Ne amplitudes were larger when task instructions emphasized accuracy over speed, and when errors were less likely (Arbel & Donchin, 2009; Falkenstein et al., 2000; Gehring et al., 1993; potentially due to increased significance of errors; Steinhauser & Yeung, 2012). In terms of behavioral correlates, its amplitudes were positively related to error correction magnitude, error correction likelihood and speed, as well as less forceful responses, and more post-error slowing (Falkenstein et al., 1995; Gehring et al., 1993; Yeung & Summerfield, 2012).

In relation to metacognitive decision, it has been asked for a long time whether its amplitudes are modulated by error awareness or confidence. However, the answer to this question is still unclear as mixed findings have been reported (Dehaene et al., 1994; Endrass et al., 2007; Gehring et al., 1993; Scheffers & Coles 2000; Nieuwenhuis et al., 2001; Steinhauser & Yeung, 2010; Wessel, 2012; Wessel et al., 2011). For example, Hewig et al. (2011) showed that the ERN/Ne was the larger for detected errors than undetected errors and correct trials, and Scheffers and Coles (2000) found that ERN/Ne amplitudes were modulated by confidence and error awareness in both correct and error trials, suggesting its relationship with error awareness or even confidence. However, Endrass et al. (2007) and Neuwenhuis et al., (2001) reported that ERN/Ne amplitudes were the same for detected and undetected errors in their tasks. Additionally, these reported associations with error awareness and confidence could result from aggregating different proportions of error detection trials within confidence / error awareness categories, as some studies pooled together correct and error trials in their analyses (e.g., Boldt & Yeung, 2015). It was suggested that this component does not directly index error awareness or confidence (but might reflect one source of evidence that contributes to them; Boldt & Yeung, 2015; Charles et al., 2013; Hewig et al., 2011; Rausch et al., 2020; Steinhauser & Yeung, 2010).

1.5.1.4 Error positivity (Pe)

The Pe is a positive slow deflection 200-400 ms after erroneous response in decision tasks (e.g., perceptual decision tasks, Flanker tasks, etc), measured at midline parieto-central sites (Gehring et al., 2012; Falkenstein et al., 1991; Hester, 2005; Yeung & Summerfield, 2012). Although its source is less well-characterized, it showed properties similar to the CPP/P3, including morphology and neural generators (Niewenhuis et al., 2001; Overbeek et al., 2005; Ridderinkhof et al., 2009; Murphy et al., 2015; Yeung & Summerfield, 2012). It has been theorized that this component reflects affective response to error, behavioural adaptation after error, and error awareness (Overbeek et al., 2005). It has also been noted that the Pe is composed of two sub-components: an early, fronto-central Pe, and a late, posterior Pe (Arbel & Donchin, 2009; Ruchsow et al., 2005; van Veen & Carter, 2002). While the early Pe could be functionally similar to the ERN/Ne (that they are both related to error detection but not necessarily error awareness), the late Pe might be more related to error awareness or affective response (Endrass, 2007; van Veen & Carter, 2002).

Studies that have contrasted the Pe and ERN/Ne have shown that the Pe typically reflects error awareness as it shows stronger amplitudes for aware errors compared to unaware errors (Di Gregorio et al., 2018; Endrass et al., 2007; Murphy et al., 2015; Overbeek et al., 2005; Ridderinkhof et al., 2009). This relationship was also extended to confidence judgment by the finding that Pe amplitudes reflected both error awareness and confidence: Its amplitudes were negatively and linearly related to confidence ratings ranging from "surely incorrect" to "surely correct" (Boldt & Yeung, 2015; Desender et al., 2019). Based on this relationship and the link between the Pe and CPP/P3, some theorized that the Pe reflects post-decisional accumulation of evidence analogous to the CPP/P3, which reflects evidence accumulation during decision formation. Note that these theories differ in terms of what evidence is accumulated (Desenders et al., 2021). Specifically, while some proposed that the Pe represents a metacognitive decision variable that accumulates "error evidence" that could be based on difference sources (Desenders et al., 2021; Murphy et al. 2015), others suggested that it represents evidence accumulation continued form the decision process (Rausch et al. 2019).

However, the relationship between the Pe and full-range confidence was only established preliminarily and has not been replicated in many studies (Boldt & Yeung, 2015; Desender et al., 2019). Particularly, the linear relationship between confidence and Pe amplitudes reported by Boldt and Yeung (2015) could potentially be confounded by their analysis that pooled correct and error trials, as different levels of confidence are associated with different proportions of correct and error responses. That is, even if Pe amplitudes only differed between correct and erroneous decisions, a linear relationship between Pe amplitudes and confidence could still be observed due to this pooling procedure. Further, a recent study has suggested that the typical measures of Pe in these studies could be biased by the practice of pre-response baseline correction (Feuerriegel et al., 2021; as mentioned above). The reason is that such a baseline often overlaps with the CPP/P3, which could also vary with error awareness and confidence, and thus contaminate the measure of Pe amplitudes.

Other suggested roles of the Pe can also be found in the literature. For example, an alternative hypothesis is that stronger Pe amplitudes represent nonspecific ambiguity/uncertainty, not necessarily error awareness (Hewig et al., 2011), or the Pe is an expression of error awareness rather than representing the process that leads to error awareness (Ridderinkhof et al., 2009).

1.5.1.5 Summary

The above section reviewed the three ERP components that could be related to metacognitive decisions. While the ERN/Ne appeared to be only modulated by detection of errors, both CPP/P3 and Pe were suggested to be related to error awareness and/or confidence. These two latter components proposed to reflect evidence accumulation in different ways: While the CPP/P3 reflects accumulation of sensory evidence, the Pe could reflect error evidence accumulation, mapping respectively onto the decision and metacognitive variables in a unified model of metacognitive decisions (Desenders et al., 2021). While this notion receives partial support from studies the relationships of CPP/P3 and Pe with confidence (e.g., Boldt et al., 2015; Herding et al., 2019), some methodological concerns have been raised. Furthermore, whether these relationships are similar in correct is yet to be clarified (Feuerriegel et al., 2021). The use of ERP measures in the current project allows investigating the temporal aspect of evidence accumulation. Particularly, whether different types of metacognitive decisions are similarly related to sensory evidence and error evidence accumulation.

1.5.3. Experimental manipulations of sensory evidence

Behavioral experiments probing the mechanism underlying metacognitive decision often involve the manipulation of stimulus properties (Shekhar & Rahnev, 2019). Particularly, qualitative patterns between these manipulations and the dependent variables of accuracy, confidence, RT of primary and secondary decisions could be diagnostic of different models (Moran et al., 2015). For example, when comparing different computational models of confidence judgment, Moran et al. (2015) established a set of empirical hurdle requirements that a plausible model should meet. This section reviews two important dimensions of stimulus properties: evidence strength and evidence variability.

1.5.3.1. Evidence strength

Studies on decision making often investigate the effects of evidence strength on decision performance as the observed patterns guide the development of theories and computational models for both primary and metacognitive decisions (Moran et al., 2015). In most studies, the strength of task-relevant evidence was manipulated. However, recent studies have shown that task-irrelevant evidence also showed systematic effects on both primary and metacognitive decisions, and models were assessed on how well they could account for these patterns (Ratcliff et al., 2018; Teodorescu et al., 2016; Turner et al., 2021).

Specifically, these studies focused on two-choice comparative judgment tasks, where evidence strength can be defined in two ways: Relative evidence refers to the magnitude difference between the two stimuli (which is task-relevant), and absolute evidence refers to the overall magnitude of the two stimuli (which is task-irrelevant).

Empirically, it has been shown that while increased relative evidence increases accuracy and shortens RT (Ratcliff et al., 2018; Teodorescu et al., 2016). In terms of secondary decision performance, increased relative evidence increases confidence in correct trials, but reduce confidence in error trials (i.e., more certain that errors occurred; Sander et al., 2016). This pattern is consistent across tasks and stimulus modalities (length discrimination task [Johnson, 1939; Festinger, 1943; Yu et al., 2015]; weight-lifting tasks [Pierrel & Murray, 1963]; extended decision tasks with visual length stimuli [Vickers & Packer, 1982; Vickers et al., 1985]; auditory tasks [Sander et al., 2016] and knowledge-based decision tasks [e.g., which of the two countries given has a larger population; Sander et al., 2016; Yu et al., 2015]; dot sampling task [Charles & Yeung, 2019]). These patterns have been well-captured by model variants that assume the balance-of-evidence hypothesis, that indeed confidence was largely determined by relative evidence (Moran et al., 2015).

On the other hand, increased absolute evidence impairs decision accuracy and shortens RT, possibly through increased internal noise or variability of stimulus representation (Ratcliff et al., 2018; Teodorescu et al., 2016). Importantly, increased absolute evidence also appear to affect metacognitive decision at least in terms of changes of mind, as Turner et al. (2021) showed that it led to slower change-of-mind RT and lower change-of-mind rates. Unlike the effects of relative evidence, these effects of absolute evidence could not be adequately explained by existing models. While these findings are currently limited to changes of mind, exploring the effects of absolute evidence might inform the mechanism underlying metacognitive decisions in general.

1.5.3.2. Evidence variability

Although less studied than evidence strength, evidence variability has also been shown to affect confidence as it represents how reliable the evidence is (Boldt et al., 2017; de Gardelle & Mamassian, 2015; Yeung & Summerfield, 2014). For example, in a color-averaging task, it has been shown that increased variability in stimulus presentation reduced confidence after primary decision performance (accuracy and RT) and relative evidence were taken into account (Boldt et al., 2017). Notably, while the effect of relative evidence on confidence was explained by decision performance, variability predicted confidence over and above primary decision performance, suggesting that variability itself could serve as a cue for confidence judgment.

However, such an effect does not appear to be universal and might be task-dependent. On one hand, consistent with the findings by Boldt et al. (2017), some found that confidence decreases with variability. For example, Spence et al. (2015, 2018) reported that wider motion range in a dot motion task did not influence accuracy, but reduced confidence. On the other hand, some studies found the opposite pattern that confidence increased with higher variability. For example, Zylbergerg et al. (2016) showed that increased variability of motion coherence led to a small decrease in accuracy but increased confidence. In addition, the effect of variability could interact with task difficulty. For example, Zylberberg et al. (2014) found that in a bar orientation detection task, increased variability of bar orientation reduced accuracy, but its effect on confidence depended on evidence strength: Higher variability increased confidence when evidence strength was low but decreased confidence when signal strength was high. It was suggested that in conditions where evidence strength was limited, variability could be mistaken as a cue to confidence (Boldt et al., 2017; Sander et al., 2016).

1.5.3.3. Summary

While relative evidence has long been manipulated to probe the mechanism underlying metacognitive decisions, some recent studies shifted the focus to other evidence properties such as absolute evidence strength and evidence variability.

Although only a small number of studies have investigated their effects and they were often limited to a single type of metacognitive decision, their effects appear to be as important as that of relative evidence, particularly in the understanding of how metacognitive decisions could be affected by internal noise and cues other than directly task-relevant signals.

1.6. The current project

This introduction chapter has provided a review on three types of metacognitive decisions, namely confidence judgment, error awareness, and changes of mind. Given their similarities, it has been proposed that they could be closely related, and evidence accumulation models have been proposed to link confidence with changes of mind and error awareness (Desender et al., 2021; van den Berg, Anandalingam et al., 2016). The connection between confidence and evidence accumulation was also supported by ERP studies that showed confidence was related to indexes of pre- and post-decisional evidence accumulation.

While these models have received preliminary support from behavioral and ERP studies, the extent to which confidence and other metacognitive decisions are based on the same sources of evidence or processes remains to be explored, given their differences on methodological, conceptual, and empirical grounds, as well as potential methodological confounds in previous ERP studies. Therefore, the current project focused on the role of confidence in metacognitive decisions and investigated: (a) how confidence is related to changes of mind, and (b) how confidence is related to pre- and post-decisional processes.

The first part of this project includes two behavioral studies aimed to answer the question: Do confidence and changes of mind change consistently in response to changes in stimulus properties? (Chapter 2, Studies 1 and 2). On a behavioral level, as confidence and changes of mind are assumed to share the same source of sensory evidence, they should change consistently with the manipulations of sensory evidence, including relative evidence, absolute evidence, and evidence variability, which were shown to affect both confidence and changes of mind in past studies.

The second part of this project includes an EEG study aimed to answer the question: How is confidence related to pre- and post-decisional processes? Previous studies have provided evidence that confidence is related to ERP indexes of pre- and post-decisional evidence accumulation, namely the CPP/P3 and Pe (Desender et al., 2021). However, such findings are mixed and potentially methodologically confounded. Moreover, the relationships between confidence and these ERP components could be in fact dependent on accuracy. Therefore, in an EEG study, these relationships were examined with the methodological confounds being controlled (Chapter 3, Study 3).

Throughout these three studies, a two-choice, comparative brightness judgment task was used as the main task paradigm, as this perceptual decision task has been used in past studies and it allows the manipulation of different stimulus properties that represent relative evidence, absolute evidence, and evidence variability (Ratcliff et al., 2018; Teodorescu et al., 2015; Turner et al., 2022). In each trial, participants were required to make a brightness judgment based on two flickering, grayscale square stimuli, and then report confidence using a full-range accuracy rating scale.

Lastly, the findings of these investigations will be summarized in Chapter 4, which discusses the role of confidence in metacognitive decision processes, as well as the limitations of the current project.

Chapter 2. Divergent effects of absolute evidence magnitude on decision accuracy and confidence in perceptual judgements

The first part of the current project investigated how different types of metacognitive judgment could be affected by changes in different sensory evidence properties, namely relative evidence strength, absolute evidence strength, and evidence variability in a dynamic luminance judgment task. The current chapter focused specifically on confidence judgment and changes of mind. Research findings are reported in the published article below.

2.1. Manuscript

Cognition 225 (2022) 105125

Contents lists available at ScienceDirect

Cognition

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

Divergent effects of absolute evidence magnitude on decision accuracy and confidence in perceptual judgements

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

[a] Melbourne School of Psychological Sciences, The University of Melbourne, Australia
[b] Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre Jülich, Germany
[c] Department of Psychology, Faculty of Human Sciences, University of Cologne, Germany
[d] Department of Neurology, University Hospital Cologne and Faculty of Medicine, University of Cologne, Germany

ARTICLE INFO

Keywords:
Decision-making
Confidence
Metacognition
Change-of-mind
Perception
Absolute evidence

ABSTRACT

Whether people change their mind after making a perceptual judgement may depend on how confident they are in their decision. Recently, it was shown that, when making perceptual judgements about stimuli containing high levels of 'absolute evidence' (i.e., the overall magnitude of sensory evidence across choice options), people make less accurate decisions and are also slower to change their mind and correct their mistakes. Here we report two studies that investigated whether high levels of absolute evidence also lead to increased decision confidence. We used a luminance judgment task in which participants decided which of two dynamic, flickering stimuli was brighter. After making a decision, participants rated their confidence. We manipulated relative evidence (i.e., the mean luminance difference between the two stimuli) and absolute evidence (i.e., the summed luminance of the two stimuli). In the first experiment, we found that higher absolute evidence was associated with decreased decision accuracy but increased decision confidence. In the second experiment, we additionally manipulated the degree of luminance variability to assess whether the observed effects were due to differences in perceived evidence variability. We replicated the results of the first experiment but did not find substantial effects of luminance variability on confidence ratings. Our findings support the view that decisions and confidence judgements are based on partly dissociable sources of information, and suggest that decisions initially made with higher confidence may be more resistant to subsequent changes of mind.

1. Introduction

The cognitive and neural processes underlying simple decisions have been studied extensively over the past decades, and performance in discrete choice tasks has been successfully accounted for using computational models (Gold & Shadlen, 2007; Ratcliff, Voskuilen, & Teodorescu, 2018; Smith & Ratcliff, 2004). The most prominent class of models are evidence accumulation models, such as the Diffusion Decision Model (DDM; Ratcliff, 1978). These models describe the decision process as a noisy accumulation of evidence towards alternative decision thresholds. For a discrete perceptual decision, such as deciding whether a cloud of dots are predominantly moving to the left or the right, these models propose that sensory evidence is sampled and integrated over time, and a decision is made when the accumulated evidence reaches a threshold in favour of a particular choice outcome.

When an incorrect decision is made, we can often rapidly detect that an error has occurred (Ullsperger, Danielmeier, & Jocham, 2014). For example, in a typical Flanker task, when judging the identity of a central letter in the presence of distracting flankers, a detected decision error is reflected in brain activity following the incorrect motor response (Scheffers & Coles, 2000). Beyond simply detecting an error, we can also rapidly change our minds and correct erroneous decisions (Resulaj, Kiani, Wolpert, & Shadlen, 2009; van den Berg et al., 2016). This capacity for fast changes of mind has been linked to metacognitive processes – specifically, those which allow us to derive a subjective sense of confidence in our decisions (Fleming & Daw, 2017; van den Berg et al., 2016). Consistent with the notion of a link between the processes underlying confidence and changes of mind, decisions initially made with high confidence are less likely to be overruled than those made with a lower degree of confidence (van den Berg et al., 2016). In light of this, it

* Corresponding author at: Melbourne School of Psychological Sciences, Parkville, VIC 3010, Australia.
E-mail address: sbode@unimelb.edu.au (S. Bode).

https://doi.org/10.1016/j.cognition.2022.105125
Received 10 June 2021; Received in revised form 8 April 2022; Accepted 11 April 2022
Available online 25 April 2022
0010-0277/© 2022 Elsevier B.V. All rights reserved.

has been proposed that decision confidence may influence the processes that determine how quickly and how often we change our minds (Turner, Feuerriegel, Andrejevic, Hester, & Bode, 2021; van den Berg et al., 2016). Beyond these initial proposals, however, this idea is yet to be systematically tested.

1.1. The effect of absolute evidence on changes of mind

A recent study investigated how variations in absolute evidence magnitude affect the speed and accuracy of decisions and subsequent changes of mind (Turner, Feuerriegel, et al., 2021). In this study, participants judged which of two flickering squares was, on average, brighter by integrating information over a short period of time. The overall (i.e., summed) luminance across the two stimuli was manipulated to investigate the effect of absolute evidence, while relative evidence (i.e., their luminance difference) was held constant. According to certain classes of evidence accumulation models (i.e., 'purely relative' models such as the DDM), the *differences* in evidence for each choice option are accumulated in the decision process (Ratcliff & Rouder, 1998). This is empirically supported by the standard finding that increasing relative evidence leads to higher accuracy and faster response times (RTs; Ratcliff et al., 2018; Teodorescu, Moran, & Usher, 2016). However, purely relative models do not predict an effect of absolute evidence on decision accuracy or RT if relative evidence is held constant (although performance differences at different absolute evidence levels could still emerge due to Weber-scaling, i.e., diminished increase in perceived brightness as luminance increases; Geisler, 1989). Nevertheless, Turner, Feuerriegel, et al. (2021) showed that, even after considering the effects of Weber-scaling, decisions are sensitive to variations in absolute evidence magnitude – a finding which purely relative models cannot account for. In particular, consistent with other studies, it was shown that increasing absolute evidence leads to faster but less accurate decisions (Ratcliff et al., 2018; Teodorescu et al., 2016; Turner, Feuerriegel, et al., 2021).

It should be noted that, to be consistent with previous studies (Peters et al., 2017; Teodorescu et al., 2016; Turner, Feuerriegel, et al., 2021), we use the term 'evidence' to refer to sensory evidence. That is, sensory information about each of the choice options. Therefore, in concrete terms, we define the 'evidence' for each choice option as their respective luminance values. Importantly, under this definition, the term 'evidence' should not be interpreted as necessarily referring to 'choice evidence'. That is, information which can be used to inform a choice. This is because the sensory evidence associated with a single choice option (i.e., the luminance of one of the squares), or indeed the overall level of absolute evidence, is by itself not informative for decision-making (at least for comparative judgements).

Turner, Feuerriegel, et al. (2021) asked the additional question of how absolute evidence magnitude affects the speed and likelihood of change-of-mind decisions. In their study, the stimuli were first presented for an initial luminance judgment and then remained on the screen for a further 1 s, allowing participants to submit a second, change-of-mind response within this time window. They reported that higher levels of absolute evidence led to slower change-of-mind RTs relative to the time of the decision. Importantly, these RT effects also remained when effects of Weber-scaling were accounted for in a follow-up experiment. This finding suggests that participants may have required a larger amount of conflicting, post-decisional sensory evidence to overrule their decisions in conditions of high absolute evidence. As decision RTs were consistently faster in higher absolute evidence conditions, and faster RTs are associated with higher levels of decision confidence (Kiani, Corthell, & Shadlen, 2014), participants may have been more confident in their decisions, despite being objectively less accurate. If this were the case, this might have led participants to wait longer and accumulate more evidence before deciding to overrule their decision. The current study aimed to examine whether confidence would increase with increased absolute evidence. This would point to a moderation effect of confidence that could ultimately drive changes of mind. We further tested more directly whether any effects of confidence would potentially translate into changes of mind by converting our confidence measure into a change-of-mind measure, following previous approaches (Charles & Yeung, 2019; Fleming, van der Putten, & Daw, 2018).

1.2. The decision-congruent evidence hypothesis

The idea that confidence may have increased with higher absolute evidence magnitude is consistent with the *decision-congruent evidence hypothesis*. This hypothesis suggests that the extent of sensory evidence in favour of the selected option primarily informs confidence judgements (Koizumi, Maniscalco, & Lau, 2015; Odegaard et al., 2018; Peters et al., 2017; Samaha & Denison, 2020; Zylberberg, Barttfeld, & Sigman, 2012). These accounts suggest that, while decisions are determined by the difference in evidence between choice options (i.e., relative evidence), confidence might be a product of a winner-takes-all process. The more evidence for the winning option, the higher the confidence in the decision, regardless of the amount of evidence for the alternative, non-chosen option (Peters et al., 2017; Zylberberg et al., 2012). Similar results have been shown experimentally via manipulations of 'positive' evidence (i.e., evidence supporting the correct response) and 'negative' evidence (i.e., evidence supporting the alternative response) in random dot motion and grating orientation judgment tasks (Koizumi et al., 2015; Odegaard et al., 2018; Samaha, Barrett, Sheldon, LaRocque, & Postle, 2016; Samaha & Denison, 2020). Specifically, when the ratio between positive and negative evidence was held constant, increased positive and negative evidence together led to increased confidence without impacting accuracy. This effect was termed the *Positive Evidence Bias* (Maniscalco et al., 2021; Samaha & Denison, 2020). Taken in relation to the findings of Turner, Feuerriegel, et al. (2021), this would mean that stronger absolute evidence (and the corresponding increase in decision-congruent evidence) may have increased participants' subjective confidence in their decisions and, in turn, made them less prone to changing their mind. As Turner, Feuerriegel, et al. (2021) did not investigate confidence, the current study was designed to directly test whether stronger absolute evidence in the same task as used by Turner and colleagues does indeed lead to increased decision confidence.

1.3. The current study

We employed a luminance discrimination task using flickering stimuli, as in Turner, Feuerriegel, et al. (2021), and manipulated both absolute and relative evidence across three levels (low, medium, and high). Stimuli were presented for a maximum of 1.5 s and disappeared when the keypress response reported the perceptual decision. Participants subsequently reported their degree of confidence in their decision on a 7-point scale ranging from "surely incorrect" to "surely correct".

In Experiment 1, we first aimed to replicate previous findings that (a) increasing relative evidence leads to increased decision accuracy and faster RTs, and (b) increasing absolute evidence leads to both lower accuracy and faster RTs (Ratcliff et al., 2018; Teodorescu et al., 2016; Turner, Feuerriegel, et al., 2021). Furthermore, we predicted that confidence in trials with correct responses would increase with stronger relative evidence, while confidence in error trials would decrease with stronger relative evidence, as shown in previous studies (Sanders, Hangya, & Kepecs, 2016). Critically, we also predicted that stronger absolute evidence would be associated with increased confidence for both correct and error trials, despite decreased decision accuracy. This is because higher absolute evidence implies stronger decision-congruent evidence for both correct and error trials. Such findings in a highly similar task to Turner, Feuerriegel, et al. (2021) would suggest that slower change-of-mind responses co-occur with a higher degree of confidence in one's decision.

It should be taken into account that the high absolute evidence stimuli (i.e., brighter pairs of squares) would likely have been perceived as being less variable over time. This is because physical luminance and perceived brightness are related via a nonlinear compressive function (Geisler, 1989). In other words, under conditions of high luminance, changes in perceived brightness with an equivalent increase in luminance are diminished. Accordingly, when there is the same amount of luminance variability, the variability in brightness over time will be perceived as more pronounced in the dimmer stimulus condition (i.e., dimmer squares will appear to flicker more than brighter squares).

Perceived stimulus variability is theorised to inform confidence judgements (Yeung & Summerfield, 2012). Consistent with this theory, some previous studies have shown that higher stimulus variability can lead to lower confidence ratings (e.g., Desender, Boldt, & Yeung, 2018; Navajas et al., 2017; Spence, Dux, & Arnold, 2016). However, the opposite has also been found (e.g., Zylberberg, Roelfsema, & Sigman, 2014; Zylberberg, Fetsch, & Shadlen, 2016), where the observed effects appear to vary by the task and type of variability manipulation. It is therefore possible that reductions in perceived brightness variability for higher absolute evidence conditions in our task might have led to higher confidence ratings, which might explain the observed effects of absolute evidence.

In Experiment 2, we examined whether decreases in stimulus variability could lead to increased confidence in our experimental design, by directly manipulating luminance variability (i.e., the distribution of luminance values between frames around the same mean), in addition to relative and absolute evidence. This experiment therefore served two purposes: (a) to replicate the general effects of Experiment 1, and (b) to directly (i.e., experimentally) test the effect of stimulus variability on confidence within our specific luminance task. While finding that reduced stimulus variability leads to substantially higher confidence ratings would not prove that this is indeed the explanation for why increased absolute evidence impacts confidence, it would nevertheless demonstrate that it is possible for stimulus variability to affect confidence in our study design. We could then speculate that this might also be a potential alternative explanation for the effects of absolute evidence on confidence in our experiments. However, if directly manipulating stimulus variability does not affect confidence in our study, a relevant contribution of stimulus variability on the current findings can be essentially ruled out. Moreover, we could include stimulus variability in our models to test whether effects of absolute evidence are reproducible at different levels of stimulus variability.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Thirty-seven university student volunteers with normal or corrected-to-normal vision were recruited. Six participants were excluded: Three failed to report confidence in more than 20% of all trials, two showed lower than 55% accuracy, and one reported the same confidence level in more than 90% of trials where confidence was reported. The final sample comprised 31 participants (mean age = 26 years, SD = 5, range 19–38 years, 17 females). This experiment was approved by the University of Melbourne ethics committee (ID: 1954641.2).

2.1.2. Experimental procedures

Before the experiment, participants gave written consent and were given task instructions. Participants were then seated in a dark testing booth 70 cm from a computer monitor. At the beginning of the experiment, participants underwent task training while the experimenter stayed in the testing booth. This procedure ensured that participants understood the task instructions correctly. Participants then completed the main experiment. After completion of the task, participants were reimbursed 20 AUD and were debriefed by the experimenter.

2.1.3. Task and stimuli

We used a luminance judgment task to examine the effects of relative and absolute evidence on perceptual decisions, RTs, and decision confidence. Participants had to decide which of two flickering squares, presented on the left and right of a central red fixation dot, was brighter (Fig. 1A). There were three levels (low, medium, high) for both relative evidence and absolute evidence, resulting in a 3 × 3 factorial design (Fig. 1B).

The two square stimuli changed in luminance with each frame refresh (i.e., every 13.3 ms at 75 Hz). For each refresh, the luminance value for each square was determined by randomly drawing values sampled from two truncated normal distributions around the predetermined means for each square, respectively. Mean luminance values for the two distributions were specified by pairs of RGB values, such that one distribution had a higher mean than the other (therefore, one stimulus appeared on average brighter than the other; Fig. 1B). Following Ratcliff et al. (2018), both distributions had a standard deviation of 25.5 and were truncated at ±1 SD from their means. The mean luminance values mapped onto relative evidence strength, defined as the difference in distribution means for the two stimuli, and absolute evidence strength, defined as the sum of the distribution means for both stimuli. The size of both stimuli was 70 × 70 pixels, and they were positioned at equal distance from the centre of the screen, separated from each other by 180 pixels. The positions of the stimuli were counterbalanced such that in half of the trials, the left stimulus was brighter, and in the other half of the trials the right stimulus was brighter. The order of the trials was randomised. Participants were instructed to respond as quickly as possible. After submitting the choice response, participants were required to indicate their confidence using a 7-point rating scale ranging from "surely incorrect" (1) to "surely correct" (7), with a midpoint rating (4) indicating they were unsure whether they were correct or incorrect (i.e., they felt they were guessing). They were again instructed to respond as fast as possible.

Participants completed a training phase before starting the experimental phase. They first practised the experimental task (see below) for 36 trials, without making confidence ratings. Instead, they received performance feedback after each trial to familiarise themselves with the judgement task. Subsequently, they practised the entire task, including confidence judgements for another 36 trials in which (as in the main experiment) no performance feedback was given. During training, a confidence rating scale was presented on the screen for a maximum of 1500 ms or until response. Participants were instructed that during the main task, this visual presentation of the scale would be removed, and only the word "confidence" would prompt the rating.

Experiment. After the two training blocks, participants started the main experiment. Each trial started with an intertrial interval lasting 500 ms. A red fixation dot followed this interval in the middle of the screen for 600 ms, and then a blank screen was presented for 200 ms. After that, the flickering squares were presented, and participants were required to make the brightness decision by pressing either the left or right key on a response pad using left and right index fingers, corresponding to which stimulus they perceived as brighter. The stimuli were presented for a maximum of 1500 ms and disappeared immediately after a response was submitted. Subsequently, after an interval of 500 ms with a blank screen, participants were asked to rate their confidence without a visual presentation of a rating scale but prompted by the word "confidence". The scale had the same properties as during training, and participants were required to press one of the seven keys on the response pad to indicate their confidence level. No confidence rating was required if the brightness judgment was "too slow" (>1500 ms RT) or "too quick" (<250 ms RT). In this case, only the respective timing feedback was presented for 1500 ms, and then the next trial began.

The experiment comprised 1008 experimental trials equally allocated across 14 blocks. Each block was followed by a self-terminated rest period. An equal number of trials from all conditions were randomly interleaved within each block.

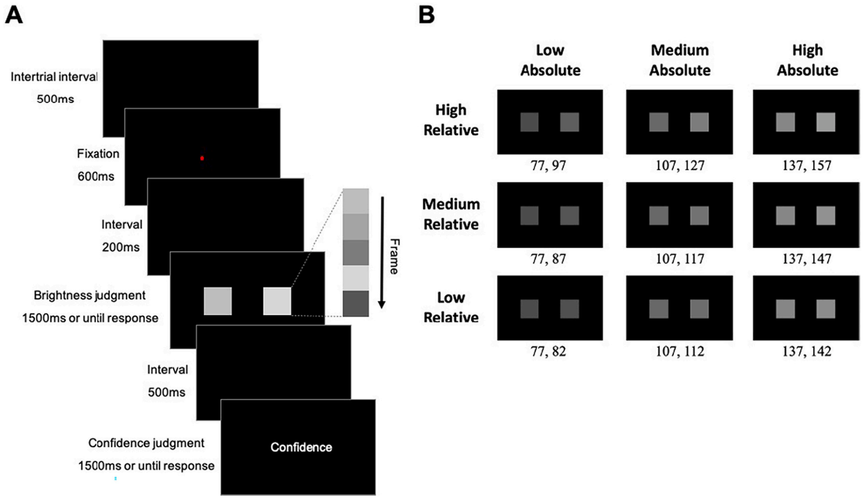

Fig. 1. Task paradigm and stimuli. (A) Paradigm. In each trial, two flickering square stimuli of different average luminance were presented. Each square changed in luminance with each frame. Participants were required to select the stimulus that appeared brighter on average and subsequently reported their decision confidence using a 7-point scale while the word "confidence" was presented on the screen. (B) Illustration of average luminance values for stimuli for all experimental conditions of Experiment 1. Luminance values were randomly sampled from normal distributions truncated one standard deviation around pre-defined means. The standard deviation of all distributions was 25.5.

2.1.4. Apparatus

Stimuli were presented on a Sony Trinitron Multiscan G420 CRT Monitor (resolution 1280 × 1024 pixels; frame rate 75 Hz) that was gamma-corrected with a ColorCAL MKII Colorimeter (Cambridge Research Systems), such that the physical luminance of the stimuli was linearly related to the RGB values. The task was programmed in MATLAB R2018b (The Mathworks) using Psychtoolbox-3 (Brainard, 1997; Kleiner et al., 2007). Participants responded using a seven-button Cedrus response pad (RB-740, Cedrus Corporation).

2.1.5. Data analysis

We used generalised linear mixed-effects models (GLMMs) to examine the effects of relative and absolute evidence on accuracy, RT, and confidence ratings. For RT and confidence ratings, we ran two separate sets of analyses: one included only correct trials and the second set included only error trials. This was done to control for the effect of accuracy, given that error trials have different RT distributions than correct trials, and confidence patterns for correct and error trials could also potentially differ across relative and absolute evidence conditions (Gajdos, Fleming, Saez Garcia, Weindel, & Davranche, 2019; Turner et al., 2021; Urai, Braun, & Donner, 2017). Additionally, by converting confidence ratings into a change-of-mind measure, we also analysed how change-of-mind frequency was affected by absolute and relative evidence strength. This was done by transforming confidence ratings into a binary variable (confidence lower than 4 as 1 [change-of-mind trials], and confidence higher or equal to 4 as 0 [trials without changes of mind]), as in Charles and Yeung (2019) and Fleming et al. (2018). This approach resulted in seven separate and independent analyses with different dependent variables: accuracy, RT (correct), RT (error), confidence (correct), confidence (error), changes of mind (correct), and changes of mind (error).

For each model, the model structure included fixed effects of relative evidence, absolute evidence, and the interaction between relative and absolute evidence, and a random intercept by participant. We also attempted to fit models with random slopes for each effect of interest. However, we found that not all models with random slopes converged across the different analyses of accuracy, RT and confidence. To be consistent across these different analyses we therefore used models without random slopes.

As in Turner, Feuerriegel, et al. (2021), for different dependent variables different distributions were assumed, and different link functions were used: Binomial distributions with a logit function were used to model accuracy and changes of mind, gamma distributions with an identity function were used to model RTs, and normal distributions with an identity function were used to model confidence. All analyses were conducted in R (version 4.0.1). GLMMs were fitted using the lme4 package (version 1.1; Bates, Maechler, Bolker, & Walker, 2015), and statistical significance of each effect was determined by likelihood ratio tests conducted using the afex package (version 0.28; Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2020). For brevity, only significant effects are reported in the results section. Complete statistical results including likelihood ratio test results for all effects and regression coefficients of the full models are reported in Supplementary Material. Code and data used for the analyses in this paper are available at https://osf.io/r8vfx/.

As shown by the results below, increases in absolute evidence led to increased confidence and faster RTs. As faster responding has been shown to contribute to higher confidence (Kiani et al., 2014), we further asked whether the effect of absolute evidence on confidence could simply be explained by faster RTs in conditions with higher absolute evidence. To answer this question, we ran post-hoc analyses in which confidence was predicted by RT and the same predictors as in the main analyses except absolute evidence, and then included absolute evidence in the model to examine whether absolute evidence could predict confidence above the effect of RT. Also similar to the main analyses, each variable was entered into the model in a forward stepwise approach, and the statistical significance of each predictor was determined by likelihood ratio tests comparing the models before and after the predictor was included.

2.2. Results

2.2.1. Accuracy and response times

First, we aimed to replicate previous findings that accuracy increases with stronger relative evidence but decreases with stronger absolute evidence (Ratcliff et al., 2018; Teodorescu et al., 2016; Turner, Feuerriegel, et al., 2021). As expected, there was a positive effect of relative evidence ($\chi^2[2] = 1433.01$, $p < .001$), a negative effect of absolute evidence ($\chi^2[2] = 485.87$, $p < .001$), and an interaction ($\chi^2[4] = 91.71$, $p < .001$; the interaction was observed because the log odds of being correct was reduced by absolute evidence more strongly when relative evidence was high; see Supplementary Fig. S1. However, this pattern was not observable in terms of proportion correct). Fig. 2A shows that the average proportion of correct decisions increased with stronger relative evidence but decreased with stronger absolute evidence. Full statistical results are presented in Supplementary Tables S1 and S2.

RTs were expected to be faster in conditions of stronger relative evidence and higher absolute evidence (Ratcliff et al., 2018; Teodorescu et al., 2016; Turner, Feuerriegel, et al., 2021). Consistently, for correct trials, there was an effect of relative evidence ($\chi^2[2] = 314.71$, $p < .001$), an effect of absolute evidence ($\chi^2[2] = 35.85$, $p < .001$), and an interaction ($\chi^2[4] = 106.25$, $p < .001$; Fig. 2B). RTs were faster in conditions with stronger relative and stronger absolute evidence, and the effects of relative evidence appeared to diminish in conditions of higher absolute evidence. When the analysis was repeated for error trials only, similar effects were found (relative evidence: $\chi^2[2] = 18.54$, $p < .001$; absolute evidence: $\chi^2[2] = 49.96$, $p < .001$; interaction: $\chi^2[4] = 20.20$, $p < .001$; Fig. 2C). Fig. 2B and C show that RTs generally became faster with stronger relative evidence. Stronger absolute evidence also led to faster RTs for low and medium relative evidence, both for correct as well as error trials. The exception was that, for the high relative evidence condition, RTs were slower with low absolute evidence in error trials, but faster in correct trials. This result pattern was also reported in a previous study and appears to be a feature of this task (Ratcliff et al., 2018). Supplementary Fig. S2 further shows that RTs were faster in conditions of higher absolute evidence across all RT quantiles, as also reported by Turner, Feuerriegel, et al. (2021). Full statistical results are presented in Supplementary Tables S3 – S6.

Taken together, these results show that increases in relative evidence are associated with faster and more accurate decisions. Moreover, these results replicate recent reports that increases in absolute evidence were associated with faster but less accurate decisions. The following section investigates the effect of absolute evidence magnitude on participants' confidence ratings directly.

2.2.2. Confidence

We predicted that confidence would increase with both stronger relative and absolute evidence for correct trials. Consistent with our prediction, there was an effect of relative evidence ($\chi^2[2] = 879.07$, $p < .001$), an effect of absolute evidence ($\chi^2[2] = 293.89$, $p < .001$), and an interaction ($\chi^2[4] = 121.55$, $p < .001$). Fig. 3A shows that mean confidence ratings increased with both relative and absolute evidence, although the effect of absolute evidence diminished as relative evidence increased. For the analysis of error trials, there was only an effect of relative evidence ($\chi^2[2] = 99.99$, $p < .001$) and an effect of absolute evidence ($\chi^2[2] = 392.15$, $p < .001$); (Fig. 3B). As expected, the direction of the relative evidence effect was reversed, with highest confidence ratings seen in lower as compared to higher relative evidence conditions. Full statistical results are presented in Supplementary Tables S7 – S10. Interestingly, even on error trials, absolute evidence magnitude was positively associated with confidence.

2.2.3. Change of mind

When confidence was transformed into a binary variable that indicates changes of mind (confidence lower than 4 indicates a change of mind), in correct trials we observed a negative effect of relative evidence ($\chi^2[2] = 259.51$, $p < .001$), a negative effect of absolute evidence ($\chi^2[2] = 38.67$, $p < .001$), and an interaction between relative and absolute evidence ($\chi^2[4] = 22.35$, $p < .001$). Fig. 4A showed that changes of mind were less likely with both stronger relative and absolute evidence, although the effect of absolute evidence diminished for stronger relative evidence. For error trials, a similar negative effect of absolute evidence was observed ($\chi^2[2] = 176.64$, $p < .001$), while relative evidence showed a positive effect ($\chi^2[2] = 96.92$, $p < .001$). There was also an interaction between relative and absolute evidence ($\chi^2[4] = 14.09$, $p = .007$). Fig. 4B showed that changes of mind were less likely with stronger absolute evidence and weaker relative evidence, and the effect of absolute evidence was diminished when relative evidence was weaker. These patterns of results are opposite to the patterns of confidence, consistent with the negative relationship between confidence and

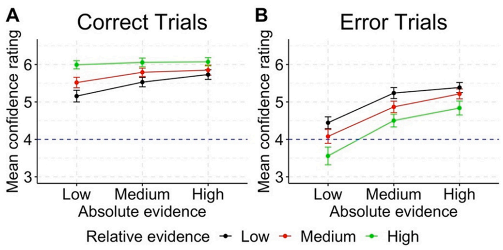

Fig. 3. Mean confidence ratings in each condition in Experiment 1. (A) Correct trials. (B) Error trials. Confidence ratings were measured on a scale ranging from 1 ("surely incorrect") to 7 ("surely correct"). The dotted line indicates the mid-point of the scale. Error bars represent SEM.

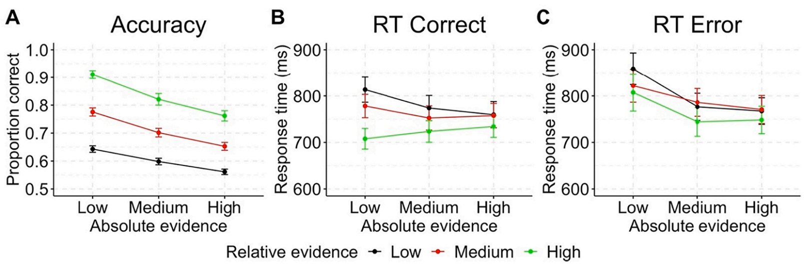

Fig. 2. Experiment 1 accuracy and response time (RT). (A) Decision accuracy (average proportion correct) in each condition. (B) Mean RTs for correct trials. (C) Mean RTs for error trials. Error bars represent standard errors of the mean (SEM).

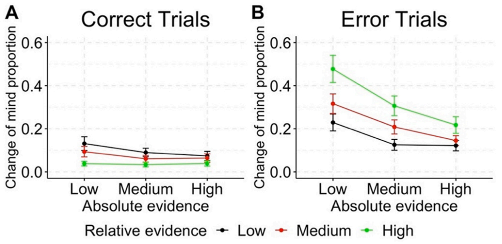

Fig. 4. Experiment 1 proportions of change-of-mind trials in each condition. (A) Correct trials. (B) Error trials. Change-of-mind trials were defined as trials with confidence ratings lower than 4, indicating that the participant believed their brightness judgement was incorrect. Error bars represent SEM.

changes of mind. Full statistical results are presented in Supplementary Tables S11 – S14.

2.2.4. The effect of absolute evidence on confidence in addition to RT

Lastly, to examine whether the effect of absolute evidence on confidence was simply due to faster RTs in higher absolute evidence conditions, we fitted models in which confidence was predicted by RT and relative evidence, and then compared model fits with a model that included the predictor of absolute evidence. When controlling for effects of RT in this way, confidence in correct trials was still predicted by relative evidence ($\chi^2[2] = 667.60$, $p < .001$) and RT ($\chi^2[1] = 1207.10$, $p < .001$), and additionally by absolute evidence ($\chi^2[2] = 270.15$, $p < .001$) as well as the interaction between relative and absolute evidence ($\chi^2[4] = 80.22$, $p < .001$). Similarly, confidence in error trials was also predicted by relative evidence ($\chi^2[2] = 125.88$, $p < .001$) and RT ($\chi^2[1] = 405.39$, $p < .001$), and additionally by absolute evidence ($\chi^2[2] = 335.72$, $p < .001$) and its interaction with relative evidence ($\chi^2[4] = 10.37$, $p = .035$). Full statistical results are presented in Supplementary Tables S15 – S18. To further visualize how confidence was related to RT and absolute evidence, we binned the data into six RT bins using RT quantiles of each participant (10%, 30%, 50%, 70%, 90%; separately for correct and error trials), and plotted mean confidence in each bin by absolute evidence in Fig. 5A and B. These figures show that increasing absolute evidence generally led to higher confidence across correct and error trials across RT bins.

In summary, these results confirm that, even though decision accuracy decreased with increasing absolute evidence, confidence increased for both correct and incorrect responses. Changes of mind likelihood results showed the opposite pattern, which could be expected from its negative relationship with confidence. Lastly, the positive effect of absolute evidence on confidence did not appear to be simply due to faster RTs in conditions with higher absolute evidence. Next, we investigated in Experiment 2 whether confidence also co-varied with stimulus variability in our task, which was introduced as an additional factor.

3. Experiment 2

3.1. Methods

3.1.1. Participants

A different sample of 35 university student volunteers with normal or corrected-to-normal vision was recruited. Six participants were subsequently excluded: Three failed to report confidence in more than 20% of all trials, one showed lower than 55% accuracy, and two reported the same confidence level in more than 90% of trials where confidence was reported. Twenty-nine participants were included in the analysis (mean age = 23, SD = 5, range 18–39 years, 25 females). This experiment was approved by the University of Melbourne ethics committee (ID: 1954641.2).

3.1.2. Experimental procedures

Experimental procedures were identical to Experiment 1, except where noted below.

3.1.3. Task and stimuli

Experiment 2 aimed to test the effect of evidence variability (i.e., the standard deviations of the distributions from which luminance values were sampled in each frame) on decision accuracy and confidence. It was a replication of Experiment 1 but included the additional factor "luminance variability". We again used three levels of absolute evidence (low, medium, high), as in Experiment 1. We only included two levels of relative evidence (low, high) because the effects of relative evidence were not of primary interest in this experiment. The mean luminance values for the low and high relative evidence conditions were in-between the values used in Experiment 1 (see Supplementary Table S19) to reduce ceiling and floor effects. We further included two levels of luminance variability (low, high), resulting in a 3 × 2 × 2 design. Evidence variability was operationalised as the variability of the luminance value distributions (standard deviation of 25.5 for high variability and 12.5 for low variability; Supplementary Fig. S3). The high variability condition was identical to Experiment 1, while the low variability condition contained only half the variability around the mean as compared to Experiment 1. The task structure, stimulus presentation, and apparatus were otherwise the same as in Experiment 1.

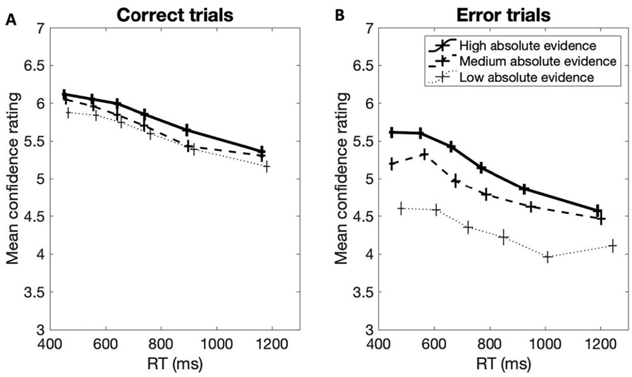

Fig. 5. Experiment 1 mean RTs and confidence ratings by absolute evidence level and RT quantile bin (with borders between bins at the 10th, 30th, 50th, 70th, and 90th RT percentiles) for (A) correct trials and (B) error trials. Horizontal and vertical error bars indicate SEM of RT and confidence, respectively. Note that the statistical effects for absolute and relative evidence cannot be seen clearly in this figure due to the division into RT quantiles and the omission of relative evidence levels (for illustrations of these effects see Figs. 2B and C).

3.1.4. Data analysis

The same GLMM approach used in Experiment 1 was used for Experiment 2, except that the models included the additional factor of variability and its interactions with relative evidence, absolute evidence, and the three-way interaction term.

3.2. Results

3.2.1. Accuracy and response times

For decision accuracy, there were an effect of relative evidence ($\chi^2[1] = 576.22$, $p < .001$), an effect of absolute evidence ($\chi^2[2] = 588.29$, $p < .001$), and an interaction between relative and absolute evidence ($\chi^2[2] = 37.02$, $p < .001$) as in Experiment 1. This interaction was again because log odds of being correct were reduced by absolute evidence more strongly when relative evidence was high (see Supplementary Fig. S4). Importantly, luminance variability and its interaction terms were not significant. As depicted in Fig. 6A and D, the accuracy of participants' responses increased with stronger relative evidence and decreased with higher absolute evidence, regardless of flicker variability. Full statistical results are presented in Supplementary Tables S20 and S21.

For RTs, when analysing data from correct trials, there was an effect of relative evidence ($\chi^2[1] = 106.53$, $p < .001$), an effect of absolute evidence ($\chi^2[2] = 44.46$, $p < .001$), and an interaction between relative and absolute evidence ($\chi^2[2] = 32.20$, $p < .001$), again replicating Experiment 1 (Fig. 6B and E). Additionally, we found an effect of luminance variability ($\chi^2[1] = 6.61$, $p = .010$), an interaction between absolute evidence and luminance variability ($\chi^2[2] = 7.13$, $p = .028$), and an interaction among all three factors ($\chi^2[2] = 6.24$, $p = .044$). This effect appears to be driven by a small dip in RTs for the low relative / medium absolute evidence condition in the high compared to the low luminance variability condition. The overall pattern of RT results, however, was highly similar between luminance variability conditions, suggesting no substantial effect of luminance variability on RT in correct trials.

Analyses of error trials again showed a similar pattern of results (relative evidence: $\chi^2[1] = 23.74$, $p < .001$; absolute evidence: $\chi^2[2] = 37.22$, $p < .001$; interaction between relative and absolute evidence: $\chi^2[2] = 14.25$, $p < .001$), and no effect of luminance variability or interaction involving luminance variability. Fig. 6C and F show that the RT effects for error trials from Experiment 1 replicated regardless of variability condition. Full statistical results are presented in Supplementary Tables S22 – S25.

Taken together, the effects of relative and absolute evidence on accuracy and RT found in Experiment 1 were replicated in Experiment 2. Importantly, we did not find strong and significant effects of luminance variability on accuracy or RT measures.

3.2.2. Confidence

When analysing correct trials, there was an effect of relative evidence ($\chi^2[1] = 261.57$, $p < .001$), an effect of absolute evidence ($\chi^2[2] = 191.90$, $p < .001$), and an interaction between relative and absolute evidence ($\chi^2[2] = 47.83$, $p < .001$). Confidence was higher in trials with stronger relative evidence and higher absolute evidence, and the effect of relative evidence was diminished in high absolute evidence conditions, replicating the pattern of results of Experiment 1. There was no significant effect of luminance variability nor significant interactions with luminance variability (Fig. 7A and C).

For analyses of error trials, we again found effects for relative and absolute evidence, but unlike in Experiment 1, despite reproducing the same overall pattern, their interaction failed to reach significance (relative evidence: $\chi^2[1] = 57.71$, $p < .001$; absolute evidence: $\chi^2[2] = 400.02$, $p < .001$; Fig. 7B and D). The absence of this interaction is most likely because we included only two levels of relative evidence, which were more similar to each other than in Experiment 1. While there was no main effect of luminance variability, an interaction was found between relative evidence and luminance variability ($\chi^2[2] = 4.35$, $p = .037$); as well as between absolute evidence and luminance variability ($\chi^2[2] = 8.47$, $p = .014$). The interaction between relative evidence and luminance variability was because higher variability was associated with increased confidence when relative evidence was high. That is, when relative evidence was high, low luminance variability was

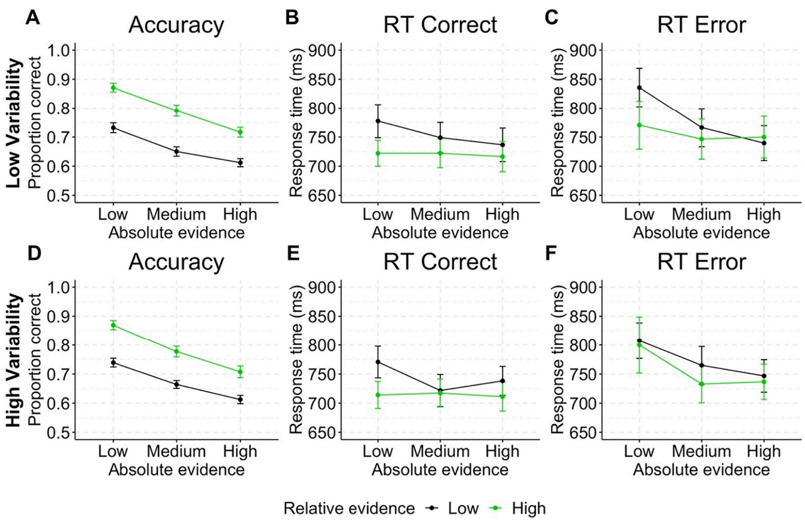

Fig. 6. Experiment 2 accuracy and response time (RT). (A, D) Decision accuracy (average proportion correct) for the low (A) and high (D) luminance variability conditions. (B, E) Mean RTs for correct trials for low (B) and high (E) luminance variability conditions. (C, F) Mean RTs for error trials for low (C) and high (F) luminance variability conditions. Error bars represent SEM.

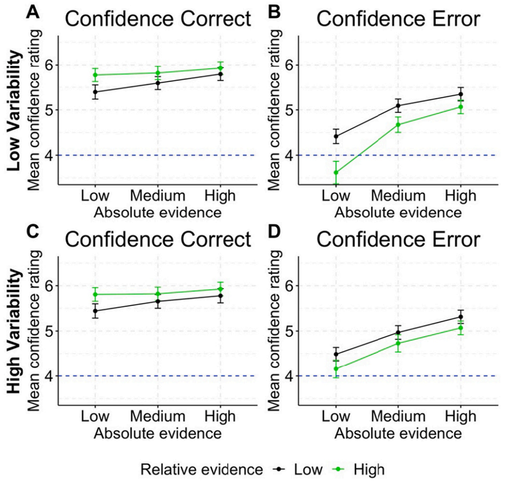

Fig. 7. Mean confidence ratings in Experiment 2. (A, C) Mean confidence for correct trials for low (A) and high (C) variability. (B, D) Mean confidence for error trials for low (B) and high (D) variability. Confidence ratings were measured on a scale ranging from 1 ("surely incorrect") to 7 ("surely correct"). The dotted line indicates the mid-point of the scale. Error bars represent SEM.

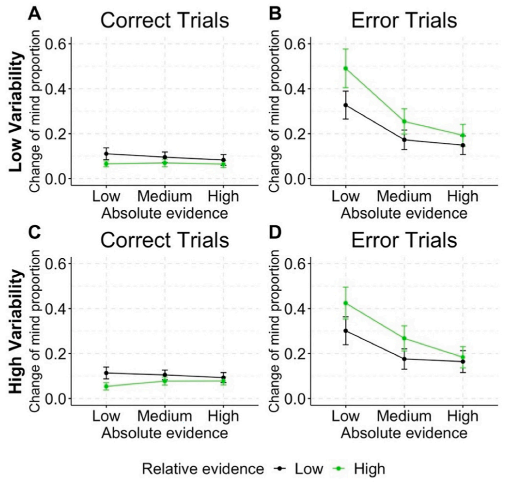

Fig. 8. Experiment 2 proportions of change-of-mind trials in each condition. (A, C) Mean proportion of change of mind for correct trials for low (A) and high (C) variability. (B, D) Mean proportion of change of mind for error trials for low (B) and high (D) variability. Change-of-mind trials were defined as trials with confidence ratings lower than 4, indicating that the participant believed their brightness judgment was incorrect. Error bars represent SEM.

associated with lower confidence judgements, despite no detected change in performance (see above). The interaction between absolute evidence and luminance variability was reflected in higher luminance variability associated with increased confidence when absolute evidence was low. Full statistical results are presented in Supplementary Tables S26 – S29.

3.2.3. Change of mind

For change-of-mind likelihood in correct trials, there was a negative effect of relative evidence ($\chi^2[2] = 36.33, p < .001$), and an interaction between relative and absolute evidence ($\chi^2[2] = 9.72, p = .008$). However, unlike in Experiment 1, there was no effect of absolute evidence (Fig. 8A and C). For error trials, there was a positive effect of relative evidence ($\chi^2[2] = 119.24, p < .001$), a negative effect of absolute evidence ($\chi^2[2] = 19.63, p < .001$) and an interaction between relative and absolute evidence ($\chi^2[2] = 125.93, p < .001$; Fig. 8B and D), as in Experiment 1. Full statistical results are presented in Supplementary Tables S30 – S33.

3.2.4. The effect of absolute evidence on confidence in addition to RT

When modelling confidence in correct response trials using predictors of RT, luminance variability and relative evidence (but not absolute evidence), we observed both the effects of relative evidence ($\chi^2[1] = 195.23, p < .001$) and RT ($\chi^2[1] = 1044.45, p < .001$). When we added absolute evidence (and interactions involving this variable) as a predictor, we also observed the effect of absolute evidence ($\chi^2[2] = 169.31, p < .001$), and an interaction between absolute evidence and relative evidence ($\chi^2[2] = 33.46, p < .001$). For error trials, confidence was predicted by relative evidence ($\chi^2[1] = 75.31, p < .001$) and RT ($\chi^2[1] = 357.34, p < .001$). Additionally, it was predicted by absolute evidence ($\chi^2[2] = 357.69, p < .001$), an interaction between luminance variability and absolute evidence ($\chi^2[2] = 6.83, p = .033$), and an interaction between luminance variability and relative evidence ($\chi^2[1] = 4.95, p = .026$). Consistent with Experiment 1, Fig. 9A and B also showed that stronger absolute evidence increased confidence across RT bins for both correct and error trials. Full statistical results are presented in Supplementary Tables S34 and S37.

Taken together, the results of Experiment 2 replicated the effects of relative and absolute evidence on decision accuracy and RT from Experiment 1. They further replicated the overall patterns of results for effects of relative and absolute evidence on confidence. The result that increasing absolute evidence led to increased confidence and faster RTs was also replicated. However, the effects on change-of-mind frequency were only replicated for error trials, but not for correct trials. This is possibly due to the fact that the effect on change of mind was rather weak as participants rarely change their mind after a correct response, and in Experiment 2 the luminance values of the low relative evidence level were higher than that of Experiment 1, in which the effect was stronger.

The luminance variability manipulation did not appear to substantially affect decision performance or confidence judgements. Only when looking at error trials specifically we observed some interactions between luminance variability and confidence. Overall, the pattern of results for high variability looked highly similar to Experiment 1. However, firstly, with lower luminance variability, confidence for error trials in high relative evidence trials was reduced. Given that these were trials in which errors were indeed committed, this means the combination of low luminance variability and strong relative evidence made it easier for participants to sense that their decision might have been wrong. Secondly, the combination of low variability and low absolute evidence also led to decreased confidence in error responses. This again indicates that participants found it somewhat easier to sense that their decision might have been wrong. However, we do not interpret the reported interaction effect as strong evidence that variability substantially influenced confidence judgements in our designs. This is because these interaction effects were rather small compared with the effects of relative and absolute evidence, and they were only observed within low ranges of relative and absolute evidence.

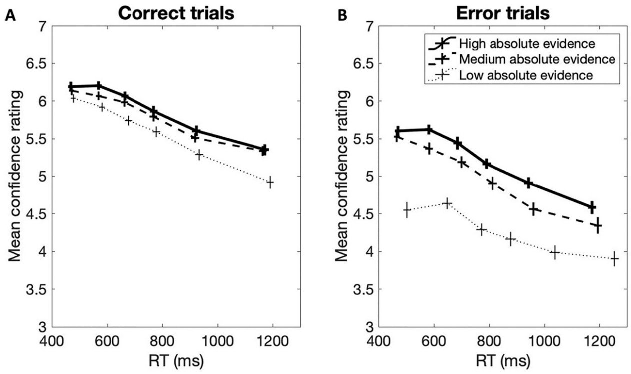

Fig. 9. Experiment 2 mean RTs and confidence ratings by absolute evidence level and RT quantile bin (with borders between bins at the 10th, 30th, 50th, 70th, and 90th RT percentiles) for (A) correct trials and (B) error trials. Horizontal and vertical error bars indicate SEM of RT and confidence, respectively. Note that the statistical effects for absolute and relative evidence cannot be seen clearly in this figure due to the division into RT quantiles and the omission of relative evidence levels (for illustrations of these effects see Fig. 6B, C, E, F).

4. Discussion

Based on the previous finding that higher levels of absolute evidence were associated with less accurate perceptual decisions and slower changes of mind (Turner, Feuerriegel, et al., 2021), and suggestions that confidence in one's decision might moderate the speed and likelihood of later changes of mind (Turner, Feuerriegel, et al., 2021; van den Berg et al., 2016), we asked whether increases in absolute evidence are associated with higher decision confidence. We used a luminance discrimination task that was highly similar to that in Turner, Feuerriegel, et al. (2021) and examined the effect of absolute evidence on decision accuracy, RTs and confidence ratings. In this task, to manipulate absolute evidence we varied the overall (summed) luminance across the two stimuli, in addition to manipulating relative evidence (i.e., their luminance difference). Experiment 1 first replicated previous findings showing that increases in absolute evidence are associated with decreased accuracy and faster RTs (Ratcliff et al., 2018; Teodorescu et al., 2016; Turner, Feuerriegel, et al., 2021). We also found that while stronger relative evidence increased confidence for correct trials and reduced confidence for error trials, absolute evidence increased confidence for both correct and error trials. We replicated these effects in Experiment 2 and did not identify any substantial effects of luminance variability manipulations on task performance or decision confidence. Our findings suggest that heuristic biases in decision confidence associated with absolute evidence magnitude may ultimately make decisions harder to be subsequently overruled, in line with recent theoretical accounts (Turner, Feuerriegel, et al., 2021; van den Berg et al., 2016).

4.1. *Why does high absolute evidence lead to decreased decision accuracy but increased confidence?*

Increasing absolute evidence led to reduced accuracy but increased confidence. While seemingly paradoxical, these divergent effects have several coherent explanations within the general framework of an evidence accumulation process.

Considering first the negative effect of absolute evidence on decision accuracy, this can be explained, at least in part, by Weber's law (Geisler, 1989; Ratcliff et al., 2018; Teodorescu et al., 2016; Turner, Feuerriegel, et al., 2021). Weber's law suggests that relative evidence is perceptually reduced when absolute evidence is increased, due to nonlinear compressive scaling of the incoming sensory information. Over and above the effects of this compressive scaling, however, increases in absolute evidence are also thought to increase noise within the evidence accumulation process (Ratcliff et al., 2018; Turner, Feuerriegel, et al., 2021). This could be explained by assuming that the variability of brightness representations scales with their mean luminance, such that more intense (i.e., brighter) stimuli are represented more variably in terms of brightness (Ratcliff et al., 2018). By assuming an input-dependent increase in noise within the decision process it is possible to account for both the speed up in initial RT and the decrease in choice consistency, which have been observed in previous studies under conditions of high absolute evidence (Ratcliff et al., 2018; Turner, Feuerriegel, et al., 2021).

Turning now to the effect of absolute evidence magnitude on decision confidence, we found that increasing absolute evidence increased confidence, for both correct and incorrect decisions, despite the simultaneous decrease in decision accuracy. Hereafter, we will consider three possible explanations for this effect.

Firstly, this finding is in line with more recent studies that showed a positive evidence bias for decision confidence (Koizumi et al., 2015; Odegaard et al., 2018; Samaha et al., 2016; Samaha & Denison, 2020). In these studies, confidence was increased by experimentally increasing positive evidence (i.e., the extent of sensory evidence for the chosen decision outcome) while maintaining the ratio between positive and negative evidence. In other words, confidence judgements appeared to be based on the absolute magnitude of decision-congruent evidence but not decision-incongruent evidence.

These observations have led to development of several models of the processes that underlie decisions and confidence judgements (Maniscalco et al., 2021; Peters et al., 2017; Zylberberg et al., 2012). Central to all these models is the assumption that decisions and confidence judgements are based on two distinct sources of sensory evidence. While perceptual decisions are determined by relative evidence, confidence involves the heuristic use of only decision-congruent evidence.

From this viewpoint, the divergence between confidence and accuracy which we observed can be simply explained. For the decision process, relative evidence drove the decision outcome, with higher levels of absolute evidence leading to a decreased signal (due to Weber-scaling) and increased variability (due to input-dependent noise), within the decision process. As a result, decisions were, on average, faster and less accurate. Coincidently, due to our absolute evidence manipulation, the amount of decision-congruent evidence was boosted in conditions of high absolute evidence, leading to an increase in confidence, irrespective of these co-occurring accuracy and RT effects. In other words, the combined effects of Weber-scaling and a positive evidence bias can account for the effects of absolute evidence on decision accuracy, RTs, and confidence.

An alternative explanation for our divergent accuracy and confidence effects is that the increase in confidence we observed in high absolute evidence trials may have been a by-product of faster RTs. Certain models, distinct from those discussed above, suggest that confidence is partly determined by the time taken to come to a decision (e.

g., Kiani et al., 2014; Zylberberg et al., 2016). As faster decision times are often associated with more reliable sources of evidence and correct decisions, RTs may inform confidence judgements. Considering the current findings, this view suggests that the increase we observed in decision confidence following increases in absolute evidence might have been due to the co-occurring speed up in response times. However, our results showed that, while higher absolute evidence led to both faster RTs and increased confidence, and faster RTs co-occurred with higher confidence ratings, RT alone could not fully explain the effect of absolute evidence on confidence. Within similar RT ranges, increased absolute evidence was still associated with increased confidence. This provides evidence that the effect of absolute evidence on confidence was not simply due to speeding of RTs.

It should further be noted that the decision-congruent evidence hypothesis has been challenged recently by Khalvati, Kiani, and Rao (2021), who suggested an alternative explanation. Participants might use only a subset of the evidence provided in each trial to make a decision, while the data analysis includes all evidence available (including evidence not used by the participants). In particular, the subset of evidence used by the participant may contain a higher proportion of decision-congruent evidence than the unused subset (and therefore the full set), which might lead to an overestimate of the weight of decision-congruent evidence. Their model further assumes that the decision process involves continuous belief updating about the relevant environmental states (i.e., task parameters, which would correspond to luminance discrepancy in our task) based on sensory observations and prior beliefs. Where the decision maker terminates the decision process early, the small amount of evidence processed would lead to low choice accuracy but high confidence (in particular for incorrect decisions). Our pattern of faster response times fits with this explanation, which means that participants might have adapted a strategy of terminating the evidence accumulation process earlier when absolute evidence was higher. However, we also note again that response time effects did not fully explain the effect of absolute evidence on confidence, which calls into question whether the model by Khalvati et al. (2021) can fully explain our findings. Future research could address this issue by using different task instructions with either an emphasis on speed or on accuracy.

4.2. How does the effect of absolute evidence on confidence translate to change-of-mind decisions?

Our findings show that increasing absolute evidence magnitude, which led to slower change-of-mind decisions in an almost identical design (Turner, Feuerriegel, et al., 2021), also increases participants' sense of confidence in their decisions. Moreover, when confidence ratings were coded as a binary change-of-mind variable, increasing absolute evidence similarly led to reduced change-of-mind frequency (except for Experiment 2, correct trials). This supports that confidence and changes of mind in our design were indeed closely related and both depended on absolute evidence. This finding is consistent with the idea that subjective feelings of confidence in one's decision may affect subsequent change-of-mind decisions (van den Berg et al., 2016). More generally, this implies that heuristic biases in confidence judgements, such as those associated with absolute evidence magnitude / the positive evidence bias (e.g., Peters et al., 2017; Zylberberg et al., 2012), or motor-related confidence biases (e.g., Fleming & Daw, 2017; Gajdos et al., 2019; Turner, Angdias, et al., 2021) may play an essential role in determining the speed and likelihood of subsequent change-of-mind decisions.

The positive association we found between absolute evidence magnitude and decision confidence may be important to consider when attempting to model the dynamics of error correction and changes of mind. For example, it was recently shown that existing change-of-mind models, based solely on the ongoing accumulation of relative evidence after a decision, cannot easily account for the effects of absolute evidence on change-of-mind likelihood or RTs (Turner, Feuerriegel, et al., 2021). In particular, these models have difficulty capturing patterns of slower change-of-mind RTs in higher absolute evidence conditions. To completely account for decision and changes-of-mind behaviour, novel modelling assumptions may need to be explored. For example, one possibility is that the decision threshold for changing one's mind may depend, at least in part, on decision confidence. That is, decisions made with higher confidence may require more significant amounts of contradictory evidence to trigger a decision reversal (Turner, Feuerriegel, et al., 2021; see also van den Berg et al., 2016). Alternatively, it is possible that the weighting of post-decisional evidence may depend on decision confidence (Braun, Urai, & Donner, 2018; Rollwage et al., 2020). In other words, the dynamics of post-decisional processing may be fundamentally altered in a confidence-depended manner.

By demonstrating that decision confidence does vary across changes in absolute evidence magnitude, the current study provides empirical backing for exploring such assumptions. Suppose future theoretical work were to incorporate confidence-related biases in specific model parameters (such as shifts in the change-of-mind threshold) within existing computational frameworks, this may yield better accounts of the dynamics underlying change-of-mind decisions, and may help to integrate insights from recent theoretical accounts that capture effects of absolute evidence and positive evidence biases (e.g., Maniscalco et al., 2021).

4.3. Limitations

Our findings should be interpreted with the following caveats in mind. Firstly, one difference between our study and Turner, Feuerriegel, et al.'s (2021) study that limits the generalizability of our confidence findings to their change-of-mind results is that in their study, stimuli remained on the screen for a short duration after the initial response was submitted, while in our study the stimuli were terminated after the response. It should be noted that it is reasonable to assume that visual processing continues for around 300 ms after a stimulus is terminated (Resulaj et al., 2009), reducing the relevance of this issue. However, it is still possible that the difference in trial structure prompted temporally different computations. Change-of-mind decisions in Turner, Feuerriegel, et al. (2021) might have been more strongly based on late-arriving post-decisional evidence (e.g., Charles & Yeung, 2019), while our confidence judgements could not be based on such information. Future studies could investigate whether such different presentation conditions might prompt participants to use incoming sensory evidence differently to compute confidence. If a change-of-mind is the end-product of a confidence computation, this could affect how confidence is reported.

An avenue for future studies is therefore to further explore the temporal dynamics of the evidence signal that contribute to the formation of confidence judgements. For example, the use of reverse correlation approaches (e.g., Charles & Yeung, 2019; Turner, Feuerriegel, Hester, & Bode, 2022; Zylberberg et al., 2012) might be useful to relate the frame-by-frame fluctuations in available evidence to confidence judgements. Using such an approach in combination with a similar brightness judgement task as used here, Turner et al. (2022) have recently shown that changes of mind can already be predicted by random fluctuations of evidence contained in the very first frames of stimulus presentation. Given the close link between change-of-mind and confidence, our findings might suggest that such early evidence could also already contribute to confidence formation; however, this needs to be tested experimentally for which much larger trial numbers are required than were available in our study.

Another limitation is that we did not include an extensive range of luminance variability conditions in Experiment 2. This means that we cannot rule out the possibility that more extensive magnitude manipulations of luminance variability might exert more substantial effects on task performance and confidence measures in our task. For example, there are many examples of stimulus variability-related effects in other decision tasks (e.g., Desender et al., 2018; Zylberberg et al., 2014), although these are not always consistent in the direction of their effects. However, given that the observed effects appeared to be very small in our study, we believe that changes in perceived variability are an unlikely explanation for the much larger and consistent effects of absolute evidence in our experiments.

5. Conclusion

Using a luminance discrimination task, we showed that stronger absolute evidence led to reduced decision accuracy, faster RTs, and increased decision confidence. This finding parallels previous findings of higher absolute evidence leading to slower changes of mind (Turner, Feuerriegel, et al., 2021) and suggests that decision confidence may moderate the speed of decision reversals in perceptual judgment tasks. Our results are compatible with recent suggestions that confidence might be driven by decision-congruent evidence, in addition to the theory that faster response time contributes to higher confidence. Finally, by demonstrating that decision confidence varies with absolute evidence magnitude, the current work provides an empirical basis for future exploration of potential confidence-related changes in post-decisional information processing (e.g., shifts in the change-of-mind threshold or changes in the weighting of evidence).

CRediT authorship contribution statement

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

Declaration of Competing Interest

All authors declare no competing interests.

Acknowledgements

The authors thank all colleagues from the Decision Neuroscience Lab, and the Institute of Neuroscience and Medicine (INM-3), Cognitive Neuroscience, for discussions and support, and Megan Peters and Brian Maniscalco for valuable discussions. This work was funded by an Australian Research Council Discovery Project Grant (DP160103353) to S.B. and R.H. Y.H.K. and H.O. were funded by a Jülich University of Melbourne Postgraduate Academy (JUMPA) scholarship. G.R.F. and P. H.W. were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 431549029 – SFB 1451.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2022.105125.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.
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*, Article e12192. https://doi.org/10.7554/eLife.12192
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357
Braun, A., Urai, A. E., & Donner, T. H. (2018). Adaptive history biases result from confidence-weighted accumulation of past choices. *Journal of Neuroscience, 38*(10), 2418–2429.
Charles, L., & Yeung, N. (2019). Dynamic sources of evidence supporting confidence judgments and error detection. *Journal of Experimental Psychology: Human Perception and Performance, 45*(1), 39–52. https://doi.org/10.1037/xhp0000583
Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science, 29*(5), 761–778.
Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review, 124*(1), 91–114. https://doi.org/10.1037/rev0000045
Fleming, S. M., van der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience, 21*(4), 617–624.
Gajdos, T., Fleming, S. M., Saez Garcia, M., Weindel, G., & Davranche, K. (2019). Revealing subthreshold motor contributions to perceptual confidence. *Neuroscience of Consciousness, 2019*(1), niz001.
Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review, 96*(2), 267–314. https://doi.org/10.1037/0033-295x.96.2.267
Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*, 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038
Khalvati, K., Kiani, R., & Rao, R. P. (2021). Bayesian inference with incomplete knowledge explains perceptual confidence and its deviations from accuracy. *Nature Communications, 12*(1), 1–16.
Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron, 84*(6), 1329–1342. https://doi.org/10.1016/j.neuron.2014.12.015
Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1–16.
Koizumi, A., Maniscalco, B., & Lau, H. (2015). Does perceptual confidence facilitate cognitive control? *Attention, Perception, & Psychophysics, 77*(4), 1295–1306. https://doi.org/10.3758/s13414-015-0843-3
Maniscalco, B., Odegaard, B., Grimaldi, P., Cho, S. H., Basso, M. A., Lau, H., & Peters, M. A. (2021). Tuned inhibition in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. *PLoS Computational Biology, 17*(3), Article e1008779.
Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E., & Bahrami, B. (2017). The idiosyncratic nature of confidence. *Nature Human Behaviour, 1*(11), 810–818. https://doi.org/10.1038/s41562-017-0215-1
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. *Proceedings of the National Academy of Sciences of the United States of America, 115*(7), E1588–E1597. https://doi.org/10.1073/pnas.1711628115
Peters, M. A. K., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., … Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature Human Behaviour, 1*. https://doi.org/10.1038/s41562-017-0139
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59
Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347–356. https://doi.org/10.1111/1467-9280.00067
Ratcliff, R., Voskuilen, C., & Teodorescu, A. (2018). Modeling 2-alternative forced-choice tasks: Accounting for both magnitude and difference effects. *Cognitive Psychology, 103*, 1–22. https://doi.org/10.1016/j.cogpsych.2018.02.002
Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature, 461*(7261), 263–266. https://doi.org/10.1038/nature0827
Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature Communications, 11*(1), 1–11.
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. *Frontiers in Psychology, 7*, 851.
Samaha, J., & Denison, R. (2020). The positive evidence bias in perceptual confidence is not post-decisional. *bioRxiv*. https://doi.org/10.1101/2020.03.15.991513
Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a statistical computation in the human sense of confidence. *Neuron, 90*(3), 499–506.
Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 141–151. https://doi.org/10.1037/0096-1523.26.1.141
Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S.. afex: Analysis of Factorial Experiments. R package version 0.28-0. https://CRAN.R-project.org/package=afex.
Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27*(3), 161–168. https://doi.org/10.1016/j.tins.2004.01.006
Spence, M. L., Dux, P. E., & Arnold, D. H. (2016). Computations underlying confidence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance, 42*(5), 671–682.
Teodorescu, A. R., Moran, R., & Usher, M. (2016). Absolutely relative or relatively absolute: Violations of value invariance in human decision making. *Psychonomic Bulletin & Review, 23*(1), 22–38. https://doi.org/10.3758/s13423-015-0858-8
Turner, W., Angdias, R., Feuerriegel, D., Chong, T. T. J., Hester, R., & Bode, S. (2021). Perceptual decision confidence is sensitive to forgone physical effort expenditure. *Cognition, 207*, Article 104525.

Turner, W., Feuerriegel, D., Andrejevic, M., Hester, R., & Bode, S. (2021). Perceptual change-of-mind decisions are sensitive to absolute evidence magnitude. *Cognitive Psychology, 124*, Article 101358. https://doi.org/10.1016/j.cogpsych.2020.101358

Turner, W., Feuerriegel, D., Hester, R., & Bode, S. (2022). An initial 'snapshot' of sensory information biases the likelihood and speed of subsequent changes of mind. *PLoS Computational Biology, 18*(1), Article e1009738.

Ullsperger, M., Danielmeier, C., & Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. *Physiological Reviews, 94*(1), 35–79. https://doi.org/10.1152/physrev.00041.2012

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice Bias. *Nature Communications, 8*, 1–11.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society, B: Biological Sciences, 367*(1594), 1310–1321.

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience, 6*, 79. https://doi.org/10.3389/fnint.2012.00079

Zylberberg, A., Fetsch, C. R., & Shadlen, M. N. (2016). The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *eLife, 5*, Article e17688.

Zylberberg, A., Roelfsema, P. R., & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness and Cognition, 27*, 246–253.

2.2. Summary of findings

Chapter 2 presented two studies in which stronger absolute evidence in a brightness judgment task disrupted the typically observed positive association between accuracy and decision confidence. That is, increased absolute evidence strength impaired decision accuracy but inflated confidence ratings. Further, when confidence ratings were converted into a measure of changes of mind (by assuming that confidence ratings lower than "guessing" would have led to changes of mind), a consistent pattern was observed: Where confidence was inflated by stronger absolute evidence, the proportions of changes of mind trials were reduced. These consistent findings suggest a close relationship between confidence and changes of mind, and support the possibility that confidence could affect the likelihood that changes of mind occur, as suggested in previous studies (van den Berg, Anandalingam et al., 2016; Turner et al., 2021).

As the effect of absolute evidence on confidence was consistent the positive evidence bias (PEB) observed in previous studies (e.g., Samaha & Denison, 2022), it can be attributed to a confirmation bias that confidence was only based on decision-congruent evidence. Taken together with the absolute evidence effect on changes of mind, it further suggests that this confirmation bias could be carried over to the change-of-mind decisions, for example, through a positive effect of confidence on evidence accumulation rate (Navajas et al., 2016; Rollwage et al., 2020; Braun et al., 2018). Mechanisms underlying the effect of confidence on changes of mind are further discussed in Chapter 4.

In summary, the first part of the current project supports a link between confidence and changes of mind, and their relationship might be explained by the evidence accumulation framework. To further explore how confidence is related to evidence accumulation, the next chapter presents an EEG study that focused on the relationships between confidence and proposed ERP indexes of evidence accumulation.

2.3. Preview of Chapter 3

Theories of confidence judgment have previously linked confidence to pre- and post-decisional evidence accumulation (reviewed in Chapter 1). Under the evidence accumulation framework, previous studies have suggested that confidence could track the amount of sensory evidence during pre-decisional processes, or the amount of evidence accumulated after the decision (Desender et al., 2021). In parallel, the ERP components of CPP/P3 and Pe have been proposed to be indexes of pre- and post-decisional evidence accumulation processes (reviewed in Chapter 1). Taken together, confidence should also be related to these ERP indexes. However, their relationships with confidence are still being debated. Particularly, a recent study (Feuerriegel et al., 2021) has suggested that these relationships between confidence and Pe could be artefactual results of using a typical pre-response baseline correction procedure, which is confounded by CPP amplitudes. With pre-stimulus baseline correction, their study suggested that CPP and Pe showed different specificity to objective accuracy (i.e., their relationships with confidence depends on whether decisions were correct).

The EEG study presented in the next chapter therefore focuses on how the amplitudes of centro-parietal potential (CPP) and error positivity (Pe) could be related to confidence ratings in general, as well as how these relationships are potentially dependent on objective and subjective accuracy. Based on previous studies, it was predicted that CPP amplitudes would only be positively related to confidence in correct decisions, while Pe amplitudes would only be negatively related to confidence in erroneous decisions. Using the same brightness judgment task as in Studies 1 and 2, where participants were required to judge which of the two stimuli was brighter, it was found evidence supporting this specific association to objective accuracy. It further examined whether the two ERPs also showed specific association to subjective accuracy (i.e., whether the relationship between confidence and these ERPs were only observed when decisions were perceived as correct or incorrect). However, results suggest that only CPP but not Pe showed this specific association. These findings are discussed in relation to recent evidence accumulation models of performance monitoring.

Chapter 3. Event Related Potential Correlates of Decision Confidence Show Specific Associations to Objective and Subjective Accuracy

3.1. Introduction

Humans are capable of making metacognitive decision, that are decisions based on one's own cognitive process (Fleming & Frith, 2014; Sanders et al., 2016). For example, people were able to detect whether their own decisions were incorrect in error detection tasks, and assign graded levels of confidence to their decisions when they were asked to report how likely they believed their decisions were correct (van den Berg, Anandalingam et al., 2016; Yeung & Summerfield, 2012). This ability is not only important for immediate behavioral adjustment (Desender et al., 2019; Wessel, 2017), but has also been associated with important cognitive functions such as learning (Drugowitsch et al., 2019) and information seeking (Desenders et al., 2018).

Many studies have been using electroencephalography (EEG) signals to investigate metacognitive processes (Yeung & Summerfield, 2012). With high temporal resolution, EEG allows the monitoring of the time course of neural processes that contribute to metacognitive decision outcomes, such as decision confidence. Specifically, event-related potentials (ERPs) have been used to investigate confidence because several ERPs have been consistently related to constructs that are closely related to confidence, including as error awareness, evidence accumulation, visibility, and stimulus discriminability (error negativity [ERN/Ne; Falkenstein et al., 1991], error positivity [Pe; Nieuwenhuis et al., 2001], P3/centro-parietal potential [CPP; Kelly & O'Connell, 2013], and late positive potential [LPP; Tagliabue et al., 2019]). For example, ERP components that occur before a decision such as P3/CPP and LPP have been related to sensory evidence accumulation and visibility (Tagliabue et al., 2019). ERP components that occur after a decision such as ERN/Ne and Pe have been related to error detection and error awareness (Di Gregorio et al., 2018; Niessen et al., 2017). Among the components mentioned above, it has been suggested that CPP and Pe closely track confidence specifically (Desender et al., 2021; Rausch et al., 2020). Empirically, both CPP and Pe amplitudes have been found to be generally related to confidence across correct and error trials. These findings have led to the proposal that they are indexes of decision variable and metacognitive decision variable, which represent the states of evidence accumulated for choice and metacognitive decisions respectively (Desender et al., 2021; Rausch et al., 2020).

3.1.1. Centro-parietal potential (CPP)

CPP (or equivalently P3; Twomey et al., 2015) is a parietal positive component leading up to and peaking at the time of decision, typically measured as a stimulus-locked component 200 to 350 ms after the presentation of a stimulus, or as response-locked component -250 to -100 ms before a response (Kelly & O'Connell, 2013). It is present for both correct and incorrect decision although its amplitudes in incorrect trials is smaller (Kelly & O'Connell, 2013; O'Connell et al., 2012), and its amplitudes and build-up rate are positively related to discriminability of the stimuli (Kelly & O'Connell, 2013; O'connell et al., 2012). Different theories have been proposed regarding what the CPP reflects. Recent studies suggest that its build-up rate tracks the accumulation of sensory evidence during decision formation (Kelly & O'Connell, 2013; Kelly et al. 2021). It has also been theorized that CPP reflects confidence (Eimer & Mazza, 2005; Hillyard et al., 1971; Herding et al., 2019; Rausch et al., 2020) and subjective visibility of stimulus (Lamy et al., 2008; Sergent et al., 2005; Tagliabue et al., 2019).

3.1.2. Error positivity (Pe)

The Pe is a centro-parietal positive deflection that occurs 200 to 600 ms after decisions, with amplitudes typically larger for erroneous decisions than correct decisions (Desender et al., 2021; Nieuwenhuis et al., 2001; O'Connell et al., 2007; Orr & Carrasco, 2011). It was initially suggested to reflect error awareness as it shows stronger amplitudes for aware errors compared to unaware errors (Di Gregorio et al., 2018; Endrass et al., 2007), but it has also been suggested that it reflects confidence as its amplitudes appear to decrease with higher confidence (Boldt & Yeung, 2015; Desender et al., 2019). Particularly, in the study by Boldt and Yeung (2015), participants completed a numerosity judgment task with confidence ratings ranging from "certainly wrong" to "certainty correct". Pe amplitudes were found to be monotonically decreasing with confidence ratings. Given the similarity between the Pe and CPP in morphological and functional characteristics, it has been proposed that the Pe also reflect a similar evidence accumulation process, which however accumulates error evidence or sensory evidence after decision, and it could be based on different sources of information and contribute to different metacognitive decisions including error detection and confidence judgment (Desender et al., 2021; Di Gregorio, Maier, & Steinhauser, 2017; Wessel, 2018; Murphy et al., 2015). Others suggest that Pe does not reflect confidence but simply post-decisional evidence accumulation, which does not necessarily correlate with confidence (Rausch et al., 2020).

3.1.3. Specific associations with objective accuracy

While many studies have shown the above ERP components to be related to confidence and/or error awareness generally (Boldt & Yeung, 2015; Desender et al., 2021; Rausch et al., 2020), it remains unclear whether their relationships could be dependent on objective accuracy (i.e., whether such relationships are only observed in correct or error trials). This is because previous studies often investigated these relationships only in correct trials (in confidence judgment studies, e.g., Rausch et al., 2020) or error trials (in error detection studies, e.g., Murphy et al., 2015), and sometimes in a mix of correct and error trials (and thus masking any potential moderating role of objective accuracy; Boldt & Yeung, 2015; Desender et al., 2019). Also, some past studies showed that the relationship between the CPP and confidence was weaker when decisions were objectively incorrect (Eimer & Mazza, 2005; Hewig et al., 2011), and some showed that the CPP only reflected evidence accumulated for correct trials but not error trials (Herding et al., 2019). Similarly for the Pe, its relationship with confidence was also stronger in objectively incorrect trials (Hewig et al., 2011).

In addition, a recent study by Feuerriegel and colleagues (2022) showed that both the relationship between confidence and CPP amplitudes and the relationship between confidence and Pe amplitudes were subjected to methodological confounds. On the one hand, the relationship between Pe amplitudes and confidence could be confounded by the practice of baseline correction with a pre-response baseline (e.g., as used in Boldt & Yeung [2015]). This is because the pre-response baseline often overlaps with the CPP, which shows amplitude differences by confidence, and these differences could then be propagated into the Pe measure when this contaminated baseline is used for correction. On the other hand, the relationship between stimulus-locked CPP amplitudes and confidence could be confounded by the negative relationship between confidence and response times. This is because low confidence trials are often coupled with slower response times, which lead to more variable latency and thus lower averaged amplitudes. In their study, which used a pre-stimulus baseline and response-locked measures to avoid these confounds, it was found that while the CPP was positively related to confidence in both correct and incorrect decisions, the Pe was only negatively related to confidence in incorrect decisions. Additionally, with current source density (CSD; Kayser & Tenke, 2006) transformed data, they found a pre-response, frontal-central component that was only related to confidence in correct decisions, which was hypothesized to be the origin of the relationship between CPP and confidence.

These specific associations to objective accuracy could potentially be attributed to different evidence accumulation processes related to correct and erroneous decision outcomes and the proposed locus of confidence judgment (Baranski & Petrusic, 1998; Yeung & Summerfield, 2012). Particularly, in an evidence accumulation framework where the CPP reflects accumulated sensory evidence (O'Connell & Kelly, 2021), it could be that evidence is effectively accumulated in correct trials but not in error trials (due to more noise in error trials). Assuming that confidence depends on the evidence accumulated during decision formation (i.e., a decision locus of confidence judgment, e.g., Kiani & Shadlen, 2009; Vickers, 1979), it follows that only in correct trials the CPP could serve as an index of confidence. In contrast, the Pe reflects error evidence accumulation, which is more likely to be effective and correspond to actual error commission likelihood in error trials than in correct trials. If confidence depends on the error evidence accumulated after the decision (i.e., a post-decisional locus of confidence judgment, e.g., Desender et al., 2021; Moran et al., 2015), it follows that only in error trials the Pe is related to confidence.

3.1.4. Specific association with subjective accuracy

In addition to objective accuracy, a decision can also be defined as subjectively correct or incorrect (i.e., whether a decision considered by the participant to be correct or incorrect). This is the case when the option to signal an erroneous decision is available, as in error detection studies, and in confidence studies where a full-range confidence scale (ranging from "surely incorrect" to "surely correct") allows participants to report how certain they are that their decisions were correct or incorrect.

Subjective accuracy in this sense has only been investigated in a few studies as only recently error detection and confidence judgment were proposed to be closely related and measured using the same scale (Boldt & Yeung, 2015; Charles & Yeung, 2019; Yeung & Summerfield, 2012), and most previous studies focused on only side of the scale (measuring certainty of being correct or certainty of being incorrect).

To provide a unifying account for both error awareness and confidence judgment, it has been assumed that decisions that are subjectively correct and incorrect lie on the same continuum, and that ERP correlates of confidence such as Pe amplitudes are monotonically decreasing with confidence levels on the full-range scale (Boldt & Yeung, 2015; Charles & Yeung, 2019; Desender et al., 2021; Yeung & Summerfield, 2012). However, empirically this monotonic relationship was only reported by Yeung and Summerfield (2012) and was not observed in other studies (Feuerriegel et al., 2022; Hewig et al., 2011). Additionally, it was also suggested that the two ends of the proposed continuum could be qualitatively different (Charles et al., 2013; Feuerriegel et al., 2022; Yeung & Summerfield, 2012).

If the specific associations with objective accuracy mentioned above could be explained by different evidence accumulation in correct and erroneous decisions, it is then possible that the relationships between the CPP and confidence, and between the Pe and confidence are also dependent on subjective accuracy, as subjective accuracy levels should also reflect different levels of accumulated sensory evidence and error evidence (Desender et al., 2021). In other words, only evidence in both subjectively and objectively correct decisions might scale with certainty of being correct (and hence the relationship between CPP and certainty of being correct). Decisions that are objectively correct but made with a sense of error, should not show the relationship between CPP and certainty of being incorrect (because of the limited amount of sensory evidence).

Likewise, only evidence in both subjectively and objectively incorrect scales certainty of being incorrect. Decisions that are objectively incorrect but made with a sense of correctness, should not show the relationship between Pe and certainty of being incorrect (because of the limited amount of error evidence). In other words, it is possible that CPP does not simply reflect confidence levels in correct trials but reflects certainty of being correct specifically. Similarly, Pe might not simply reflect confidence levels in error trials, but certainty of being incorrect specifically.

Preliminary support for these hypothesized relationships was reported by Feuerriegel and colleagues (2022), who showed that CPP amplitudes correlated positively with certainty of being correct and Pe amplitudes positively correlated with certainty of being incorrect. However, as they included both correct and error trials in their analyses, it is unclear whether such specific associations would be found within correct and error trials respectively.

3.1.5. The current study

The current study aimed to examine (1) whether CPP amplitudes are only related to confidence in correct trials but not in error trials, and (2) whether Pe amplitudes are only related to confidence in error trials but not correct trials. A speeded brightness judgment task used in a previous study by Ko et al. (2022) was used in this study. In this task, participants were required to make a comparative judgment based on the brightness of two grayscale, flickering squares, and then report confidence on a full-range confidence ratings scale. This task was used as it allowed the manipulation of sensory evidence and could induce a wide range of confidence ratings in the previous study (Ko et al., 2022).

We first predicted that when confidence was measured using a full-range scale, CPP amplitudes should only be positively related to confidence in correct trials, and Pe amplitudes should only be negatively related to confidence in error trials. Secondly, as we found evidence for the hypothesised relationships between the two components and confidence (i.e., specific associations with objective accuracy), we examined further whether CPP and Pe amplitudes reflected the graded certainty of being correct and incorrect (i.e., specific association with subjective accuracy). Specifically, we predicted that in correct trials, CPP would be positively related to certainty of being correct but not related to certainty of being incorrect. On the other hand, in error trials, Pe would be negatively related to graded certainty of being incorrect but not graded certainty of being correct.

3.2. Methods

3.2.1. Participants

We recruited 36 university student volunteers with normal or corrected-to-normal vision. We excluded two participants who failed to report confidence in more than 20% of all trials, three participants for overall accuracy lower than 55%, one for reporting the same confidence level in more than 90% of trials where confidence was reported (using the same exclusion criteria as in Ko et al., 2022) and two for excessively noisy EEG data. The final sample included 28 participants (mean age 26, age SD = 6, age range = 18-39, 16 females). This study was approved by the Human Ethics Committee of the Melbourne School of Psychological Sciences (ID 1954641.2).

3.2.2. Task and stimuli

Participants completed a speeded luminance discrimination task in which two flickering squares were presented, and they were required to choose the stimulus that appeared brighter to them on average. The luminance level of the two flickering squares changed with each frame refresh (i.e., every 13.3 ms at 75 Hz). At each screen refresh, the luminance levels of the stimuli were randomly sampled from a pair of truncated normal distributions, which had mean RGB values used in a previous study (Ko et al., 2022; specified in Figure 3.1B. As one distribution had a higher mean than the other, one stimulus appeared on average brighter than the other. The truncated distributions had the same parameters (standard deviation of 25.5, truncated at ±1 SD from their means) as in Ratcliff et al. (2018) and Ko et al. (2022). The difference between the means were varied as a manipulation of relative sensory evidence strength, and the overall mean values were varied as a manipulation of absolute sensory evidence strength. Both stimuli had a size of 70 × 70 pixels and were positioned with equal distance from the center of the screen, separated from each other by 180 pixels. Stimuli positions were counterbalanced such that the brighter square was on the left in half of the trials, and one the right in the other half of the trials. The order of the trials was also randomized. Participants were instructed to respond as quickly as possible.

After the response for brightness judgment, participants were required to rate their confidence level in their brightness judgment using a 7-point scale ranging from “surely incorrect” (1) to “surely correct” (7), with a midpoint rating (4) that indicated that they were unsure whether the brightness judgment was correct or incorrect (i.e., they felt they were guessing). They were also instructed to respond as quickly as possible for this confidence rating.

The effects of these manipulations on accuracy, RT and confidence were investigated in a previous study (Ko et al., 2022), and they were used again in the current study to induce changes in confidence. The same task was used in the previous study and only the timing parameters were adjusted in this study for EEG recording (Figure 3.1A). Additionally, the fixation dot color also changed from white to red before stimulus presentation to signal the onset of stimuli, and it was presented together with the stimuli to keep participants focused on the center of the screen.

Before the main experiment, participants completed a training block of 36 trials without making confidence ratings. Instead, they received performance feedback after each trial to familiarize themselves with the brightness judgment task. Subsequently, they completed another training block of 36 trials where both brightness and confidence judgment were required. No performance feedback was given in this training and a confidence rating scale was presented on the screen for 1,500 ms. Participants were instructed that during the main experiment, the visual presentation of the scale would be removed, and the confidence judgment would only be prompted by the word "confidence".

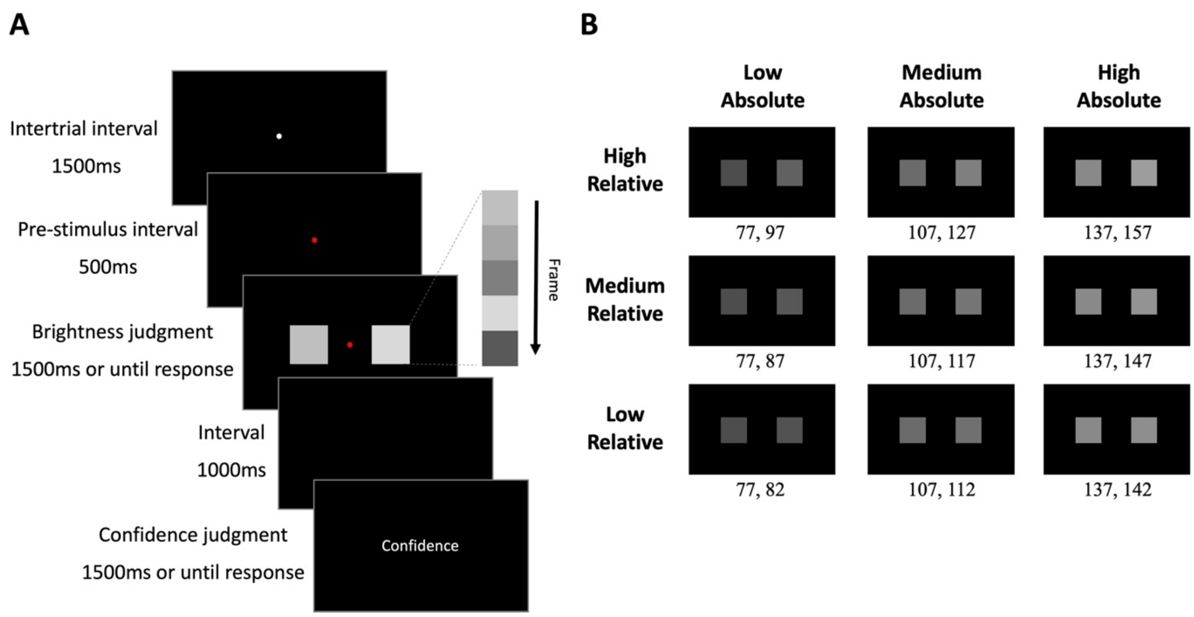

Figure 3.1. Task paradigm and stimuli. (A) Paradigm. In each trial, two flickering square stimuli of different average luminance were presented. Each square changed in luminance with each frame. Participants were required to select the stimulus that appeared brighter on average and subsequently reported their decision confidence using a 7-point scale while the word "confidence" was presented on the screen. (B) Illustration of average luminance values (in RGB) for stimuli for all experimental conditions of Experiment 1. Luminance values were randomly sampled from normal distributions truncated one standard deviation around pre-defined means. The standard deviation of the grayscale values for all distributions was 25.5.

3.2.3. Experimental procedures

After giving written consent and receiving task instructions, participants were seated in a dark testing booth, 70 cm from a computer monitor. They completed task training while the experimenter stayed in the testing booth. This ensured that participants understood the task instructions correctly. Participants then completed the main experiment alone. They were reimbursed 30 AUD and debriefed by the experimenter after the experiment.

3.2.3.1. Experiment

Each trial of the experiment started with an intertrial interval with white fixation dot for 1500 ms. The white fixation dot then turned to red in the pre-stimulus interval lasting 500 ms, which was to signal the upcoming stimuli. After that, the flickering squares were presented, and the brightness judgment was required. Participants were required to press either the left of right key on the response pad using left and right index fingers, corresponding to the stimulus that they perceived to be brighter. The stimuli were presented for a maximum of 1,500 ms and disappeared immediately after a response. Subsequently, after an interval of 1,000 ms with a blank screen, a confidence judgment was required. Participants were required to rate their confidence level when prompted by the word "confidence". The scale had the same properties as during training, and participants were required to press one of the seven keys on the response pad to indicate their confidence level. No confidence rating was required if the brightness judgment was "too slow" (>1,500 ms RT) or "too quick" (<250 ms RT). In this case, only the respective timing feedback was presented for 1,500 ms, and then the next trial began. The experiment comprised 1,008 experimental trials equally allocated across 14 blocks. Each block was followed by a self-terminated rest period. An equal number of trials from all conditions were randomly interleaved within each block.

3.2.3.2. Apparatus

A Sony Trinitron Multiscan G420 CRT Monitor (resolution 1280 x 1024 pixels; frame rate 75 Hz) that was gamma-corrected with a ColorCAL MKII Colorimeter (Cambridge Research Systems) was used. The task was programmed in MATLAB R2018b (The Mathworks) using Psychtoolbox-3 (Brainard, 1997; Kleiner et al., 2007). Participants responded using a seven-button Cedrus response pad (RB-740, Cedrus Corporation).

3.2.4. EEG data processing

EEG data were recorded using a Biosemi Active Two system, with 64 electrodes and a sampling rate of 512Hz. Six external electrodes were additionally included: two behind the left and right mastoids (for referencing), two at the outer canthi, and one above and one below the right eye (for measuring eye movement).

EEG data pre-processing was conducted using EEGLab (v2019_1; Delorme & Makeig, 2004). Raw data were first re-referenced to the average of linked mastoids. Excessively noisy sections of data were excluded based on visual inspection. Excessively noisy channels were identified based on visual inspection and were excluded from further processing and were later interpolated. Independent Component Analysis (ICA) was then used to identify and remove artefacts. To obtain clearer independent components, ICA was applied to a copy of the dataset that was high-pass filtered at 1Hz and low-pass filtered at 30Hz. The results of the ICA were then copied to a dataset that was high-pass filtered at 0.1Hz and low-pass filtered at 30Hz, and subsequent processing was based on this dataset (as done by Feuerriegel et al., 2018). Independent components identified as generated by eye blinks, eye movements, and muscle movement (based on visual inspection and the IClabel algorithm; Pion-Tonachini et al., 2019) were then removed. Channels that were identified earlier as noisy were then interpolated using spherical spline interpolation (median number of channels interpolated = 3, range = 0 - 10).

3.2.4.1. Epoching and baseline correction

To avoid two main methodological confounds associated with the stimulus-locked measure of CPP, and pre-response baseline correction for the Pe, we followed Feuerriegel et al. (2022) to use response-locked measures and pre-stimulus baseline for both ERP components. The data were first segmented using the time window of -100 to 2200 ms relative to stimulus onset, and baseline corrected to the 100 ms pre-stimulus interval before stimulus onset. From this data, response-locked epochs were derived using the time window of -500 to 700 ms relative to choice response. Epochs with amplitudes exceeding ±150 μV at any scalp channel were excluded from further analyses. Across participants, the number of epochs included in the final analyses ranged from 450 to 984, with a median of 808. Single trial amplitudes of the components of interest were then measured according to the following definitions. Based on previous work (Steinemann et al., 2018; Feuerriegel et al., 2021), we measured CPP mean amplitudes as the average amplitudes from -130 to -70ms relative to response at electrode Pz. We measured Pe mean amplitudes as the average amplitudes from 300 to 400 ms relative to the response at electrode Pz. The Pe time window was defined as the time range where the amplitude difference between correct and error trials emerged and remained prominent. The resulting time window was highly similar to the ones used in previous studies (Gehring et al., 2012; Falkenstein et al., 2000).

3.2.5. Behavioural data analyses

For analyses of task performance and confidence ratings, we first examined the effects on relative evidence and absolute evidence on accuracy, RT (correct), RT (error), confidence (correct), confidence (error). This was done to test whether the pattern of results established in our earlier work (Ko et al., 2022) were replicated here and to confirm that the experimental manipulations led to the desired spread of choice behaviour, giving rise to a wide spectrum of subsequent confidence ratings. We fitted generalised linear mixed models with relative evidence, absolute evidence, their interaction and a random intercept for participants to predict these dependent variables in separate models. This approach was identical to that of Ko et al. (2022). We used binomial distributions with a logit function for accuracy, gamma distribution with an identity function for RTs, and normal distribution with an identity function for confidence*. Additionally, we also investigated how confidence was related to accuracy and RTs.*

3.2.6. Analyses of ERP component amplitudes

For ERP component amplitudes analyses, we fitted linear mixed models with confidence as the predictor and a random intercept for participants to predict single-trial amplitudes of each component of interest (i.e., the CPP and Pe components). Random slopes were not included because not all models with random slopes converged, and therefore omitting random slopes allowed for using the same model structure across analyses. Also, to examine whether specific association with objective accuracy reported by Feuerriegel et al. (2021) could be observed, analyses were run separately for trials with correct responses and errors.

Our first prediction was that CPP amplitudes would only be positively related to confidence in correct trials, and that Pe amplitudes would only be negatively related to confidence in error trials. We tested this hypothesis using models where the amplitudes of the respective components were predicted by the full-range confidence ratings (from “surely incorrect” to “surely correct”), similar to the approach in Feuerriegel et al.

(2021). When significant relationships were found, trend analyses were conducted to examine how the components changed between levels of confidence ratings (i.e., whether a linear trend was observed).

To investigate whether the amplitudes of the CPP and Pe differed by binary subjective accuracy, we further coded the trials based on confidence ratings as subjectively correct trials (collapsed across trials with confidence ratings > 4), subjectively incorrect trials (collapsed across trials with confidence ratings < 4). In addition, we specified a third category as guessing trials in which participants were unsure about their confidence (confidence ratings = 4). These three categories were then compared. This analysis allowed us to detect any difference due to subjective accuracy, which might not be captured by the analyses with full-range confidence ratings due to small number of trials in some confidence categories. The comparisons among three categories also allowed our results to be comparison with other studies that used a binary or trichotomized scales. When significant effects were found, post-hoc pairwise comparisons with Holm-Bonferroni correction were conducted.

Our second prediction was that CPP amplitudes would only be positively related to certainty of being correct in correct trials, and that Pe amplitudes would only be negatively related to certainty of being incorrect in error trials To examine whether certainty of being correct/incorrect predicted CPP and Pe amplitudes, we fitted models with the same structure (confidence predicting ERP amplitudes, separately for correct and error trials), but within subsets of trials that were rated as subjectively incorrect (confidence < 4) or subjectively correct (confidence > 4). Guessing trials were excluded from these analyses as they might include trials with no clear error awareness or confidence judgment.

All analyses were conducted in R (version 4.0.1). GLMMs were fitted using the lme4 package (version 1.1; Bates et al., 2015), statistical significance of each effect was determined by likelihood ratio tests conducted using the afex package (version 0.28; Singmann et al., 2017) in a stepwise forward approach, where each effect of interest was enter into the model and models before and after an effect was included were compared. Post hoc comparisons and trend analysis were conducted using the emmeans package (version 1.4.8; Lenth et al., 2020). Complete statistical results including likelihood ratio test results for all effects and regression coefficients of the full models are reported in Appendix A. The effect of confidence on alternative ERP measures, and the effects of relative and absolute evidence on ERP measures were also examined and reported in Appendices B and C. As in Feuerriegel et al. (2022), we also explored whether specific associations to subjective accuracy would be observed when objectively correct and error trials were pooled, however, as the results are highly similar to the results reported here, they are reported in Appendix D.

3.3. Results

3.3.1. Behavioural data analysis

3.3.1.1. Effects of relative and absolute evidence on accuracy, RTs, and confidence

We first examine the behavioural data as in Ko et al. *(2022), to ensure that the same patterns were found in terms of* how the manipulation of relative and absolute evidence affected accuracy, RTs, and confidence. Consistent with the previous study, accuracy was increased by stronger relative evidence and reduced by stronger absolute evidence (Figure 3.2A; *ps* $< .001$ for all main effects and interaction). RTs for both correct and error trials were faster when absolute evidence was increased and when relative evidence was increased (*ps* $< .01$ for all main effects and interactions except that for error trials RT, where relative evidence did not have a significant effect; Figure 3.2B-C, where 3.2C shows relative did not reduce RT in error trials). Confidence was increased with stronger relative evidence and absolute evidence in correct trials (*ps* < .001; Figure 3.3), and spread across a wide range as expected.

Except for the relative evidence effect on RT in error trials, all the effects of relative evidence and absolute evidence on accuracy, RT (correct), RT (error), confidence (correct), confidence (error) were replicated. Statistical results are reported in the Appendix A.

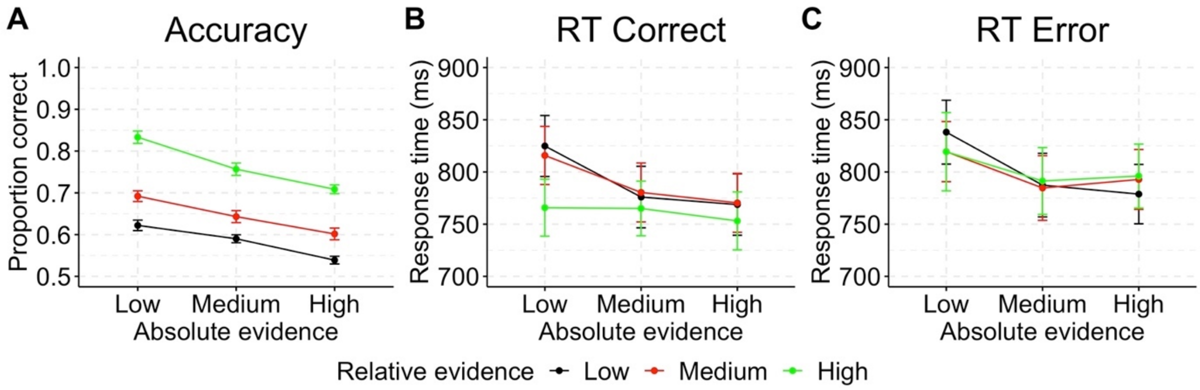

Figure 3.2. Accuracy and response time (RT) for different combinations of relative and absolute evidence levels. (A) Decision accuracy (average proportion correct) in each condition. (B) Mean RTs for correct trials. (C) Mean RTs for error trials. Error bars represent standard errors of the mean (SEM).

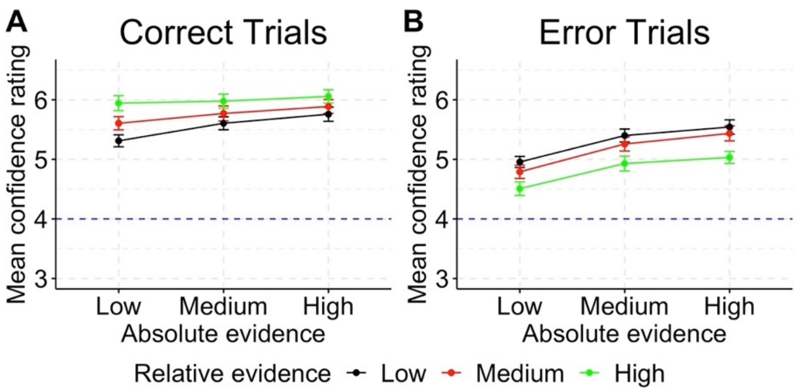

Figure 3.3. Mean confidence ratings for different combinations of relative and absolute evidence levels. (A) Correct trials. (B) Error trials. Confidence ratings were measured

on a scale ranging from 1 ("surely incorrect") to 7 (surely correct). The dotted line indicates the mid-point of the scale. Error bars represent SEM.

3.3.1.2. Accuracy and RTs across confidence levels, and confidence distributions

We additionally examined how confidence ratings varied with accuracy and RT, in order to confirm that participants used the rating scale to meaningfully report the confidence level. Overall mean proportion correct was 66.53% (SE = 0.91%) and mean response time was 784 ms (SE = 28ms). Figure 3.4 shows error rates in each confidence level, confidence distributions for correct and error trials, and response times in each confidence level. Figure 3.4A shows that confidence increased as error rates decreased, suggesting that participants reported confidence in a way that correlated with their objective decision accuracy. Figure 3.4B shows that the confidence distributions for both correct and error trials were negatively skewed, showing that participants tended to report higher confidence, although previous studies more commonly reported positively skewed distribution for error trials (e.g., Boldt & Yeung, 2015). Figure 3.4C shows that response times were negatively correlated with certainty of being correct (when confidence > 4) in both correct and error trials (mean r for correct trials = -.201; mean r for error trials = -.167 ps < .001), but not related to certainty of being incorrect (mean r for correct trials = -.066, p = .254; mean r for error trials = .022 p = .617). These patterns show that while accuracy increased with higher confidence levels in general, RT showed different relationships for certainty of being correct and certainty of being incorrect. Full statistical results are reported in Tables A1 - A10.

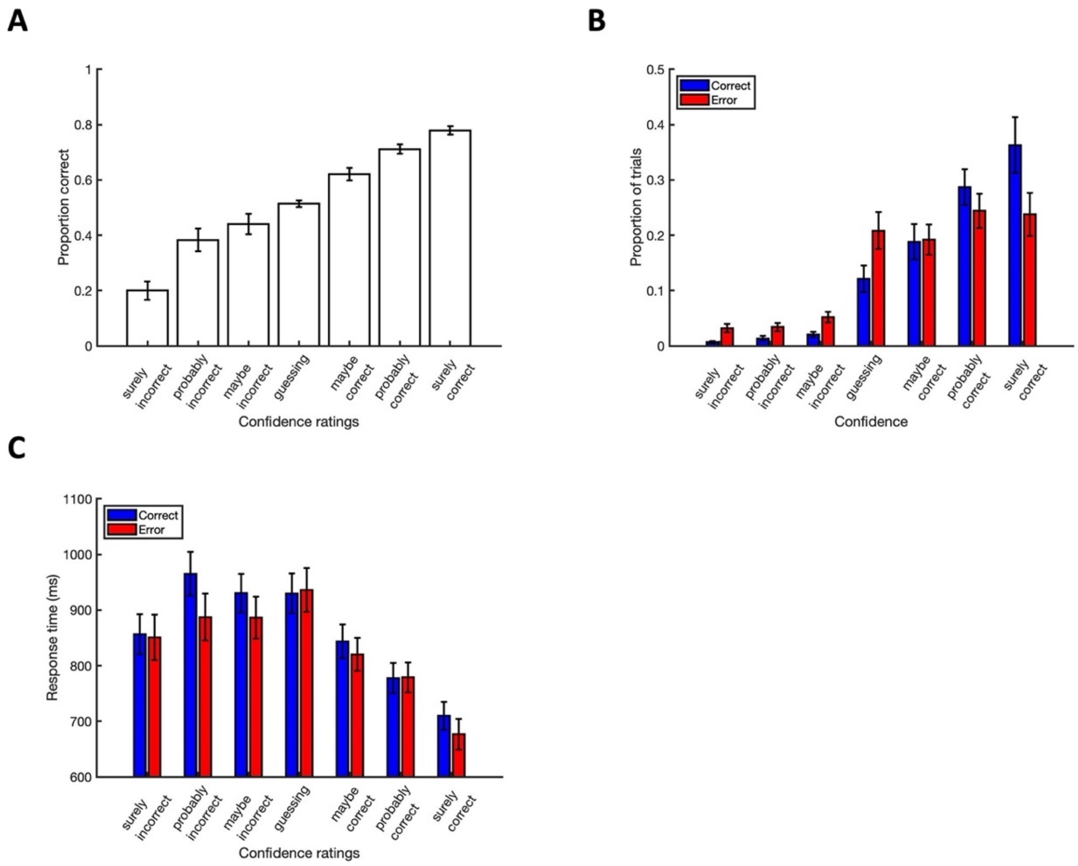

Figure 3.4. Primary task performance by confidence ratings. (A) Proportion correct across confidence levels. (B) Confidence distributions for correct and error trials. (C) Mean RT by confidence levels between correct and error trials. Error bars represent SEM.

3.3.2. EEG data analysis

3.3.2.1. CPP and Pe by correct ad error trials

Before the main analyses related to confidence, we first examined whether accuracy predicted the amplitudes of CPP and Pe as the accuracy effects were reported in previous studies (e.g., Desender et al., 2021). ERP waveforms of CPP and Pe for correct and error trials are presented in Figure 3.5. The difference in CPP amplitudes for correct compared to error trials was not statistically significant ($\chi^2 = 2.62$, $p = .106$), whereas Pe amplitudes in error trials were significantly larger than in correct trials ($\chi^2 = 3.94$, $p = .047$). The absence of accuracy effect for CPP could be due to the fact that CPP reflects confidence rather than error commission, and confidence in error trials was similar to confidence in correct trials (as shown by the negatively skewed distribution in Figure 3.4B). Full statistical results are reported in Tables A11 - A12.

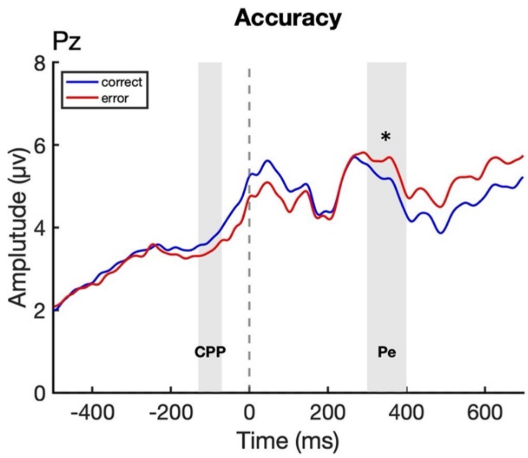

Figure 3.5. Group mean ERP waveforms for correct and error trials. Shaded areas show the time windows of the CPP (-130 to 70 ms) and the Pe (300 to 400 ms). Pe amplitudes were larger for error than correct trials.
Note. **p* <.05 ***p* <.01 ****p* <.001

3.3.2.2. Specific association to objective accuracy

In this main analysis, we examined how the CPP and Pe amplitudes were related to confidence. We first tested whether CPP amplitudes would only be positively related to confidence in correct trials, with the regression model where CPP amplitudes were predicted by full-range confidence ratings. Figure 3.7 shows the waveforms by accuracy and all confidence levels. In line with our hypothesis, the model for correct trials showed that CPP was only positively related to confidence (χ^2 = 32.52, p < .001) but the model for error trials did show a significant effect (χ^2 = 4.96, p = .549). Trend analysis showed a significant linear trend for correct trials (p = .006), suggesting that CPP amplitudes increased with higher confidence (see Figure 3.7A). This finding was however contrary to the finding by Feuerriegel et al. (2021), which showed that CPP was positively related to both confidence in correct and error trials.

We then tested whether Pe amplitudes would only be negatively related to confidence in error trials. Also as predicted, the model for error trials showed that Pe was negatively related to confidence (χ^2 = 32.18, p <.001), and trend analysis showed a significant linear trend (p <. 001), consistent with the waveforms and amplitudes plotted in Figure 3.7B. Unexpectedly, Pe also varied with confidence in correct trials (χ^2 = 18.98, p = .004). However, trend analysis showed no linear trend (p = .915) but a quartic trend (p = .023). The waveforms are shown in Figure 3.7A. Pe amplitudes did not increase linearly with confidence, instead, it fluctuated across confidence levels and the guessing level showed the largest amplitudes. Other than this unexpected finding, we found that CPP and Pe amplitudes were indeed only related to confidence in correct and error trials respectively. Full statistical results are reported in Tables A13 - A16.

3.3.2.3. Binary subjective accuracy and guessing

After testing how CPP and Pe were related to full-range confidence (as done by Feuerriegel et al., 2022), we then conducted additional analyses with binary categories of subjective accuracy, which allowed us to examine whether subjectively correct and incorrect trials overall differed in these two components. We coded as subjectively incorrect (collapsed across ratings < 4), guessing (ratings = 4), and subjectively correct (collapsed across ratings > 4), and then repeated the analyses separately for objectively correct and error trials. Figure 3.6 showed the waveforms by subjectively correct, subjective incorrect, and guessing trials. Only within correct trials, subjectively correct trials had higher CPP amplitudes than subjectively incorrect trials (χ^2 = 13.51, p <.001). Conversely, only within error trials, subjectively incorrect trials had higher Pe amplitudes than subjectively correct trials (χ^2 = 17.99, p <.001). This is consistent with the above analysis with full-range confidence ratings.

However, there was no significant difference in Pe amplitudes for objectively correct trials (χ^2 = 0.89, p = .345). This suggests that the relationship between the Pe and full-range confidence ratings in correct trials does not imply that the Pe differed between subjectively correct and incorrect trials. Instead, that relationship was likely due to guessing trials having larger amplitudes. This was confirmed by repeating the same analysis with guessing trials included. When comparing guessing trials, subjectively correct trials, and subjectively incorrect trials, confidence again had a significant effect on Pe in objectively correct trials ($\chi^2 = 9.89$, $p = .007$) and pairwise comparison showed that guessing trials indeed had larger amplitudes than subjectively correct trials ($p = .007$; but not subjectively incorrect trials, $p = .623$).

For other analyses with binary subjective accuracy as the predictor, inclusion of guessing trials did not change the patterns of statistical significance. Post-hoc pairwise comparisons showed that, for the Pe in error trials, guessing trials showed similar amplitudes as subjectively incorrect trials, and they were both different from subjectively correct trials ($ps < .01$). However, for the CPP in correct trials, guessing trials showed similar amplitudes as subjectively incorrect trials, and they were both different from subjectively correct trials ($ps < .001$). For CPP in error trials, no effect of confidence was found, as in other analyses. These results suggest that guessing trials did not separate clearly from both subjectively correct and subjectively incorrect trials, potentially due to the fact that guessing could include a mix of trials where participants inclined to report certainty of being correct or incorrect (but reported guessing). Full statistical results are reported in Tables A17 – A24.

In summary, by trichotomizing the confidence scale, we found results that were consistent with the analysis with the full-range scale. It also clarified that guessing trials contributed to the relationship between confidence and Pe in correct trials, but Pe did not differentiate subjectively correct and incorrect trials.

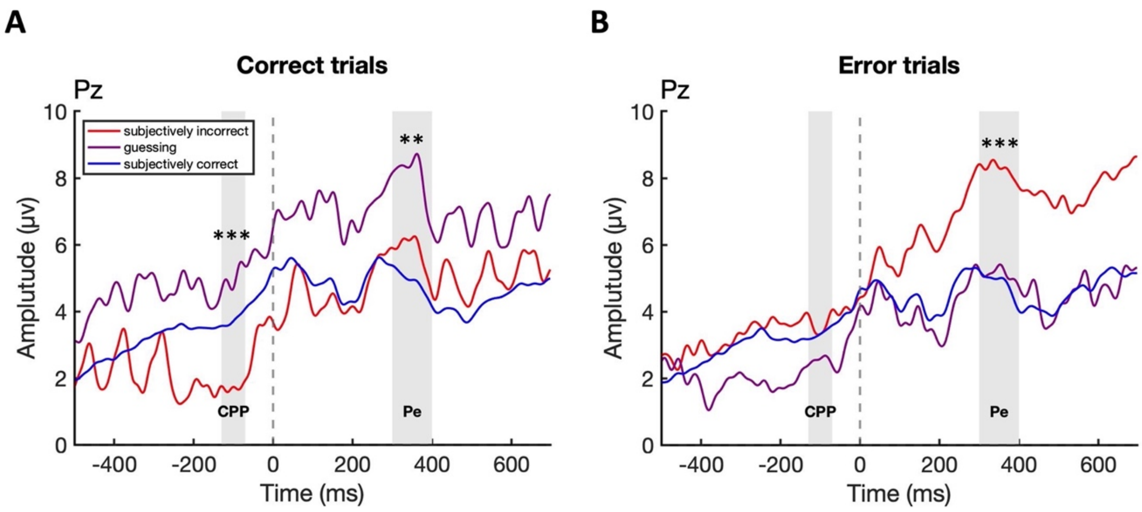

Figure 3.6. (A-B) Group mean ERP waveforms by subjectively correct, subjective incorrect, and guessing trials, separated by correct and error trials. Shaded areas show the time windows of the CPP (-130 to 70 ms) and the Pe (300 to 400 ms). CPP amplitudes were larger for correct trials rated as subjectively correct than correct trials rated as subjectively incorrect. Pe amplitudes were larger for error trials rated as subjectively incorrect than error trials rated as subjectively correct. Pe amplitudes were also larger for correct trials rated as guessing than correct trials rated as correct.
Note. **p* <.05 ***p* <.01 ****p* <.001

3.3.2.4. Specific associations to subjective accuracy

We also predicted that CPP amplitudes would only be positively related certainty of being correct and Pe amplitudes would only be negatively related to certainty of being incorrect. To test these predictions, we ran the same analyses as above but limiting the scale ranges to ratings of 1-3 (certainty of being incorrect) and ratings of 5-7 (certainty of being correct).

For the CPP in correct trials, when only trials with high confidence ratings of 5-7 were included, a significant positive, linear effect of certainty of being correct (χ^2 = 24.04, p < .001) was found. This relationship was however absent when the range was limited to low ratings of 1-3 (χ^2 = 0.95, p = .621), suggesting that CPP did not scale with graded certainty of being incorrect. For CPP in error trials, as in the analyses above, CPP amplitudes were not related to confidence in error trials, regardless of the scale ranges (high range of 5-7: χ^2 = 0.00 p =.998; low range of 1-3: χ^2 = 0.16, p =.923). These results are consistent with our prediction that CPP should positively related to certainty of being correct but not certainty of being incorrect, and this is true for correct decision only.

Unexpectedly, for Pe in correct trials, when only trials with high confidence ratings of 5-7 were included, a significant effect of certainty of being correct was found (χ^2 = 7.05, p = .029). This effect, however, did not follow a linear trend (p = .122) as a follow-up trend analysis showed, but was driven by the fact that "surely correct" had a higher amplitudes than "probably correct" (p = .022). It should also be noted that this direction of effect was inconsistent with the negative relationship between confidence and Pe as commonly reported in previous studies (Boldt & Yeung, 2015; Desender et al., 2021). When only correct trials with low confidence ratings of 1-3 were included, there was no significant effect of certainty of being incorrect (χ^2 =1.90, p = .387). For Pe in error trials, neither certainty of being correct (χ^2 = 3.66, p =.161) or certainty of being incorrect (χ^2 = 0.97, p = .615) had an effect. Contrary to our prediction that Pe amplitudes could be positively related to certainty of being incorrect in error trials only, we did not find Pe amplitudes in error trials related to certainty of either direction. Full statistical results are reported in Tables A25 – A32.

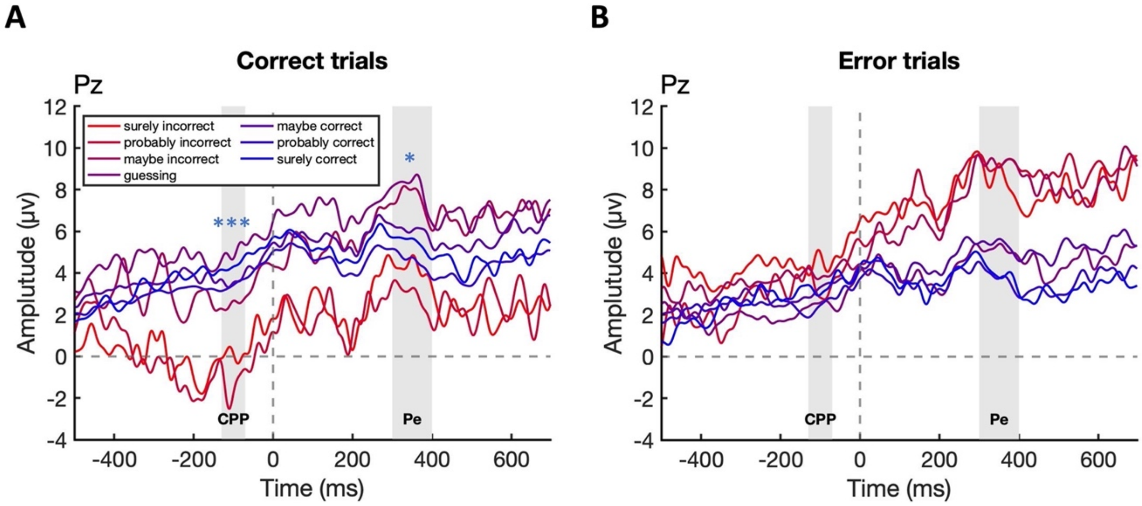

Figure 3.7. (A-B) Group mean of stimulus-locked ERP waveforms by all confidence levels separated by correct and error trials. Shaded areas show the time windows of the CPP (-130 to 70 ms) and the Pe (300 to 400 ms). Only CPP amplitudes in correct trials showed a positive relationship with certainty of being correct. The relationship between Pe amplitudes in correct trials and certainty of being correct was significant but did not follow a linear trend. *Note*. **p* <.05 ***p* <.01 ****p* <.001. Blue asterisks indicate the p value for the relationship with certainty of being correct.

3.4. Discussion

In the current study, we asked whether CPP and Pe are related to confidence only in correct and incorrect decisions respectively, that is, whether the relationships between CPP and Pe and confidence show specific associations to objective decision accuracy. We further asked whether their relationships with confidence were only driven by certainty of being correct and certainty of being incorrect respectively, that is, whether these relationships show specific associations to subjective decision accuracy. To answer these two questions, we measured ERPs during a luminance judgment task with confidence ratings, and we found support for the hypothesized specific associations to objective accuracy, but only partial evidence for the hypothesized specific associations to subjective accuracy (for the CPP but not the Pe).

3.4.1. Specific associations with objective accuracy

As reported by Feuerriegel et al. (2022), we replicated that confidence ratings were related to Pe amplitudes only in error trials and related to CPP amplitudes only in correct trials. These results were supported by our analyses with full-range confidence ratings, and when the confidence scale was trichotomized.

For the CPP, this pattern of results is consistent with previous findings that confidence in correct trials was positively related to CPP amplitudes (Feuerriegel et al., 2022; Herding et al., 2019; Rausch et al., 2020). As confidence is assumed to reflect the subjective likelihood that a decision is correct, which should be positively related to the amount of sensory evidence accumulated according to the decision-locus models, it is congruent that CPP amplitudes, as an index of sensory evidence accumulation (Kelly & O'Connell, 2013), were positively related to confidence ratings. However, this relationship should only be observed when sensory evidence is effectively accumulated (such that the amount of sensory evidence is actually related to the likelihood of being correct). As this is not the case for error trials (sensory evidence in error trials is likely noisy), the same relationship was not observed in error trials.

For the Pe, our pattern of results is consistent with those reporting Pe amplitudes in error trials were lower when confidence was high (Feuerriegel et al. 2021; Hewig et al., 2011). Considering Pe as an index of post-decisional error evidence accumulation (Desender et al., 2021), an account similar to that of our CPP results is possible. That is, this relationship between Pe and confidence in error trials reflects that error evidence was more effectively accumulated in a way that its amount was indeed positively related to error likelihood (and hence negatively related to confidence). This also explains why in correct trials such a relationship was not observed: Error evidence accumulated in correct trials was likely due to noise, and therefore was not related to actual error likelihood (or confidence ratings).

However, we also found an unexpected result: Correct trials rated as guessing showed larger Pe amplitudes than subjectively correct trials (while correct trials rated as subjectively incorrect did not show even larger Pe amplitudes). There are two potential interpretations of this finding. First, the pattern that correct trials rated as guessing showed larger Pe amplitudes suggests that, Pe in correct trials might not be completely ineffective as previously suggested: Even in correct trials where error evidence was limited, guessing trials still involve some amount of error evidence. However, such error evidence in correct trials could be derived from sources and not reliably related to actual error likelihood, e.g., a sense of uncertainty or response conflict.

A second interpretation of this finding is that, correct trials rated as subjectively incorrect involved qualitatively different processes that limited error evidence accumulation (such that Pe amplitudes in these trials were not larger than that in guessing trials). In fact, some correct trials rated as subjectively incorrect to a larger extent ("surely incorrect" and "probably incorrect") showed descriptively lower amplitudes not only in the Pe time window, but throughout the epoch time window (Figure 3.7A), suggesting that in this case stimulus processing could be qualitatively different from the pre-decisional stage (e.g., because correct trials in these two categories were driven by premature responding, which involved no sensory or error evidence accumulation, but led to the reports of "surely incorrect" and "probably incorrect"). In contrast, the correct trials rated as "maybe incorrect", which did not show lower amplitudes throughout the epoch, showed descriptively higher amplitudes than those trials rated as subjectively correct. Therefore, it is possible that some error evidence accumulation might have still occurred in correct trials, but was disrupted in trials where a strong sense of certainty of being incorrect was involved.

3.4.2. Specific association with subjective accuracy

Regarding the specificity to subjective accuracy, we showed that the CPP was only related to different levels of certainty of being correct but not certainty of being incorrect. This supports the account that CPP does not reflect a general confidence level, but specifically certainty of being correct. The absence of relationship with certainty of being incorrect suggests that decisions perceived to be incorrect likely involve ineffective sensory evidence accumulation, even though these subjectively incorrect decisions turned out to be correct.

In contrast, we did not observe the parallel pattern for Pe. Contrary to our prediction, while Pe was not linearly related to certainty of being correct, it was also not related to certainty of being incorrect. However, taken together with the analysis with binary subjective accuracy, it shows that while Pe did reflect subjective accuracy, but was not sensitive to the graded certainty of being incorrect. This suggests that the same amount of error evidence could be accumulated regardless of the certainty levels, unlike the relationship between certainty of being correct and CPP amplitudes in correct trials. Therefore, these patterns of results suggest that error awareness and confidence judgment might not lie on the same linear scale. Even though in behavioral terms participants were able to signal certainty of being incorrect in a graded fashion (that was linearly related to proportions of correct trials), the Pe appeared to be a binary, all-or-none signal that reflects whether error awareness has occurred. This is analogous to some accounts that error detection is at least partly all-or-none (Charles et al., 2013; Janssen et al. 2016; Spinelli et al., 2021). This possibility is also partly supported by the pattern in our data that RTs were similar across different levels of certainty of being incorrect. It should however be noted that as trials that were rated as incorrect were small in number (Figure 3.4B), this notion remains to be tested.

3.4.3. Connection with the evidence accumulation framework

Taken together, inconsistent with the proposal that the Pe reflects a general metacognitive variable (Desender et al. [2021]) and previously findings supporting this proposal (Boldt & Yeung [2015]; Desender et al. [2019]; Scheffer & Coles [2000]), we found no monotonic relationship between the Pe and full-range confidence. Therefore, if the CPP and Pe are taken to be the decision variable and metacognitive decision variable, then confidence appeared to be related to them in a more complex way than what was assumed in the evidence accumulation framework (Desender et al., 2021).

The evidence accumulation framework proposed by Desender et al. (2021) suggests that decision and confidence are both driven by evidence accumulation processes. Specifically, a decision variable is assumed to accumulate noisy sensory evidence and give rise to a decision once the amount of sensory evidence reaches a boundary. This is then followed by a similar process in which a metacognitive variable accumulates error evidence. Critically, the model suggests that the metacognitive variable gives rise to confidence judgment and error awareness depending on criterion setting. For example, error awareness occurs when the accumulated error evidence is against the decision and its amount exceeds a criterion, and confidence emerges when the opposite is true. Confidence and error awareness thus lie on the same continuum, and both depend on the state of the metacognitive variable, which was proposed to be reflected by the Pe (Desender et al, 2021).

However, the current finding regarding the relationship between the Pe and confidence is not in line with the framework. As discussed by Feuerriegel et al. (2022), Desender and colleagues' claim that the Pe should reflect both error awareness and confidence (2021) was incompatible with their findings that Pe amplitudes were only related to certainty of being incorrect in error trials. Using a pre-stimulus baseline as in Feuerriegel et al. (2022), the current study similarly suggests that Pe is not sensitive to certainty of being correct is therefore inconsistent with the proposal by Desender et al. (2021). However, while the Pe might not serve as a general index of full-range confidence, it might reflect error evidence accumulation, that is limited in objectively correct trials, and error trials rated as correct. This is because Pe amplitudes still differentiated error trials rated as incorrect from error trials rated as guessing or correct, potentially because error evidence accumulation was effective in the former case but not in the latter. Taken together with the finding that confidence in correct trials was related to CPP amplitudes, this would then imply that confidence and error awareness are not two sides of the same coin: Confidence could emerge during decision formation, while error awareness might be based more on post-decisional processes.

While the current findings are broadly consistent with Feuerriegel et al. (2022), there are however some differences. First, whereas they reported that Pe amplitudes were related to certainty of being incorrect (measured on a continuous scale), the current study did not find such an effect. Further, Pe appeared to be an all-or-none signal of error awareness in error trials as similar Pe amplitudes across different levels of certainty of being incorrect were observed. This inconsistency could however be due to different scales used between studies and that guessing trials were excluded in the current analyses. As only few studies measured graded error awareness, it remains to be tested in the future whether error awareness is indeed binary or graded. Second, while Feuerriegel et al. (2022) found that the CPP was related to confidence in both correct and error trials, the current study found it to be related to confidence only in correct trials. This difference could be attributed to the difference between task paradigms. It is possible that errors in their task were easier to detect because detected errors involved the absence of sensory evidence accumulation, whereas in the current task detected and undetected errors had similarly noisy sensory evidence accumulation (as the current paradigm was shown to induce a confidence bias through noise; Ko et al., 2022). Additionally, the fact that the Pe as a response-locked component was baseline corrected using a pre-stimulus baseline might have increased the noise in its measure, contributing to the difference between our findings.

Lastly, although the current study does not support the claim that Pe indexes a metacognitive variable as Desender and colleagues (2021) suggested, it does not reject the suggestion that a metacognitive variable exists. The idea that decisions and confidence are supported by two separate variables has been previously put forward (Fleming & Daw, 2017). Without assuming that they have respective evidence accumulation processes, serial or parallel structures, or the relationship between decision and metacognitive variables, it has been shown that the presence of a metacognitive variable explains a range of empirical findings in both confidence and error awareness studies (Fleming & Daw, 2017). This suggests that metacognitive variable remains a useful construct for understanding metacognitive decisions, although its neural implementations could be more complex than the Pe.

3.4.4. Limitation and future directions

One limitation in the current study was that the number of trials rated as incorrect (ratings < 4) was much lower than the number of high confidence trials (ratings > 4), especially when only correct trials were considered (Tables 3.1 & 3.2). This could have contributed to the unexpected ERP patterns of correct trials rated as incorrect, as discussed above. Therefore, for the analyses involving certainty of being incorrect in correct trials (and to a lesser extent in error trials), the absence of effects could potentially be due to underpowered statistical tests and noise due to low trial numbers. While this is a typical pattern in a range of perceptual tasks (decisions are less likely to be rated as incorrect than correct), future studies could employ tasks where a larger number of trials rated as incorrect occur, for example, go/no-go tasks (Murphy et al., 2015) or instruct participant to detect error with lower threshold (Steinhauser & Yeung, 2010).

Another limitation was that the observed relationships might not be due to the proposed mechanisms, but correlation between the components. For example, participants could have determined their confidence levels before indicating a decision with a response even though they were only required to do so after an interval (Baranski & Petrusic, 1994). This could still lead to amplitude differences in the Pe which was observed after the response and the Pe might reflect continued processing of sensory evidence (Rausch et al., 2020). However, while this could be possible in correct trials where the CPP and Pe show similar amplitude differences (Figure 3.7A), error trials showed that the amplitude difference in the Pe time window only appeared after response, thus suggesting that Pe is likely to reflect evidence from a different source rather than only sensory evidence. To avoid mixing the accumulation processes of sensory evidence and error evidence in future studies, stimulus masked such that sensory evidence is unlikely to continue after decision.

Lastly, the relationships between confidence ratings and ERP components were analysed across trials from all stimulus conditions, due to the low number of trials in each condition. As the CPP and Pe could vary with both confidence ratings and stimulus conditions with different relative and absolute evidence strength (Appendix C), the reported relationships between confidence ratings and ERP measures could have been confounded by the effects of stimulus conditions. Additionally, different stimulus conditions could have led to stimulus-offset potentials of different amplitudes, which could have further overlapped and confounded with the Pe. In future studies where the number of trials is sufficient, one could take into account of both stimulus conditions and confidence ratings to separate the two effects. For example, this could be done by examining trials within the same stimulus condition but with different confidence ratings (Tagliabue et al., 2019). To avoid the overlapping between stimulus-offset potentials and the Pe, future studies could employ tasks where stimuli remain on screen throughout the trial, or stimulus offset and decision are separated by an interval.

Table 3.1 Means and standard deviations of trial counts by confidence ratings and objective accuracy

	Surely incorrect	Probably incorrect	Maybe incorrect	Guessing	Maybe correct	Probably correct	Surely correct
Correct	4±7	8±14	13 ±15	77±82	122±110	186 ±114	233±170
Error	11±13	11±15	16 ±15	82±58	61±47	81±60	77±67

Table 3.2 Means and standard deviations of trial counts by subjective and objective accuracy

	Subjectively incorrect	Guessing	Subjectively correct
Correct	25±29	77±82	541±103
Error	38±30	67±58	219±67

3.4.5. Conclusion

In summary, we found that the two proposed correlates of confidence, the CPP and Pe, showed respective positive and negative relationships with confidence ratings in a luminance judgment task, thus supporting the hypothesis that both of these components exhibit specific associations with objective decision outcomes. We further showed that CPP amplitudes in particular were only positively related to certainty of being correct, but there was no evidence that Pe amplitudes were only related to certainty of being incorrect. While the current findings could be explained by an evidence accumulation framework, they suggest that confidence might be largely dependent on how much sensory evidence is accumulated, and error awareness emerges through an error evidence accumulation process after decision. This is inconsistent with the recent suggestion that confidence and error awareness could be considered as different expressions of metacognitive decision that is determined by a common metacognitive variable.

Chapter 4. General discussion

4.1. Summary of the current project

Metacognitive decisions regarding subjective decision accuracy, including confidence judgment, error awareness, and change-of-mind decisions have been traditionally studied as different areas of research. However, they have been recently proposed to be closely related and might be explained by a common mechanism based on evidence accumulation, and confidence is a common construct in these proposals (Desenders et al., 2021; van den Berg, Anandalingam et al., 2016).

The current project therefore focused on the role of confidence in metacognitive decisions and investigated: (a) how confidence is related to changes of mind, and (b) how confidence is related to pre- and post-decisional processes. Particularly, Chapter 2 (Studies 1 and 2) investigated behaviorally how confidence and changes of mind varied with stimulus properties including relative evidence strength, absolute evidence strength, and evidence variability. Chapter 3 (Study 3) investigated how ERP components related to sensory and error evidence accumulation, namely the CPP and Pe, could be linked to confidence differentially.

This final chapter first summarizes the key research findings, then discusses these findings in relation to relevant cognitive processes, as well as their implications to models of metacognitive decision, and lastly discusses the methodological limitations of the current project and future research directions.

4.2. Main research findings

4.2.1. Behavioral findings

First, on the behavioral level, when stimulus properties including relative evidence and absolute evidence were varied, confidence ratings and the proportions of change-of-mind trials were found to change in largely consistent ways (Studies 1 & 2).

In correct trials, stronger relative evidence led to increased confidence, and reduced proportions of change-of-mind trials, and stronger absolute evidence similarly increased confidence and reduced proportions of change-of-mind trials. In contrast, in error trials, stronger absolute evidence increased confidence and reduced proportions of change-of-mind trials, and stronger relative evidence led to reduced confidence and higher proportions of change-of-mind trials. Therefore, the current findings showed that confidence and changes of mind were consistently affected by sensory evidence strength. These key findings are summarized in Table 4.1.

4.2.2. EEG findings

Second, on the neural level, the relationships between confidence and ERP components were examined (Study 3). The main finding was that these relationships were dependent on objective accuracy. CPP amplitudes were only positively related to full-range confidence (and binary subjective accuracy) in correct decisions, while Pe amplitudes were only negatively related to full-range confidence (and binary subjective accuracy) in erroneous decisions. Further, when the full-range confidence scale was divided into measures of certainty, only certainty of being correct in objectively correct decisions was related to the CPP and Pe: It was positively related to CPP amplitudes and nonlinearly related to Pe amplitudes. Such findings remained largely unchanged when alternative measures of the same ERP components were used (Appendix B) or when only subjective accuracy but not objective accuracy was considered (Appendix D). In an exploratory analysis on the effects of relative and absolute evidence (Appendix C), it was found that stronger absolute evidence reduced both Pe and CPP amplitudes in error trials, but only when relative evidence was strong. The key findings are summarized in Tables 4.2 – 4.4.

Table 4.1. Summary of the effects of relative and absolute evidence on confidence ratings and change-of-mind trial proportions.

	Relative evidence		**Absolute evidence**	
	Confidence	**Change of mind**	**Confidence**	**Change of mind**
Correct	increase	decrease	increase	decrease
Error	decrease	increase	increase	decrease

Table 4.2. Summary of the relationships between ERP component amplitudes and confidence measures.

	Response-locked CPP				**Pe**			
	Full-range	**Binary**	**Correct certainty**	**Error certainty**	**Full-range**	**Binary**	**Correct certainty**	**Error certainty**
Correct	positive	positive	positive	n.s.	nonlinear	n.s.	nonlinear	n.s.
Error	n.s.	n.s.	n.s.	n.s.	negative	negative	n.s.	n.s.

Note. n.s. no significant relationship.

Table 4.3. Summary of the relationships between alternative ERP component amplitudes and confidence measures.

	Stimulus-locked CPP			Late Pe			ERN/Ne		
	Binary	**Correct certainty**	**Error certainty**	**Binary**	**Correct certainty**	**Error certainty**	**Binary**	**Correct certainty**	**Error certainty**
Correct	positive	positive	n.s.	nonlinear	n.s.	positive	negative	n.s.	n.s.
Error	n.s.	n.s.	n.s.	negative	n.s.	n.s.	n.s.	n.s.	n.s.

Note. n.s. no significant relationship.

Table 4.4. Summary of the effects of relative and absolute evidence on ERP component amplitudes.

	Response-locked CPP			Pe			Late Pe		
	Relative evidence	**Absolute evidence**	**Interaction**	**Relative evidence**	**Absolute evidence**	**Interaction**	**Relative evidence**	**Absolute evidence**	**Interaction**
Correct	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	negative	n.s.
Error	n.s.	n.s.	negative[a]	n.s.	negative[a]	negative[a]	n.s.	negative	negative[a]

Note. n.s. no significant relationship.

[a]Negative relationship between absolute evidence strength and ERP amplitudes, only when relative evidence was strong

4.3. Theoretical implications

This section discusses the implications of the current findings. It first discusses (a) how increased absolute evidence increased confidence while impairing accuracy, and (b) the possibility that increased absolute evidence reduced changes of mind through increased confidence. This section then moves on to discuss (c) how confidence in correct and erroneous decisions could be differentially related to CPP and Pe amplitudes, and (d) how these ERP findings map onto existing evidence accumulation models of metacognitive decisions.

4.3.1. How did stronger absolute evidence increase confidence?

Chapter 2 has suggested that the positive effect of absolute evidence on confidence was analogous to the positive evidence bias (PEB) observed in previous studies. The PEB (or more recently termed, "high-intensity-high-confidence" effect; Shehkar & Rahnev, 2022) was originally found in earlier studies employing dot motion tasks where positive evidence (i.e., evidence supporting the correct response) was experimentally increased, while the ratio between positive and negative evidence (i.e., evidence supporting other responses) was maintained (Koizumi et al., 2015; Odegaard et al., 2018; Peters et al., 2017; Samaha & Denison, 2022). In the condition with stronger positive evidence, confidence was consistently found to be increased while decision accuracy remained unchanged.

Although the dot motion task is the most commonly used paradigm for inducing the PEB, similar biases were later shown in comparative judgment paradigms involving perceptual and value-based decisions (Folke et al., 2016; Sepulveda et al., 2020). For example, Sepulveda et al. (2020) showed that in a numerosity task that required participants to decide which of the two visually presented boxes contained more dots, the summed dot number of the two boxes predicted confidence beyond the dot number difference between the boxes. The same pattern was also observed in value-based decision tasks involving the choice between multiple food items assigned with different preference values (Folke et al., 2016; Sepulveda et al., 2020). Given that the task used in the current project was also a comparative judgment task and produced similar effects, the effect of absolute evidence on confidence could be considered to be of the same nature as the PEB.

The PEB provides an uncommon case where confidence is dissociated with decision accuracy, serving as a hurdle phenomenon that computational models of confidence judgment attempted to explain (Shekhar & Rahnev, 2022). One major assumption that has been incorporated into models of confidence judgment is the decision-congruent (or response-congruent) evidence hypothesis discussed in Chapter 2, which suggests that while a decision is based on sensory evidence difference between the selected and nonselected option (i.e., balance-of-evidence), confidence is based on only sensory evidence congruent with the selected option (Zylberberg et al., 2012; Peters et al., 2017; Maniscalco et al., 2021). As increase in positive evidence (and absolute evidence in the current task paradigm) implies increase in decision-congruent evidence, this hypothesis provides a direct account for the PEB, as supported by previous modelling works (Zylberberg et al., 2012; Peters et al., 2017; Maniscalco et al., 2021). Thus, in line with this account, stronger absolute evidence should have increased decision-congruent evidence, and thereby increased confidence.

4.3.1.1. Why is confidence based on decision-congruent evidence?

Previous studies have discussed why confidence could be estimated in such an apparently suboptimal way that confidence does not track decision accuracy (Mazor, 2021; Navajas et al., 2016). One possibility is that such confirmation bias is a form of heuristic that allows efficient estimation of confidence in the natural environment (Maniscalco et al., 2016). As the natural environment often contains more than two choice alternatives, maintaining the evidence supporting each of the alternatives could be resource-demanding. Instead, by discarding such evidence and basing confidence on decision-congruent evidence only, confidence could still be adequately estimated (albeit less accurate under a lab condition where an ideal estimate of confidence requires taking into account the balance-of-evidence between two alternatives). This is aligned with the idea that a confidence estimation strategy adaptive for detection tasks is employed for estimating confidence in discrimination tasks, altohugh confidence estimated in this way corresponds to the likelihood that a stimulus is present vs. absent, rather than the likelihood of a stimulus being different from other stimuli (Maniscalco et al., 2016; Mazor, 2021).

This notion that the PEB originates from decision making in the natural environment is further supported by simulation studies. For example, Miyoshi and Lau (2020) showed that, under conditions where representations of stimulus strength were differentially variable for different stimulus categories (e.g., by assuming that target-present trials involve higher variability than target-absent trials), the use of decision-congruent in confidence could actually lead to more accurate confidence ratings than the use of balance-of-evidence. As such large variance difference is likely present in the natural environment, it was suggested that basing confidence on decision-congruent evidence could be adaptive. Consistently, Webb et al. (2021) showed that neural network models that were trained with stimuli of variable contrast levels (corresponding to higher variability), could naturally produce the PEB similarly to human participants. This finding accords with the argument that the presence of the PEB stems from applying confidence judgment rule from a more variable environment to a more controlled environment.

However, others have suggested that the over-reliance on decision-congruent evidence might be more dependent on the decision-making agent. For example, some have suggested such strategy serves to maintain self-consistency in decision making (Peters et al., 2017), to minimize cognitive dissonance (Navajas et al., 2016), and to allocate attentional resources to focus on evidence that matches one's behavioral goal (Sepulveda et al., 2020).

This explanation is consistent with previous studies showing that the commitment to a decision could bias further processing of the stimuli that the decision was based on (Stocker & Simoncelli, 2007).

4.3.1.2. Does the PEB involve pre- and/or post-decisional processes?

Another question that has been asked about the PEB is whether this bias involve pre- and/or post-decisional processes (Samaha & Denison, 2022). As the PEB is driven by sensory evidence, it is reasonable to assume that it occurs during the pre-decisional stage. In fact, with the PEB induced in a dot motion task, Samaha and colleagues (2022) showed that the PEB was also observed when choice and confidence were reported simultaneously (where the time for post-decisional processing was limited), although such finding did not completely rule out a post-decisional account (as in such condition one could still internally commit to a choice and then a confidence level; Baranski & Petrusic, 1994; Desender et al., 2021). In support of the pre-decisional account, psychophysical kernel analyses also showed that decision-congruent evidence at early stimulus presentation contributed to confidence (Zylberberg et al., 2012; Mazor, 2021).

Notably, it has also been suggested that such confirmation bias could extend to the post-decisional stage, thus influencing confidence reported later and changes of mind (Navajas et al., 2016; Rollwage et al., 2020). In the current project, additionally analyses were conducted to examine the relationships between absolute evidence and the ERP components of the CPP and Pe, in order to identify the neural processes related to the absolute evidence effect. However, these analyses showed that stronger absolute evidence reduced both the CPP and Pe in error trials, suggesting that both sensory and error evidence accumulation before and after erroneous decisions were reduced (Appendix C). Particularly, reduced sensory evidence accumulation due to stronger absolute evidence could be explained by the fact that stronger absolute evidence reduced perceived relative evidence, which was assumed to be directly relevant to sensory evidence accumulation in comparative judgment tasks (discussed in Chapter 2). On the other hand, reduced error evidence accumulation due to stronger absolute evidence corresponded to the positive effect of absolute evidence on confidence. This suggests that the bias observed affected post-decisional processes, and might reduce error evidence accumulation.

However, it should be noted that the effect of absolute evidence on ERP amplitudes did not correspond completely with their effects on confidence. First, confidence changes due to absolute evidence were mostly found at all levels of relative evidence, and for both correct and error trials, but ERP changes were observed within a limited subset of trials (error trials where relative evidence was strong). This is potentially because the absolute evidence effect involve not only the processes reflected by these ERPs (i.e., sensory and error evidence accumulations), but other aspects such as metacognitive bias, and decision-congruent evidence was not measured in the current studies (Rollwage et al., 2020; Samaha et al., 2022). Second, the negative effect of absolute evidence on CPP amplitudes in error trials was however not consistent with the fact that stronger absolute evidence increased confidence (given that confidence and CPP amplitudes were positively related), but this could be due to the possibility that sensory evidence in error trials does not inform confidence (discussed in Chapter 4).

4.3.2. How did increased confidence translate into lower change-of-mind frequency?

4.3.1.1. Confidence and changes of mind share the same stream of evidence

Chapter 2 showed that stronger absolute evidence not only increased confidence, but also reduced changes of mind. This pattern of results is also consistent with the previous finding reported by Turner et al. (2021), which showed that changes of mind in a highly similar task were also reduced by stronger absolute evidence. These converging findings between the current studies and Turner et al. (2021) suggest that confidence and changes of mind are closely related, to the extent that they are not only negatively related when task difficulty was high in general (e.g., due to reduced relative evidence), but also when a bias was induced due to absolute evidence strength. This consistency is in support of the proposal that both confidence and changes of mind are based on the same stream of stimulus-based, sensory evidence (Rollwage et al., 2020; van den Berg, Anandalingam et al., 2016).

In the study by van den Berg, Anandalingam et al. (2016), participants completed a dot motion task where they moved a handle to indicate choice and confidence simultaneously, and changes of mind were measured as changes of movement trajectories. Their psychophysical kernel analysis showed that while initial confidence and choice were explained by balance-of-evidence and RT, changes of mind were explained additionally by post-decisional evidence accumulation even when the stimulus was absent. This showed that stimulus-based evidence that did not contribute to initial choice and confidence (i.e., late-arriving evidence in the processing pipeline) was continuously processed and affected changes of mind. Specifically, changes of mind occurred when evidence supporting one choice alternative changed to support another choice alternative.

Using a brightness judgment task similar to the current studies, Turner et al. (2022) replicated the same psychophysical kernel analysis results, and additionally demonstrated that even evidence from early stimulus presentation was also related to changes of mind: Choice options supported by stronger momentary evidence initially were less likely to be reversed. They suggested that it was potentially because such early evidence biased sensory evidence accumulation. Taken together, these previous studies suggest that changes of mind and confidence share the same stream of sensory evidence, but the occurrence of changes of mind occur depend on whether late-arriving evidence is in conflict with earlier evidence. This provides an explanation for the consistent findings between the current studies and the findings reported by Turner et al. (2021): Stimulus-based, sensory evidence contributes to confidence, which serves as the basis of changes of mind.

4.3.1.2. The effect of confidence on changes of mind

Considering that confidence could be the basis of changes of mind, what could be an underlying process that connects confidence and changes of mind? While this is beyond the scope of the current project, this section discusses the possible mechanisms in the evidence accumulation framework. Specifically, confidence could affect changes of mind in two aspects: change-of-mind evidence accumulation rates and change-of-mind criteria.

As discussed above, van den Berg and colleagues (2016) suggested that confidence emerges during decision formation based on the state of the decision variable, which then receives late-arriving sensory evidence and determines whether changes of mind occur. Particularly, decisions made with higher confidence (due to more accumulated sensory evidence) required more conflicting, late-arriving evidence to be overruled, compared with decisions made with lower confidence (equivalent to an effect on the starting point of the change-of-mind evidence accumulation process; Rollwage et al., 2020).

Beyond the effect on the starting point, confidence could also affect the change-of-mind evidence accumulation rates. In the current studies, both relative and absolute evidence strength consistently affected confidence and changes of mind. As relative and absolute evidence were respectively linked to drift rates and drift rate variability in previously proposed evidence accumulation models (Ratcliff et al., 2018; Turner et al., 2021), it is possible that earlier drift rates and drift rate variability could extend to the change-of-mind evidence accumulation process. However, this idea that post-decisional processes are simply extension of pre-decisional processes appeared to be insufficient, at least when this was applied to explain the effect of absolute evidence on changes on mind (Turner et al., 2021). Such findings might suggest that in addition to the sensory evidence accumulation, change-of-mind evidence accumulation process could involve additional sources of (error) evidence (Desender et al., 2021; Murphy et al. 2015; Stone et al., 2022; Ullsperger et al., 2010), or drift rate changes dependent on confidence (Braun et al., 2018; Rollwage et al., 2020).

Additionally, the relationship between confidence and changes of mind could be explained by change-of-mind criteria. For example, under conditions where confidence is high due to liberal confidence criteria, change-of-mind criteria could also be more liberal, leading to a higher change-of-mind likelihood. While this idea has not yet been tested, two streams of studies suggest that this mechanism is possible. First, the PEB has been shown to be related to more liberal criteria for reporting high confidence, meaning that a smaller amount of evidence is required to give a high confidence response (Samaha et al., 2022), suggesting that absolute evidence might not only affect drift rate variability (Ratcliff et al., 2018; Turner et al., 2021), but also confidence criteria. If higher confidence results from more liberal criteria (instead of accumulated a large amount of evidence), this might imply that in such case changes of mind might as well occur easily. Second, studies have reported cases in which confidence could affect subsequent decisions (Desender et al., 2019; Overhoff et al., 2021; van den Berg, Zylberberg et al., 2016), although the direction of effects could be task-specific. Typically, in tasks where only one decision is made in each trial, lower confidence likely leads to more conservative decision criteria / more caution for the following decision (Desender et al., 2019). However, in a task where participants were required to make a pair of decisions sequentially, it was found that the decision threshold that determines speed-accuracy tradeoff of the second decision was dependent on the confidence level in the first decision (van den Berg, Zylberberg et al., 2016). Particularly, as task performance in this particular task was contingent on both decisions being correct and accuracy feedback was not available until the end of trial, participants prioritized accuracy over speed in making the second decision if they were confident in their first decisions. Nevertheless, these studies generally suggest that confidence could be considered as an internal feedback signal that influences the decision criteria for the following decision.

In summary, the current behavioral findings suggest that absolute evidence increased confidence by increasing decision-congruent evidence, and such increase in confidence might in turn affect post-decisional processes including drift rates and change-of-mind criteria for changes of mind.

4.3.3. ERP correlates of metacognitive decisions

4.3.3.1. Confidence (certainty of being correct) was only related to the CPP amplitudes

Single-stage evidence accumulation models of confidence judgment suggest that confidence emerges during decision formation, based on the amount of sensory evidence accumulated (Kiani & Shadlen, 2009; Vickers, 1979; Lee et al., 2022; Rahnev, 2022). Consistently, Study 3 showed that CPP amplitudes, which was proposed to be an index of sensory evidence accumulation, was a pre-decisional correlate of decision confidence. Likely because the amount of sensory evidence in errors and subjectively incorrect trials was noisy, such correlate was only observed for correct trials rated as correct (i.e., certainty of being correct), but did not differ across confidence levels. It therefore also explains why even though error trials in the current study showed variability in certainty of being correct, CPP amplitudes were not related to certainty of being correct in error trials. Confidence was however not related in the same way to the Pe, which was proposed to be a post-decisional correlate of confidence (Desender et al., 2021). Instead, Pe amplitudes in correct trials were higher for guessing trials than subjectively correct and incorrect trials, suggesting that confidence (i.e., certainty of being correct) was not meaningfully related to post-decisional error evidence accumulation. Therefore, the current findings are consistent with a single-stage model of confidence (Lee et al., 2022; Rahnev, 2022), but inconsistent with the claim that post-decisional accumulation of sensory or error evidence determines confidence (Desender et al., 2021; Moran et al., 2015).

4.3.3.2. Changes of mind could be related to both CPP and Pe amplitudes

In contrast, subjectively correct and subjectively incorrect trials differed in terms of both CPP amplitudes in correct trials and Pe amplitudes in error trials. Assuming that subjectively incorrect trials would have led to changes of mind, this suggests that, on the one hand, changes of mind occur when the amount of sensory evidence accumulated was low (in initially correct trials), consistent with the previous finding that weaker early sensory evidence could predict more frequent changes of mind (Turner et al., 2022). On the other hand, changes of mind could also occur when the amount of error evidence accumulated was high (in error trials). This is consistent with the previous findings that changes of mind and confidence depend on the amount of evidence accumulated before and after decision (Charles & Yeung, 2019).

These respective patterns for correct and error trials also suggest a difference between changes of mind in correct and changes of mind in error trials, where the latter could be associated with error awareness. In correct trials rated as incorrect, changes of mind could be driven by the lack of sensory evidence, e.g., due to premature responding, motor slips, or lack of response preparation, and thus one could report changes of mind without post-decisional error evidence accumulation. In error trials rated as incorrect, changes of mind could be driven by post-decisional error evidence accumulation. Specifically, error awareness in trials where errors actually occurred could involve more error evidence due to error-specific, processes such as orienting responses (Wessel, 2017). It also provides a potential explanation for the common finding that changes of mind were often corrective (Stone et al., 2022), as additional error evidence in error trials could lead to more changes of mind, compared with correct trials.

Additionally, certainty of being incorrect was not related to Pe amplitudes, suggesting that the same error evidence accumulation process was involved regardless of levels of certainty. Assuming that certainty of being incorrect is related to error awareness, this finding could be explained by the possibility that error awareness is binary. Unlike confidence, the graded nature of error awareness has not been well established as most error monitoring studies employed binary accuracy ratings (Wessel, 2012). Evidence suggesting that error awareness could be graded comes from studies investigating neural correlates of error detection. For example, Gehring et al. (1993) showed that ERN amplitudes were related to a range of parameters (response force, error correction proportion, correct response RT in the next trial) in a graded fashion. Scheffer and Coles (2000) showed that ERN amplitudes decreased monotonically with confidence ratings. A ERP component similar to the ERN, the feedback-related negativity (FRN), was also modulated by error feedback that reflected a large or small discrepancy to ideal response (Luft et al., 2014). In contrast, Janssen et al. (2016) showed that outcome monitoring activities from medial frontal cortex showed binary patterns even in response to prediction errors in different sizes. Similarly, Charles et al. (2013) also reported that the ERN occurred depending on binary visibility report instead of gradual manipulation of stimulus masking time. Given that the ERN might reflect processes that contribute to error awareness and the Pe, e.g., response conflict (Charles et al., 2013; Ullsperger et al., 2010), it is possible that the Pe similarly reflects binary error awareness. However, it should be noted that the number of trials rated as incorrect was low in Study 3, which might have contributed to the non-significant results regarding certainty of being incorrect and Pe amplitudes. In summary, the current findings suggest that on the neural level error awareness appears to be binary, as Pe amplitudes were invariant across regarding certainty levels of being incorrect.

4.3.4. What can explain the asymmetry between metacognitive decisions in correct and erroneous decisions?

The current studies showed that metacognitive decisions in correct and erroneous decisions were different in terms of ERP correlates. This suggests that error commission could involve qualitatively different cognitive processes than those of correct decisions (Rabbitt, 1966; Wessel et al., 2017).

In fact, previous studies on metacognitive decisions have often reported asymmetrical patterns between correct and erroneous decisions. In confidence studies, it has been typically found that when stimulus discriminability is varied, retrospective confidence ratings in correct and error trials showed a folded-X pattern (but not when choice and confidence were reported simultaneously; Desender et al., 2020; Kepecs et al., 2008; Rausch et al., 2020; Sander et al., 2016): Confidence in correct trials increased with discriminability, but confidence in error trials decreased with discriminability. Also, confidence in correct trials is invariant to the interval between the primary decision and confidence report (i.e., inter-judgment time), while confidence in error trials appears to decay with longer inter-judgment time. These findings suggest that post-decisional processes contribute to lower confidence ratings following erroneous decisions.

Similarly, changes of mind patterns also differ between correct and error trials. In addition to overall higher change-of-mind frequency in error trials (Stone et al., 2022), changes of mind in error trials also occurred most often at intermediate task difficulty level (as low task difficulty produces few errors, but high task difficulty prevents decision from being reversed by conflicting evidence), while changes of mind in correct trials decreased monotonically with lower difficulty (Albantakis et al., 2012). These findings suggest that post-decisional processing differ between correct and error trials, as on average post-decisional evidence tend to contradict sensory evidence more in error than in correct trials (Yu et al., 2015).

In error monitoring studies, both ERN and Pe amplitudes have been found to be elevated for error trials compared with correct trials, and error signaling occurs more often in error trials, suggesting error-specific processing following error commission (e.g., Steinhauser & Yeung, 2010). In a recent study by Overhoff and colleagues (2022), it was also found that full-range confidence was related to RT and response force in correct and error decisions respectively, suggesting the involvement of different response parameters informing metacognitive judgments in correct and erroneous decisions.

While different explanations have been proposed to explain these asymmetrical patterns, one common assumption is that error trials might involve qualitatively different processes from early to later processing stages, e.g., inhibition of ongoing cognitive processes, automatic orienting responses, and behavioral adjustments such as post-error slowing (Wessel et al., 2017). Within the evidence accumulation framework, these processes could be involved in a distinct error evidence accumulation process (Desender et al., 2020; Ullsperger et al., 2010; Yu et al., 2015), leading to the current observation that increased Pe amplitudes appear to be specific to error trials rated as incorrect.

4.3.5. Implications to models of metacognitive judgment

The above section discussed how absolute evidence affected confidence and changes of mind, as well as how the relationships between confidence and ERP indexes of evidence accumulation differed between correct and error trials. Based on these findings, this section discusses some implications for future theory development, regarding (a) the nature of evidence accumulated, and (b) the decision locus of confidence.

4.3.5.1. Dissociation between choice and confidence

A model of metacognitive decisions must explain metacognitive decisions together with the primary decision. The current behavioural finding that accuracy could be dissociated from confidence posed a hurdle for several common model structures, e.g., evidence accumulation models that assume confidence and changes of mind could be explained by balance-of-evidence (Vickers & Packer, 1982) or post-decisional sensory evidence accumulation (Moran et al., 2015; Resulaj et al., 2009; van den Berg, Anandalingam et al., 2016). In contrast, this dissociation is in support of models that assume separate sources of evidence for choice and metacognitive decisions, e.g., the unified model of metacognitive decisions that assumes metacognitive decisions are based on error evidence that is distinct from sensory evidence (Desender et al., 2021).

The distinction between continued accumulation of sensory evidence and error evidence has been noted in the literature (or the *sensory/response* vs. *accuracy* reference frame of evidence accumulation; Desender et al., 2021). The current findings are congruent with the error evidence hypothesis that the Pe reflects error evidence accumulation (Steinhauser et al., 2010; Murphy et al., 2015), as it could account for the observed pattern that error trials rated as incorrect showed clearly increased Pe amplitudes than error trials rated as correct, while their CPP amplitudes were similar. This pattern might suggest that post-decisional accumulation departs from pre-decisional accumulation for error trials rated as incorrect, which should not be expected if post-decisional accumulation is simply a continuation of sensory evidence accumulation.

However, it is possible that error evidence and sensory evidence are correlated or both processes exist (Shekhar & Rahnev, 2022). For example, when examining the effects of absolute evidence on CPP and Pe amplitudes in error trials, the negative effect of increased absolute evidence on amplitudes appeared to maintain from the CPP to Pe time windows, suggesting that in such condition both sensory evidence as well as error evidence accumulations were reduced (Appendix C). While this exploratory finding does not clarify how the two types of evidence are correlated, previous studies have suggested different proposals on how the decision variable is transformed into a metacognitive variable (e.g., by incorporating Gaussian or logarithmic noise; Shekhar & Rahnev, 2022).

4.3.5.2. Decision locus of confidence

A key assumption that is incorporated in models of metacognitive decisions is that metacognitive decisions are based on a metacognitive variable that accumulates sensory/error evidence (e.g., Desender et al., 2021). This assumption is however inconsistent with the current findings that confidence appeared to be instead dependent on the decision variable that accumulates sensory evidence only, while changes of mind and error awareness could be related to error evidence in addition to sensory evidence. Correspondingly, future theory might consider a mechanism where confidence emerges during decision formation, and changes of mind and error awareness require post-decisional evidence, consistent with existing models in these two areas (Murphy et al., 2015; Steinhauser & Yeung, 2010; Stone et al., 2022; Ullsperger et al., 2010).

While such proposal would be against the models that suggest that confidence depends on post-decisional accumulation (Boldt & Yeung, 2015; Desender et al., 2021; Moran et al., 2015), recently it has been suggested that post-decisional accumulation is not a generally necessary feature for explaining confidence (but could be more important in some cases, e.g., when decisions are speeded and stimulus is presented continuously after response; Lee et al., 2022; Shekhar & Rahnev, 2022). This however does not preclude the possibility that confidence could be modified in post-decisional stages (as suggested by Lee et al., 2022).

In the context where relative and absolute evidence are manipulated, this tentative proposal would suggest that, during decision formation, both relative and absolute evidence contribute to the primary decision, but absolute evidence predominantly contributes to confidence. Furthermore, absolute evidence could continue to affect error evidence accumulation, which leads to the refinement of confidence, and sometimes changes of mind and error awareness.

4.4. Methodological limitations

In addition to the theoretical implications, several methodological limitations of the current studies should also be considered given that they could be related to the divergent findings between the current and previous studies. Particularly, this section focuses on (a) the difference between simultaneous and sequential measures of confidence, (b) evidence availability in task design, (c) emphasis on response speed, and (d) the low frequency of change-of-mind trials.

4.4.1. Simultaneous vs. sequential measures of confidence

In the current task paradigm, the primary decision and confidence were measured sequentially (e.g., as in Charles & Yeung, 2019; Petrusic & Baranski, 2003), as opposed to simultaneously with one response (e.g., as in Burk et al., 2014; Van den Berg, Anandalingam et al., 2016). Given that this methodological difference could lead to different confidence patterns (Desender et al., 2020, 2021; Petrusic & Baranski, 2003), the sequential measure of choice and confidence could have contributed to the biased confidence patterns observed in the current studies.

Particularly, when the decision and confidence reports are separated by a time interval (i.e., inter-judgment time; IJT), post-decisional processing is more likely to contribute to confidence report (Desender et al., 2020; Moran et al., 2015; Yu et al., 2015), whereas post-decisional processing is likely to be limited with simultaneous report (Samaha et al., 2022). Previous studies on the effect of inter-judgment time on confidence have shown mixed findings. For example, Desender et al. (2020) showed that longer IJT had positive effects on metacognitive accuracy: Longer IJT increased confidence in correct trials but reduced confidence in error trials, suggesting that unbiased stimulus processing occurs during IJT and leads to more accurate confidence judgment, particularly when stimuli remain present. However, others have suggested that processing during IJT could be biased. For example, Yu et al. (2015) showed that longer IJT reduced confidence in error trials but did not affect confidence in correct trials, suggesting that post-decisional processing involves decay of evidence supporting erroneous decisions. More relevant to the current studies, Navajas et al. (2016) also suggested that post-decisional processing could be subject to the confirmation bias underlying the PEB. This suggests that the absolute evidence effect on confidence observed in the current studies might be partially attributed to the sequential measurement.

Additionally, stimuli in the current task were terminated after the primary decision, leading to a post-decisional interval without stimulus presentation. This absence of stimuli could have also strengthened the bias in confidence ratings. It has been suggested that information that is incorporated into confidence could be task dependent (Stone et al., 2022). For example, some studies that allowed continued presentation of stimuli, have shown that under such condition post-decision accumulation of sensory evidence was facilitated (e.g., Charles & Yeung, 2019, Moran et al., 2015). Notably, when additional information is relevant and the use of such evidence is encouraged (e.g., continued presentation of the same stimulus in the decisional stage), metacognitive sensitivity might increase (Charles & Yeung, 2019; Desender et al., 2020; Turner et al., 2022). On the other hand, absence of stimuli might have limited effect on metacognitive sensitivity (Desender et al., 2019). In the current study where a confidence bias was induced, absence of stimuli might lead to the loss of metacognitive sensitivity (as post-decisional processes are likely biased; Rollwage et al., 2020).

It should however be noted that the sequential measurement and termination of stimuli do not fully explain the absolute evidence effect, as previous studies on the PEB showed similar effects with the simultaneous measure of decision and confidence (e.g., Samaha et al., 2022). The current measurement approach therefore could have at most amplified the PEB effect.

4.4.2. Measure of changes of mind based on converted confidence ratings

In Studies 1 and 2, confidence ratings were converted into a binary change-of-mind, by assuming that confidence ratings lower than "guessing" would have led to changes of mind (as done in e.g., Charles & Yeung, 2019). This converted measure based on confidence ratings however could have confounded the finding that changes of mind and confidence exhibited similar patterns in response to both relative and absolute evidence, as it is based on several assumptions.

For example, it assumes that measuring them together would not alter how participants respond. This assumption might not be valid as requiring accuracy report could affect primary and metacognitive performance (Double & Birney, 2019; Grützmann et al., 2014; Porth et al., 2022). It also assumes that metacognitive decisions could be converted, e.g., accuracy ratings lower than guessing implies changes of mind, while changes of mind could involve additional processes than confidence judgment. Additionally, by merging their measurements, it abandoned some features specific to different types of metacognitive decisions (discussed in Chapter 1). For example, it would not be reasonable to simultaneously measure primary decisions with changes of mind, but simultaneous report of choice and confidence is common in confidence studies. Also, change-of-mind paradigms usually involve continued presentation of stimuli, which was not incorporated in the current paradigm. The current measurement approach therefore only examined changes of mind indirectly. However, it should be noted that the convergence between the current findings and the study by Turner et al. (2021), which involved a highly similar paradigm, partially supports that the validity of the converted measure. Future studies investigating the relationship between confidence and changes of mind could incorporate measures of both (e.g., as in van den Berg, et al., 2016).

4.4.3. Speed emphasis on primary and secondary tasks

The task used in the current project involved instructing participants to respond as quickly as possible for both primary and secondary decisions. This might have influenced both primary and secondary decision processes given that previous studies have reported different effects of speed and accuracy emphasis (Baranski & Petrusic, 1994; 1998; Falkenstein et al., 2000; Rafiei & Rahnev, 2021; Steinhauser & Yeung, 2010; Summerfield & Yeung, 2012).

In error monitoring studies, Steinhauser and Yeung (2012) and Steinhauser et al. (2008) tested the effect of primary task speed/accuracy trade-off on error signalling and found that people tended to signal more errors with higher speed pressure. Likewise, with speed emphasis on primary decision, number of changes of mind increased (Resulaj et al., 2009). Additionally, primary decision RT and changes of mind are also negatively related (Albantakis et al., 2012). Confidence was also lower and was more likely to involve post-decisional processing when speed was emphasized (Baranski & Petrusic, 1998; Desender et al., 2021). When accuracy was stressed, confidence could be readily judged before primary decision response and error awareness was less likely to occur (Baranski & Petrusic, 1998).

Generally, speed emphasis reduces the threshold for primary decision, such that errors occur more often due to premature responding and more errors are detected (Baranski & Petrusic, 1994; Charles et al., 2013; Desender et al., 2022). This is consistent with the common finding that high time pressure increases metacognitive sensitivity (Moran et al., 2015). In terms of ERP measures, earlier studies have found that speed pressure impaired performance monitoring, as reflected by reduced ERN/Ne and Pe amplitudes (Arbel & Donchin, 2009; Falkenstein et al., 2000; Gehrin et al., 1993). However, when task instructions required speeding response but did not reduce significance of errors, Pe amplitudes were increased by speeded responses, as in such case more error evidence could be present (Steinhauser & Yeung, 2012).

Taken together, these findings suggest that speed pressure could alter both primary and metacognitive decision processes. These effects of speeded responses could have influenced the findings of the current studies where the primary decision was speeded. For example, it could explain the smaller ERN/Ne than previous studies (Falkenstein et al., 2000; Appendix B) and the fact that changes of mind occurred comparatively more often than typical perceptual decision tasks (e.g., Albantakis et al., 2012). Confidence ratings in the current studies were also speeded as participants were instructed to rate as quickly as possible. Based on previous studies investigating the effect of inter-judgment time (Desender et al., 2020; Yu et al., 2015), making metacognitive judgment faster likely leads to lower metacognitive sensitivity, potentially due to smaller difference in post-decisional evidence between correct and error trials. This instruction reduced the criterion for reporting errors, and could have led to reduced metacognitive accuracy (Boldt et al., 2017; Moran et al., 2015; Pleskac & Busemyer, 2010). Additionally, time pressure might also induce effects other than only changing the decision criterion, such as reduced evidence accumulation, as evidence during decision formation might not be equally used in primary decisions (Calder-Travis et al., 2020; Carsten et al., 2022; Rae et al., 2014), and the same idea might apply to metacognitive decisions. As the current studies only involved speed emphasis without a comparison with accuracy emphasis, it is entirely unclear how speed pressure might have impacted the current findings.

4.4.4. Low number of changes of mind and aware errors

The current studies have a rather low number of trials with changes of mind or error awareness. As discussed above, in typical perceptual tasks, changes of mind occurred rarely (see Chapter 1; ~10% of all trials in current studies). Similarly, error monitoring studies also sometimes suffered from the lack of error awareness trials (e.g., Scheffer et al. [2000], Wessel et al. [2012]). Although this could be increased by giving instructions that, e.g., encourage participants to change their mind (Turner et al., 2022), or use more liberal criteria for reporting errors (Steinhauser & Yeung, 2010), these manipulations might influence the mechanism underlying spontaneously occurring changes of mind or error awareness. As insufficient trials could lead to underpower statistical analyses (particularly when further exclusion criteria are used e.g., in ERP studies), future studies could include a larger number of trials with multiple testing sessions (e.g., van dan Berg et al., 2016).

4.5. Future directions

4.5.1. What constitutes error evidence?

As discussed above, the current studies suggest that the post-decisional process accumulates error evidence. However, two aspects of this concept are still to be clarified. Current proposals involving the error evidence hypothesis have not yet specified clearly what constitutes error evidence. Early studies defined that error evidence is evidence that indicate an error has occurred. It was initially assumed to be generated by response conflict or internal correction response (Steinhauser & Yeung, 2012), but then also extended to include information from a wide range of sources (Desender et al., 2022; Stone et al., 2022), including information from different stimulus properties (Bolt et al., 2017; Rausch et al., 2018; Zylbergerg et al. 2012) to motor execution (Fleming et al., 2015; Resulaj et al., 2009; Pereira et al., 2020; Turner et al., 2021). While this conceptualization (or termed a multi-cue model; Boldt et al., 2017) accommodates different empirical findings where confidence or metacognitive decisions in general could be influenced by factors other than sensory evidence, what exactly contributes error evidence appears less clear. Existing studies have suggested these different factors could be task-specific (Hagura et al., 2022; Bolt et al., 2017) and perhaps individual-specific (de Gardelle & Mamassian, 2015; Moreira et al., 2018; Navajas et al., 2018). In future studies, manipulations that specifically affect error evidence but not sensory evidence, e.g., instructions on criteria for error detection (Steinhauser & Yeung, 2010) or even confidence and changes of mind could be useful for investigating how different cues are selectively involved, represented, and integrated into metacognitive decisions (Desender et al., 2021).

4.5.2. How is error evidence accumulation related to sensory evidence accumulation?

The relationship between error evidence accumulation and sensory evidence accumulation is so far unclear. Although the current studies could be explained by the error evidence hypothesis, previous studies also provided strong support that post-decisional accumulation of sensory evidence occurred (e.g., van den Berg, Anandalingam et al., 2016). Therefore, it is still unclear whether both processes exist, and how they are potentially related in terms of correlation and mechanisms. While the current project could not further clarify their relationship, future studies could aim to specify this relationship.

One proposal by Desender and colleagues (2021) is that post-decisional accumulation of sensory evidence precedes error evidence accumulation, corresponding to the finding that the CPP, indexing sensory evidence accumulation, was sometimes found to persist shortly after the decision, and that the Pe, indexing error evidence accumulation, occurs much later. In terms of mechanism, while not under the evidence accumulation framework, Fleming and Daw (2017) suggested that the decision variable feeds into the metacognitive variable. This idea that the decision variable together with other sources of information contribute to the metacognitive variable broadly accords with the proposal that additional information enters the post-decisional process after response (Stone et al., 2022; Ullsperger et al., 2011).

It has been suggested that post-decisional accumulation involves error evidence that is qualitatively different from sensory evidence (Desender et al., 2021). This is supported by previous studies showing that speed pressure on primary decision leads to more error evidence and larger Pe amplitudes in error trials (Steinhauser & Yeung, 2012; Desender et al., 2020, 2021). This could result from the fact that the requirement to give speeded responses lower the response criterion for primary decision, which also leads to more conflict and thus error evidence. As speed pressure is assumed to affect response criterion but not the rate of sensory evidence accumulation, it shows that error evidence could be independently manipulated. Consistently, criterion for error signaling could be manipulated by monetary incentive, without impacting primary decision performance (Steinhauser & Yeung, 2010). Taken together, error evidence accumulation appears to be at least partially dissociated from sensory evidence accumulation.

4.5.3. Can a unified model account for confidence, changes of mind, and error awareness?

The current studies have attempted to clarify how confidence could be related to other metacognitive decisions. While the current findings cannot adjudicate whether they are outcomes of the exact same process or metacognitive variable (Desender et al., 2021), they could still be linked to the same mechanism. As discussed, one possible mechanism could be one in which a pre-decisional process leads to the primary decision and an initial sense of confidence, which in turn serves as the basis for error evidence accumulation and leads to changes of mind and error awareness (Stone et al., 2022).

This proposal, however, remains to be validated by future experimental and computational works. Experimentally, future work could employ decision tasks with independent measures of different metacognitive decisions, instead of the converted measures used in the current and previous studies. Computationally, future studies could test whether confidence could directly impact different parameters of evidence accumulation models (Braun et al., 2018; Rollwage et al., 2020; van den Berg, Anandalingam et al., 2016), for example, how confidence might lead to more conservative change-of-mind criterion, or changes in error evidence accumulation rate. This approach might also benefit from using a wider range of task paradigms, e.g., those that induce a confidence bias, as previous model comparisons have shown that model performance differed considerably when tested with different tasks.

4.6. Conclusion

Recent research has extensively applied evidence accumulation models to explain how metacognitive processes emerge and behave under various experimental conditions. This allowed the decomposition of metacognitive performance into its underlying mechanistic components and thus advanced the understanding of metacognitive decisions. Given the central role of confidence in this general framework, the current project firstly investigated the effects of sensory evidence strength on confidence judgment and changes mind, and secondly investigated the relationships between confidence and certainty are related to ERP components linked to evidence accumulation processes.

By showing that confidence and changes of mind varied consistently in response to stimulus properties such as relative and absolute evidence strength, the current studies suggest that confidence might moderate the effect of sensory evidence on changes of mind. On the neural level, it was shown that confidence in correct and error trials were respectively associated with CPP and Pe amplitudes, suggesting confidence in correct decisions is related to pre-decisional sensory evidence accumulation, while changes of mind could involve post-decisional error evidence accumulation. Taken together, current findings converge on the notion that confidence emerges during decision formation and could, with the contribution from post-decisional evidence, serve as a basis of changes of mind.

References

Albantakis, L., Deco, G., 2011. Changes of mind in an attractor network of decision-making. PLoS Comput. Biol. 7 (6), e1002086.

Alexander, W. H., & Brown, J. W. (2010). Computational models of performance monitoring and cognitive control. *Topics in Cognitive Science*, 2(4), 658–677.

Arbel, Y., & Donchin, E. (2009). Parsing the componential structure of post-error ERPs: A principal component analysis of ERPs following errors. *Psychophysiology*, 46(6), 1179– 1189.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics*, *55*(4), 412-428.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 929-945.

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. (1995). " Where did I go wrong?" A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(6), 1312-1322.

Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 26). New York: Academic Press.

Boldt, A., De Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Rxperimental Psychology: Human Perception and Performance*, *43*(8), 1520-1531.

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, *35*(8), 3478-3484.

Boldt, A., De Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(8), 1520-1531.

Brainard, D. H. (1997) The Psychophysics Toolbox, *Spatial Vision, 10*, 433-436.

Braun, A., Urai, A. E., & Donner, T. H. (2018). Adaptive history biases result from confidence-weighted accumulation of past choices. *Journal of Neuroscience*, *38*(10), 2418-2429.

Burk, D., Ingram, J. N., Franklin, D. W., Shadlen, M. N., & Wolpert, D. M. (2014). Motor effort alters changes of mind in sensorimotor decision making. *PloS one*, *9*(3), e92681.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280*(5364), 747–749.

Carp, J., & Compton, R. J. (2009). Alpha power is influenced by performance errors. *Psychophysiology*, *46*(2), 336-343.

Chapman, S., Colvin, L. E., & Cosentino, S. (2020). Translational aspects of the multidisciplinary study of metacognition. *Translational Issues in Psychological Science*, *6*(1), 26-31.

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

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

Clarke, F., Birdsall, T. & Tanner, W. 1959 Two types of ROC curves and definition of parameters. *J. Acoust. Soc*. Am. 31, 629–630. (doi:10.1121/1.1907764)

Cohen, M. X., Van Gaal, S., Ridderinkhof, K. R., & Lamme, V. (2009). Unconscious errors enhance prefrontal-occipital oscillatory synchrony. *Frontiers in Human Neuroscience*, 54.

Coles, M. G., Scheffers, M. K., & Holroyd, C. B. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological psychology*, *56*(3), 173-189.

Compton, R. J., Arnstein, D., Freedman, G., Dainer-Best, J., & Liss, A. (2011). Cognitive control in the intertrial interval: evidence from EEG alpha power. *Psychophysiology*, *48*(5), 583-590.

Crites, S. L., Jr., Devine, J. V., Lozano, D. I., & Moreno S. (1998). Event-related potentials and serial position effects in a visual probe recognition task. *Psychophysiology*, 35, 293– 304.

De Bruijn, E. R., Hulstijn, W., Verkes, R. J., Ruigt, G. S., & Sabbe, B. G. (2004). Drug-induced stimulation and suppression of action monitoring in healthy volunteers. *Psychopharmacology*, *177*(1), 151-160.

de Gardelle, V., & Mamassian, P. (2015). Weighting mean and variability during confidence judgments. *PloS one*, *10*(3), e0120870.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for er ror-detection and compensation. *Psychological Science*, 5(5), 303–305.

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS biology*, *5*(10), e260.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, *29*(5), 761-778.

Desender, K., Boldt, A., Verguts, T., & Donner, T. H. (2019). Confidence predicts speed-accuracy tradeoff for subsequent decisions. *Elife*, *8*, e43499.

Desender, K., Donner, T. H., & Verguts, T. (2021). Dynamic expressions of confidence within an evidence accumulation framework. *Cognition*, *207*, 104522.

Desender, K., Ridderinkhof, K. R., & Murphy, P. R. (2021). Understanding neural signals of post-decisional performance monitoring: An integrative review. *Elife*, *10*, e67556.

Desender, K., Vermeylen, L. & Verguts, T. Dynamic influences on static measures of metacognition. *Nat Commun* **13**, 4208 (2022). https://doi.org/10.1038/s41467-022-31727-0

Desmet, C., Fias, W., Hartstra, E., & Brass, M. (2011). Errors and conflict at the task level and the response level. *Journal of Neuroscience*, *31*(4), 1366-1374.

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.

Di Gregorio, F., Maier, M. E., & Steinhauser, M. (2020). Are errors detected before they occur? Early error sensations revealed by metacognitive judgments on the timing of error awareness. *Consciousness and Cognition*, *77*, 102857.

Donchin, E., Ritter, W., & McCallum, W. C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. *Event-related brain potentials in man*, 349-411.

Donchin, E. (1981). Surprise!… surprise?. *Psychophysiology*, *18*(5), 493-513.

Dotan, D., Meyniel, F., & Dehaene, S. (2018). On-line confidence monitoring during decision making. *Cognition*, *171*, 112-121.

Double, K. S., & Birney, D. P. (2019). Reactivity to measures of metacognition. *Frontiers in Psychology*, 2755.

Drugowitsch, J., Mendonça, A. G., Mainen, Z. F., & Pouget, A. (2019). Learning optimal decisions with confidence. *Proceedings of the National Academy of Sciences*, *116*(49), 24872-24880.

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E. J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, *74*(2), 454-465.

Eimer, M., & Mazza, V. (2005). Electrophysiological correlates of change detection. *Psychophysiology*, *42*(3), 328-342.

Endrass, T., Reuter, B., & Kathmann, N. (2007). ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. *European Journal of Neuroscience*, *26*(6), 1714-1720.

Evans, S., & Azzopardi, P. (2007). Evaluation of a 'bias-free' measure of awareness. *Spatial Vision*, *20*(1), 61-78.

Fairhurst, M. T., Travers, E., Hayward, V., & Deroy, O. (2018). Confidence is higher in touch than in vision in cases of perceptual ambiguity. *Scientific Reports*, *8*(1), 1-9.

Falkenstein,M., Hohnsbein, J., Hoormann, J., Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalogr. *Clin. Neurophysiol. 78*(6), 447–455. http://dx.doi.org/10.1016/0013-4694(91)90062-9.

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (2001). Changes of error-related ERPs with age. *Experimental brain research*, *138*(2), 258-262.

Falkenstein, M., Hohnsbein, J., Hoorman, J., & Blanke, L. (1989). Error processing in choice reaction tasks with focused and crossmodal divided attention. An ERP study. Poster presented at the Ninth International Conference on Event-Related Potentials of the Brain, May 23–June 3, Noordwijk, the Netherlands.

Falkenstein, M., Koshlykova, N. A., Kiroj, V. N., Hoormann, J., & Hohnsbein, J. (1995). Late ERP components in visual and auditory Go/Nogo tasks. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *96*(1), 36-43.

Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and Cognition*, *9*(2), 288-307.

Festinger, L. (1943). Studies in decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *Journal of Experimental Psychology, 32*(4), 291-306.

Feuerriegel, D., Churches, O., Coussens, S., & Keage, H. A. (2018). Evidence for spatiotemporally distinct effects of image repetition and perceptual expectations as measured by event-related potentials. *NeuroImage*, *169*, 94-105.

Feuerriegel, D., Murphy, M., Konski, A., Mepani, V., Sun, J., Hester, R., & Bode, S. (2022). Electrophysiological correlates of confidence differ across correct and erroneous perceptual decisions. *NeuroImage*, *259*, 119447.

Fiehler, K., Ullsperger, M., & Von Cramon, D. Y. (2005). Electrophysiological correlates of error correction. *Psychophysiology*, *42*(1), 72-82.

Fitzgerald, L. M., Arvaneh, M., & Dockree, P. M. (2017). Domain-specific and domain-general processes underlying metacognitive judgments. *Consciousness and Cognition*, *49*, 264-277.

Flavell J (1979). Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. *Am Psychol*, 34(10), 906.

Fleming, S. M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1), nix007.

Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91-114.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1338-1349.

Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: computation, biology and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1280-1286.

Fleming, S. M., & Frith, C. D. (2014). Metacognitive neuroscience: An introduction. In *The Cognitive Neuroscience of Metacognition* (pp. 1-6). Springer, Berlin, Heidelberg.

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*, 443.

Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-specific disruption of perceptual confidence. *Psychological Science*, *26*(1), 89-98.

Fleming, S. M., Van Der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*, *21*(4), 617-624.

Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. (2016). Explicit representation of confidence informs future value-based decisions. *Nature Human Behaviour*, *1*(1), 1-8.

Ford, J. M. (1999). Schizophrenia: The broken P300 and beyond. *Psychophysiology*, 36(6), 667–682.

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, *10*(4), 843-876.

Gehring, W. J., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1990). The error-related neg ativity: An event-related brain potential accompanying errors [Abstract]. *Psychophysiology, 27*, S34.

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385–390.

Gehring, W. J., & Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitor ing. *Nature Neuroscience, 3*(5), 516–520.

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

Gherman, S., & Philiastides, M. G. (2015). Neural representations of confidence emerge from the process of decision formation during perceptual choices. *Neuroimage*, *106*, 134-143.

Gherman, S., & Philiastides, M. G. (2018). Human VMPFC encodes early signatures of confidence in perceptual decisions. *Elife*, *7*, e38293.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506-528.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, *30*, 535-574.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1, pp. 1969-2012). New York: Wiley.

Grützmann, R., Endrass, T., Klawohn, J., & Kathmann, N. (2014). Response accuracy rating modulates ERN and Pe amplitudes. *Biological psychology*, *96*, 1-7.

Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neuroscience & Biobehavioral Reviews*, *55*, 88-97.

Guggenmos, M. (2021). Measuring metacognitive performance: type 1 performance dependence and test-retest reliability. *Neuroscience of Consciousness*, *2021*(1), niab040.

Hagura, N., Esmaily, J., & Bahrami, B. (2022). Does decision confidence reflect effort?. *Psyarxivx*.

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

Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). *Detection theory: A user's guide*. Routledge.

Hillyard, S. A., Squires, K. C., Bauer, J. W., & Lindsay, P. H. (1971). Evoked potential correlates of auditory signal detection. *Science*, *172*(3990), 1357-1360.

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.

Henmon, V. A. C. (1911). The relation of the time of a judgment to its accuracy. *Psychological Review*, *18*(3), 186-201.

Hester, R., Foxe, J. J., Molholm, S., Shpaner, M., and Garavan, H. (2005). Neural mechanisms involved in error processing: a com- parison of errors made with and without awareness. *Neuroimage,* 27, 602–608.

Hewig, J., Coles, M. G., Trippe, R. H., Hecht, H., & Miltner, W. H. (2011). Dissociation of Pe and ERN/Ne in the conscious recognition of an error. *Psychophysiology*, *48*(10), 1390-1396.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Re inforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*(4), 679–709.

Holroyd, C. B., Larsen, J. T., & Cohen, J. D. (2004). Context dependence of the event-re lated brain potential associated with reward and punishment. *Psychophysiology, 41*(2), 245–253.

Jentzsch, I., & Dudschig, C. (2009). Short article: Why do we slow down after an error? Mechanisms underlying the effects of posterror slowing. *Quarterly Journal of Experimental Psychology*, *62*(2), 209-218.

Janssen, D. J., Poljac, E., & Bekkering, H. (2016). Binary sensitivity of theta activity for gain and loss when monitoring parametric prediction errors. *Social Cognitive and Affective Neuroscience*, *11*(8), 1280-1289.

Johnson, D. M. (1939). Confidence and speed in the two-category judgment. *Archives of Psychology, 34*, 1-53.

Kayser, J., & Tenke, C. E. (2006). Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. *Clinical Neurophysiology*, *117*(2), 348-368.

Kelly, S. P., Corbett, E. A., & O’Connell, R. G. (2021). Neurocomputational mechanisms of prior-informed perceptual decision-making in humans. *Nature Human Behaviour*, *5*(4), 467-481.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *Journal of Neuroscience*, *33*(50), 19434-19441.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227-231.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329-1342.

King, J. A., Korb, F. M., von Cramon, D. Y., & Ullsperger, M. (2010). Post-error behavioral adjustments are facilitated by activation and suppression of task-relevant and task-irrelevant information processing. *Journal of Neuroscience*, *30*(38), 12759-12769.

Kleiner M, Brainard D, Pelli D (2007). "What's new in Psychtoolbox-3?" Perception 36 ECVP Abstract Supplement.

Koizumi, A., Maniscalco, B., & Lau, H. (2015). Does perceptual confidence facilitate cognitive control?. *Attention, Perception, & Psychophysics*, *77*(4), 1295-1306.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovich, & E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (pp. 289-325). New York: Cambridge University Press.

Kramer, A. F., Wickens, C. D., & Donchin, E. (1985). Processing of stimulus properties: Evidence for dual-task integrality. *Journal of Experimental Psychology: Human Perception and Performance, 11*(4), 393–408. https://doi.org/10.1037/0096-1523.11.4.393

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation. *Science*, 197, 792–795.

Kunimoto, C. (2001). Confidence and accuracy of near threshold discrimination responses. *Conscious. Cogn*.,10, 294–340. doi:10.1006/ccog.2000.0494

Laming, D. (1968). *Information Theory of Choice-Reaction Times*. London: Academic.

Laming, D. (1979). Choice reaction performance following an error. *Acta Psychologica*, *43*(3), 199-224.

Lamy, D., Salti, M., & Bar-Haim, Y. (2009). Neural correlates of subjective awareness and unconscious processing: an ERP study. *Journal of Cognitive Neuroscience*, *21*(7), 1435-1446.

Lee, D. G., & Pezzulo, G. (2021). Evidence or Confidence: What Really Accumulates During a Decision?. *bioRxiv*.

Lenth, R., H. Singmann, J. Love, P. Buerkner, and M. Herve. (2020). emmeans: Estimated marginal means, aka least-squares means, https://cran.r-project.org/web/packages/emmeans /index.html

Logan, G. D., & Crump, M. J. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science*, *330*(6004), 683-686.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know?. *Organizational Behavior and Human Performance*, *20*(2), 159-183.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art to 1980. *Decision making and change in human affairs*, 275-324.

Luck, S. J. (2014). *An Introduction to the Event-related Potential Technique*. MIT press.

Luttrell, A., Briñol, P., Petty, R. E., Cunningham, W., & Díaz, D. (2013). Metacognitive confidence: A neuroscience approach. *Revista de Psicología Social*, *28*(3), 317-332.

Luu, P., Flaisch, T., & Tucker, D. M. (2000). Medial frontal cortex in action monitoring. *Journal of Neuroscience, 20*(1), 464–469.

Luu, P., & Pederson, S. M. (2004). The anterior cingulate cortex: Regulating actions in context. In M. I. Posner (Ed.), *Cognitive Neuroscience of Attention* (pp. 232–244). New York: Guilford Press.

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science, 14*(1), 47–53.

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation. *Clinical Neurophysiology, 115*(8), 1821–1835.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Magliero, A., Bashore, T., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology, 21*, 171-186.

Maier, M., Steinhauser, M., & Hübner, R. (2008). Is the error-related negativity amplitude related to error detectability? Evidence from effects of different error types. *Journal of Cognitive Neuroscience*, *20*(12), 2263-2273.

Marco-Pallarés, J., Camara, E., Münte, T. F., & Rodríguez-Fornells, A. (2008). Neural mechanisms underlying adaptive actions after slips. *Journal of Cognitive Neuroscience*, *20*(9), 1595-1610.

Maniscalco, B., Charles, L., & Peters, M. A. (2022). Optimal Metacognitive Decision Strategies in Signal Detection Theory. Psyarxiv.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422-430.

Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: meta-d′, response-specific meta-d′, and the unequal variance SDT model. In *The Cognitive Neuroscience of Metacognition* (pp. 25-66). Springer, Berlin, Heidelberg.

Maniscalco, B., Peters, M. A., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception, & Psychophysics*, *78*(3), 923-937.

Maniscalco, B., Odegaard, B., Grimaldi, P., Cho, S. H., Basso, M. A., Lau, H., & Peters, M. A. (2021). Tuned inhibition in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. *PLoS Computational Biology*, *17*(3), e1008779.

Mazor, M. (2021). *Self-Modelling in Inference about Absence* (Doctoral dissertation, UCL (University College London)).

McWilliams, A., Bibby, H., Steinbeis, N., David, A. S., & Fleming, S. (2022). Age-related decreases in global metacognition are independent of local metacognition and task performance. *Psyarxiv*.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*(3), 159-163.

Metcalfe, J., & Schwartz, B. L. (2016). The ghost in the machine: Self-reflective consciousness and the neuroscience of metacognition. In J. Dunlosky & S. K. Tauber (Eds.), The Oxford Handbook of Metamemory (pp. 407-437).

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian probability: From neural origins to behavior. *Neuron*, *88*(1), 78–92.

Miyapuram, K. P., & Pammi, V. C. (2013). Understanding decision neuroscience: a multidisciplinary perspective and neural substrates. *Progress in Brain Research*, *202*, 239-266.

Miyoshi, K., & Lau, H. (2020). A decision-congruent heuristic gives superior metacognitive sensitivity under realistic variance assumptions. *Psychological Review*, *127*(5), 655-671.

Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive Psychology*, *78*, 99-147.

Moreira, C. M., Rollwage, M., Kaduk, K., Wilke, M., & Kagan, I. (2018). Post-decision wagering after perceptual judgments reveals bi-directional certainty readouts. *Cognition*, *176*, 40-52.

Moreno-Bote, R. (2010). Decision confidence and uncertainty in diffusion models with partially correlated neuronal integrators. *Neural Computation*, *22*(7), 1786-1811.

Murphy, P. R., Robertson, I. H., Harty, S., & O'Connell, R. G. (2015). Neural evidence accumulation persists after choice to inform metacognitive judgments. *Elife*, *4*, e11946.

Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of Learning and Motivation* (Vol. 26, pp. 125-173). Academic Press.

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews*, *28*(4), 441-448.

Nieuwenhuis, S., Ridderinkhof, K. R., 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*(5), 752–760.

Navajas, J., Bahrami, B., & Latham, P. E. (2016). Post-decisional accounts of biases in confidence. *Current Opinion in Behavioral Sciences*, *11*, 55-60.

Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E., & Bahrami, B. (2017). The idiosyncratic nature of confidence. *Nature Human Behaviour*, *1*(11), 810-818.

Norman, E., Pfuhl, G., Sæle, R. G., Svartdal, F., Låg, T., & Dahl, T. I. (2019). Metacognition in psychology. *Review of General Psychology*, *23*(4), 403-424.

Norman, E., & Price, M. C. (2015). Measuring consciousness with confidence ratings. *Behavioural Methods in Consciousness Research*, 159-180.

Notebaert, W., Houtman, F., Van Opstal, F., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: an orienting account. *Cognition*, *111*(2), 275-279.

Nunez Castellar, E., Kuhn, S., Fias, W., and Notebaert, W. (2010). Outcome expectancy and not accuracy deter- mines posterror slowing: ERP sup- port. *Cogn. Affect. Behav. Neurosci.* 10, 270–278.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience, 15*(12), 1729.

O'Connell, R. G., & Kelly, S. P. (2021). Neurophysiology of human perceptual decision-making. *Annual Review of Neuroscience*, *44*, 495-516.

Odegaard, B., Grimaldi, P., Cho, S. H., Peters, M. A., Lau, H., & Basso, M. A. (2018). Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. *Proceedings of the National Academy of Sciences*, *115*(7), E1588-E1597.

Orr, J. M., & Carrasco, M. (2011). The role of the error positivity in the conscious perception of errors. *Journal of Neuroscience*, *31*(16), 5891-5892.

Ouyang, G., Sommer, W., & Zhou, C. (2015). Updating and validating a new framework for restoring and analyzing latency-variable ERP components from single trials with residue iteration decomposition (RIDE). *Psychophysiology*, *52*(6), 839-856.

Overbeek, T. J., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *Journal of Psychophysiology*, *19*(4), 319-329.

Overhoff, H., Ko, Y. H., Feuerriegel, D., Fink, G. R., Stahl, J., Weiss, P. H., ... & Niessen, E. (2021). Neural correlates of metacognition across the adult lifespan. *Neurobiology of Aging*, *108*, 34-46.

Patterson, J. V., Pratt, H., & Starr, A. (1991). Event-related potential correlates of the seri al position effect in short-term memory. Electroencephalography and Clinical *Neurophysiology, 78*, 424–437.

Pavone, E. F., Marzi, C. A., & Girelli, M. (2009). Does subliminal visual perception have an error-monitoring system?. *European Journal of Neuroscience*, *30*(7), 1424-1431.

Peirce, C S., & Jastrow, J. "On Small Differences of Sensation." Memoirs of the National Academy of Sciencesfor 1884 3 (1885): 75-83. Reprinted in American Contributions to Mathematical Statistics in the Nineteenth Century, vol. 2, edited by Stephen M. Stigler. New York: Arno, 198.

Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S., ... & Millán, J. D. R. (2020). Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proceedings of the National Academy of Sciences*, *117*(15), 8382-8390.

Persaud, N., McLeod, P. & Cowey, A. 2007 Postdecision wagering objectively measures awareness. *Nat. Neurosci. 10*, 257–261. doi:10.1038/nn1840

Peters, M. A. (2022). Confidence in Decision-Making. In *Oxford Research Encyclopedia of Neuroscience*.

Peters, M. A., & Lau, H. (2015). Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *Elife*, *4*, e09651.

Peters, M. A., Thesen, T., Ko, Y. D., Maniscalco, B., Carlson, C., Davidson, M., ... & Lau, H. (2017). Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature human behaviour*, *1*(7), 1-8.

Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgments. *Psychonomic Bulletin & Review*, *10*(1), 177-183.

Pfister, R., & Foerster, A. (2022). How to measure post-error slowing: The case of pre-error speeding. *Behavior Research Methods*, *54*(1), 435-443.

Pierrel, R., & Murray, C. S. (1963). Some relationships between comparative judgment, confidence, and decision-time in weight-lifting. *The American Journal of Psychology*, *76*(1), 28-38.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181-197.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864-901.

Polich. (2012). Neuropsychology of P300. In S. J. Luck & E. Kappenman (Eds.),*Oxford handbook of event-related potential components.* New York: Oxford University Press.

Porth, E., Mattes, A., & Stahl, J. (2022). The influence of error detection and error significance on neural and behavioral correlates of error processing in a complex choice task. *Cognitive, Affective, & Behavioral Neuroscience*, *22*(6), 1231-1249.

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. *Nature Neuroscience*, *19*(3), 366-374.

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

Rabbitt, P. M. A. (1968). Three kinds of error-signalling responses in a serial choice task. *Quarterly Journal of Experimental Psychology*, *20*(2), 179-188.

Rabbitt, P. M. A., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology*, *29*(4), 727-743.

Rafiei, F., & Rahnev, D. (2021). Qualitative speed-accuracy tradeoff effects that cannot be explained by the diffusion model under the selective influence assumption. *Scientific reports*, *11*(1), 1-19.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59-108.

Ratcliff, R., & McKoon, G. (2018). Modeling numerosity representation with an integrated diffusion model. *Psychological Review*, *125*(2), 183-217.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, *116*(1), 59-83.

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychological Review*, *120*(3), 697-719.

Ratcliff, R., Voskuilen, C., & Teodorescu, A. (2018). Modeling 2-alternative forced-choice tasks: Accounting for both magnitude and difference effects. *Cognitive Psychology*, *103*, 1-22.

Rausch, M., Hellmann, S., & Zehetleitner, M. (2018). Confidence in masked orientation judgments is informed by both evidence and visibility. *Attention, Perception, & Psychophysics*, *80*(1), 134-154.

Rausch, M., Zehetleitner, M., Steinhauser, M., & Maier, M. E. (2020). Cognitive modelling reveals distinct electrophysiological markers of decision confidence and error monitoring. *NeuroImage*, *218*, 116963.

Rahnev, D. (2020). Confidence in the real world. *Trends in Cognitive Sciences*, *24*(8), 590-591.

Rahnev, D. (2021). Visual metacognition: Measures, models, and neural correlates. *American Psychologist*, *76*(9), 1445-1453.

Rahnev, D., Desender, K., Lee, A. L., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., ... & Zylberberg, A. (2020). The confidence database. *Nature Human Behaviour*, *4*(3), 317-325.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changing your mind: a computational mechanism of vacillation. *Nature*, *461*(7261), 263-266.

Ridderinkhof, K. R., Ramautar, J. R., & Wijnen, J. G. (2009). To PE or not to PE: A P3-like ERP component reflecting the processing of response errors. *Psychophysiology*, *46*(3), 531-538.

Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature Communications*, *11*(1), 1-11.

Rouault, M., McWilliams, A., Allen, M. G., & Fleming, S. M. (2018). Human metacognition across domains: insights from individual differences and neuroimaging. *Personality Neuroscience*, *1*.

Ruchsow, M., Spitzer, M., Gron, G., Grothe, J., & Kiefer, M. (2005). Error processing and impulsiveness in normals: Evidence from event-related potentials. *Cognitive Brain Research, 24*(2), 317–325.

Samaha, J., & Denison, R. (2022). The positive evidence bias in perceptual confidence is unlikely post-decisional. *Neuroscience of Consciousness*, *2022*(1), niac010.

Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a statistical computation in the human sense of confidence. *Neuron, 90*(3), 499-506.

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

Scheffers, M. K., Coles, M. G., Bernstein, P., Gehring, W. J., & Donchin, E. (1996). Event-related brain potentials and error-related processing: An analysis of incorrect responses to go and no-go stimuli. *Psychophysiology*, *33*(1), 42-53.

Schmidt, R. A., & Gordon, G. B. (1977). Errors in motor responding,“rapid” corrections, and false anticipations. *Journal of Motor Behavior*, *9*(2), 101-111.

Schwartz, B. L., & Díaz, F. (2014). Quantifying human metacognition for the neurosciences. In *The cognitive neuroscience of metacognition* (pp. 9-23). Springer, Berlin, Heidelberg.

Sepulveda, P., Usher, M., Davies, N., Benson, A. A., Ortoleva, P., & De Martino, B. (2020). Visual attention modulates the integration of goal-relevant evidence and not value. *Elife*, *9*, e60705.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, *8*(10), 1391-1400.

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, *18*(4), 186-193.

Shekhar, M., & Rahnev, D. (2021). The nature of metacognitive inefficiency in perceptual decision making. *Psychological Review*, *128*(1), 45-70.

Shekhar, M., & Rahnev, D. (2022). How do humans give confidence? A comprehensive comparison of process models of metacognition. Psyarxiv.

Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology*, *7*, 218.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S.. afex: Analysis of Factorial Experiments. R package version 0.28-0. https://cran.r-project.org/%20package=afex.

Smith, S. M., & Krajbich, I. (2021). Mental representations distinguish value-based decisions from perceptual decisions. *Psychonomic Bulletin & Review*, *28*(4), 1413-1422.

Spence, M. L., Dux, P. E., & Arnold, D. H. (2016). Computations underlying confidence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(5), 671-682.

Spence, M. L., Mattingley, J. B., & Dux, P. E. (2018). Uncertainty information that is irrelevant for report impacts confidence judgments. *Journal of Experimental psychology: Human Perception and Performance*, *44*(12), 1981-1994.

Spinelli, G., Tieri, G., Pavone, E. F., & Aglioti, S. M. (2018). Wronger than wrong: Graded mapping of the errors of an avatar in the performance monitoring system of the onlooker. *NeuroImage*, *167*, 1-10.

Steinemann, N. A., O'Connell, R. G., & Kelly, S. P. (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications*, *9*(1), 1-13.

Steinhauser, M., & Kiesel, A. (2011). Performance monitoring and the causal attribution of errors. *Cognitive, Affective, & Behavioral Neuroscience*, *11*(3), 309-320.

Steinhauser, M., Maier, M., & Hübner, R. (2008). Modeling behavioral measures of error detection in choice tasks: response monitoring versus conflict monitoring. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 158-176.

Steinhauser, M., & Yeung, N. (2010). Decision processes in human performance monitoring. *Journal of Neuroscience*, *30*(46), 15643-15653.

Steinemann, N. A., O'Connell, R. G., & Kelly, S. P. (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. *Nature Communications*, *9*(1), 1-13.

Stocker, A. A., & Simoncelli, E. (2007). A Bayesian model of conditioned perception. *Advances in Neural Information Processing Systems*, *20*.

Stone, C., Mattingley, J. B., & Rangelov, D. (2022). On second thoughts: changes of mind in decision-making. *Trends in Cognitive Sciences*.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187-1188.

Sutton, S., Ruchkin, D. S., Munson, R., Kietzman, M. L., & Hammer, M. (1982). Event-related potentials in a two-interval forced-choice detection task. *Perception & Psychophysics*, *32*(4), 360-374.

Sun, S., Zhen, S., Fu, Z., Wu, D. A., Shimojo, S., Adolphs, R., ... & Wang, S. (2017). Decision ambiguity is mediated by a late positive potential originating from cingulate cortex. *NeuroImage*, *157*, 400-414.

Squires, K. C., Squires, N. K., & Hillyard, S. A. (1975). Decision-related cortical potentials during an auditory signal detection task with cued observation intervals. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 268.

Tagliabue, C. F., Mazzi, C., Bagattini, C., & Savazzi, S. (2016). Early local activity in temporal areas reflects graded content of visual perception. *Frontiers in Psychology*, *7*, 572.

Tagliabue, C. F., Veniero, D., Benwell, C. S., Cecere, R., Savazzi, S., & Thut, G. (2019). The EEG signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience. *Scientific Reports*, *9*(1), 1-12.

Tanaka, H., Masaki, H., Takasawa, N., & Yamazaki, K. (2002). The relationship between the error-related negativity (ERN) and mental resource allocation in the error—related processes. *Japanese Journal of Physiological Psychology and Psychophysiology*, *20*, 29-37.

Teodorescu, A. R., Moran, R., & Usher, M. (2016). Absolutely relative or relatively absolute: violations of value invariance in human decision making. *Psychonomic Bulletin & Review*, *23*(1), 22-38.

Tops, M., Koole, S. L., & Wijers, A. A. (2013). The Pe of perfectionism. *Journal of Psychophysiology*.

Turner, W., Feuerriegel, D., Andrejević, M., Hester, R., & Bode, S. (2021). Perceptual change-of-mind decisions are sensitive to absolute evidence magnitude. *Cognitive Psychology*, *124*, 101358.

Turner, W., Feuerriegel, D., Hester, R., & Bode, S. (2022). An initial 'snapshot'of sensory information biases the likelihood and speed of subsequent changes of mind. *PLoS Computational Biology*, *18*(1), e1009738.

Twomey, D. M., Murphy, P. R., Kelly, S. P., & O'Connell, R. G. (2015). The classic P300 encodes a build-to-threshold decision variable. *European Journal of Neuroscience, 42*(1), 1636-1643.

Ullsperger, M. (2017). Neural bases of performance monitoring. In T. Egner (Ed.), *The Wiley handbook of cognitive control* (pp. 292–313). Wiley Blackwell. https://doi.org/10.1002/9781118920497.ch17

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure and Function*, *214*(5), 629-643.

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 1-11.

Webb, T. W., Miyoshi, K., So, T. Y., Rajananda, S., & Lau, H. (2021). Performance-optimized neural networks as an explanatory framework for decision confidence. *bioRxiv*.

Winter, C. J., & Peters, M. A. (2022). Variance misperception under skewed empirical noise statistics explains overconfidence in the visual periphery. *Attention, Perception, & Psychophysics*, *84*(1), 161-178.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931-959.

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310-1321.

Yeung, N., & Summerfield, C. (2014). Shared mechanisms for confidence judgements and error detection in human decision making. In *The Cognitive Neuroscience of Metacognition* (pp. 147-167). Springer, Berlin, Heidelberg.

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General*, *144*(2), 489-510.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550-592.

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.

van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence is the bridge between multi-stage decisions. *Current Biology*, *26*(23), 3157-3168.

van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the an terior cingulate cortex. *Journal of Cognitive Neuroscience, 14(*4), 593–602.

Vickers, D. (1979). Decision process in visual perception. New York: Academic Press

Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on RT, accuracy and confidence in a unidimensional discrimination task. *Acta Psychologica*, *50*(2), 179-197.

Vickers, D., Smith, P., Burt, J., & Brown, M. (1985). Experimental paradigms emphasising state or process limitations: II effects on confidence. *Acta Psychologica*, *59*(2), 163-193.

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the “error negativity” spe cific to errors? *Biological Psychology, 51*(2–3), 109–128.

Wessel, J. R. (2012). Error awareness and the error-related negativity: evaluating the first decade of evidence. *Frontiers in Human Neuroscience*, *6*, 88.

Wessel, J. R. (2018). An adaptive orienting theory of error processing. *Psychophysiology*, *55*(3), e13041.

Wessel, J. R., Danielmeier, C., & Ullsperger, M. (2011). Error awareness revisited: accumulation of multimodal evidence from central and autonomic nervous systems. *Journal of Cognitive Neuroscience*, *23*(10), 3021-3036.

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*, 79.

Zylberberg, A., Fetsch, C. R., & Shadlen, M. N. (2016). The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *Elife*, *5*.

Zylberberg, A., Roelfsema, P. R., & Sigman, M. (2014). Variance misperception explains illusions of confidence in simple perceptual decisions. *Consciousness and Cognition*, *27*, 246-253.

.

Appendix A Mixed-effects model results for Chapter 3 main analyses (likelihood ratio tests and regression coefficients)

Mixed-effects models results of behavioural data analysis

Table A1 Likelihood Ratio Tests Results for Predicting Accuracy (Log Odds of Being Correct) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	$\chi2$	p
Rel	2	746.23	<.001***
Abs	2	225.62	<.001***
Rel × Abs	4	26.62	<.001***

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table A2 Regression Coefficients for Predicting Accuracy (Log Odds of Being Correct) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	0.73	0.04	17.08	<.001***
Low Rel	-0.38	0.02	-20.91	<.001***
Med Rel	-0.12	0.02	-6.24	<.001***
Low Abs	0.26	0.02	13.20	<.001***
Med Abs	-0.02	0.02	-1.21	.228
Low Rel × Low Abs	-0.10	0.03	-3.63	<.001***
Med Rel × Low Abs	-0.05	0.03	-1.80	.073
Low Rel × Med Abs	0.05	0.03	1.82	.069
Med Rel × Med Abs	0.01	0.03	0.32	.749

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table A3 Likelihood Ratio Tests Results for Predicting Response Time (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	χ2	p
Rel	2	64.28	<.001***
Abs	2	98.51	<.001***
Rel × Abs	4	38.80	<.001***

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table A4 Regression Coefficients for Predicting Response Time (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	790.47	2.92	270.88	<.001***
Low Rel	8.07	1.70	4.73	<.001***
Med Rel	8.91	1.64	5.42	<.001***
Low Abs	21.16	1.82	11.65	<.001***
Med Abs	-5.40	1.99	-2.72	.007**
Low Rel × Low Abs	13.26	2.10	6.33	<.001***
Med Rel × Low Abs	4.62	2.46	1.88	.060
Low Rel × Med Abs	-7.21	2.13	-3.38	.001**
Med Rel × Med Abs	-3.81	2.08	-1.83	.067

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table A5 Likelihood Ratio Tests Results for Predicting Response Time (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	$\chi 2$	p
Rel	2	3.21	.201
Abs	2	28.92	<.001***
Rel × Abs	4	14.08	.007**

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table A6 Regression Coefficients for Predicting Response Time (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	814.86	3.19	255.27	<.001***
Low Rel	3.19	2.43	1.31	.190
Med Rel	3.67	2.51	1.46	.144
Low Abs	19.47	2.15	9.07	<.001***
Med Abs	-13.11	2.16	-6.07	<.001***
Low Rel × Low Abs	15.34	2.84	5.40	<.001***
Med Rel × Low Abs	2.35	3.12	0.75	.451
Low Rel × Med Abs	-2.93	2.71	-1.08	.280
Med Rel × Med Abs	-4.86	3.00	-1.62	.106

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table A7 Likelihood Ratio Tests Results for Predicting Confidence (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	$\chi 2$	p
Rel	2	494.24	<.001***
Abs	2	200.91	<.001***
Rel × Abs	4	54.75	<.001***

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table A8 Regression Coefficients for Predicting Confidence (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	5.77	0.11	52.91	<.001***
Low Rel	-0.21	0.01	-17.82	<.001***
Med Rel	-0.02	0.01	-1.43	.154
Low Abs	-0.15	0.01	-13.17	<.001***
Med Abs	0.02	0.01	1.50	.132
Low Rel × Low Abs	-0.10	0.02	-6.02	<.001***
Med Rel × Low Abs	-0.00	0.02	-0.09	.931
Low Rel × Med Abs	0.03	0.02	1.82	.068
Med Rel × Med Abs	0.00	0.02	0.26	.795

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table A9 Likelihood Ratio Tests Results for Predicting Confidence (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	$\chi2$	p
Rel	2	128.07	<.001***
Abs	2	239.46	<.001***
Rel × Abs	4	3.95	.413

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table A10 Regression Coefficients for Predicting Confidence (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	5.10	0.10	51.71	<.001***
Low Rel	0.20	0.02	9.38	<.001***
Med Rel	0.06	0.02	2.82	.005**
Low Abs	-0.36	0.02	-15.14	<.001***
Med Abs	0.12	0.02	5.46	<.001***
Low Rel × Low Abs	0.01	0.03	0.32	.751
Med Rel × Low Abs	-0.04	0.03	-1.25	.212
Low Rel × Med Abs	-0.02	0.03	-0.60	.549
Med Rel × Med Abs	-0.01	0.03	-0.27	.791

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Mixed-effects models results of ERP data analysis: Accuracy, full-range confidence, and subjective accuracy

Table A11 Regression Coefficients for Predicting CPP Amplitudes from Accuracy

Parameters	Estimate	SE	z	p
Intercept	3.57	0.54	6.62	<.001***
Correct	-0.15	0.09	-1.62	.106

Note. Intercept represents the estimate for error.
*p <.05 **p <.01 ***p <.001

Table A12 Regression Coefficients for Predicting Pe Amplitudes from Accuracy

Parameters	Estimate	SE	z	p
Intercept	5.29	0.64	8.29	<.001***
Correct	0.22	0.11	1.99	.047*

Note. Intercept represents the estimate for error.
*p <.05 **p <.01 ***p <.001

Table A13 Regression Coefficients for Predicting CPP Amplitudes (Correct Trials) from Full-Range Confidence

Parameters	Estimate	SE	z	p
Intercept	2.87	0.61	4.73	<.001***
Surely incorrect	-0.94	1.17	-0.80	.422
Probably incorrect	-1.49	0.83	-1.80	.073
Maybe incorrect	-0.90	0.69	-1.32	.188
Guessing incorrect	1.11	0.39	2.86	.004**
Maybe correct	-0.01	0.35	-0.02	.985
Probably incorrect	0.75	0.32	2.35	.019*

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table A14 Regression Coefficients for Predicting CPP Amplitudes (Error Trials) from Full-Range Confidence

Parameters	Estimate	SE	z	p
Intercept	3.28	0.56	5.83	<.001***
Surely incorrect	0.13	0.79	0.16	.870
Probably incorrect	-0.37	0.75	-0.49	.622
Maybe incorrect	-0.61	0.63	-0.97	.331
Guessing incorrect	0.72	0.39	1.85	.065
Maybe correct	0.21	0.39	0.53	.599
Probably incorrect	0.14	0.36	0.40	.688

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table A15 Regression Coefficients for Predicting Pe Amplitudes (Correct Trials) from Full-Range Confidence

Parameters	Estimate	SE	z	p
Intercept	5.21	0.72	7.28	<.001***
Surely incorrect	0.04	1.40	0.03	.975
Probably incorrect	-1.42	1.00	-1.43	.153
Maybe incorrect	1.52	0.82	1.85	.065
Guessing incorrect	0.95	0.46	2.03	.042*
Maybe correct	-0.38	0.42	-0.89	.371
Probably incorrect	-0.74	0.38	-1.92	.055

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table A16 Regression Coefficients for Predicting Pe Amplitudes (Error Trials) from Full-Range Confidence

Parameters	Estimate	SE	z	p
Intercept	6.16	0.69	8.90	<.001***
Surely incorrect	0.51	0.95	0.54	.587
Probably incorrect	1.73	0.90	1.91	.056
Maybe incorrect	1.27	0.75	1.69	.091
Guessing incorrect	0.48	0.47	1.02	.307
Maybe correct	-0.58	0.47	-1.24	.215
Probably incorrect	-1.35	0.43	-3.14	.002**

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table A17 Regression Coefficients for Predicting CPP Amplitudes (Correct Trials) from Binary subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	2.73	0.61	4.47	<.001***
Subjectively incorrect	-1.00	0.27	-3.68	<.001***

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table A18 Regression Coefficients for Predicting CPP Amplitudes (Error Trials) from Binary subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	3.13	0.58	5.42	<.001***
Subjectively incorrect	-0.19	0.25	-0.76	.445

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table A19 Regression Coefficients for Predicting Pe Amplitudes (Correct Trials) from Binary subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	5.19	0.71	7.31	<.001***
Subjectively incorrect	0.31	0.33	0.94	.345

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table A20 Regression Coefficients for Predicting Pe Amplitudes (Error Trials) from Binary subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	6.17	0.67	9.21	<.001***
Subjectively incorrect	1.28	0.30	4.25	<.001***

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table A21 Regression Coefficients for Predicting CPP Amplitudes (Correct Trials) from Trichotomized Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	3.21	0.57	5.61	<.001***
Subjectively incorrect	-1.43	0.38	-3.76	<.001***
Guessing	0.89	0.29	3.14	.002**

Note. Intercept represents the estimate for subjective correct.
*p <.05 **p <.01 ***p <.001

Table A22 Regression Coefficients for Predicting CPP Amplitudes (Error Trials) from Trichotomized Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	3.42	0.56	6.14	<.001***
Subjectively incorrect	-0.47	0.34	-1.40	.163
Guessing	0.57	0.30	1.92	.055

Note. Intercept represents the estimate for subjective correct.
*p <.05 **p <.01 ***p <.001

Table A23 Regression Coefficients for Predicting Pe Amplitudes (Correct Trials) from Trichotomized Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	5.53	0.69	8.07	<.001***
Subjectively incorrect	0.03	0.46	0.06	.955
Guessing	0.62	0.34	1.82	.069

Note. Intercept represents the estimate for subjective correct.
*p <.05 **p <.01 ***p <.001

Table A24 Regression Coefficients for Predicting Pe Amplitudes (Error Trials) from Trichotomized Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	6.26	0.68	9.25	<.001***
Subjectively incorrect	1.13	0.41	2.79	.005**
Guessing	0.32	0.36	0.90	.366

Note. Intercept represents the estimate for subjective correct.
*p <.05 **p <.01 ***p <.001

Mixed-effects models results of ERP data analysis: Certainty of being correct and certainty of being incorrect

Table A25 Regression Coefficients for Predicting CPP Amplitudes (Correct Trials) from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	3.58	0.58	6.22	<.001***
Maybe correct	-0.85	0.21	-4.09	<.001***
Probably correct	-0.04	0.17	-0.22	.824

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table A26 Regression Coefficients for Predicting CPP Amplitudes (Error Trials) from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	3.30	0.59	5.62	<.001***
Maybe correct	0.01	0.31	0.05	.963
Probably correct	0.01	0.27	0.03	.978

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table A27 Regression Coefficients for Predicting Pe Amplitudes (Correct Trials) from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	4.82	0.65	7.39	<.001***
Maybe correct	-0.14	0.25	-0.55	.586
Probably correct	-0.40	0.20	-1.94	.052

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table A28 Regression Coefficients for Predicting Pe Amplitudes (Error Trials) from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	4.88	0.63	7.70	<.001***
Maybe correct	0.66	0.37	1.77	.077
Probably correct	-0.04	0.32	-0.11	.910

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table A29 Regression Coefficients for Predicting CPP Amplitudes (Correct Trials) from Certainty of Being Incorrect

Parameters	Estimate	SE	z	p
Intercept	1.48	1.05	1.41	.167
Maybe incorrect	-0.55	1.15	-0.48	.634
Probably incorrect	-0.31	0.90	-0.34	.733

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Table A30 Regression Coefficients for Predicting CPP Amplitudes (Error Trials) from Certainty of Being Incorrect

Parameters	Estimate	SE	z	p
Intercept	3.06	0.77	3.96	.001**
Maybe incorrect	0.14	0.82	0.17	.865
Probably incorrect	0.15	0.77	0.20	.841

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Table A31 Regression Coefficients for Predicting Pe Amplitudes (Correct Trials) from Certainty of Being Incorrect

Parameters	Estimate	SE	z	p
Intercept	5.63	1.06	5.31	<.001***
Maybe incorrect	0.34	1.33	0.25	.802
Probably incorrect	-1.26	1.06	-1.19	.235

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Table A32 Regression Coefficients for Predicting Pe Amplitudes (Error Trials) from Certainty of Being Incorrect

Parameters	Estimate	SE	z	p
Intercept	7.70	1.10	7.02	<.001***
Maybe incorrect	-0.78	1.00	-0.78	.434
Probably incorrect	0.89	0.93	0.96	.340

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Appendix B Relationships between additional ERP measures and confidence

B.1. Background

While Chapter 3 showed how the response-locked CPP and Pe were related to confidence ratings, previous studies have also suggested that other ERP measures are related to confidence judgment. Specifically, the CPP and Pe have been defined differently in the literature and the component of error negativity (ERN/Ne) has also been considered important to performance monitoring. Therefore, this following analyses aimed to extend the previous investigation to these additional ERP measures, in order to explore how findings in the main analyses could be generalized.

B.1.1. Stimulus-locked CPP

Based on the assumption that the CPP is closely related to stimulus processing, this component was also measured relative to stimulus onset. In this case, CPP was measured from central and parietal EEG channels 300 ms after stimulus onset (Kelly & O'Connell, 2013) and is also termed P3 or late positive potential (LPP; Sun et al., 2017). While some studies investigated both response-locked and stimulus-locked CPP (Kelly & O'Connell, 2013), many only focused on stimulus-locked measures (Herding et al., 2019; Sun et al., 2017; Rausch et al., 2020). Although both measures were assumed to reflect the same component, which could be similarly related to confidence, different measures might lead to different relationship with confidence empirically. Specifically, Feuerriegel et al. (2022) suggested that the relationship between stimulus-locked CPP and confidence could be confounded by RT. Therefore, the following analysis examined whether stimulus-locked CPP amplitudes would show similar results as response-locked CPP amplitudes. This would clarify if such confound suggested by Feuerriegel et al. (2022) was involved in the previous analyses.

B.1.2. Late Pe

It has been proposed that the Pe is composed of two sub-components: early Pe (measured from around 200 to 450 ms after response) and late Pe (measured from around 400 to 600 ms after response; Moreau et al., 2022; Ruchsow et al., 2005). Functionally, while the early Pe was similar to ERN/Ne, the late Pe was more related to error awareness (Endrass et al., 2007; Ruchsow et al., 2005). As the Pe time window in Chapter 3 ranged from 300 to 400 ms relative to response, this time window might not fully capture the later subcomponent (and also the early Pe, which is not focused here as it is less relevant to error awareness). It is therefore possible that the later component might show different relationships with confidence. Based on previous findings on Pe, late Pe was expected to also differ between subjectively correct and subjectively incorrect trials. However, it might also show relationships with certainty of being correct or incorrect given its stronger relationship with error awareness.

B.1.3. ERN/Ne

Another ERP component related to performance monitoring is the ERN/Ne. Past studies have provided different proposals about what the ERN/Ne reflects (see Chapter 1), e.g., response conflict or conflict between error response and a corrective tendency (Scheffers & Coles, 2000; Di Gregorio et al., 2018). While past studies have shown that ERN/Ne amplitudes were larger for error trials, mixed findings have been reported regarding its relationships with error awareness and confidence. For example, Hewig et al. (2011) showed errors rated as incorrect (i.e., detected errors) showed larger ERN/Ne amplitudes than errors rated as guessing, which also showed larger amplitudes than errors rated as correct (i.e., undetected errors). This difference between detected and undetected errors were also found in other studies (Scheffers & Coles, 2000; Steinhauser & Yeung, 2010; Maier et al., 2008). On the other hand, Hewig et al. (2011) found no modulation by confidence in correct trials and Rausch et al. (2020) replicated this finding.

However, these studies used a more typical pre-response baseline correction procedure which could have biased the measurement of response-locked components due to amplitude differences in the baseline (as discussed in Chapter 3). In a visual discrimination task where pre-stimulus baseline correction was used, no ERN/Ne difference was found between detected and undetected errors (Pavone et al., 2009). Therefore, it could be expected that, in the current dataset where pre-stimulus baseline correction was used, ERN/Ne in error trials would not be modulated by confidence. However, as no previous studies have used a similar pre-stimulus baseline correction procedure and investigated correct trials, it is unclear how ERN/Ne in correct trials would be modulated.

B.2. Method

B.2.1. ERP measures

B.2.1.1. Stimulus-locked CPP

Previous studies measured stimulus-locked CPP at Pz from around 300 to 800 ms after stimulus onset (Tagliabue et al., 2019; Sun et al., 2017; depending on the morphology of the waveforms shorter time windows were also used, e.g., 300 to 400 ms [Del Cul et al., 2007]). Considering RTs in most trials in the current study were above 600ms, stimulus-locked CPP was defined as the averaged amplitude recorded at Pz from 400 to 600 ms after stimulus onset.

B.2.1.2. Late Pe

Previous studies have reported late Pe at Pz in the time intervals ranging from 250 to 750 ms (Endrass et al., 2007; Ruchsow et al., 2005; Moreau et al., 2022; Tops et al., 2013). Following these studies and using a time window not overlapping with that of the Pe in Chapter 3, the late Pe was defined as the averaged amplitudes recorded at Pz from around

400 to 600 ms. Note that to be consistent with the term used in Chapter 3, the label *Pe* is reserved to refer to the component as defined in Chapter 3 (300 to 400 ms relative to response).

B.2.1.3. ERN/Ne

Previous studies measuring ERN/Ne at FCz from around -40 to 200 ms after response (Falkenstein et al., 2001; Endrass et al., 2007; Boldt & Yeung, 2015]). Following these studies and considering that the ERN/Ne appeared to peak at around response in the current study, the ERN/Ne was defined as the averaged amplitude recorded at FCz from -50 to 50 ms in response-locked epochs.

B.2.2. Data analysis

The same processed dataset and linear mixed-effects model analysis approach as in Chapter 3 was used for the following analyses. First, the above additional ERP measures of stimulus-locked CPP amplitudes, ERN/Ne amplitudes, and late Pe amplitudes in correct and error trials were predicted by a trichotomized subjective accuracy measure ("subjectively incorrect" [confidence ratings < 4], "guessing" [confidence ratings = 4], and "subjectively correct" [confidence ratings > 4]). In separate sets of mixed-effects models, the same ERP measures were then predicted by the graded measures of certainty recoded from confidence ratings: certainty of being correct (confidence > 4) and certainty of being in correct (confidence < 4). As in Chapter 3, ERP waveforms by subjective accuracy and all confidence levels were plotted. Full statistical results are reported in Tables B1 - B18, Section B.5.

B.3. Results

B.3.1. Stimulus-locked CPP

To examine how the relationships between response-locked CPP and confidence measures could be extended to stimulus-locked CPP, the effects of subjective accuracy and certainty on stimulus-locked CPP amplitudes were tested.

First, for stimulus-locked CPP amplitudes, subjective accuracy in correct trials showed a significant effect ($p < .001$). Pairwise comparisons showed that it was driven by the fact that subjectively incorrect showed lower amplitudes than guessing trials ($p = .011$) and subjectively correct trials ($p < .001$). However, such effect was not observed in error trials ($p = .503$). Stimulus-locked waveforms by subjective accuracy, separated for correct and error trials, are shown in Figure B1.

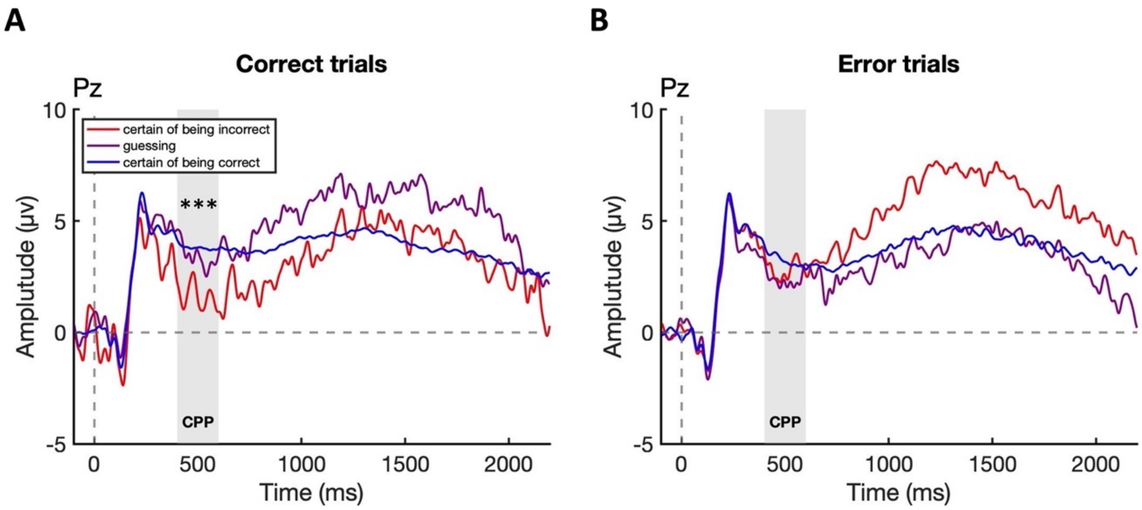

Figure B1. (A, B) Stimulus-locked waveforms at Pz by subjective accuracy, separated for correct and error trials. Stimulus-locked CPP amplitudes differed by subjective accuracy in correct trials but not in error trials.
Note. *p <.05 **p <.01 ***p <.001.

Further, certainty of being correct also had an effect on stimulus-locked CPP amplitudes in correct trials ($p < .001$) and trend analysis showed a linear effect ($p < .001$), suggesting that its amplitudes increased with higher certainty of being correct. However, certainty of being incorrect did not predict CPP amplitudes ($p = .069$). For error trials, CPP amplitudes were neither predicted by certainty of being correct ($p = .353$) or certainty of being incorrect ($p = .433$). Stimulus-locked waveforms at Pz by all levels of confidence ratings, separated for correct and error trials, are shown in Figure B2. Overall, the patterns of results of stimulus-locked CPP are consistent with the results based on response-locked CPP amplitudes.

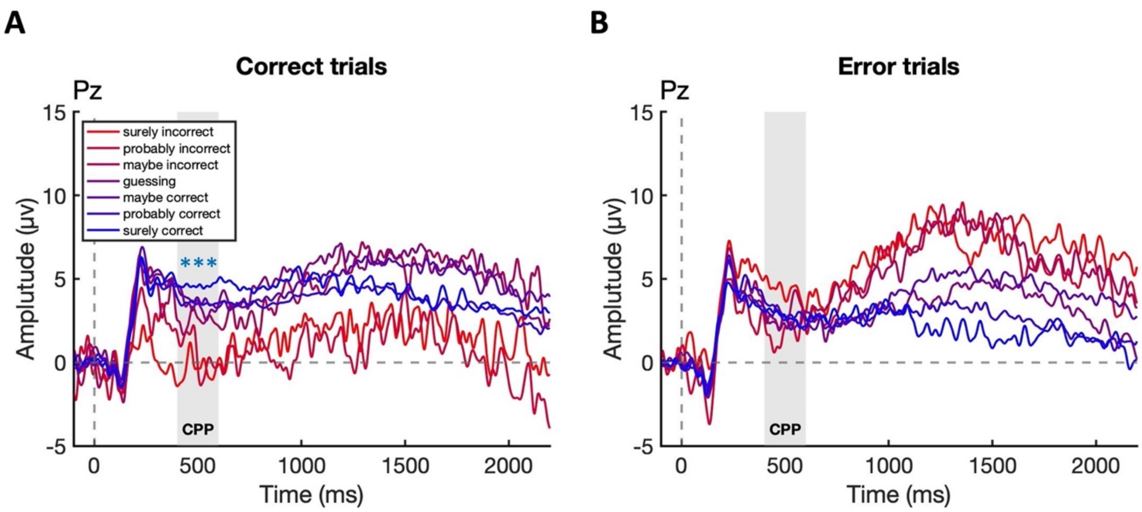

Figure B2. (A, B) Stimulus-locked waveforms at Pz by all confidence rating levels, separated for correct and error trials. Shaded areas show the time windows of the stimulus-locked CPP (400 to 600 ms). Stimulus-locked CPP only differ by certainty of being correct in correct trials. Error bars indicate SEM.

Note. *p <.05 **p <.01 ***p <.001. Blue asterisks indicate the p value for the relationship with certainty of being correct.

B.3.2. Late Pe

As Pe defined in Chapter 3 was related to confidence only in error trials, the following analysis examine whether this finding extends the late Pe measure. Consistent with the findings based on Pe amplitudes, late Pe amplitudes also differed by subjective accuracy for both correct (p = .011) and error trials (p < .001), as shown by Figure B3. The effect in correct trials, however, was driven by the difference between guessing trials and subjectively correct trials (p = .010) without a linear trend (p = .399). On the other hand, the effect in error trials was driven by the pattern that both subjectively incorrect trials and guessing trials had larger amplitude than subjectively correct trials (ps < .01), with a linear trend (p < .001).

Waveforms of all confidence levels are presented in Figure B4. In error trials, late Pe amplitudes were related to certainty of being correct (p = .016) and a linear trend was observed (p = .004), suggesting that lower certainty of being correct was related to larger Pe amplitudes. Late Pe amplitudes were however, not related to certainty of being incorrect in error trials, or any certainty in correct trials. Therefore, the late Pe showed the same patterns of results as the Pe except for the above-mentioned linear trend, which was absent for the Pe.

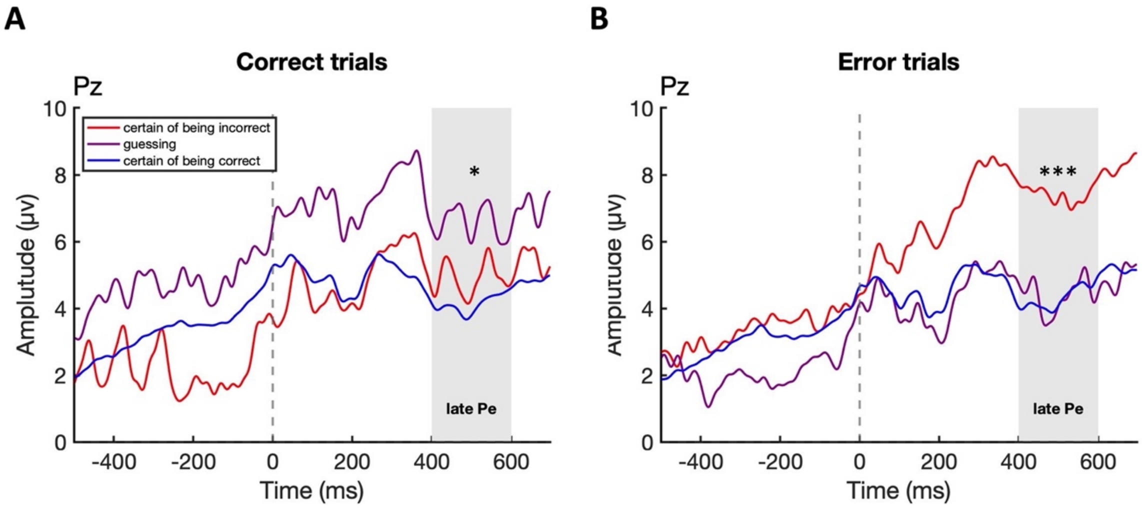

Figure B3. (A, B) Response-locked waveforms at Pz by subjective accuracy, separated for correct and error trials. Shaded areas show the time windows of the late Pe (400 to 600 ms). Late Pe amplitudes differed by subjective accuracy in both correct ad error trials.
Note. *p <.05 **p <.01 ***p <.001.

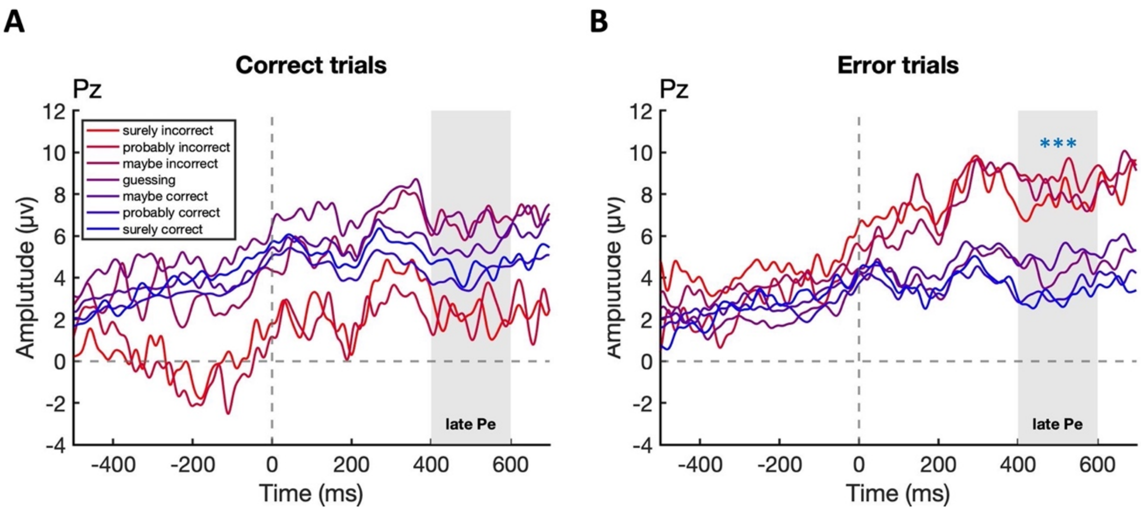

Figure B4. (A, B) Response-locked waveforms at Pz by all confidence rating levels, separated for correct and error trials. Shaded areas show the time windows of the late Pe (400 to 600 ms). Late Pe amplitudes differed by certainty of being correct in error trials. Error bars indicate SEM.

Note. *p <.05 **p <.01 ***p <.001. Blue asterisks indicate the p value for the relationship with certainty of being correct.

B.3.3. ERN/Ne

When ERN/Ne amplitudes were predicted by subjective accuracy, a significant effect was found in correct trials (p = .006) but not in error trials (p = .101). Post-hoc comparisons showed that the effect in correct trials was due to the fact that subjectively incorrect trials showed lower amplitudes than guessing trials (p = .004), and subjectively correct trials (p = .036). However, ERN/Ne amplitudes were not predicted by certainty of being correct or certainty of being incorrect, in either correct or error trials. Response-locked waveforms at FCz by subjective accuracy, as well as by all levels of confidence ratings, separated for correct and error trials, are shown in Figures B5 and B6.

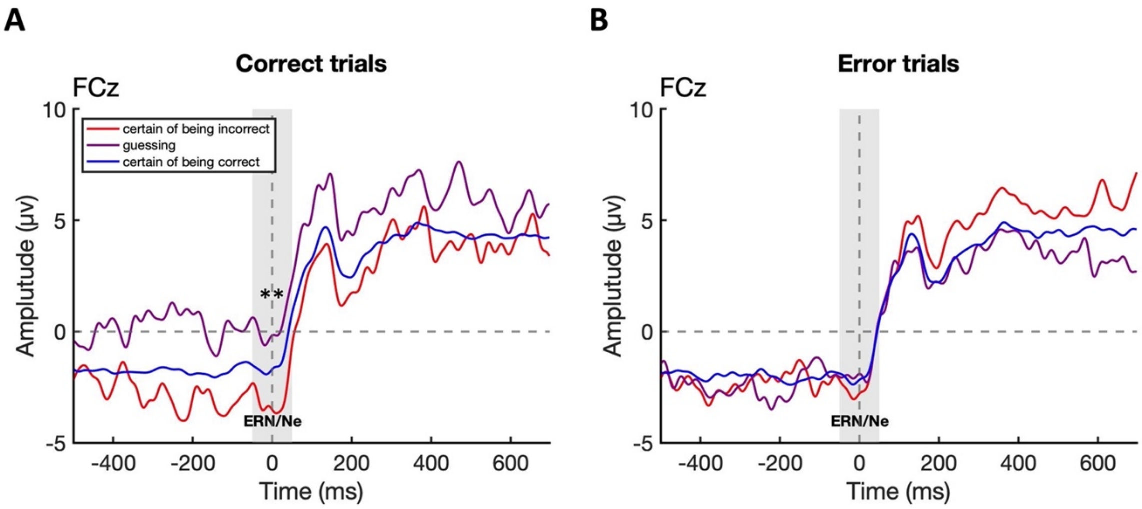

Figure B5. (A, B) Response-locked waveforms at FCz by subjective accuracy, separated for correct and error trials. Shaded areas show the time windows of the ERN/Ne (-50 to 50 ms). ERN/Ne differed by subjective accuracy in correct trials but not in error trials.
Note. *p <.05 **p <.01 ***p <.001.

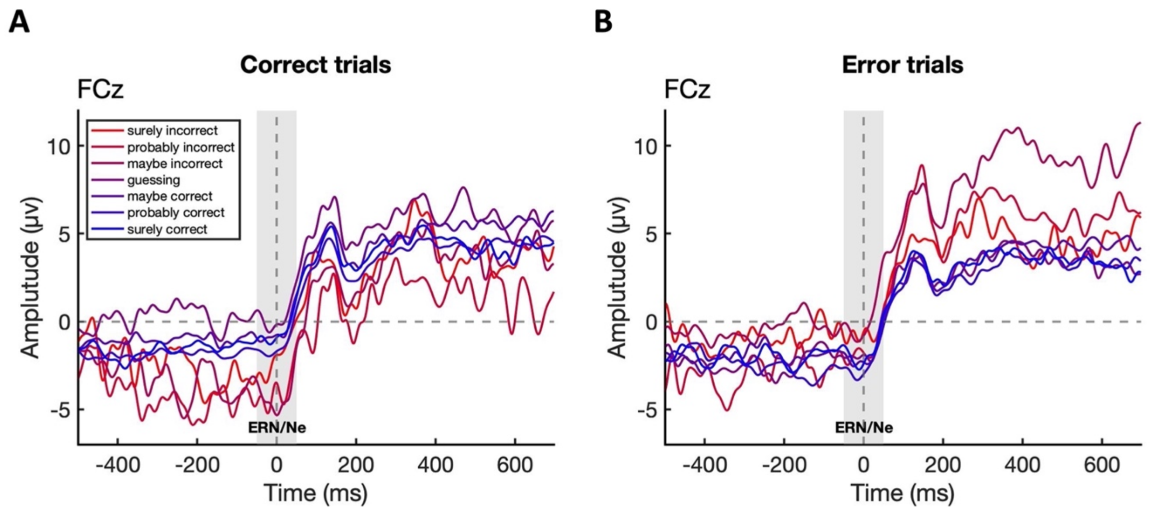

Figure B6. (A, B) Response-locked waveforms at FCz by all confidence rating levels, separated for correct and error trials. Shaded areas show the time windows of the ERN/Ne (-50 to 50 ms). ERN/Ne amplitudes were not modulated by certainty of being correct or incorrect. Error bars indicate SEM.

B.4. Summary

B.4.1. Stimulus-locked CPP

Compared with response-locked CPP amplitudes (discussed in Chapter 3), stimulus-locked CPP amplitudes showed the same patterns of being modulated by subjective accuracy in correct trials, as well as certainty of being correct in correct trials. This suggests that both measures are similarly related to confidence despite the suggestion that RT could confound this relationship for stimulus-locked CPP (Feuerriegel et al., 2022). In fact, this is consistent with the results reported by Feuerriegel and colleagues, as they also found the stimulus-locked CPP amplitudes were related to confidence in correct trials but not in error trials.

B.4.2. Late Pe

In terms of subjective accuracy and certainty, the late Pe showed similar patterns as the Pe (discussed in Chapter 3): The late Pe amplitudes were negatively related to subjective accuracy in error trials, while in correct trials the amplitudes were larger for guessing trials than subjectively correct trials. The only differences between the Pe and late Pe measures were that, first, in error trials only the late Pe amplitudes were negatively related to certainty of being correct, and second, in correct trials the late Pe amplitudes were not related to certainty of being correct as the Pe.

While the overall findings suggest that the Pe and late Pe measures in the current dataset are highly similar, the first inconsistent finding suggests that the late Pe could be more sensitive to the graded differences in certainty of being correct in error trials. This may be because late Pe is more closely related to confidence than the early Pe (Endrass et al., 2007; Ruchsow et al., 2005). Indeed, the Pe time window in the current dataset was close to the time window of some early Pe measures, e.g., 300 to 450 ms [Moreau et al., 2022]). Considering the long response deadline (1500 ms) for confidence ratings in the task, it is also possible that the accumulation of error evidence continued beyond the measurement window of Pe (300 to 400 ms relative to response), such that more graded differences emerged only at a later stage (potentially because of more sources of information entering the error evidence accumulation process), captured by the late Pe measurement window (400 to 600 ms). This would then suggest that the post-decisional metacognitive process involves binary error awareness in the early stage, but then develops more fine-grained evaluation. This also suggests that even in cases where subjective certainty of being correct was observed in error trials, error evidence could still be negatively related to accuracy ratings, rather than simply driven by noise. Due to the inconsistency between the measures of Pe and late Pe, this hypothesis remains to be tested in future studies. It should however be noted that even if such relationship is true, the Pe still appears to be more sensitive to error awareness, as its amplitudes for subjectively incorrect trials were considerably larger compared with subjectively correct trials.

The second finding that the late Pe amplitudes in correct trials were not related to certainty of being correct as the Pe could be attributed to the noisy accumulation of error evidence. As discussed in Chapter 3, it could be assumed that error evidence is accumulated ineffectively in correct trials. This noise could have driven the relationship between the Pe and subjective accuracy as well as certainty of being correct in correct trials, which did not sustain in the later time window of the late Pe.

B.4.3. ERN/Ne

ERN/Ne amplitudes were modulated by subjective accuracy only in correct trials but not error trial. The null finding in error trials is inconsistent with previous studies that found subjectively incorrect trials had larger ERN using a pre-response baseline (e.g., Scheffers & Coles, 2000), but consistent with a study that did not find such an effect using a pre-stimulus baseline (Pavone et al., 2009). It is therefore possible that it could be explained by these different baseline correction procedures. For example, as errors rated as correct might have larger CPP overlapping with baseline compared with errors rated as incorrect, the true difference between subjectively correct and incorrect trials in ERN/Ne might be reduced by correction with the pre-response baseline.

The null effect of subjective accuracy on the ERN/Ne in error trials was previously interpreted as evidence showing that the ERN/Ne reflects unconscious error detection, such that similar ERN/Ne amplitudes were observed across subjective accuracy levels (Pavone et al., 2009). However, the current analysis also showed an effect of subjective accuracy in correct trials, which was only rarely investigated in previous studies (Scheffers & Coles, 2000). Two interpretations could be considered. First, if ERN/Ne reflects response conflict (Scheffers & Coles, 2000; Di Gregorio et al., 2018), then it suggests that the larger ERN/Ne amplitudes for correct trials rated as incorrect involve larger conflict than correct trials rated as correct, where the source of conflict could be related to the difference between subjective and objective accuracy. Although using a pre-response baseline, Scheffers and Coles (2000) also similarly reported that ERN/Ne amplitudes were the largest for correct trials rated as subjectively incorrect, even when compared with error trials rated as subjectively incorrect, showing that ERN/Ne amplitudes might not be only dependent on error awareness. Second, such effect might reflect differences due to stimulus processing, as such differences emerged before response (Figure 3.7A, Chapter 3).

It should be noted that the ERN/Ne was not as large as in previous studies (e.g., Scheffers & Coles, 2000). Reduced ERN/Ne amplitudes could be due to the speeded nature of the task paradigm (Arbel & Donchin, 2009; Gehrin et al., 1993) or pre-stimulus baseline correction (Pavone et al., 2009).

Overall, the results showed largely similar findings as reported in Chapter 3, thus strengthening the claim that the CPP is specific to confidence in correct decisions and Pe/late Pe is specific to confidence in error decisions, even when alternative measures were used. Additionally, the ERN/Ne showed modulation by subjective accuracy only in correct trials, which could be due to larger response conflict when correct trials were rated as incorrect.

B.5. Linear mixed-effects model results (regression coefficients)

Mixed-effects models results of ERP data analysis: Subjective accuracy

Table B1 Regression Coefficients for Predicting Stimulus-locked CPP amplitudes (Correct Trials) from Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	2.96	0.50	5.88	<.001***
Subjectively incorrect	-1.12	0.31	-3.62	<.001***
Guessing	0.37	0.23	1.58	.115

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table B2 Regression Coefficients for Predicting Stimulus-locked CPP amplitudes (Error Trials) from Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	3.00	0.46	6.48	<.001***
Subjectively incorrect	-0.31	0.27	-1.15	.248
Guessing	0.17	0.24	0.71	.479

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table B3 Regression Coefficients for Predicting Late Pe amplitudes (Correct Trials) from Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	0.00	0.01	0.10	.924
Subjectively incorrect	-0.02	0.01	-1.58	.114
Guessing	-0.00	0.01	-0.17	.867

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table B4 Regression Coefficients for Predicting Late Pe amplitudes (Error Trials) from Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	0.01	0.01	2.08	.042*
Subjectively incorrect	-0.00	0.01	-0.40	.692
Guessing	0.00	0.01	0.12	.904

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table B5 Regression Coefficients for Predicting ERN/Ne amplitudes (Correct Trials) from Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	4.76	0.70	6.82	<.001***
Subjectively incorrect	-0.04	0.48	-0.08	.938
Guessing	0.66	0.36	1.85	.065

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table B6 Regression Coefficients for Predicting ERN/Ne amplitudes (Error Trials) from Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	5.64	0.67	8.42	<.001***
Subjectively incorrect	1.15	0.43	2.71	.007**
Guessing	0.17	0.38	0.46	.642

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Mixed-effects models results of ERP data analysis: Certainty of being correct and certainty of being incorrect

Table B7 Regression Coefficients for Predicting Stimulus-locked CPP amplitudes (Correct Trials) from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	3.58	0.52	6.87	<.001***
Maybe correct	-0.75	0.17	-4.36	<.001***
Probably correct	-0.18	0.14	-1.29	.198

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table B8 Regression Coefficients for Predicting Stimulus-locked CPP amplitudes (Error Trials) from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	3.16	0.49	6.41	<.001***
Maybe correct	-0.35	0.26	-1.37	.169
Probably correct	0.04	0.22	0.19	.852

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table B9 Regression Coefficients for Predicting Late Pe amplitudes (Correct Trials) from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	0.02	0.00	3.61	.001**
Maybe correct	-0.00	0.01	-0.03	.974
Probably correct	-0.00	0.00	-0.76	.449

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table B10 Regression Coefficients for Predicting Late Pe amplitudes (Error Trials) from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	0.01	0.01	2.49	.021*
Maybe correct	0.01	0.01	0.94	.347
Probably correct	-0.00	0.01	-0.51	.609

Note. Intercept represents the estimate for surely correct. *p <.05 **p <.01 ***p <.001

Table B11 Regression Coefficients for Predicting ERN/Ne amplitudes (Correct Trials) from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	4.10	0.65	6.26	<.001***
Maybe correct	-0.06	0.26	-0.21	.833
Probably correct	-0.13	0.21	-0.59	.554

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table B12 Regression Coefficients for Predicting ERN/Ne amplitudes (Error Trials) from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	4.35	0.62	7.04	<.001***
Maybe correct	1.00	0.39	2.57	.010*
Probably correct	0.01	0.34	0.03	.979

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table B13 Regression Coefficients for Predicting Stimulus-locked CPP amplitudes (Correct Trials) from Certainty of Being incorrect

Parameters	Estimate	SE	z	p
Intercept	1.32	0.65	2.02	.052
Maybe incorrect	-1.16	0.89	-1.31	.192
Probably incorrect	-0.41	0.71	-0.57	.568

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Table B14 Regression Coefficients for Predicting Stimulus-locked CPP amplitudes (Error Trials) from Certainty of Being incorrect

Parameters	Estimate	SE	z	p
Intercept	2.83	0.64	4.40	<.001***
Maybe incorrect	0.78	0.64	1.22	.222
Probably incorrect	-0.16	0.60	-0.26	.793

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Table B15 Regression Coefficients for Predicting Late Pe amplitudes (Correct Trials) from Certainty of Being incorrect

Parameters	Estimate	SE	z	p
Intercept	-0.01	0.02	-0.73	.467
Maybe incorrect	0.02	0.03	0.70	.485
Probably incorrect	-0.03	0.02	-1.33	.186

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Table B16 Regression Coefficients for Predicting Late Pe amplitudes (Error Trials) from Certainty of Being incorrect

Parameters	Estimate	SE	z	p
Intercept	0.00	0.01	0.29	.776
Maybe incorrect	-0.00	0.02	-0.17	.868
Probably incorrect	-0.02	0.02	-1.22	.225

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Table B17 Regression Coefficients for Predicting ERN/Ne amplitudes (Correct Trials) from Certainty of Being incorrect

Parameters	Estimate	SE	z	p
Intercept	4.46	1.03	4.31	<.001***
Maybe incorrect	-0.56	1.37	-0.41	.685
Probably incorrect	-0.61	1.09	-0.56	.578

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Table B18 Regression Coefficients for Predicting ERN/Ne amplitudes (Error Trials) from Certainty of Being incorrect

Parameters	Estimate	SE	z	p
Intercept	7.05	1.10	6.38	<.001***
Maybe incorrect	-0.89	1.05	-0.85	.397
Probably incorrect	1.46	0.98	1.49	.136

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Appendix C Effects of relative and absolute evidence on ERP measures

C.1. Background

Given that the CPP and Pe are assumed to reflect sensory and error evidence accumulation, the manipulations of relative and absolute evidence strength were likely to modulate the amplitudes of these two components. The following analyses explore whether such effects would be observed.

C.1.1. CPP

As discussed in Chapter 1, CPP was proposed to reflect accumulation of sensory evidence, and specifically task-relevant evidence (Kelly et al., 2021). Therefore, a direct effect of relative evidence could be expected, such that CPP amplitudes would be larger when relative evidence was stronger.

In terms of the effect of absolute evidence, two potential effects could be expected. First, as stronger absolute evidence is assumed to reduced perceived relative evidence (discussed in Chapter 2), stronger absolute evidence might lead to weaker CPP amplitudes in general, or modulate the effect of relative evidence on CPP, such that the effect was stronger when absolute evidence was weak. On the other hand, given that some studies suggest that stimulus-locked CPP could reflect visual awareness of stimuli (Tagliabue et al., 2016, 2019), stimuli with stronger absolute evidence could have appeared more visible and thus increased stimulus-locked CPP amplitudes. However, such proposal was based on stimuli with intensity close to detection threshold (Tagliabue et al., 2016, 2019) and might not extend to the highly visible stimuli in the current study.

C.1.2. Pe

The Pe was proposed to reflect the accumulation of post-decisional sensory evidence, or alternatively, error evidence (Desender et al., 2021; Moran et al., 2015; Rausch et al., 2020). If Pe simply reflects sensory evidence accumulation, then the same predictions made for the CPP could also be made for the Pe: The Pe should increase with stronger relative evidence and potentially with weaker absolute evidence, for both correct and error trials. However, such prediction would be at least partially inconsistent with the previous findings that Pe amplitudes increased with lower confidence (Boldt & Yeung, 2015), which was associated with weaker relative evidence and weaker absolute evidence in the current behavioural data.

On the other hand, as it has been suggested in the literature and the analyses in Chapter 3 that Pe is likely to be error-specific, it could be considered to reflect error evidence accumulation (discussed in Chapter 3). Although it is unclear how error evidence is related to relative evidence and absolute evidence, the effects of these manipulations on the likelihood of errors being rated as incorrect (analysis on changes of mind, reported in Chapter 2) suggest the following predictions. Given that stronger absolute evidence reduced the likelihood of errors being rated as incorrect, less error evidence should be accumulated when absolute evidence was strong. In contrast, as stronger relative evidence increased the likelihood of errors being rated as incorrect, more error evidence should be accumulated. Therefore, it was predicted that the Pe in error trials was stronger when absolute evidence was weak and when relative evidence was strong. Pe amplitudes in correct trials, which should reflect limited error evidence accumulation, however, should not be affected.

C.2. Method

The same processed dataset, measures, and linear mixed-effects model analysis approach in Appendix B was used for the following analyses. It involved examining how relative evidence and absolute evidence affected the amplitudes of stimulus-locked CPP, response-locked CPP, Pe, and late Pe. Linear mixed-effects models with random intercept for participants were used to test the effects of relative evidence, absolute evidence, and their interaction. Full statistical results are presented in Tables C1 - C16, Section C.5..

C.3. Results

C.3.1. Stimulus-locked and response-locked CPP

No effects on the stimulus-locked CPP measure reached significance, but response-locked CPP in error trials showed an interaction between relative and absolute evidence (p = .001). This interaction was due the pattern that CPP amplitudes were not affected by absolute evidence at low relative evidence level but negatively related to absolute evidence when relative evidence was high. Post-hoc comparisons showed that at high relative evidence, CPP amplitudes were lower for high absolute evidence than low absolute evidence (p =.007). However, at medium relative evidence, CPP amplitudes were lower for low absolute evidence than high absolute evidence (p = .031). Response-locked waveforms at Pz by absolute evidence levels, separated for relative evidence levels, correct and error trials, are shown in Figure C1.

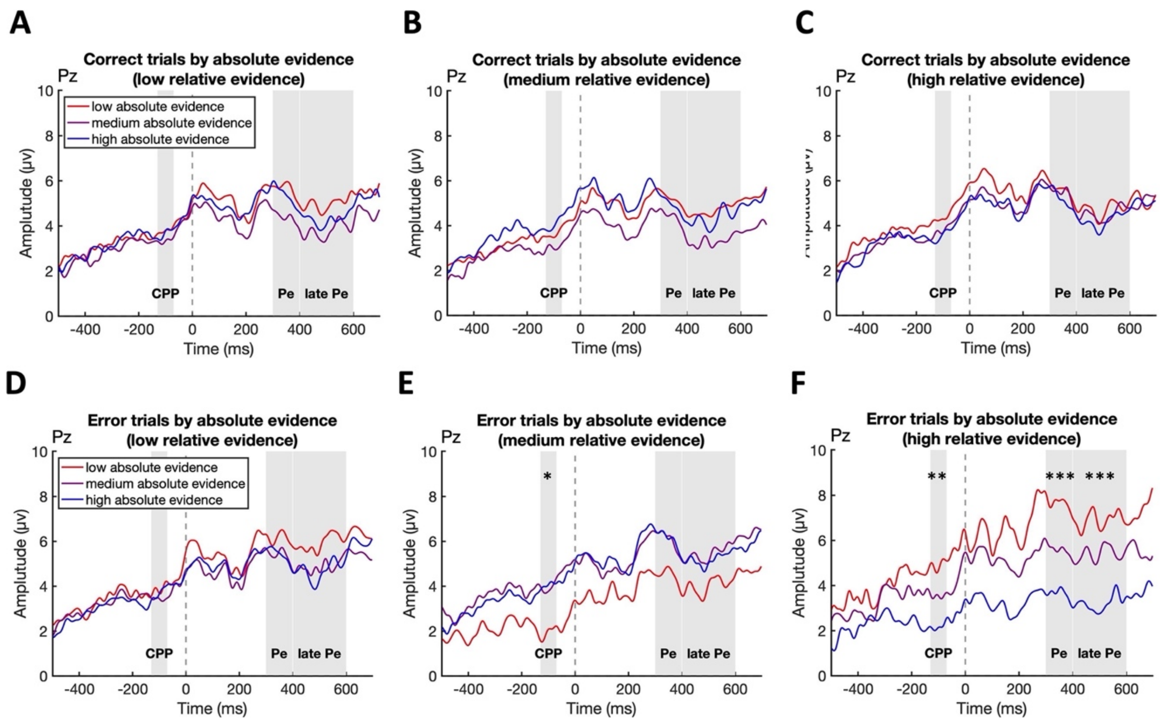

Figure C1. (A-C) Response-locked waveforms in error trials by absolute evidence levels, separated for relative evidence levels in correct trials. (D-F) Response-locked waveforms in error trials by absolute evidence levels, separated for relative evidence levels in error trials. Shaded areas show the time windows of the Pe (300 to 400 ms) and the late Pe (400 to 600

ms). CPP, Pe, and late Pe amplitudes showed interactions between relative and absolute evidence in error trials, while such effects were not observed for correct trials.
Note. **p* <.05 ***p* <.01 ****p* <.001

C.3.2. ***Pe and late Pe***

Both Pe and late Pe showed similar patterns of results. In error trials, they both showed an negative effect of absolute evidence with a linear trend (*p*s < .05), as well as an interaction between relative evidence and absolute evidence (*p*s < .01). Post-hoc comparisons showed that this was due to significant negative relationships between absolute evidence and Pe/late Pe amplitudes only at high relative evidence level (*ps* < .001; Figure C1). Additionally, in correct trials, only late Pe showed a main effect of absolute evidence (*p* = .036). However, this was driven by the difference between low and medium absolute evidence levels (*p* = .035). Response-locked waveforms at Pz by absolute evidence, separated for correct and error trials, are shown in Figure C2.

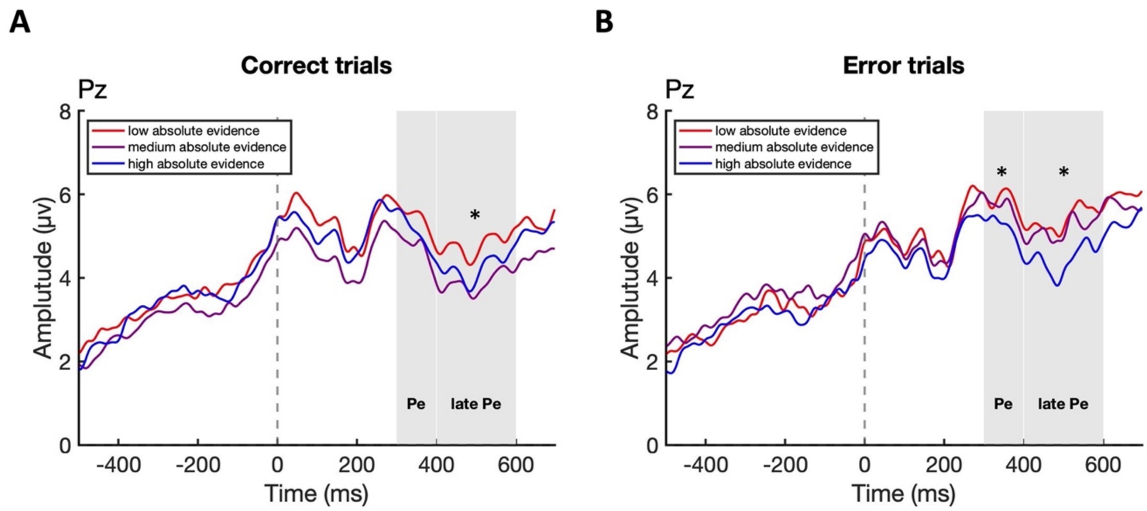

Figure C2. (A, B) Response-locked waveforms at Pz in error trials by absolute evidence levels. Shaded areas show the time windows of the Pe (300 to 400 ms) and the late Pe (400 to 600 ms). Pe amplitudes differed by absolute evidence levels in error trials only, while late Pe amplitudes differed by absolute evidence levels in both correct and error trials.
Note. **p* <.05 ***p* <.01 ****p* <.001

C.4. Summary

C.4.1. Stimulus-locked and response-locked CPP

Instead of the expected positive effect of relative evidence and negative effect of absolute evidence, it was found that only response-locked CPP amplitudes in error trials were affected by an interaction between relative and absolute evidence, which was driven by the pattern that stronger absolute evidence reduced CPP only when relative evidence was strong. This suggests that absolute evidence might have reduced sensory evidence as stronger absolute evidence reduced perceived relative evidence (perceived brightness difference; discussed in Chapter 2), but this effect was only limited to high level of relative evidence. At medium relative evidence, stronger absolute evidence was however related to lower CPP amplitudes. These opposite directions of effect might suggest that in error trials sensory evidence accumulation was less stable (particularly at when task difficulty was higher), that sensory evidence accumulation did not show the expected effect.

This further suggests that the CPP is more closely related to subjective sensory evidence rather than objective stimulus strength. On the one hand, the absence of effect in correct trials is however consistent with the previous findings that showed stronger relative evidence did not increase the amount of evidence accumulated before response (Kelly & O'Connell, 2013). In other words, in correct trials, a fixed amount of sensory evidence was accumulated, regardless of the objective stimulus strength. On the other hand, the fact that CPP amplitudes in error trials were affected by absolute evidence strength suggests that error trials involved different amounts of evidence accumulated, which was supressed when objective stimulus strength was weak due to increased absolute evidence.

The fact that stimulus-locked CPP amplitudes were not significantly predicted by relative and absolute evidence might be because in the current task stimuli were presented for a long period and therefore the stimulus-locked CPP did not show a clear build-up pattern that reflects evidence accumulation more clearly.

C.4.2. Pe and late Pe

In terms of the effects of relative and absolute evidence, late Pe again showed similar patterns of results as Pe: Both measures in error trials showed an interaction between relative and absolute evidence driven by their amplitudes being reduced with stronger absolute evidence, but only at high level of relative evidence. One interpretation is that effective accumulation of error evidence requires both high relative evidence and low absolute evidence. When relative evidence was low or medium, error evidence accumulation was limited (and error awareness rarely occurred) and thus cannot be affected by absolute evidence. Only when relative evidence was high, and evidence accumulation could effectively occur, lower absolute evidence allowed more error evidence to be accumulated compared with higher absolute evidence. Considering absolute evidence as noise in the decision process (Ratcliff et al., 2018; Turner et al., 2021), this would suggest that increased noise impaired the error awareness.

The late Pe also showed an additional, negative effect of absolute evidence in correct trials. As discussed in Chapter 3, despite the prediction that the Pe would only reflect error evidence after erroneous decisions, some error evidence could also be available after correct decisions. If that is true, the effect of absolute evidence in correct trials could be explained similarly as in error trials, that lower absolute evidence allowed more error evidence to be accumulated.

C.5. Linear mixed-effects model results (likelihood ratio tests and regression coefficients)

Mixed-effects models results of ERP data analysis: Stimulus-locked CPP and response locked CPP

Table C1 Likelihood Ratio Tests Results for Predicting Stimulus-locked CPP amplitudes (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	$\chi 2$	p
Rel	2	4.56	.102
Abs	2	4.97	.083
Rel × Abs	4	9.19	.057

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C2 Regression Coefficients for Predicting Stimulus-locked CPP amplitudes (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	3.58	0.48	7.46	<.001***
Low Rel	-0.11	0.13	-0.90	.368
Med Rel	-0.14	0.12	-1.12	.263
Low Abs	-0.14	0.12	-1.21	.228
Med Abs	-0.13	0.12	-1.09	.278
Low Rel × Low Abs	-0.29	0.18	-1.63	.103
Med Rel × Low Abs	-0.17	0.17	-1.03	.305
Low Rel × Med Abs	0.17	0.18	0.96	.335
Med Rel × Med Abs	-0.07	0.17	-0.40	.687

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C3 Likelihood Ratio Tests Results for Predicting Stimulus-locked CPP amplitudes (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	χ2	p
Rel	2	0.71	.700
Abs	2	1.86	.395
Rel × Abs	4	7.35	.118

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C4 Regression Coefficients for Predicting Stimulus-locked CPP amplitudes (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	3.08	0.45	6.78	<.001***
Low Rel	0.13	0.17	0.78	.438
Med Rel	-0.10	0.18	-0.57	.567
Low Abs	-0.23	0.19	-1.18	.238
Med Abs	0.22	0.18	1.23	.220
Low Rel × Low Abs	0.15	0.25	0.60	.551
Med Rel × Low Abs	-0.56	0.26	-2.16	.031*
Low Rel × Med Abs	-0.35	0.24	-1.46	.144
Med Rel × Med Abs	0.23	0.25	0.93	.353

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C5 Likelihood Ratio Tests Results for Predicting Response-locked CPP amplitudes (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	χ2	p
Rel	2	3.81	.149
Abs	2	3.79	.151
Rel × Abs	4	3.48	.480

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C6 Regression Coefficients for Predicting Response-locked CPP amplitudes (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	3.68	0.55	6.72	<.001***
Low Rel	-0.14	0.15	-0.92	.357
Med Rel	-0.14	0.15	-0.93	.353
Low Abs	0.22	0.15	1.50	.134
Med Abs	-0.27	0.15	-1.80	.072
Low Rel × Low Abs	-0.07	0.21	-0.31	.753
Med Rel × Low Abs	-0.15	0.21	-0.71	.477
Low Rel × Med Abs	0.05	0.22	0.23	.819
Med Rel × Med Abs	-0.17	0.21	-0.81	.417

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C7 Likelihood Ratio Tests Results for Predicting Response-locked CPP amplitudes (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	$\chi 2$	p
Rel	2	0.61	.737
Abs	2	0.88	.645
Rel × Abs	4	18.17	.001**

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C8 Regression Coefficients for Predicting Response-locked CPP amplitudes (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	3.44	0.55	6.25	<.001***
Low Rel	0.16	0.21	0.75	.454
Med Rel	-0.10	0.22	-0.46	.646
Low Abs	0.13	0.24	0.55	.585
Med Abs	0.07	0.22	0.32	.748
Low Rel × Low Abs	0.08	0.31	0.26	.794
Med Rel × Low Abs	-1.13	0.32	-3.52	<.001***
Low Rel × Med Abs	-0.38	0.30	-1.29	.199
Med Rel × Med Abs	0.31	0.31	1.01	.310

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Mixed-effects models results of ERP data analysis: Pe and late Pe

Table C9 Likelihood Ratio Tests Results for Predicting Pe amplitudes (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	$\chi 2$	p
Rel	2	3.27	.195
Abs	2	4.93	.085
Rel × Abs	4	2.06	.724

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C10 Regression Coefficients for Predicting Pe amplitudes (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	5.04	0.65	7.82	<.001***
Low Rel	-0.03	0.19	-0.16	.875
Med Rel	-0.25	0.18	-1.42	.157
Low Abs	0.34	0.18	1.92	.054
Med Abs	-0.34	0.18	-1.89	.059
Low Rel × Low Abs	0.12	0.26	0.46	.648
Med Rel × Low Abs	0.04	0.25	0.15	.884
Low Rel × Med Abs	-0.26	0.26	-0.98	.326
Med Rel × Med Abs	-0.08	0.25	-0.33	.740

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C11 Likelihood Ratio Tests Results for Predicting Pe amplitudes (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	$\chi 2$	p
Rel	2	0.11	.948
Abs	2	6.05	.048*
Rel × Abs	4	16.93	.002**

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C12 Regression Coefficients for Predicting Pe amplitudes (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	5.58	0.65	8.61	<.001***
Low Rel	-0.01	0.25	-0.05	.960
Med Rel	-0.08	0.26	-0.29	.773
Low Abs	0.59	0.29	2.07	.039*
Med Abs	-0.02	0.27	-0.06	.950
Low Rel × Low Abs	-0.09	0.37	-0.25	.802
Med Rel × Low Abs	-1.19	0.39	-3.06	.002**
Low Rel × Med Abs	-0.34	0.35	-0.95	.343
Med Rel × Med Abs	0.19	0.37	0.51	.613

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C13 Likelihood Ratio Tests Results for Predicting Late Pe amplitudes (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	$\chi 2$	p
Rel	2	1.09	.579
Abs	2	6.64	.036*
Rel × Abs	4	3.37	.498

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C14 Regression Coefficients for Predicting Late Pe amplitudes (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	4.29	0.66	6.55	<.001***
Low Rel	-0.05	0.19	-0.27	.789
Med Rel	-0.13	0.19	-0.69	.487
Low Abs	0.45	0.18	2.43	.015*
Med Abs	-0.35	0.19	-1.89	.058
Low Rel × Low Abs	0.17	0.27	0.63	.528
Med Rel × Low Abs	0.06	0.26	0.25	.806
Low Rel × Med Abs	-0.08	0.27	-0.29	.775
Med Rel × Med Abs	-0.34	0.26	-1.28	.202

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C15 Likelihood Ratio Tests Results for Predicting Late Pe amplitudes (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	df	$\chi 2$	p
Rel	2	0.59	.746
Abs	2	8.03	.018*
Rel × Abs	4	13.63	.009**

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table C16 Regression Coefficients for Predicting Late Pe amplitudes (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	SE	z	p
Intercept	5.04	0.64	7.82	<.001***
Low Rel	0.01	0.27	0.02	.982
Med Rel	-0.20	0.27	-0.73	.466
Low Abs	0.67	0.30	2.22	.027*
Med Abs	0.06	0.28	0.21	.832
Low Rel × Low Abs	0.05	0.39	0.13	.894
Med Rel × Low Abs	-1.09	0.41	-2.68	.007**
Low Rel × Med Abs	-0.41	0.37	-1.10	.272
Med Rel × Med Abs	0.10	0.39	0.27	.789

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Appendix D Relationships between ERP measures and certainty of being correct/incorrect regardless of objective accuracy

D.1. Method and results

In Chapter 3, the relationships between confidence and ERP measures were examined first in data subsets defined by objective accuracy, and then in data subsets defined by the combination of objective and subjective accuracy. To complement the results in Chapter 3, and to perform comparable analyses with Feuerriegel and colleagues (2022), the same analyses were repeated in data subsets defined by subjective accuracy only and reported here. Full statistical results are presented in Tables D1 – D6, Section D.3.. Waveforms separated by subjective accuracy and all confidence levels are presented in Figurer D1.

CPP amplitudes were predicted by subjective accuracy (p = .001), and pairwise comparisons showed significantly lower CPP amplitudes for subjectively incorrect trials than guessing trials (p = .001) and subjectively correct trials (p = .005). When CPP amplitudes in correct trials were predicted by certainty of being incorrect, no significant effect was found (p = .919). CPP amplitudes were however significantly predicted by certainty of being correct (p <.001). Trend analysis show a significant linear relationship (p < .001).

Pe amplitudes were predicted by subjective accuracy (p < .001), and pairwise comparisons showed significantly lower Pe amplitudes for subjectively correct trials compared with guessing trials and subjectively incorrect trials (ps < .001). Pe amplitudes were however not predicted by certainty of being correct (p = .159) or certainty of being incorrect (p = .810).

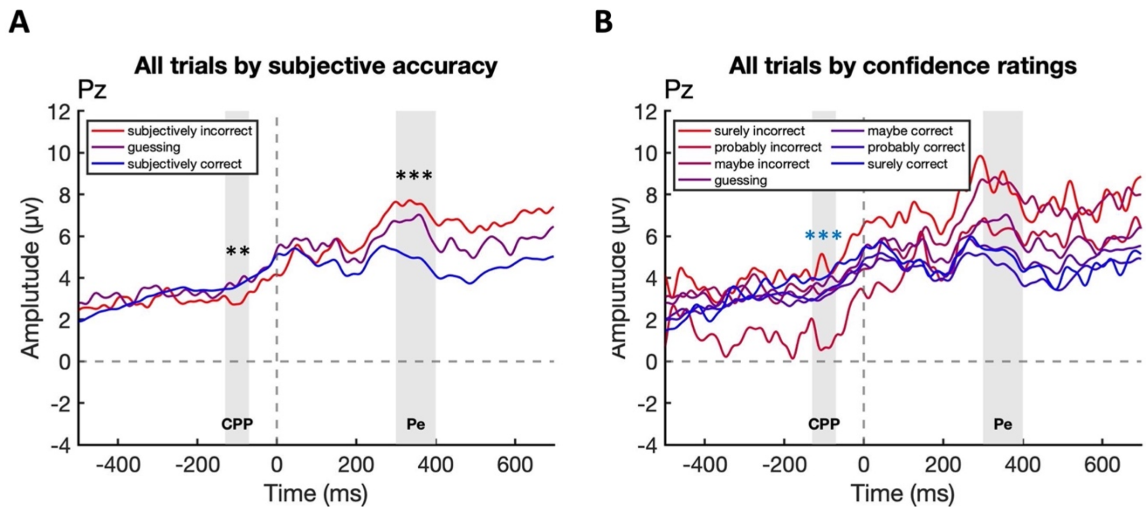

Figure D1. Group mean ERP waveforms at Pz by subjective accuracy and all confidence rating levels, pooling correct and error trials. (A) Group mean ERP waveforms by subjectively correct, subjective incorrect, and guessing trials, pooling correct and error trials. CPP amplitudes showed a positive relationship with subjective accuracy. Pe amplitudes showed a negative relationship with subjective accuracy. (B) Group mean of stimulus-locked ERP waveforms by all confidence levels, pooling correct and error trials. Only CPP amplitudes showed a positive relationship with certainty of being correct. Shaded areas show the time windows of the CPP (-130 to 70 ms) and the Pe (300 to 400 ms).

Note. *p <.05 **p <.01 ***p <.001. Black asterisks indicate the p values for the relationship with subjective accuracy. Blue asterisks indicate the p value for the relationship with certainty of being correct.

D.2. Summary

Overall, the results of this analysis approach mirrored the main findings in Chapter 3. Only certainty of being correct was significantly related to response-locked CPP amplitudes, but other relationships between certainty of being correct/incorrect and the two ERP measures were not significant. Therefore, the current results are largely consistent with Feuerriegel et al. (2022), except that Pe amplitudes were not related to certainty of being incorrect.

D.3. Linear mixed-effects model results

Table D1 Regression Coefficients for Predicting CPP Amplitudes from Trichotomized Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	3.39	0.55	6.19	<.001***
Subjectively incorrect	-0.89	0.25	-3.58	<.001***
Guessing	0.64	0.20	3.15	.002**

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table D2 Regression Coefficients for Predicting Pe Amplitudes from Trichotomized Subjective Accuracy

Parameters	Estimate	SE	z	p
Intercept	5.97	0.66	9.07	<.001***
Subjectively incorrect	0.69	0.30	2.32	.021*
Guessing	0.41	0.24	1.68	.092

Note. Intercept represents the estimate for subjectively correct.
*p <.05 **p <.01 ***p <.001

Table D3 Regression Coefficients for Predicting CPP Amplitudes from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	3.54	0.57	6.20	<.001***
Maybe correct	-0.66	0.17	-3.81	<.001***
Probably correct	-0.04	0.14	-0.28	.777

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table D4 Regression Coefficients for Predicting Pe Amplitudes from Certainty of Being Correct

Parameters	Estimate	SE	z	p
Intercept	4.87	0.63	7.67	<.001***
Maybe correct	0.10	0.21	0.46	.647
Probably correct	-0.32	0.17	-1.85	.065

Note. Intercept represents the estimate for surely correct.
*p <.05 **p <.01 ***p <.001

Table D5 Regression Coefficients for Predicting CPP Amplitudes from Certainty of Being incorrect

Parameters	Estimate	SE	z	p
Intercept	2.60	0.70	3.72	.001**
Maybe incorrect	-0.06	0.67	-0.10	.924
Probably incorrect	-0.16	0.59	-0.26	.792

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Table D6 Regression Coefficients for Predicting Pe Amplitudes from Certainty of Being incorrect

Parameters	Estimate	SE	z	p
Intercept	6.96	0.85	8.14	<.001***
Maybe incorrect	-0.33	0.80	-0.41	.686
Probably incorrect	-0.12	0.71	-0.17	.861

Note. Intercept represents the estimate for surely incorrect.
*p <.05 **p <.01 ***p <.001

Appendix E Supplementary material of Chapter 2

Supplementary Material

Divergent effects of absolute evidence magnitude on decision accuracy and confidence in perceptual judgements

Yiu Hong Ko[1,2,3], Daniel Feuerriegel[1], William Turner[1], Helen Overhoff[1,2,3], Eva Niessen[2,3], Jutta Stahl[3], Robert Hester[1], Gereon R. Fink[2,4], Peter H. Weiss[2,4], & Stefan Bode[1*]

[1] Melbourne School of Psychological Sciences, The University of Melbourne, Australia

[2] Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Centre Jülich, Germany

[3] Department of Psychology, Faculty of Human Sciences, University of Cologne

[4] Department of Neurology, University Hospital Cologne and Faculty of Medicine, University of Cologne

Experiment 1

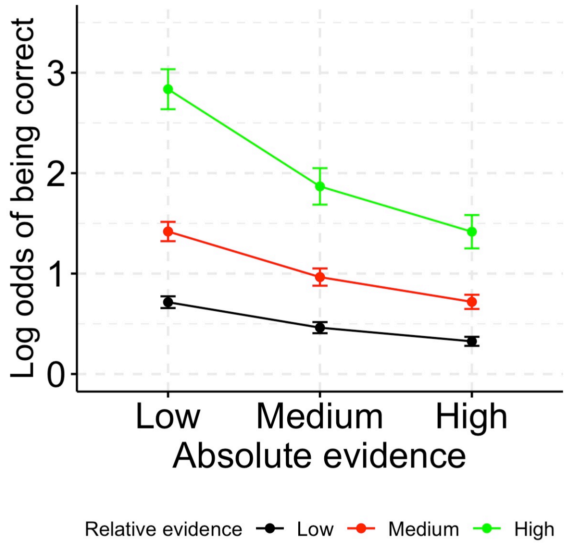

Figure S1. Experiment 1 mean log odds of being correct in each condition. Error bars represent SEM.

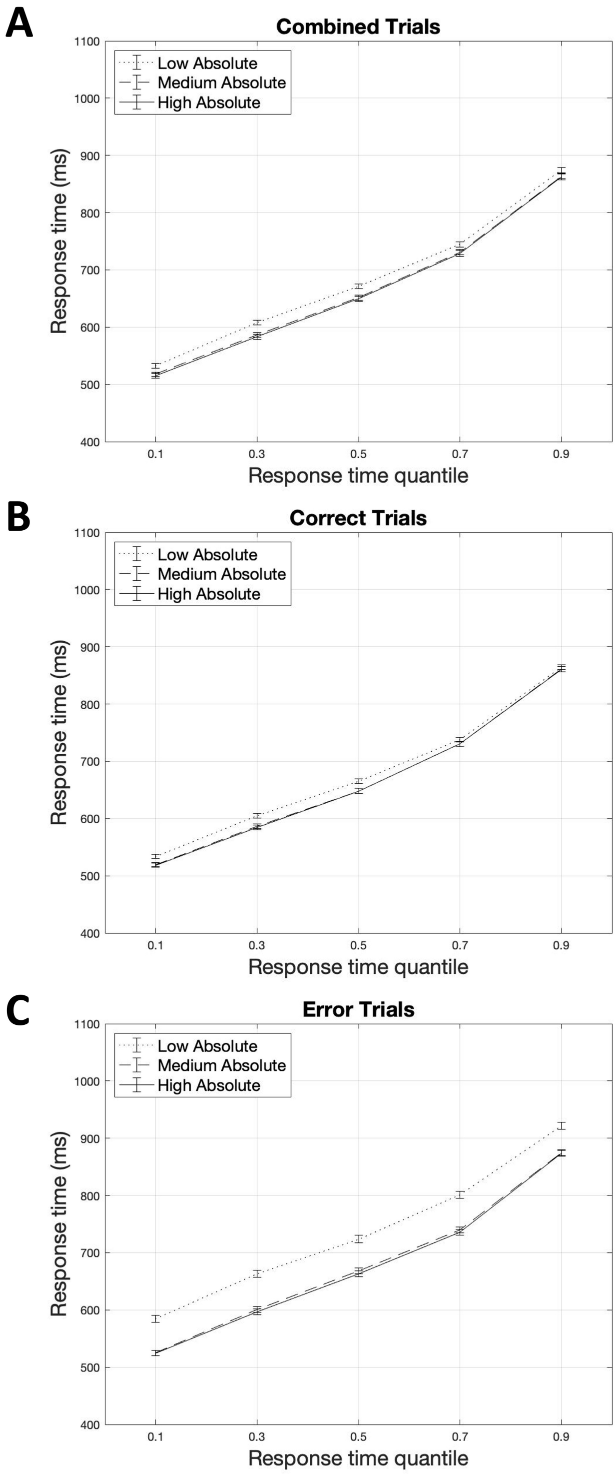

Figure S2. Experiment 1 response time quantiles across absolute evidence levels (collapsing across relative evidence levels). (A) Correct and error trials combined. (B) Correct trials. (C) Error trials. Error bars represent SEM.

Accuracy (Log Odds of Being Correct)

Table S1
Experiment 1 Likelihood Ratio Tests Results for Predicting Accuracy (Log Odds of Being Correct) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	2	1433.01	<.001***
Abs	2	485.87	<.001***
Rel × Abs	4	91.71	<.001***

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table S2
Experiment 1 Regression Coefficients for Predicting Accuracy (Log Odds of Being Correct) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	1.11	0.07	16.47	<.001***
Low Rel	-0.60	0.02	-32.09	<.001***
Med Rel	-0.10	0.02	-4.97	<.001***
Low Abs	0.43	0.02	19.56	<.001***
Med Abs	-0.09	0.02	-4.50	<.001***
Low Rel × Low Abs	-0.22	0.03	-7.72	<.001***
Med Rel × Low Abs	-0.06	0.03	-2.16	.031*
Low Rel × Med Abs	0.05	0.03	2.04	.041*
Med Rel × Med Abs	0.02	0.03	0.92	.356

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Response Time (Correct Trials)

Table S3
Experiment 1 Likelihood Ratio Tests Results for Predicting Response Time (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	2	314.71	<.001***
Abs	2	35.85	<.001***
Rel × Abs	4	106.25	<.001***

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table S4
Experiment 1 Regression Coefficients for Predicting Response Time (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	765.12	2.84	269.20	<.001***
Low Rel	24.69	1.42	17.43	<.001***
Med Rel	6.31	1.14	5.55	<.001***
Low Abs	11.16	1.53	7.31	<.001***
Med Abs	-6.39	1.32	-4.83	<.001***
Low Rel × Low Abs	19.47	1.89	10.31	<.001***
Med Rel × Low Abs	4.06	1.30	3.12	.002**
Low Rel × Med Abs	-1.88	1.45	-1.29	.195
Med Rel × Med Abs	-4.33	1.66	-2.61	.009**

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Response Time (Error Trials)

Table S5
Experiment 1 Likelihood Ratio Tests Results for Predicting Response Time (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	2	18.54	<.001***
Abs	2	49.96	<.001***
Rel × Abs	4	20.20	<.001***

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table S6
Experiment 1 Regression Coefficients for Predicting Response Time (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	811.14	7.68	105.63	<.001***
Low Rel	12.87	2.62	4.92	<.001***
Med Rel	2.11	2.67	0.79	.427
Low Abs	24.84	2.75	9.02	<.001***
Med Abs	-5.51	2.86	-1.93	.054
Low Rel × Low Abs	21.36	3.70	5.77	<.001***
Med Rel × Low Abs	-10.33	3.76	-2.75	.006**
Low Rel × Med Abs	-14.42	3.59	-4.02	<.001***
Med Rel × Med Abs	7.49	3.78	1.98	.048*

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Confidence (Correct Trials)

Table S7
Experiment 1 Likelihood Ratio Tests Results for Predicting Confidence (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	*df*	$\chi2$	*p*
Rel	2	879.07	<.001***
Abs	2	293.89	<.001***
Rel × Abs	4	121.55	<.001***

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table S8
Experiment 1 Regression Coefficients for Predicting Confidence (Correct Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	5.74	0.12	49.75	<.001***
Low Rel	-0.28	0.01	-23.65	<.001***
Med Rel	-0.02	0.01	-2.10	.036*
Low Abs	-0.19	0.01	-16.72	<.001***
Med Abs	0.05	0.01	4.31	<.001***
Low Rel × Low Abs	-0.14	0.02	-8.25	<.001***
Med Rel × Low Abs	-0.01	0.02	-0.63	.526
Low Rel × Med Abs	0.01	0.02	0.73	.463
Med Rel × Med Abs	0.02	0.02	1.29	.195

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Confidence (Error Trials)

Table S9
Experiment 1 Likelihood Ratio Tests Results for Predicting Confidence (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	*df*	*χ2*	*p*
Rel	2	99.99	<.001***
Abs	2	392.15	<.001***
Rel × Abs	4	9.30	.054

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table S10
Experiment 1 Regression Coefficients for Predicting Confidence (Error Trials) from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	4.77	0.13	35.48	<.001***
Low Rel	0.25	0.03	9.93	<.001***
Med Rel	-0.02	0.03	-0.70	.486
Low Abs	-0.58	0.03	-18.52	<.001***
Med Abs	0.14	0.03	5.23	<.001***
Low Rel × Low Abs	0.01	0.04	0.18	.855
Med Rel × Low Abs	-0.02	0.04	-0.41	.681
Low Rel × Med Abs	0.08	0.03	2.34	.019*
Med Rel × Med Abs	-0.02	0.04	-0.63	.531

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Change of Mind (Log Odds of Confidence Lower Than 4) in Correct Trials

Table S11
Experiment 1 Likelihood Ratio Tests Results for Predicting Change of Mind (Log Odds of Confidence Lower Than 4) in Correct Trials from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	2	259.51	<.001***
Abs	2	38.67	<.001***
Rel × Abs	4	22.35	<.001***

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table S12
Experiment 1 Regression Coefficients for Predicting Change of Mind (Log Odds of Confidence Lower Than 4) in Correct Trials from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	3.58	0.29	12.44	<.001***
Low Rel	-0.53	0.04	-12.81	<.001***
Med Rel	-0.14	0.04	-3.31	.001**
Low Abs	-0.26	0.04	-6.31	<.001***
Med Abs	0.13	0.04	3.02	.003**
Low Rel × Low Abs	-0.17	0.05	-3.19	.001**
Med Rel × Low Abs	-0.08	0.06	-1.40	.162
Low Rel × Med Abs	-0.04	0.06	-0.60	.550
Med Rel × Med Abs	0.07	0.06	1.14	.255

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Change of Mind (Log Odds of Confidence Lower Than 4) in Error Trials

Table S13
Experiment 1 Likelihood Ratio Tests Results for Predicting Change of Mind (Log Odds of Confidence Lower Than 4) in Error Trials from Relative Evidence, Absolute Evidence, and Their Interactions

Predictor	*df*	$\chi2$	*p*
Rel	2	96.92	<.001***
Abs	2	176.64	<.001***
Rel × Abs	4	14.09	.007**

Note. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Table S14
Experiment 1 Regression Coefficients for Predicting Change of Mind (Log Odds of Confidence Lower Than 4) in Error Trials from Relative Evidence, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	1.60	0.19	8.26	<.001***
Low Rel	0.43	0.04	9.59	<.001***
Med Rel	-0.04	0.04	-0.95	.344
Low Abs	-0.61	0.05	-12.38	<.001***
Med Abs	0.09	0.05	2.04	.042*
Low Rel × Low Abs	0.00	0.06	0.04	.970
Med Rel × Low Abs	-0.03	0.06	-0.49	.623
Low Rel × Med Abs	0.19	0.06	3.16	.002**
Med Rel × Med Abs	-0.04	0.06	-0.62	.536

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence.
*p <.05 **p <.01 ***p <.001

Confidence (Correct Trials)

Table S15
Experiment 1 Likelihood Ratio Tests Results for Predicting Confidence (Correct Trials) from Relative Evidence, RT, Absolute Evidence, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	2	667.60	<.001***
RT.cmc	1	1207.10	<.001***
Abs	2	270.15	<.001***
Rel × Abs	4	80.22	<.001***

Note. Rel: Relative evidence; Abs: Absolute evidence;
RT.cmc: Response time (cluster-mean centered).
*p <.05 **p <.01 ***p <.001

Table S16
Experiment 1 Regression Coefficients for Predicting Confidence (Correct Trials) from Relative Evidence, RT, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	5.75	0.12	49.77	<.001***
Low Rel	-0.24	0.01	-20.99	<.001***
Med Rel	-0.01	0.01	-1.27	.205
RT.cmc	-0.00	0.00	-35.22	<.001***
Low Abs	-0.17	0.01	-15.92	<.001***
Med Abs	0.04	0.01	3.72	<.001***
Low Rel × Low Abs	-0.11	0.02	-6.81	<.001***
Med Rel × Low Abs	-0.00	0.02	-0.26	.793
Low Rel × Med Abs	0.01	0.02	0.53	.598
Med Rel × Med Abs	0.02	0.02	0.97	.330

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence; RT.cmc: Response time (cluster-mean centered).
*p <.05 **p <.01 ***p <.001

Confidence (Error Trials)

Table S17
Experiment 1 Likelihood Ratio Tests Results for Predicting Confidence (Error Trials) from Relative Evidence, RT, Absolute Evidence, and Their Interactions

Predictor	*df*	*χ2*	*p*
Rel	2	125.88	<.001***
RT.cmc	1	405.39	<.001***
Abs	2	335.72	<.001***
Rel × Abs	4	10.37	.035*

Note. Rel: Relative evidence; Abs: Absolute evidence; RT.cmc: Response time (cluster-mean centered).
*p <.05 **p <.01 ***p <.001

Table S18
Experiment 1 Regression Coefficients for Predicting Confidence (Error Trials) from Relative Evidence, RT, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	4.77	0.13	35.46	<.001***
Low Rel	0.27	0.02	11.09	<.001***
Med Rel	-0.01	0.03	-0.45	.654
RT.cmc	-0.00	0.00	-20.39	<.001***
Low Abs	-0.52	0.03	-17.07	<.001***
Med Abs	0.12	0.03	4.72	<.001***
Low Rel × Low Abs	0.04	0.04	1.09	.275
Med Rel × Low Abs	-0.03	0.04	-0.85	.393
Low Rel × Med Abs	0.06	0.03	1.75	.079
Med Rel × Med Abs	-0.01	0.04	-0.24	.807

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence; RT.cmc: Response time (cluster-mean centered).
*p <.05 **p <.01 ***p <.001

Experiment 2

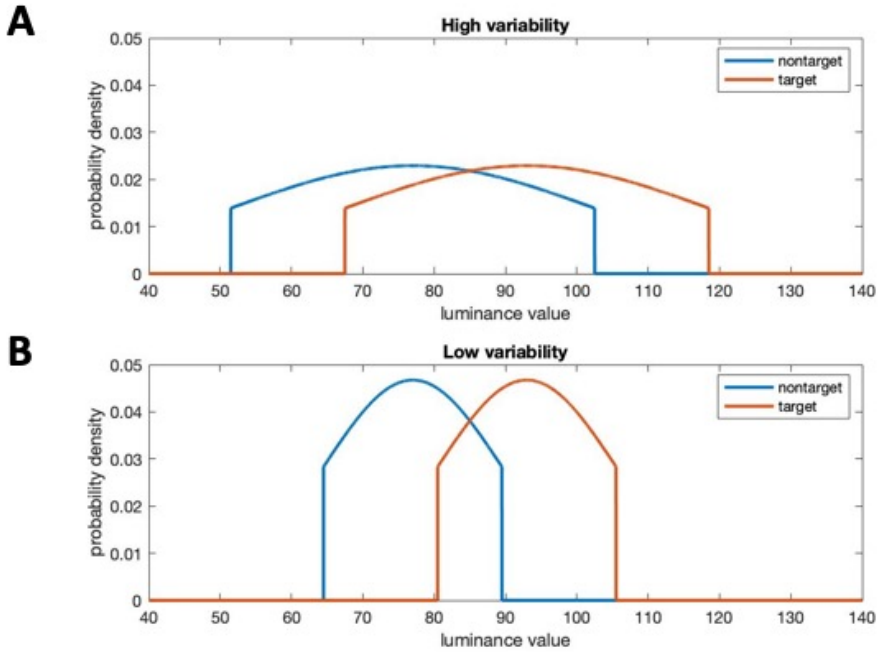

Figure S3. Example of luminance value distributions in (A) high variability conditions and (B) low variability conditions. Specifically, this example demonstrates the distributions when relative evidence is high and absolute evidence is low.

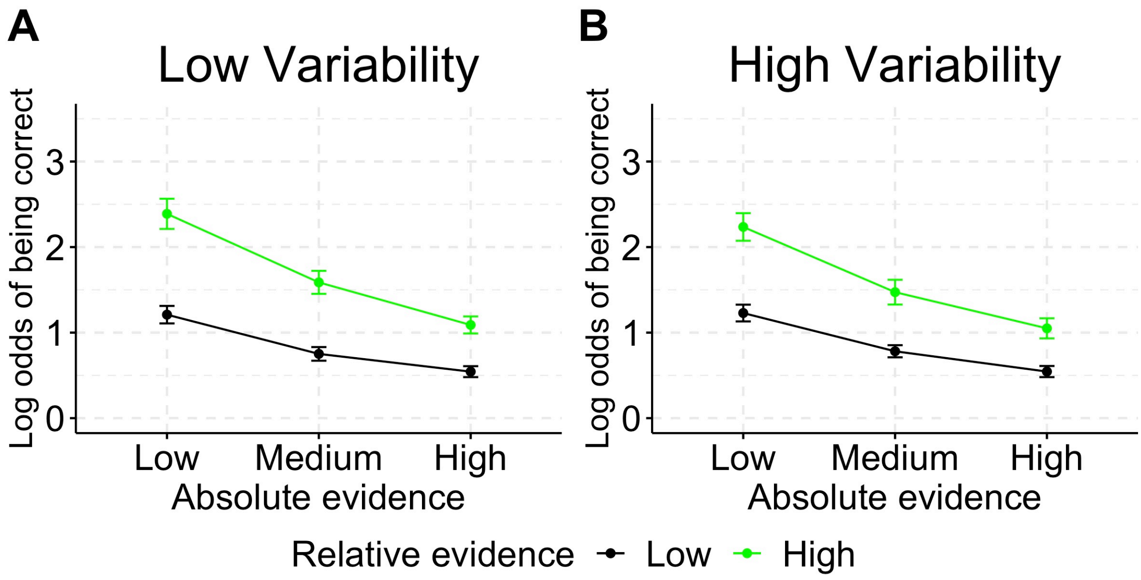

Figure S4. Experiment 2 mean log odds of being correct in each condition. Error bars represent SEM.

Table S19
Mean luminance values for all experimental conditions of Experiment 2

		Low absolute evidence	Medium absolute evidence	High absolute evidence
High luminance variability	High relative evidence	77, 93	107, 123	137, 153
	Low relative evidence	77, 85	107, 115	137, 145
Low luminance variability	High relative evidence	77, 93	107, 123	137, 153
	Low relative evidence	77 85	107 115	137 145

Note. While mean luminance values were identical across high and low luminance variability conditions, high luminance conditions had standard deviations of 25.5, and low luminance conditions had standard deviations of 12.5.

Accuracy (Log Odds of Being Correct)

Table S20
Experiment 2 Likelihood Ratio Tests Results for Predicting Accuracy (Log Odds of Being Correct) from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	1	576.22	<.001***
Abs	2	588.29	<.001***
Var	1	0.78	.378
Rel × Abs	2	37.02	<.001***
Rel × Var	1	2.77	.096
Abs × Var	2	0.25	.882
Rel × Abs × Var	2	0.53	.767

Note. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Table S21
Experiment 2 Regression Coefficients for Predicting Accuracy (Log Odds of Being Correct) from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	1.19	0.08	15.27	<.001***
Low Rel	-0.35	0.01	-23.61	<.001***
Low Abs	0.47	0.02	21.31	<.001***
Med Abs	-0.08	0.02	-4.05	<.001***
Low Var	0.01	0.01	0.88	.378
Low Rel × Low Abs	-0.12	0.02	-5.35	<.001***
Low Rel × Med Abs	0.02	0.02	0.75	.456
Low Rel × Low Var	-0.02	0.01	-1.67	.096
Low Abs × Low Var	-0.01	0.02	-0.41	.680
Med Abs × Low Var	0.01	0.02	0.47	.638
Low Rel × Low Abs × Low Var	0.01	0.02	0.29	.770
Low Rel × Med Abs × Low Var	-0.01	0.02	-0.71	.475

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Response Time (Correct Trials)

Table S22
Experiment 2 Likelihood Ratio Tests Results for Predicting Response Time (Correct Trials) from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	1	106.53	<.001***
Abs	2	44.46	<.001***
Var	1	6.61	.010*
Rel × Abs	2	32.20	<.001***
Rel × Var	1	0.37	.541
Abs × Var	2	7.13	.028*
Rel × Abs × Var	2	6.24	.044*

Note. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Table S23
Experiment 2 Regression Coefficients for Predicting Response Time (Correct Trials) from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	741.98	1.71	435.14	<.001***
Low Rel	14.11	1.04	13.54	<.001***
Low Abs	12.68	1.11	11.39	<.001***
Med Abs	-6.00	1.31	-4.57	<.001***
Low Var	3.52	1.09	3.22	.001**
Low Rel × Low Abs	10.44	1.33	7.85	<.001***
Low Rel × Med Abs	-7.55	1.42	-5.33	<.001***
Low Rel × Low Var	0.84	1.08	0.77	.440
Low Abs × Low Var	0.06	1.21	0.05	.964
Med Abs × Low Var	4.51	1.30	3.48	.001**
Low Rel × Low Abs × Low Var	-1.89	1.15	-1.64	.100
Low Rel × Med Abs × Low Var	4.80	1.25	3.83	<.001***

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Response Time (Error Trials)

Table S24
Experiment 2 Likelihood Ratio Tests Results for Predicting Response Time (Error Trials) from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	1	23.74	<.001***
Abs	2	37.22	<.001***
Var	1	0.04	.844
Rel × Abs	2	14.25	<.001***
Rel × Var	1	0.16	.685
Abs × Var	2	3.90	.143
Rel × Abs × Var	2	3.37	.185

Note. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Table S25
Experiment 2 Regression Coefficients for Predicting Response Time (Error Trials) from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	784.84	4.05	193.66	<.001***
Low Rel	12.86	2.20	5.86	<.001***
Low Abs	23.00	2.82	8.16	<.001***
Med Abs	-6.64	2.33	-2.85	.004**
Low Var	0.51	2.15	0.24	.812
Low Rel × Low Abs	14.74	2.64	5.59	<.001***
Low Rel × Med Abs	-4.74	2.59	-1.83	.068
Low Rel × Low Var	1.06	2.27	0.47	.641
Low Abs × Low Var	7.99	2.77	2.89	.004**
Med Abs × Low Var	-4.44	2.68	-1.65	.098
Low Rel × Low Abs × Low Var	-0.21	2.73	-0.08	.939
Low Rel × Med Abs × Low Var	5.29	2.79	1.89	.058

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Confidence (Correct Trials)

Table S26
Experiment 2 Likelihood Ratio Tests Results for Predicting Confidence (Correct Trials) from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	1	261.57	<.001***
Abs	2	191.90	<.001***
Var	1	2.72	.099
Rel × Abs	2	47.83	<.001***
Rel × Var	1	0.45	.501
Abs × Var	2	2.45	.294
Rel × Abs × Var	2	1.15	.561

Note. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Table S27
Experiment 2 Regression Coefficients for Predicting Confidence (Correct Trials) from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	5.73	0.14	40.43	<.001***
Low Rel	-0.12	0.01	-16.22	<.001***
Low Abs	-0.12	0.01	-12.04	<.001***
Med Abs	-0.01	0.01	-0.54	.593
Low Var	-0.01	0.01	-1.65	.099
Low Rel × Low Abs	-0.07	0.01	-6.79	<.001***
Low Rel × Med Abs	0.02	0.01	1.97	.048*
Low Rel × Low Var	-0.00	0.01	-0.67	.501
Low Abs × Low Var	-0.01	0.01	-1.00	.316
Med Abs × Low Var	-0.01	0.01	-0.60	.550
Low Rel × Low Abs × Low Var	0.00	0.01	0.11	.909
Low Rel × Med Abs × Low Var	-0.01	0.01	-1.00	.318

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
$*p < .05$ $**p < .01$ $***p < .001$

Confidence (Error Trials)

Table S28
Experiment 2 Likelihood Ratio Tests Results for Predicting Confidence (Error Trials) from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	1	57.71	<.001***
Abs	2	400.02	<.001***
Var	1	0.85	.355
Rel × Abs	2	2.65	.265
Rel × Var	1	4.35	.037*
Abs × Var	2	8.47	.014*
Rel × Abs × Var	2	3.79	.150

Note. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Table S29
Experiment 2 Regression Coefficients for Predicting Confidence (Error Trials) from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	4.78	0.13	35.64	<.001***
Low Rel	0.14	0.02	7.61	<.001***
Low Abs	-0.52	0.03	-17.88	<.001***
Med Abs	0.08	0.03	3.27	.001**
Low Var	-0.02	0.02	-0.92	.355
Low Rel × Low Abs	0.03	0.03	1.16	.245
Low Rel × Med Abs	0.00	0.03	0.20	.845
Low Rel × Low Var	0.04	0.02	2.09	.037*
Low Abs × Low Var	-0.08	0.03	-2.73	.006**
Med Abs × Low Var	0.06	0.03	2.46	.014*
Low Rel × Low Abs × Low Var	0.05	0.03	1.90	.057
Low Rel × Med Abs × Low Var	-0.02	0.03	-0.83	.408

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Change of Mind (Log Odds of Confidence Lower Than 4) in Correct Trials

Table S30
Experiment 2 Likelihood Ratio Tests Results for Predicting Change of Mind (Log Odds of Confidence Lower Than 4) in Correct Trials from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	1	36.33	<.001***
Abs	2	5.26	.072
Var	1	0.27	.601
Rel × Abs	2	9.72	.008**
Rel × Var	1	0.08	.780
Abs × Var	2	0.49	.784
Rel × Abs × Var	2	0.33	.846

Note. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Table S31
Experiment 2 Regression Coefficients for Predicting Change of Mind (Log Odds of Confidence Lower Than 4) in Correct Trials from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	3.55	0.27	13.30	<.001***
Low Rel	-0.19	0.03	-6.05	<.001***
Low Abs	-0.10	0.04	-2.29	.022*
Med Abs	0.03	0.04	0.77	.439
Low Var	0.02	0.03	0.53	.599
Low Rel × Low Abs	-0.13	0.04	-3.12	.002**
Low Rel × Med Abs	0.05	0.04	1.22	.224
Low Rel × Low Var	0.01	0.03	0.28	.779
Low Abs × Low Var	0.03	0.04	0.66	.510
Med Abs × Low Var	-0.02	0.04	-0.51	.607
Low Rel × Low Abs × Low Var	-0.02	0.04	-0.53	.595
Low Rel × Med Abs × Low Var	0.00	0.04	0.04	.971

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Change of Mind (Log Odds of Confidence Lower Than 4) in Error Trials

Table S32
Experiment 2 Likelihood Ratio Tests Results for Predicting Change of Mind (Log Odds of Confidence Lower Than 4) in Error Trials from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	1	51.07	<.001***
Abs	2	179.30	<.001***
Var	1	0.00	.976
Rel × Abs	2	3.62	.164
Rel × Var	1	3.66	.056
Abs × Var	2	7.92	.019*
Rel × Abs × Var	2	7.20	.027*

Note. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Table S33
Experiment 2 Regression Coefficients for Predicting Change of Mind (Log Odds of Confidence Lower Than 4) in Error Trials from Relative Evidence, Absolute Evidence, Luminance Variability, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	1.57	0.18	8.81	<.001***
Low Rel	0.24	0.03	7.18	<.001***
Low Abs	-0.61	0.05	-12.22	<.001***
Med Abs	0.08	0.05	1.75	.081
Low Var	0.00	0.03	0.03	.976
Low Rel × Low Abs	0.07	0.05	1.45	.147
Low Rel × Med Abs	0.01	0.05	0.30	.767
Low Rel × Low Var	0.06	0.03	1.92	.055
Low Abs × Low Var	-0.14	0.05	-2.81	.005**
Med Abs × Low Var	0.07	0.05	1.58	.113
Low Rel × Low Abs × Low Var	0.12	0.05	2.53	.011*
Low Rel × Med Abs × Low Var	-0.03	0.05	-0.55	.582

Note. Intercept represents the estimate for high relative and high absolute evidence. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability.
*p <.05 **p <.01 ***p <.001

Confidence (Correct Trials)

Table S34
Experiment 2 Likelihood Ratio Tests Results for Predicting Confidence (Correct Trials) from Relative Evidence, Luminance Variability, RT, Absolute Evidence, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	1	195.23	<.001***
Var	1	1.08	.299
RT.cmc	1	1044.45	<.001***
Abs	2	169.31	<.001***
Rel × Var	1	0.24	.622
Rel × Abs	2	33.46	<.001***
Var × Abs	2	1.81	.405
Rel × Var × Abs	2	0.54	.765

Note. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability; RT.cmc: Response time (cluster-mean centered).
*p <.05 **p <.01 ***p <.001

Table S35
Experiment 2 Regression Coefficients for Predicting Confidence (Correct Trials) from Relative Evidence, Luminance Variability, RT, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	5.73	0.14	40.43	<.001***
Low Rel	-0.10	0.01	-14.00	<.001***
Low Var	-0.01	0.01	-1.04	.299
RT.cmc	-0.00	0.00	-32.72	<.001***
Low Abs	-0.11	0.01	-10.95	<.001***
Med Abs	-0.01	0.01	-1.18	.238
Low Rel × Low Var	-0.00	0.01	-0.49	.622
Low Rel × Low Abs	-0.05	0.01	-5.55	<.001***
Low Rel × Med Abs	0.01	0.01	1.13	.258
Low Var × Low Abs	-0.01	0.01	-1.06	.287
Low Var × Med Abs	-0.00	0.01	-0.23	.815
Low Rel × Low Var × Low Abs	-0.00	0.01	-0.07	.942
Low Rel × Low Var × Med Abs	-0.01	0.01	-0.61	.541

Note. Intercept represents the estimate for high relative evidence, high absolute evidence, and high luminance variability. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability; RT.cmc: Response time (cluster-mean centered).
*p <.05 **p <.01 ***p <.001

Confidence (Error Trials)

Table S36
Experiment 2 Likelihood Ratio Tests Results for Predicting Confidence (Error Trials) from Relative Evidence, Luminance Variability, RT, Absolute Evidence, and Their Interactions

Predictor	*df*	$\chi 2$	*p*
Rel	1	75.31	<.001***
Var	1	0.66	.418
RT.cmc	1	357.34	<.001***
Abs	2	357.69	<.001***
Rel × Var	1	4.95	.026*
Rel × Abs	2	5.87	.053
Var × Abs	2	6.83	.033*
Rel × Var × Abs	2	4.66	.097

Note. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability; RT.cmc: Response time (cluster-mean centered).
*p <.05 **p <.01 ***p <.001

Table S37
Experiment 2 Regression Coefficients for Predicting Confidence (Error Trials) from Relative Evidence, Luminance Variability, RT, Absolute Evidence, and Their Interactions

Parameters	Estimate	*SE*	*z*	*p*
Intercept	4.78	0.13	35.64	<.001***
Low Rel	0.16	0.02	8.70	<.001***
Low Var	-0.01	0.02	-0.81	.418
RT.cmc	-0.00	0.00	-19.14	<.001***
Low Abs	-0.47	0.03	-16.80	<.001***
Med Abs	0.07	0.02	2.90	.004**
Low Rel × Low Var	0.04	0.02	2.23	.026*
Low Rel × Low Abs	0.05	0.03	1.88	.060
Low Rel × Med Abs	0.00	0.02	0.06	.951
Low Var × Low Abs	-0.07	0.03	-2.43	.015*
Low Var × Med Abs	0.06	0.02	2.25	.024*
Low Rel × Low Var × Low Abs	0.06	0.03	2.04	.042*
Low Rel × Low Var × Med Abs	-0.02	0.02	-0.67	.502

Note. Intercept represents the estimate for high relative evidence, high absolute evidence, and high luminance variability. Rel: Relative evidence; Abs: Absolute evidence; Var: Luminance variability; RT.cmc: Response time (cluster-mean centered).
*p <.05 **p <.01 ***p <.001