University of Tartu
Institute of Psychology

Alan Voodla

ZOOMING INTO NEURAL MECHANISMS OF DECISION CONFIDENCE THROUGH THE LENS OF EEG

Master's thesis

Supervisor: Andero Uusberg

Running head: Decision confidence and error-related EEG potentials

Tartu 2019

Zooming Into Neural Mechanisms Of Decision Confidence Through The Lens Of EEG

Abstract

Neural mechanisms responsible for feelings of certainty in reasoning and decision-making remain unclear. This thesis attempts to address this problem by examining the role of error-related EEG potentials (error-related negativity - ERN, error positivity - Pe) in decision confidence. The amplitude of these potentials has been shown to correlate with error detection and confidence ratings in simple perceptual decisions. In order to test whether this pattern holds in more complex decisions, we investigated activity changes in ERN and Pe in response to manipulations of decision confidence in an arithmetic reasoning task. In an EEG experiment, 49 participants had to quickly respond whether an equation (e.g. 9 * 7 = 65) is correct or incorrect and then report their decision confidence. Task difficulty and response fluency were varied as manipulations of confidence. The results indicated that ERN and Pe did not mediate the effect of task difficulty on confidence. Response fluency impacted confidence only for simple decisions, and this effect was partially mediated by ERN. These results suggest that Pe could be an index of implicit cognitive control, whereas ERN tracks decision confidence in simple decisions and is susceptible to response fluency manipulations.

Keywords: decision confidence, mental arithmetics, ERN, Pe, fluency, task difficulty

Elektroentsefalograafiline perspektiiv kindlustunde neuraalsetele mehhanismidele

Kokkuvõte

Mõtlemise ja otsustamisega kaasneva kindlustunde kujunemist kirjeldavad neuraalsed mehhanismid on siiani ebaselged. Üks viis nendele valgust heita, on küsida mis rolli mängivad vea-monitoorimisega seotud EEG sündmuspotentsiaalid (vea-negatiivsus - ERN ja vea-positiivsus - Pe) kindlustunde kujunemises. Varasemates uuringutes on leitud, et nende potentsiaalide amplituud korreleerub kindlustundega lihtsates tajulistes otsustustes. Kas sarnane muster kehtib ka keerukamat mõtlemist eeldavate otsuste tegemisel? Kas kontrollitud manipulatsioone kindlustundes saab seletada muutustega ERNi või Pe aktivatsioonimustrites? Nendele küsimustele vastamiseks viisime läbi EEG eksperimendi, kus 49 katseisikut pidid vastama, kas arvutiekraanile kuvatud võrrandite vastused (näiteks 9 * 7 = 65) on õiged või valed ning raporteerima oma kindlustunnet. Kindlustunde mõjutamiseks varieeriti ülesande raskusastet ja vastuse voolavust. Tulemustest selgus, et ülesande raskusaste mõju kindlustundele ei ole medieeritud ei ERNi ega Pe poolt. Lihtsa raskusastmega võrrandite puhul medieeris ERN osaliselt voolavuse mõju kindlustundele. Analüüsist järeldub, et Pe võiks indekseerida implitsiitset kognitiivset kontrolli, mis reguleerib käitumist sõltumata kogetud kindlustundest. ERN seostub kindlustundega vaid lihtsate otsutuste puhul ning on tundlik vastuse voolavuse mõjutustele.

Märksõnad: kindlustunne, peastarvutamine, ERN, Pe, voolavus, ülesande raskusaste

Introduction

Decision confidence is like a nose - although always present in our day-to-day experiences we remain unaware of it unless it is not explicitly referenced. And just like a nose has an important function within the respiratory system, decision confidence also has an important function regulating reasoning and adaptive behaviour (Luttrell, Briñol, Petty, Cunningham, & Díaz, 2013). Research with an aim to understand decision confidence dates back to the seminal years of experimental psychology (e.g. Peirce & Jastrow, 1884) and has been more recently re-established mainly within the study of perceptual decision-making (Yeung & Summerfield, 2012) and abstract reasoning (Griffin & Tversky, 1992; Shynkaruk & Thompson, 2006). From the perceptual decision-making perspective, there have been attempts to establish neural correlates of decision confidence for 2-option forced-choice tasks (e.g. Flanker task, Stroop task, dot count task). This research supports the conclusion that decision confidence can be predicted from electroencephalographic event-related potentials (ERPs) previously associated with error detection (Boldt & Yeung, 2015; Selimbeyoglu, Keskin-Ergen, & Demiralp, 2012). These results suggest that a similar pattern should be observable for more complex decision-making/reasoning tasks, which have been mostly studied in the field of abstract reasoning (e.g. deductive/inductive inference or numeric reasoning). Surprisingly, as the two research strands have developed relatively separately (Cruz, Arango-Muñoz, & Volz, 2016), the prediction that error-related ERPs could underlie decision confidence has not been extended to more complex reasoning problems.

This thesis is an attempt to synthesize the perceptual decision-making and abstract reasoning perspectives and answer the question, whether error-related EEG potentials mediate experimentally induced changes in decision confidence in a complex reasoning task. In the following, an overview of decision confidence, its determinants and neural correlates of error monitoring is given, which forms the background for defining an experimental study in which the theoretical predictions about the relationship between error-related EEG potentials and decision confidence were tested.

Decision confidence

Decision confidence can be defined as a belief about the validity of our own thoughts, knowledge or performance that relies on a subjective feeling (Grimaldi, Lau, & Basso, 2015). This definition states that decision confidence relies on a subjective feeling which has been termed feeling of rightness (FOR) for high confidence (Thompson, Turner, & Pennycook, 2011) or feeling of error (FOE) for low confidence (Gangemi, Bourgeois-Gironde, & Mancini, 2015). The term stems from metamemory research where study participants have to indicate their certainty about how well they performed in a memory test, which might often be at odds with the actual result (Koriat & Bjork, 2005). In a similar fashion, decision confidence is often measured as confidence rating on a discrete or continuous scale after a participant has made a decision in a reasoning or perceptual judgement task (Grimaldi et al., 2015).

This definition also implies a distinction between object- and metalevel psychological processes (Flavell, 1979). Object-level processes refer to psychological activity that results in phenomena such as perceptions, decisions, emotions, and memories which generally represent bodily or environmental states. For example, thinking that the correct answer to the product of 15 and 7 is 105 results from object-level processing. Meta-level processes are a kind of re-representation - they represent states of object-level processes and are therefore crucial for cognitive control and self-regulation (Botvinick, Braver, Barch, Carter, & Cohen, 2001). Continuing the previous example, being confident that the provided answer of 105 is correct, is a result of meta-level processing. These kinds of metarepresentations presumably express underlying feelings for the accuracy of a given judgement and therefore play a regulatory role in adaptive decision-making. For example, FOR is taken to be a starting or stopping rule for analytic reasoning (Thompson et al., 2011) in the sense that if a reasoner feels that she is correct in her answer (high FOR), she will no longer try to re-evaluate her decision or find flaws in it, which might result in overconfidence and sub-optimal action. On the other hand FOE results in low confidence and if it is not well calibrated, a person might be susceptible to overthinking and anxiety.

Metacognitive accuracy

The accuracy of our confidence judgements has been subject to controversy. The basic question from this perspective is, are our confidence estimates a good guide to the quality of our decisions? For example - can we detect if we make errors in Flanker[1] or similar forced-choice task? This relationship between object-level performance and its metacognitive evaluation has been termed metacognitive accuracy (Fleming & Dolan, 2012). The general finding seems to be that humans are reliable detectors of their own errors in many perceptual decision tasks and they often exhibit compensatory behaviour after a mistake has been committed (Yeung & Summerfield, 2012). However, the opposite trend has emerged in reasoning tasks, where participants are asked to evaluate their confidence after solving more complex problems. For example, Thompson and Shynkaruk (2006) asked whether people are accurate in their evaluations of confidence in solving deductive reasoning problems (e.g Premise 1: Some elephants are ill-mannered. Premise 2: Some ill-mannered beings are smart. Conclusion: Therefore some elephants are smart.). Participants of the experiment, having seen the problem, had to respond, whether the conclusion is valid, and evaluate their confidence about their decision. The results indicated that confidence ratings were often biased when compared to actual correct answers.

An implication of this is that the computation of confidence for complex and simple decisions might rely on different brain processes. Although decision confidence might be accurate for simple perceptual judgments, its computation is influenced by heuristic factors which might not be well calibrated for more complex problems. In the study by Thompson and Shynkaruk (2006), it was also detected that the confidence ratings were influenced by believability of the conclusion, perceived difficulty of the problem and the time taken to make a decision. More recently it has been shown that the ease by which the answer comes to mind (response fluency) also has an effect on FOR and decision confidence in more complex decision problems (Thompson et al., 2011). These results raise a question about the differences in computation of confidence in simple and complex decisions.

An approach to investigate confidence computation differences in simple and complex decisions is through electroencephalographic analysis, which provides a window to underlying neural mechanisms of these processes (Yeung & Summerfield, 2012). If the same mechanism underlies decision confidence for simple and difficult decision-making, then a similar pattern of neural correlates should be observed for their computation. However, if, as the behavioural level analysis suggests, decision confidence computation for complex decisions relies on different neural mechanisms than for simple decisions, a different neural pattern should be observed. Error-related EEG potentials provide a way to address this hypothesis by offering a window into the neural mechanisms that are known to be involved in driving decision confidence during simple perceptual decisions.

Error-related EEG potentials

Error-related EEG potentials stem from studies of error-monitoring in perceptual decision-making (Yeung, Botvinick, & Cohen, 2004). The general idea is to measure test participants' neuronal activity (via EEG, fMRI or other brain imaging technique) during responding in a speeded forced choice 2-option task (for example, evaluating which one of the two displayed batches has more dots in it) and then compare the activity under correct and incorrect decisions. A major finding in this paradigm has been the identification of error-related EEG potentials - error-related negativity (ERN) and error-positivity (Pe) (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993).

ERN is characterized as a fast negative spike in the EEG signal within 0-150 ms after committing an erroneous response (van Veen & Carter, 2006). The most likely neural origin for ERN has been located to the anterior cingulate cortex (ACC) (Yeung et al., 2004). The functional role ACC activity reflected in ERN in cognition is an active research interest. According to three main theories, ERN reflects a mismatch between expected and the actual representation; it is a signature of two incompatible streams of information and thirdly, and it indicates a negative feedback signal for learning (van Veen & Carter, 2006). The three theories are not necessarily incompatible. It is therefore likely that the ERN can reflect contributions from different functional processes.

Following ERN, within 200-400 ms after a commission of an error, another error-related positive potential (Pe) has been identified (van Veen & Carter, 2006). Although its functional role is more ambiguous than in the case of ERN, it has been most clearly associated with conscious processing of error (as opposed to unconscious processing for ERN) (Overbeek, Nieuwenhuis, & Ridderinkhof, 2005). In a more recent account of ERN and Pe, it has been argued that ERN is most clearly associated with error correction whereas Pe is an index of error detection (Yeung & Summerfield, 2012).

Since perceptual choice tasks indicate that humans are rather well-calibrated in evaluating their response accuracy, it makes sense to ask whether ERN and Pe would be predictive for decision confidence in these tasks. This hypothesis was confirmed for ERN in Flanker task (Scheffers & Coles, 2000), in which ERN's amplitude covaried with a rating of perceived accuracy for a given response. A second study (Selimbeyoglu et al., 2012) also supports the conclusion that ERN and Pe might be neural indexes of decision confidence - in a modified Flanker task and a visual discrimination task where task difficulty was manipulated, Pe, as well as ERN were observed to correlate with decision certainty. In another study, participants had to evaluate the number of dots and report their confidence on a 6 point scale in each trial. It was observed that both ERN and Pe were on average distinguishable for 6 levels of reported confidence - higher amplitude Pe was associated with higher confidence ratings (Boldt & Yeung, 2015). Interestingly, this association was independent of objective accuracy. This study suggests that Pe, rather than ERN is the most robust neural index of subjective decision confidence, although for both a clear relationship should be observable.

The present study

In summary, in simple perceptual decision tasks, a relationship between error-related EEG potentials (ERN, and Pe) and decision confidence has been established. This suggests that an implicit reading of decision confidence for these kinds of tasks could be made based on EEG activity (Selimbeyoglu et al., 2012). However, the conclusions are limited in scope because of the simplicity of the decision-making task and the correlational design upon which these conclusions are founded. The current study aims to fill these gaps by testing whether the ERN/Pe relates to decision confidence in a more complex mental arithmetic task and within an experimental design, which allows for explicit manipulations of decision confidence. We assume that if ERN and Pe truly represent neural responses that underlie decision confidence, they should mediate the effects of systematic manipulations of confidence on confidence ratings. If, however evidence for mediation is not detected, it could be argued that different neural mechanisms underlie computation of decision confidence.

More precisely, we designed an EEG experiment, in which participants had to quickly respond whether an equation (e.g. 9 * 7 = 65) is correct or incorrect and then report their decision confidence. Response fluency and task difficulty were used as manipulations of decision confidence, which we expected to have a robust effects based on previous literature (Cruz et al., 2016; Thompson et al., 2013) and a pilot study. We further hypothesized that more difficult equations will elicit higher ERN and Pe compared to less difficult decisions. We also hypothesized that responses with lower fluency will elicit higher ERN and Pe than disfluent responses.

Method

The primary objective of the study design was to test whether the effects of difficulty and fluency on confidence are mediated by error-related EEG potentials. In order to do this, the experiment had a 2x2 within-subject factorial design with decision confidence and error-related EEG potentials as the dependent variables (DV). A priori power analysis for repeated-measures ANOVA with 2 factors indicated that 80% power for an alpha level of 0.05 for a large effect size would be provided by a sample size of 52. Response fluency and task difficulty were both treated as 2 level independent variables (IV).

Participants

55 volunteers (age between 18 and 65, 24 males) registered to the study through an online invitation distributed in Facebook and mailing lists of University of Tartu. Prior coming to the experiment, participants were asked to fill in two adapted questionnaires measuring Big-5 personality (Konstabel et al., 2017) and Need for cognitive closure (NFCC) traits (Redi, 2017; Webster & Kruglanski, 1994). As compensation, participants were given personality feedback and an opportunity to win a 20-euros worth bookshop gift card. Students of psychology could also earn course credits. The project was approved by the Research Ethics Committee of the University of Tartu (application 286/T-3). The basic research question and main analyses were pre-registered in *Open Science Foundation* online-environment, the registration document and other public study-related files are retrievable from https://osf.io/cj2rf/. 6 participants had to be excluded from the main analyses due to technical issues in EEG recording. The final sample consisted of 49 participants.

Experimental task and stimuli

We designed a mental arithmetic task and implemented it with Psychopy (Peirce et al., 2019). This task was chosen, because it simultaneously satisfies many important criteria - it is suitable for objective measurement of accuracy, it requires symbolic as opposed to perceptual processing, it suggests simple ways for the implementation of confidence manipulations and it is applicable for gathering data over many trials, which is a requirement for an EEG study.

In this task, a participant who is sitting behind a computer is required to quickly respond whether equations (e.g. 6 x 8 = 42) displayed for 2 seconds are correct or incorrect by clicking on a respective button on the screen (Figure 1). After giving a response, the participant is presented with a visual-analogue scale for 4 seconds to indicate her decision confidence at the time of responding. Each trial is initiated when a participant clicks on a respective button on the screen. Before each equation, a number is displayed for 0.5 seconds which represents an answer to the following equation and functions as a prime producing low fluency when the answer is incorrect and high fluency when the answer is correct.

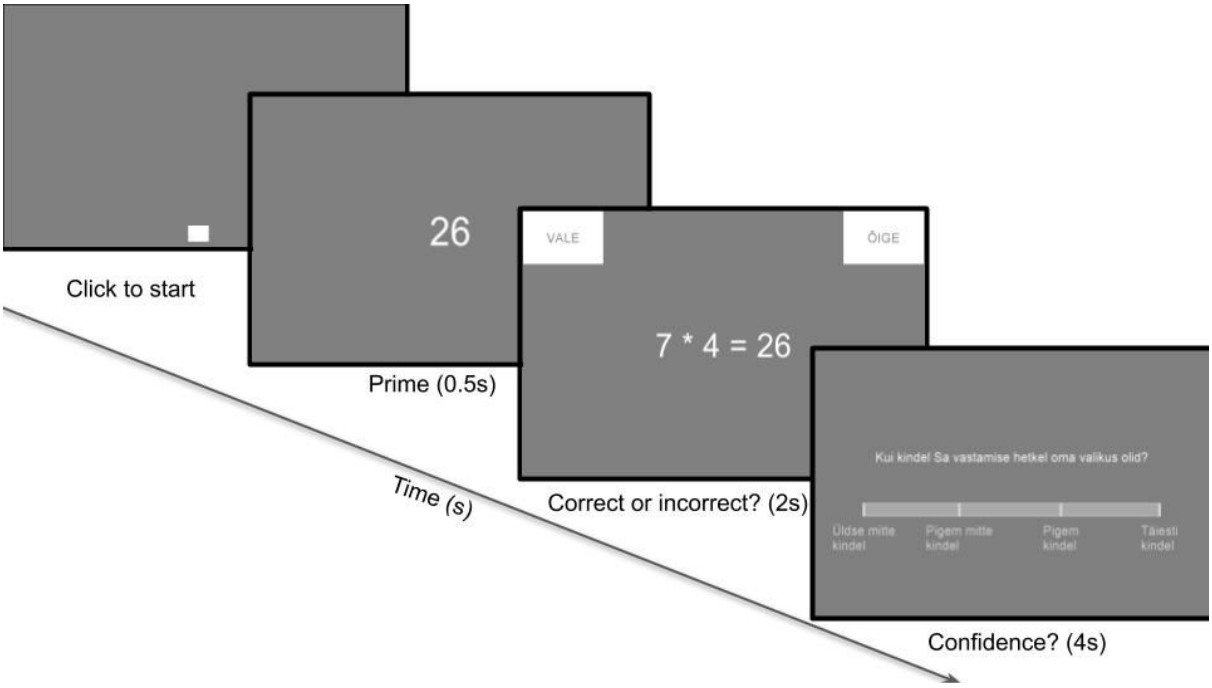

Figure 1. Arithmetic task trial structure. The choice options at the top of the "Correct or incorrect?" screen are "INCORRECT" and "CORRECT". The question above the confidence scale translates as "At the time of responding, how certain were you in Your choice?".

Stimuli

Each participant responded to a total of 160 two-multiplicand equations that were divided into 4 blocks so that each block would have equal distributions of difficulty and fluency. The equations were generated by picking numbers from 4, 6, 7, 8, and 9 for both multiplicands for the easy condition. For hard condition, one multiplicand was selected from a number from 16 to 19 and the other was either 4, 6, 7, 8 or 9. For half of the trials, the displayed answers were correct and incorrect for the other half. The incorrect answers were computed by adding or subtracting 2 from the correct answer so that the incorrect answer would begin with the same number as the correct one. After each block, participants were given feedback of correct answers in counts for the previous block and an amplified result of how well they did in comparison to a fictional study from the USA. The aim of this feedback was to boost participant motivation. Between the 4 blocks, participants reported their emotional state and tiredness on visual-analogue scales.

Procedure

The experimental procedure for each participant took on average about 1 hour and 15 minutes. The experiments took place in the Psychology Institute of University of Tartu. Participants who arrived at the experiment room were requested to sit behind an LCD monitor and asked to sign an informed consent which had also been sent to them earlier. Each participant was introduced to the general setup of the experiment and the basic idea of EEG. The EEG cap was then fit, electrode holders were filled with conductive gel and the electrodes were attached. After setting up the electrodes, each participant was positioned to sit 80 cm from the monitor. They were then instructed to solve computation problems within limited time and high accuracy. Participants were shown how to respond with a mouse click and the visual-analogue confidence evaluation scale was explained to them with a reminder that they should report the confidence they felt during giving a response. During the experiment, the participants were left alone in the room. They were instructed to take breaks between 4 experimental blocks and that after the second block, the experimentator will come in to check whether EEG electrodes are properly connected. The experiment ended with participants filling a short follow-up questionnaire about their opinion of the task, research question and answering strategies. Participants were then briefed about the aims of the project and provided with a possibility to wash their hair.

Design and Measures

We compared the confidence and error-related EEG potentials in two levels of task difficulty which has been previously shown to have a robust effect on decision confidence (Cruz et al., 2016; Griffin & Tversky, 1992). In the experiment, task difficulty was manipulated by changing one multiplicand for half of the equations to a double-digit number (e.g. $16 * 7 = 102$ as opposed to $6 * 7 = 42$ for easy condition). For manipulation of fluency, we implemented an element of priming into the experimental task that involved presenting the answer to the following equation for 0.5 seconds right before the equation was displayed. Since half of the answers for the equations were correct and half incorrect, the primed number was expected to either facilitate (when prime was directing to a correct answer - e.g. 42 before 6 * 7 = 42) or impede (when prime was directing to an incorrect answer e.g. 40 before 6 * 7 = 40) giving the correct answer. Although fluency has been shown to have an effect on decision confidence (Thompson et al., 2013), its effect has been observed to depend on task difficulty (Cruz et al., 2016). To validate the effectiveness of manipulations, we also recorded whether the provided answers were correct and what was the response speed for each answer.

Dependent variables

Decision confidence was measured on a visual-analogue scale from 0-100 with 4 labels at the points of 0, 33, 66 and 100 (Figure 1). The question above the scale asked, “At the time of responding, how certain were You in Your choice?”. The labels of the scale were named starting from 0: “Very unconfident”, “Rather unconfident”, “Rather confident” and “Very confident”.

To capture error-related EEG potentials (ERN, Pe) we used BioSemi ActiveTwo system with 32 electrodes referenced to two electrodes placed on earlobes. To eliminate eye movement noise from the signal, four electrodes were placed bipolarly on each participant vertical and horizontal ocular locations. BioSemi Actiview software was used to record the EEG signal with DMS/DRL reference scheme and 512 Hz sampling rate. EEG preprocessing was conducted in EEGLAB (Brunner, Delorme, & Makeig, 2013)

All scalp channels were re-referenced to linked earlobes. Independent Component Analysis was used to computationally remove ocular artifacts. An ICA mixing matrix was obtained by running the Infomax algorithm on a training data consisting of 4-second epochs (-2000 to 2000 ms relative to the response) of 1 Hz high-pass filtered data cleaned of noisy channels and epochs using EEGLAB automatic algorithms. Components corresponding to blinks and eye movements were identified visually and removed from the mixing matrix before using it to reconstruct the unfiltered continuous data. The ICA-pruned data were then low-pass filtered at 30 Hz and cut into epochs from -400 to 1000 ms relative to the response. The mean voltage between -400 and -200 was subtracted as a baseline. Segments with artifacts were automatically excluded based on the ±100μV threshold criterion. If any single channel would have been exclusively responsible for removing more than 2% of trials, the channel was removed. All removed channels were later spherically interpolated. After preprocessing, on average 67.1% (SD = 15.1%) EEG data in each design cell was retained for each participant.

Following existing literature and visual inspection of the current waveforms, we defined ERN as the mean voltage at Cz and Fz between 0 and 175 ms and Pe as mean voltage at Pz and Cz between 175-325 ms after a response was given.

Data analysis

R environment for statistical computing (R Core Team, 2013) with RStudio interface (RStudio Team, 2015) was used to conduct all the statistical analyses. The analyses were separated into behavioural manipulation validation and main analyses. The general aim of the behavioural manipulation validations was to demonstrate that the experimental manipulations were indeed effective. For all the statistical tests computed, the significance level was set to 0.05, Greenhouse-Geisser correction was applied for repeated-measures ANOVAs when the homogeneity of variances assumption was violated. For the main analyses, we conducted a multilevel mediation analysis, which allows determining whether the effect of an independent variable (IV) on a dependent variable (DV) can be explained by a third intervening variable M (Vuorre & Bolger, 2018). In the context of this study, this translates to whether the effects of fluency (IV) and difficulty (IV) on confidence (DV) are mediated by ERN or Pe (M).

Results

Manipulation validation analyses

The aim of the first analyses was to make sure that fluency and difficulty had indeed influenced decision confidence and response times and accuracy as expected. A mixed-model logistic regression with participant as a random factor indicated that fluency ($\beta = 0.21$), difficulty ($\beta = -0.83$) and their interaction ($\beta = 0.57$) had a statistically significant effect on response accuracy (all $p < 0.001$). Responses were more accurate on easy trials ($p < 0.001$), for which high fluency also had a facilitating effect. For hard trials, the fluency effect was reversed, so that low fluency condition resulted in more accurate responses (Figure 2A).

For response times, repeated-measures ANOVA showed significant main effects of fluency ($F(1, 47) = 39.43$, $p < 0.001$, $\eta^2 = 0.05$) and difficulty ($F(1, 47) = 78.15$, $p < 0.001$, $\eta^2 = 0.30$) as well as their interaction ($F(1, 47) = 75.87$, $p < 0.001$, $\eta^2 = 0.05$). This indicates, that responding times were slower for difficult trials and high fluency condition accelerated responding only for easy trials (Figure 2B).

Similarly for confidence ratings, there were significant main effects of difficulty ($F(1, 47) = 232{,}38$, $p < 0.001$, $\eta^2 = 0.48$), of fluency ($F(1, 47) = 17.45$, $p < 0.001$, $\eta^2 = 0.01$) as well as their interaction and interaction ($F(1,47) = 35.71$, $p < 0.001$, $\eta^2 = 0.02$). Confidence was lower in hard trials and fluency manipulation was effective only in easy trials (Figure 2C).

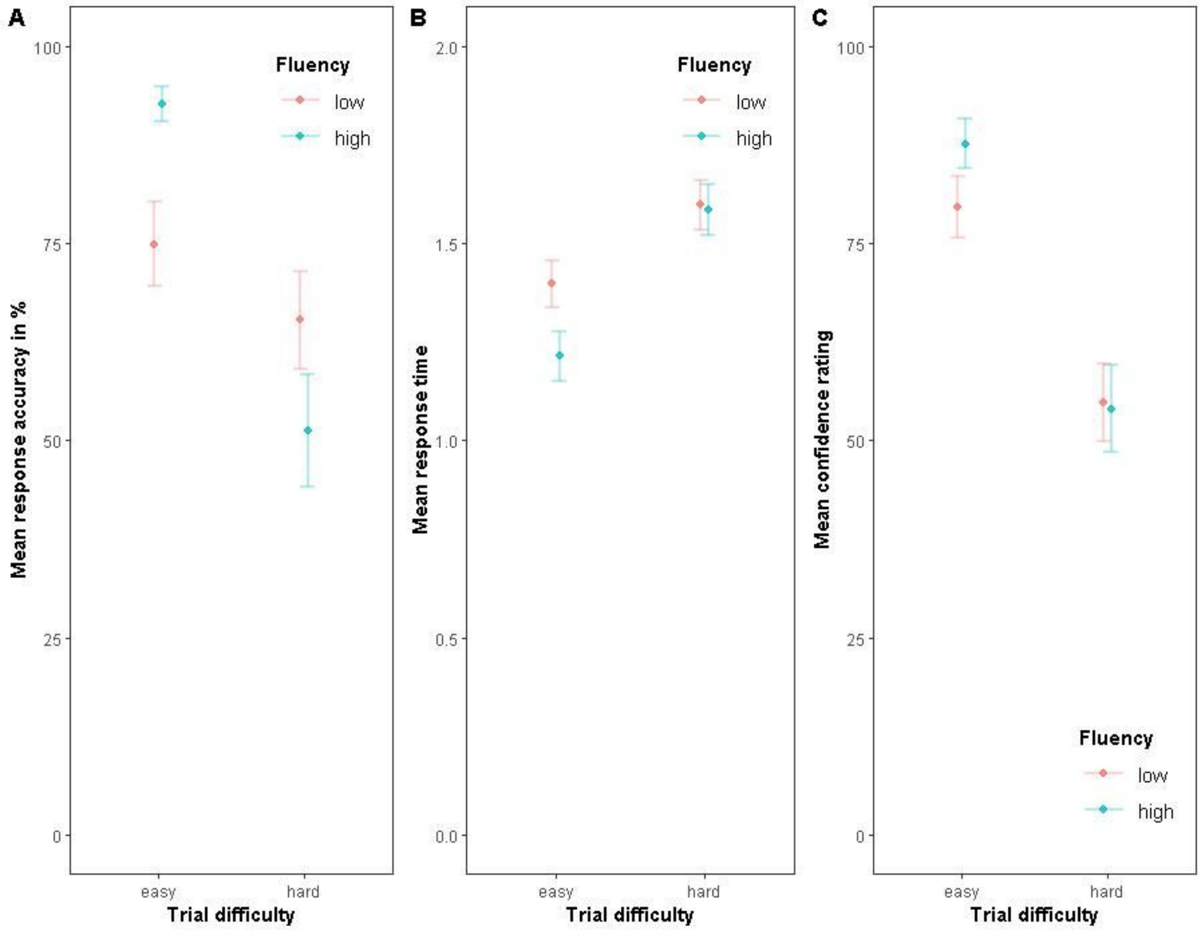

Figure 2. Mean response accuracy (A), response times (B) and confidence ratings (C) on y-scale compared in two levels of trial difficulty (on x-scale) and fluency (colour of the lines).

Main analyses

For estimating the total variation in confidence ratings which can be explained by error-related potentials, we conducted repeated-measures correlation analyses separately for ERN and Pe with *rmcorr* R package (Bakdash & Marusich, 2017). To measure the strength of association, *rmcorr* provides a repeated-measures correlation coefficient (r_{rm}) which is similar to Pearson r (it takes values from 0-1, with 0 representing no association and 1 perfect linear relationship), except that it is more suitable for repeated measures design, when there is a need to take into account between-subject variation. Results indicated that ERN had a small positive association with confidence ratings ($r_{rm}(5108) = 0.05$, $p < 0.001$) whereas Pe had a small negative association with confidence ratings ($r_{rm}(5108) = 0.03$, $p < 0.05$).

Single-trial mediation analyses

We ran four multilevel mediation models with R package *bmlm* (Vuorre & Bolger, 2018) to determine, whether the experimental manipulations of difficulty and fluency on confidence ratings would be mediated by error-related ERPs. The *bmlm* package allows estimating Bayesian mediation models for data with repeated-measures experimental design. Each mediation model was estimated with 4 Markov Chain Monte Carlo chains with 10 000 iterations for each chain. In each model, confidence ratings were treated as the dependent variable, difficulty (easy/ hard) and fluency (high/ low) as the independent variables and ERN or Pe as the mediators. To quantify the effects, we use non-standardized path coefficients (me for indirect effect and c` for the direct effect of the independent variable on the dependent variable). An effect is detected if the 95% credible intervals for the coefficient do not contain 0. We also use pme (proportion effect mediated) which indicates the proportion of total effect explained by the mediating variable ranging from 0 to 1 (Vuorre & Bolger, 2018).

The aim of the first two mediation models was to see whether the effect of difficulty on decision confidence would be mediated by ERN and Pe. We found that ERN did not mediate the effect of difficulty on confidence (indirect effect = -0.05, 95% CI = [-0.21, 0.01], proportion mediated effect = 0.00, 95% CI = [0.00, 0.01]). We also found that the effect of difficulty on confidence ratings was not mediated by Pe (indirect effect = -0.02, 95% CI = [-0.08, 0.03], proportion mediated effect = 0.00, 95% CI = [0.00, 0.00]). In this model, however, an effect of difficulty on Pe was detected (a = 0.95, 95% CI=[0.28, 1.62]). This indicates that although the effect of difficulty on decision confidence was not mediated by ERN nor Pe, the difficulty manipulation had an effect on Pe but not ERN (Figure 3).

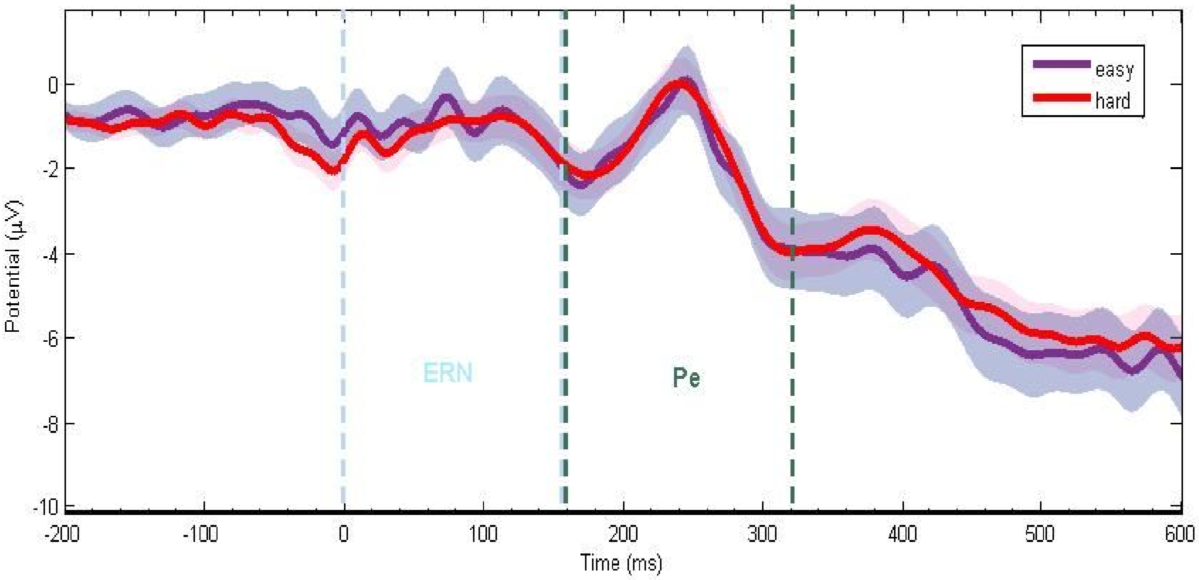

Figure 3. Differences in ERN and Pe for two levels of task difficulty. The plot captures the effect of the path a in the first and second mediation model. The shaded area denotes the standard error of the mean.

The next analysis was concerned with whether the effects of fluency on decision confidence are mediated by error-related potentials. As the manipulation validations indicated that fluency did not have an effect on decision confidence in the hard trials, we conducted the mediation analyses only for the easy trials.

Mediation analysis for easy trials indicated that the effect of fluency on confidence was partially mediated by ERN (indirect effect = 0.15, 95% CI = [0.01, 0.03], proportion mediated effect = 0.02, CI = [0.00, 0.04]). The direct effect remained significant after taking ERN into account (c' = 7.96, 95% CI = [5.75, 10.15]) (Figure 4). Although the effect was small, this result corresponds to the idea that the fluency effect on decision confidence is partially mediated by ERN.

The coefficient for path a which denotes the effect on fluency on ERN in easy task condition also excluded zero (a = 1.4, 95% CI = [0.55, 2.2]). This demonstrates that the fluency manipulation was effectively distinguished in ERN (Figure 5), which in turn has an effect on decision confidence (b = 0.11, 95% CI = [0.015, 0.2]) (Figure 4).

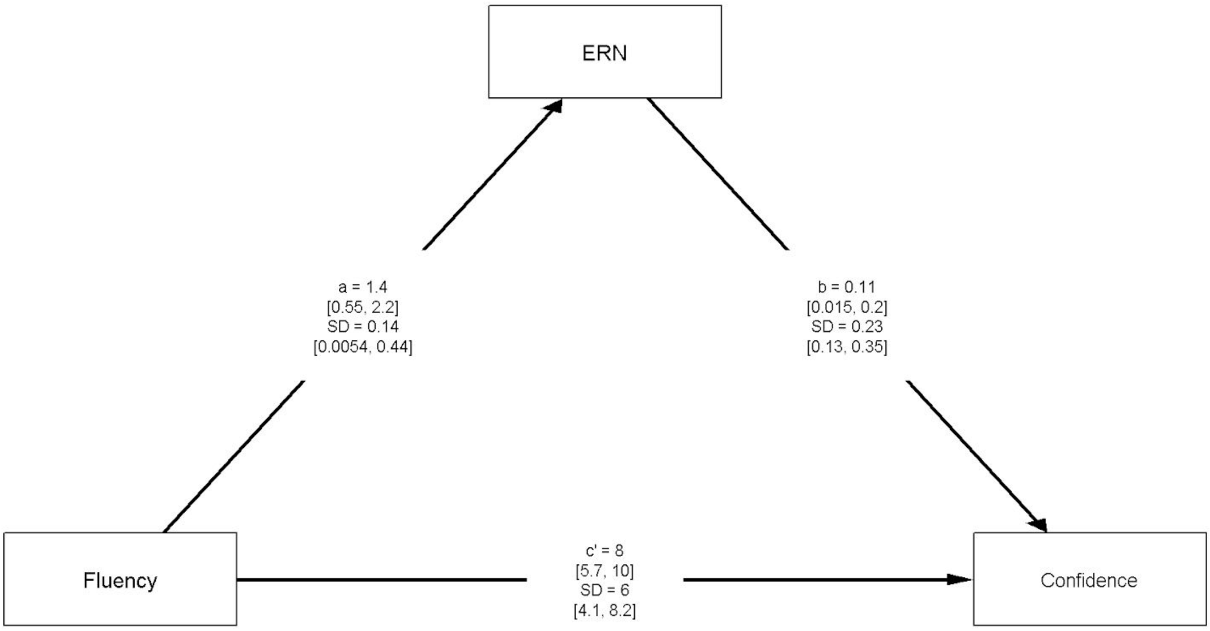

Figure 4. Path diagram for the mediation model for ERN for easy trials. Point estimates (a, b, c`) describe direct effects with associated 95% credible intervals. "SD" shows the effect's standard deviation which indicates the degree of variation between participants.

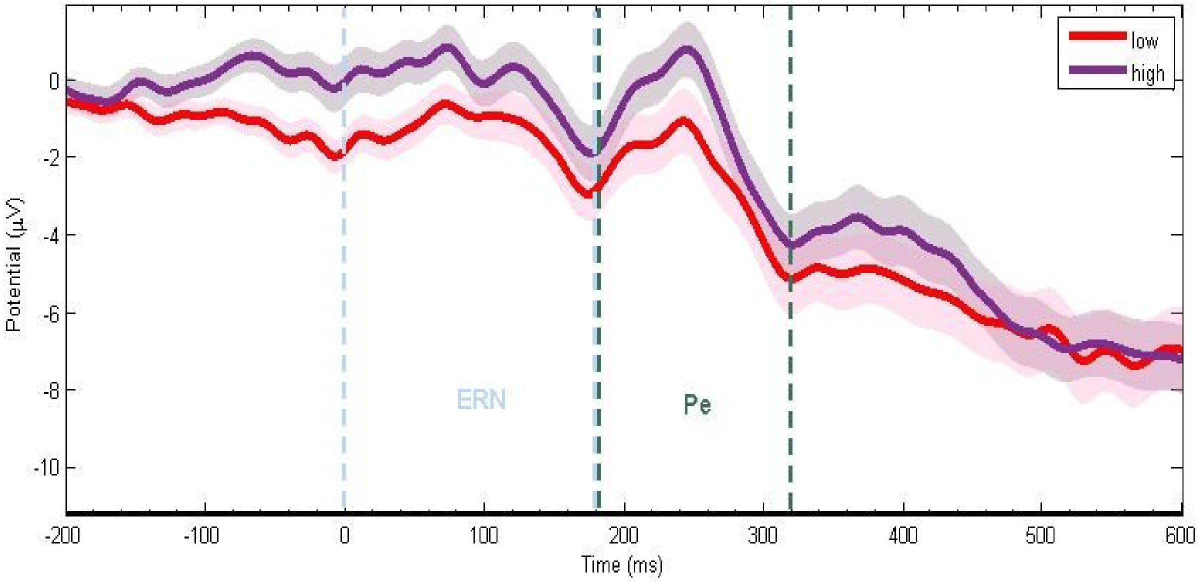

Figure 5. Differences in ERN and Pe for high and low fluency trials for easy task condition. The plot captures the effect of the path a in the third and fourth mediation model. The shaded area denotes the standard error of the mean.

In the final mediation model, our aim was to see whether the effect of fluency on confidence ratings in easy task condition is mediated by Pe. Results indicated that fluency had a direct effect on confidence (c` = 7.98, 95% CI = [5.91, 10.18]), when taking Pe into account, and it was not mediated by Pe (indirect effect = 0.01 , 95% CI = [-0.13, 0.26], proportion mediated effect = 0.00, CI = [-0.02, 0.03]). For this model, no direct effect of fluency on Pe was detected (a = 0.80, 95% CI = [-0.17, 1.83]), the differences in Pe for two levels of fluency in are visually displayed on Figure 5.

Discussion

The aim of the current thesis was to understand the relation of error-related EEG potentials to decision confidence. To do this, we set up an arithmetic reasoning experiment with simultaneous EEG recording, in which we could systematically manipulate two known determinants of decision confidence: task difficulty and fluency. Our goal was to see whether the effects of these manipulations on confidence ratings would be mediated by error-related EEG potentials (ERN, Pe).

We first confirmed that manipulation of task difficulty indeed had an expected effect on decision confidence, so that confidence ratings were on average lower in hard condition and higher in easy condition. The effect of task difficulty was further validated by analyses of reaction times and response accuracy - easy equations were solved on average more accurately and with less time than hard equations. For fluency, an effect on confidence was detected only in low but not in hard condition. As expected, confidence ratings on easy trials were higher in high fluency condition and lower in low fluency condition. This dynamic was also supported by response times and accuracy.

We also found a reversed effect of fluency on accuracy for hard trials, so that low fluency resulted in higher accuracy. A way to explain this effect is to appeal to general heuristic rules that participants tended to use. One of the heuristics for hard trials (e.g. 16 * 7 = 114) was to see whether the product of the second digit in the first number and the second multiplicand (e.g. 6 * 7 = 42) ends with the same number as the solution of the equation (114). If this is not the case, then a mistake is made very rarely if one responds “incorrect”. However, if the numbers match (e.g. the answer would be 112), then the correct answer is more uncertain since there are more possible options for an answer ending with the final number of the product (e.g. 102). The use of this heuristic was also reported by many participants in the questionnaire following the experiment.

We set up 4 hypotheses, to quantify the extent to which ERN and Pe mediate the effect of fluency and difficulty on confidence ratings. Our first pair of hypotheses were that if ERN and Pe underlie the computation of decision confidence, a manipulation of task difficulty on confidence should be mediated by ERN and Pe. We found no support for these hypotheses as no indirect effect for ERN nor Pe on decision confidence were detected. The second pair of hypotheses were concerned with whether the fluency effect on confidence would be mediated by ERN and Pe. As our manipulation validation indicated no effect of fluency in hard task condition, we ran two mediation models for only easy trials. Results showed that evidence for partial mediation of fluency by ERN on confidence was detected. For Pe, however, there was no evidence for a mediated relationship.

The fact that task difficulty effect on confidence was not mediated by error-related EEG potentials has implications for the theory about how metacognitive representations (e.g. confidence) contribute to adaptive decision-making. A proposed role for metarepresentations is to allocate mental resources (e.g working memory) for more flexible behaviour when necessary (Botvinick et al., 2001; Thompson et al., 2011). This function, also called “cognitive control”, has been associated with the anterior cingulate cortex (ACC) on the neuronal level, which has also been localized as the source of ERN and Pe (Botvinick et al., 2001; Shackman et al., 2011). The results from this study challenge the claim that ACC is always responsible for generating signals that allocate cognitive control resources (Gangemi et al., 2015; Shenhav, Botvinick, & Cohen, 2013). More specifically, if ACC is responsible for the monitoring for the recruitment of cognitive control, we should have observed a pattern in ERN and Pe in which hard problems elicit larger amplitude ERNs and Pes than easy problems. The data from this study, however, indicates that this is not the case for ERN, which is insensitive to task difficulty and does not mediate its effect on decision confidence.

An effect of difficulty on Pe was detected, which indicates that Pe, which has also been associated with error-awareness (Yeung & Summerfield, 2012) is also sensitive to task difficulty. But this activity does not in turn manifest in confidence ratings, which generates a puzzle: on the one hand, difficulty has an effect on both - decision confidence and Pe, but they seem to be unrelated in the sense that although Pe indexes difficulty, it does not in turn contribute to confidence ratings. This indicates that although the effect of difficulty on confidence is not mediated by ERN nor Pe, it is possible that Pe could be tracking cognitive control processes that remain implicit in the sense that they do not contribute to the feeling of confidence. These processes may thus be inaccessible for verbal reports (Shea et al., 2014), but could nevertheless have an adaptive regulatory function over behavioural responses. A remarkable example of this is a neurological patient, with damaged ACC, who did not report increased mental effort with increased task difficulty, although she clearly allocated more resources to the task when it was harder to solve (Naccache et al., 2005).

To sum up, the results from the first two mediation analyses demonstrate a dissociation between error-related EEG potentials and confidence ratings, which indicates that a simple theory, according to which ERN and Pe underlie computation of decision confidence is not supported. The results are more aligned with a model in which Pe is an implicit index of task difficulty that is not accessible for verbal reports.

The second pair of mediation analyses with fluency as the independent variable suggest a more nuanced picture in which error-related EEG potentials relate to confidence ratings but only for simple decisions. Validation analyses indicated that fluency only had an effect on decision confidence in easy trials. A similar difference in the effectiveness of fluency manipulation on confidence has also been previously observed in a mathematical reasoning task (Cruz et al., 2016). The authors explain this by arguing that fluency only influences feelings of error up to a point where the effort to solve the problem disrupts the link between fluency and confidence. Based on this, it could be argued that the neural system susceptible to the effect of fluency is the previously introduced metacognitive error-monitoring system associated with ACC activity and ERN/Pe (Yeung et al., 2004; Yeung & Summerfield, 2012). This hypothesis is in line with the evidence from the mediation models with fluency as the independent variable. The model indicated that the effect of fluency on decision confidence was partially mediated by ERN. As expected - ERN amplitude was more negative in high fluency condition which in turn was manifested in lower confidence ratings. This is consistent with the idea that for simple decisions confidence is an outcome of a pattern-matching error-detection system which is subject to fluency effects, whereas for more complex decisions this mechanism is no longer solely responsible for confidence computations.

These results can be taken as a support for a theory, in which ERN contributes to decision confidence in simple decisions. Moreover, the results suggest, that the error-monitoring system associated with ACC can be influenced by manipulation of answer fluency. Although Pe shows a similar pattern in two conditions of fluency (Figure 5), this effect was not found to be significant.

Contributions, limitations and further questions

Understanding of neural mechanism for decision confidence has important consequences for our understanding of metacognition and cognitive control. A relevant practical example is a study in which brain-computer interface which provided participants feedback about ERN and Pe resulted in 21% more accurate responses in a visual discrimination task (Parra, Spence, Gerson, & Sajda, 2003). The current approach in which we tried to untangle the behavioural level effects on confidence by theoretically relevant neural activity patterns contributes toward understanding our reasoning and decision-making in multiple ways.

First, the results suggest, that at the current state of knowledge, the implicit measurement of decision confidence through error-related EEG potentials is more tractable for relatively simple decisions for which error is clearly definable as a representation conflict. This does not necessarily imply that confidence in more difficult decisions cannot be implicitly measured. Rather the concept of difficulty warrants more rigorous behavioural level analysis, which could be then adequately mapped onto relevant neural mechanisms (Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017). Currently, difficulty can be thought of as an emergent property in reasoning, which could arise for various reasons (too much information, not enough time, too many necessary operations, etc). This notion is therefore highly variable for different tasks and the question “What makes a problem difficult?” warrants a detailed analysis in each setting, in order to capture this on the neurobiological level. It is possible that a similar study design would benefit from a more specific manipulation of confidence since the current result indicate that the partial mediation effect of ERN is relatively small compared to the total effects of difficulty and fluency.

The main contributions of the current study are the discovery that fluency effect on confidence is partially mediated by ERN for simple decisions and that Pe tracks task difficulty but does not mediate confidence ratings. These results set constraints on the development of metacognitive theory of decision-making and reasoning, by arguing that error-processing in the brain works in different ways for easy and difficult problems. The effect of fluency on ERN also provides a window into manipulating ERN for various further research questions. For example, it would be important to replicate this result in a simple perceptual task, to see, whether the effect of fluency would be more pronounced.

Another question inspired from these results is about the relationship between negative affect and confidence. It has been shown that negative affect is associated with greater ERN and Pe (Hajcak, McDonald, & Simons, 2004; Shackman et al., 2011). Accordingly, ERN and Pe could be interpreted as signals that convey information about uncertainty which in the case of high probability of error is experienced as a negative feeling (FOE) and should result in low ratings of confidence. However, as the data from the current study suggest, even if negative affect is manifested in confidence ratings, it is not to be associated with ERN or Pe. It could also be the case that although higher Pe is associated with negative affect, this affective experience does not contribute to confidence ratings. Therefore in future studies, it would be interesting to see if Pe would mediate difficulty effects on negative affect. This avenue of research should be further investigated, as it is often taken for granted that negative noetic feelings result in low confidence ratings.

Finally, the study extends the relationship between decision confidence and error-related EEG potentials to arithmetic reasoning task. This is important, because it shows the validity of the conclusion, that ERN and Pe contribute to confidence in different decision-making tasks. To extend these claims, it would be beneficial to address the ecological validity of the decision-making task and consider more nuanced manipulations of decision confidence in future studies.

The current thesis provides evidence that ACC activity is related to confidence in simple pattern-matching decisions, but not in decisions which require more complex reasoning. Error-related negativity is supported to be an index of decision confidence for simple decisions whereas Pe can be thought of as a manifestation of implicit cognitive control. As these are *post-hoc* explanations, further experiments should be designed to test them.

Acknowledgements

I want to thank my supervisor Andero for being a bottomless source of inspiration and feedback during the research process. I am grateful for all the smiles and support by my partner Kata, my family and friends.

References

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, 456.

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

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624.

Brunner, C., Delorme, A., & Makeig, S. (2013). Eeglab–an open source matlab toolbox for electrophysiological research. *Biomedical Engineering/Biomedizinische Technik*.

Cruz, A. L. F., Arango-Muñoz, S., & Volz, K. G. (2016). Oops, scratch that! Monitoring one's own errors during mental calculation. *Cognition*, *146*, 110–120.

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. *Electroencephalography and Clinical Neurophysiology*, *78*(6), 447–455.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906.

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.

Gangemi, A., Bourgeois-Gironde, S., & Mancini, F. (2015). Feelings of error in reasoning—in search of a phenomenon. *Thinking & Reasoning*, *21*(4), 383–396.

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.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, *24*(3), 411–435.

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.

Hajcak, G., McDonald, N., & Simons, R. F. (2004). Error-related psychophysiology and negative affect. *Brain and Cognition*, *56*(2), 189–197.

Konstabel, K., Lönnqvist, J.-E., Leikas, S., Velázquez, R. G., Qin, H., Verkasalo, M., &

Walkowitz, G. (2017). Measuring single constructs by single items: Constructing an even shorter version of the “Short Five” personality inventory. *PloS One*, *12*(8), e0182714.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one’s knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 187.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, *93*(3), 480–490. https://doi.org/10.1016/j.neuron.2016.12.041

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.

Naccache, L., Dehaene, S., Cohen, L., Habert, M.-O., Guichart-Gomez, E., Galanaud, D., & Willer, J.-C. (2005). Effortless control: executive attention and conscious feeling of mental effort are dissociable. *Neuropsychologia*, *43*(9), 1318–1328.

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.

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2003). Response error correction-a demonstration of improved human-machine performance using real-time EEG monitoring. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *11*(2), 173–177.

Peirce, C. S., & Jastrow, J. (1884). *On small differences in sensation*.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 1–9.

Redi, K. (2017). *Vajadus kognitiivse selguse järele küsimustiku adapteerimine ja seosed sotsiaalse soovitavusega* (Thesis, Tartu Ülikool). Retrieved from https://dspace.ut.ee/handle/10062/57298

R Core Team. (2013). *R: A language and environment for statistical computing*.

RStudio Team. (2015). RStudio: integrated development environment for R. RStudio. *Inc., Boston, MA*, *14*.

Scheffers, M. K., & Coles, M. G. (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.

Selimbeyoglu, A., Keskin-Ergen, Y., & Demiralp, T. (2012). What if you are not sure? Electroencephalographic correlates of subjective confidence level about a decision. *Clinical Neurophysiology*, *123*(6), 1158–1167.

Shackman, A. J., Salomons, T. V., Slagter, H. A., Fox, A. S., Winter, J. J., & Davidson, R. J. (2011). The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nature Reviews Neuroscience*, *12*(3), 154.

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.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240.

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition*, *34*(3), 619–632.

Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107–140.

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition*, *128*(2), 237–251.

van Veen, V., & Carter, C. S. (2006). Error detection, correction, and prevention in the brain: a brief review of data and theories. *Clinical EEG and Neuroscience*, *37*(4), 330–335.

Vuorre, M., & Bolger, N. (2018). Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience. *Behavior Research Methods*, *50*(5), 2125–2143.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049.

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.

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.

Non-exclusive licence to reproduce thesis and make thesis public