Cognitive Changes in Conjunctive Rule-Based Category Learning: An ERP Approach

Rahel Rabi[a], Marc F. Joanisse[b], Tianshu Zhu[b], and John Paul Minda[b]

[a]The Rotman Research Institute of Baycrest Centre

[b]The University of Western Ontario

Address correspondence to

John Paul Minda
Department of Psychology
The University of Western Ontario
London, ON N6A 5C2
jpminda@uwo.ca

WORD COUNT: 5833

Abstract

When learning rule-based categories, working memory is needed to maintain the currently active rule in memory, update rule information following feedback, and to select a new rule if necessary. Prior research has demonstrated that conjunctive rules are more complex than unidimensional rules and place greater demands on working memory. To better understand the cognitive processes underlying complex rule-based category learning, event-related potentials (ERPs) were recorded while participants performed a conjunctive rule-based category learning task with trial-by-trial feedback. In line with prior research, correct categorization responses resulted in a larger stimulus-locked late positive complex compared to incorrect responses, indexing the updating of rule information in memory. As well, incorrect trials elicited a more pronounced feedback-locked P300 compared to correct trials, suggesting a participant's confidence in their rule-based strategy. Among strong learners only, differential processing of easy and hard categorization stimuli was examined. A large late positive slow wave emerged for difficult compared to easy categorization stimuli, suggesting differential processing of category items even though strong learners performed quite well on the conjunctive category set. Overall, the findings suggest that ERP can be used to better understand the cognitive processes involved in rule-based category learning.

Keywords: category learning; event-related potentials; rule learning; computational model

Event-Related Brain Potential Correlates of Conjunctive Rule-Based Category Learning

Categorization is a core cognitive process that is used on a daily basis to help organize the world around us. Individuals are continually acquiring new categories through the process of hypothesis testing, where rules are formulated and tested to determine whether they can be used to determine category membership. The process of hypothesis testing encompasses key cognitive abilities like working memory and selective attention. In learning explicit rules, individuals must generate hypotheses regarding the possible rule, maintain candidate rules in working memory, switch from one rule to another, and update this information in working memory. During this hypothesis testing process, the types of categorization rules formulated and used can vary in complexity. For example, younger children might find it easy to learn to categorize shapes by using a simple, single-dimensional rule, based on the number of sides the shape has. Older children may be able to learn to use a conjunctive rule to categorize shapes based on the number of sides and the number and measurements of the angles. Likewise, adults may rely on simpler rules when carrying out tasks like organizing files based on urgency, and rely on more complex rules when performing tasks like doing the laundry (e.g., sorting clothes based on colour and washing procedure) or driving (e.g., determining when to drive based on the colour of the traffic light and the presence of pedestrians).

Fundamental Mechanisms Involved in Category Learning

Behavioural and neuropsychological studies have demonstrated the underlying cognitive mechanisms involved in category learning. For example, working memory processes are necessary for learning rule-based categories. Rule-based category learning has been shown to be impaired in individuals with frontal lobe damage (Schnyer et al., 2009), children (Huang-Pollock, Maddox, & Karalunas, 2011; Minda, Desroches, & Church, 2008; Rabi, Miles, & Minda, 2015; Rabi & Minda, 2014), older adults (Maddox, Pacheco, Reeves, Zhu, & Schnyer, 2010; Rabi & Minda, 2016), and cognitively depleted adults (Miles & Minda, 2011; Minda & Rabi, 2015). The reason for this appears to stem from decreased working memory abilities in these individuals, which is an important cognitive process required for rule-based learning (Minda & Miles, 2010). Accordingly, category learning studies involving concurrent tasks have shown that participants display reduced category learning performance when learning a rule-based category set while performing a concurrent task that interferes with working memory (Miles & Minda, 2011; Minda et al., 2008; Waldron & Ashby, 2001).

Along the same lines, Zeithamova and Maddox (2006) examined whether performing a concurrent task would differentially interfere with learning a simple versus a more complex rule-based category set. In the simple, single-dimensional rule condition, participants needed to formulate a rule based on one dimension while ignoring the other. In the conjunctive rule condition, participants had to attend to and integrate two different featural dimensions to arrive at the correct rule, which requires more working memory capacity than the single-dimensional rule condition. Interestingly, Zeithamova and Maddox (2006) demonstrated that both single-dimensional and conjunctive rule-based category learning was disrupted by performing a concurrent task, with participants in the conjunctive condition relying more so on single-dimensional rules, rather than conjunctive rules, to make categorization decisions. All in all, it is clear that working memory is a key process required when learning any type of rule-based category set.

Functional neuroimaging studies have also examined the link between working memory and rule-based category learning (Ashby & O'Brien, 2005; Poldrack & Foerde, 2008; Smith & Grossman, 2008). The brain regions thought to be involved in working memory and selective attention processes, such as the prefrontal cortex and parietal regions, have been shown to be involved when participants learn rule-based categories (Ashby & Ennis, 2006; Grossman et al., 2002; Seger & Miller, 2010). Jiang et al., (2007) and Li et al., (2009) showed that the prefrontal cortex conveyed the crucial conjunctions between key features in determining category membership. Nomura and Reber (2008) tested the prediction that the medial temporal lobe plays a key role in successful rule-based categorization. Findings revealed that successful rule-based categorization was associated with increased activity in the anterior medial temporal lobe (Seger, 2008). Nomura and Reber (2008) suggested that the medial temporal lobe acts together with the prefrontal cortex structures and the head of the caudate to identify verbalizable rules for categorization. In line with these findings, Filoteo and colleagues (2005) showed that differential activation was observed between those participants who learned a single-dimensional rule-based category set compared to those who did not in frontal and parietal regions known to be involved in working memory. Additionally, the head of the caudate was found to be associated with incorrect responding in participants who successful learned the rule. The involvement of this structure could either be related to processing negative feedback or in initiating a switch between potential category rules. In support of Filoteo and colleagues' (2005) findings, Monchi, Petrides, Petre, Worsley, and Dagher (2001) also demonstrated that the caudate is differentially activated on incorrect versus correct trials when participants performed the Wisconsin Card Sorting Test (a task thought to measure rule-based category learning and rule switching). Moreover, the basal ganglia interacts with the frontal cortex in corticostriatal loops, and exerts the function of strategy selection in categorization tasks (Seger, 2008). Cincotta and Seger (2007) and Merchant et al., (1997) suggested that the basal ganglia plays a critical part in learning that involves trial and error. Differential activation was seen in corticostriatal loops in subjects' progress from novices to experts in categorization. The anterior caudate determines the rate of learning; greater activation is associated with more rapid learning (Williams & Eskandar, 2006) and greater sensitivity to prediction error (Haruno & Kawato, 2006).

Research involving event related potentials (ERPs) may help to further clarify the cognitive processes involved in rule-based category learning. While ERP research on category learning is limited, some researchers have begun to examine the ERP correlates of rule-based category learning. Morrison, Reber, Bharani, and Paller (2015) monitored brain activity using ERP as participants learned, via feedback, to sort Gabor patches that varied in spatial frequency and orientation. Morrison and colleagues used a single-dimensional rule-based category set, with the rule depending on the frequency of the bands in the Gabor patches. Furthermore, participants had to test various rules, inhibit the incorrect rule corresponding to the orientation dimension, and update this information in memory. ERP findings revealed a differential correct/incorrect response in positive parietal potentials only, with participants displaying larger correct/incorrect difference performing the task more accurately. These positive potentials were thought to represent a Late Positive Complex (LPC), which has been associated with working memory (Kok, 2001; Polich, 2007). Morrison and colleagues suggested that the differential LPC responses for correct/incorrect responding reflected the engagement of the neural system responsible for making rule-based decisions. With regards to feedback processing, Morrison and colleagues observed a P300 response on incorrect trials. The P300 is an ERP component elicited in the process of decision making, and it has been associated with responses to unexpected information (Polich, 2007). During the rule-based categorization task participants are forming hypotheses about the rule to use, and when those expectations are violated by negative feedback, participants have to re-evaluate their choice of strategy. Morrison and colleagues also considered the alternative conclusion that the P300 effect may reflect memory updating. According to this viewpoint, the rule maintained in memory must be changed as a result of negative feedback.

Rationale for the Present Research

Although there is considerable evidence for the role of working memory in rule-based category learning, only a small number of studies have examined this relationship with ERP techniques. The previously described studies asked participants to learn a single dimensional rule but almost nothing is known about how these results will generalize to more complicated, two-dimensional rules. The current ERP study will examine the event-related brain potential correlates of complex, conjunctive rule-based category learning. In line with the findings of Morrison et al., we predict that a differential correct/incorrect response will be found in positive parietal potentials, indicative of an LPC, and because the LPC is thought to be an index of memory access and updating (Polich, 2007), correct responses should result in a larger LPC compared to incorrect responses, representing the updating of rule information in memory. Also in line with Morrison et al., we predict a larger P300 for incorrect compared to correct trials. The P300 is an index of violations of expectancy and so incorrect responses should influence a participant's confidence in the explicit rule they are choosing to use.

The current study will use what is known about ERP signatures of category learning to identify the mechanisms by which individuals converge on the correct categorization rule. In comparison to the single-dimensional rule-based task used by Morrison and colleagues (2015), the use of a more complex conjunctive rule-based task in the current study will enable us to examine a wider range of ERP effects. That is, in addition to examining the ERPs associated with stimulus and feedback processing, the present study will also allow us to look at how specific stimuli are processed differently based on their predicted classification according to one of two easier, suboptimal rules or a more difficult, optimal rule. While on the surface it may appear that a participant has learned and correctly applied a conjunctive rule, ERP measures may reveal that the participant is processing individual stimuli differently based on their difficulty level (i.e., whether a stimulus belonging to one category set shares a dimensional value with stimuli belonging to another category set). The ERP data can therefore provide valuable insights into the process of category learning that are not available strictly through analyses of behavioural (accuracy and RT) data.

Finally, the present study fits several computational models to the behavioral data at successive points in the category learning trajectory in order to understand how strategies may change with time. This is particularly informative for the conjunctive rule category set, as we predict that participants will first acquire the easier, suboptimal rules and will then shift to the more complicated and optimal conjunctive rule. This transition may not occur for all participants and is expected to rely on executive function abilities like working memory.

Methods

Participants

Thirty-eight undergraduate students (14 males and 21 females; mean age = 20.10 years, *SD* = 3.37) were recruited from the University of Western Ontario. The data from three participants were not included in the analyses because two of them performed at chance on the categorization task and the third participant had poor quality ERP recordings. All participants reported normal or corrected to normal vision and fluency in English. Participants received either course credit or $20 for their participation in the study.

Materials

The category structure used in this study was identical to the Zeithamova and Maddox (2006) study. For the category learning task, participants classified sine-wave gratings that varied in spatial frequency and orientation. There were 40 Category A and 40 Category B stimuli. The 80 stimuli were generated by sampling from four bivariate normal distributions. Three were assigned to Category A and one to Category B. The distribution parameters are displayed in Table 1. A scatterplot of the stimuli and the optimal rule is presented in Figure 1. The required that participants respond Category B when the spatial frequency was high and the orientation was steep and to respond Category A otherwise. We used the PsychoPy package (Peirce, 2007) to generate sine wave gratings corresponding to each coordinate sampled from the distributions above. For both category sets, sine wave grating frequency was calculated as f = 0.25 + (x_f/50) cycles per stimulus and orientation was calculated as $o = x_o \times (\pi/20)$ degrees. See Zeithamova and Maddox (2006) for more details regarding the category structure used in the current study.

Procedure

Category learning task. In the category learning task, participants were told that they would be seeing a blurred image on the screen and their job was to determine whether that image belonged to Category A or Category B. Responses were made using a button box labeled 'A' and 'B' for Categories A and B, respectively. Participants were told that they would receive feedback after every response, and that they should use this feedback to help them learn to make as many correct responses as possible. The experiment was programmed and presented Stimuli were presented electronically using the E-Prime 2.0 software (Psychology Software Tools, 2012) via a 17-inch CRT monitor.

The manner in which the stimuli were presented on each trial is displayed in Figure 2. On each trial, participants saw a fixation cross, followed by a sine wave grating in the center of the screen and an A and B in the upper left and upper right corner of the screen. Upon making a response, the sine wave grating disappeared, and feedback was delivered in the center of the screen (either a checkmark or an X). If a participant took longer than 2500 ms to respond, no feedback was presented, and no response was recorded for that trial. In between each trial, a blank screen appeared for 750 ms. Participants were presented with six blocks of the 80 stimuli; 480 trials in total. Within a block, the order of presentation of all 80 stimuli was randomized for each participant.

The participant took a short break between blocks. Participants completed 5 practice trials to familiarize themselves with the task prior to data collection. Participants were warned in advance that they would have 2.5 seconds to make a response, resulting in very few 'no response' trials. Participants were also advised to blink during the blank screen following feedback, to avoid excessive blinking during stimulus and feedback presentation which would contaminate EEG recordings.

EEG recording and preprocessing. EEG data were recorded at 32 scalp sites placed in the international 10-20 orientation using BioSemi ActiveTwo Ag/AgCl electrodes embedded in a custom elastic cap (BioSemi, Amsterdam, The Netherlands). Electro-oculogram (EOG) activity was recorded from active electrodes placed above, beside, and beneath the left eye, and beside the right eye. An additional active electrode (CMS – common mode sense) and a passive electrode (DRL – driven right leg) were used to comprise a feedback loop for amplifier reference. Two additional electrodes were placed at the left and right mastoids for offline re-reference. All EEG electrode impedances were maintained below 20 kΩ. All bioelectric signals were digitized on a PC using ActiView software (BioSemi) at a rate of 512 Hz with a bandpass of 0.1-100Hz and a 60 Hz notch filter.

Offline analysis was performed using EEGLAB v. 13.4.3b (Delorme & Makeig, 2004) and ERPLAB v. 4.0.3.1 (Lopez-Calderon & Luck, 2014). All data were re-referenced to the mean left/right mastoid electrodes, and bandpass filtered with cutoffs of 0.1 Hz and 30 *Hz*. The trials were epoched from 200 ms prior to the onset of the target stimulus to 800 ms after the onset of the target stimulus, and baseline corrected to the 200 ms pre-stimulus onset. The data were segmented into stimulus-locked and feedback-locked epochs. Trials containing eye blinks and other nonocular artifacts (EEG activity exceeding ±75 microvolts at any electrode) were discarded.

Results

Behavioural Analyses

Category Learning. The average proportion correct was obtained by calculating the mean proportion correct in each block for each participant and then averaging across participants The resulting learning curve is shown in Figure 3. A repeated measures ANOVA revealed a main effect of block, F (4, 126) = 37.98, p < .001, $\eta^2_{partial}$ = .525, [Greenhouse-Geisser corrected], indicating that learning occurred across blocks. The categorization performance of participants began at 60% correct during the first block, and ended with a categorization performance of 80% by the final block. For certain ERP analyses, only strong learners were considered, classified as participants whose response pattern over the last three learning blocks was best fit by a two-dimensional model (see below). This group consisted of 20 participants, whose performance began at 63% during the first block, and ended with a categorization performance of 85% by the final block (see Figure 3).

Computational modeling. For insight into the response strategies used by our participants, we fit decision bound models to each block of each participant's data (for details see (Ashby, 1992; Maddox & Ashby, 1993; Zeithamova & Maddox, 2006). Two unidimensional models were fit to each observer's responses, which assume that the participant sets a criterion based on one of the stimulus dimensions, either the frequency or orientation of the lines. In both unidimensional models, the intercept of the decision bound was allowed to vary. We also fit a class of two-dimensional models to the data which assume that the participant correctly based their categorization decision on both dimensions (i.e., the optimal, conjunctive rule). In one version of this model, the slope and intercept were allowed to vary and in another version of the model, only the intercept was allowed to vary. Finally, we fit two guessing models, that assumed no dimensional strategy (one assumed that participants randomly responded A or B with equal probability for each response and the other assumed unequal probability). Examples of the block by block strategy analysis can be seen in Figure 4. We fit these models to each subject's data by maximizing the log likelihood. Parameters for each model were estimated using the maximum likelihood method and the relative fit of the models were compared using the Bayesian Information Criterion (*BIC*, where $BIC = r \ln(N) - 2 \ln L$; r is the number of free parameters, *N* is the number of trials being fit, and *L* is the likelihood of the model given the data). *BIC* is a measure of goodness of fit, which penalizes a model for extra free parameters. To find the best model to account for each participant's responses, a *BIC* value is computed for each model, and the model associated with the smallest *BIC* value is chosen. A learner was classified as a participant using one of the two strategies (unidimensional, two-dimensional), aside from guessing. The optimal conjunctive rule yields the highest accuracy (close to 100%). Note that applying a unidimensional strategy could also result in good performance (i.e., accuracy of up to 80%), although this performance would not be as high as those using the optimal strategy. We assumed that both single-dimensional and two-dimensional strategy users would rely on hypothesis testing to test different rules, and would rely on working memory to update information based on the feedback received.

In the strategy analysis, learners were classified as anyone fit by a single-dimensional or two-dimensional model in their last block of learning. As mentioned in the methods section, only two participants were excluded from this analysis because their strategy performance indicated that they were guessing during their final block. Among the remainder of the participants, 9 were best fit by a unidimensional model (i.e., either a frequency or orientation strategy) and 26 were best fit by a two-dimensional model. See Table 2 for a complete list of the proportion of participants fit by each type of strategy per learning block.

ERP analyses

The data from the most central 13 electrodes were included in the analyses (see Figure 5). Fewer than 25% of trials were excluded for any given participant (i.e., non-random responders). There were 480 trials total.

Category learning. Similar to Morrison et al. (2015), correct/incorrect subtractions were performed to examine category learning. The dependent measure in our analyses was the mean amplitude of the epoch from 300 ms to 600 ms post-stimulus onset at parietal scalp sites (i.e., the P300 component), P7, P3, Pz, P4 and P8. We averaged the parietal electrodes for the analysis of variance. As shown in Figure 6, a late positive parietal ERP (300-600 ms) was predictive of correct categorization, $F(1, 34) = 5.03$, $p = .03$, $\eta^2_{partial} = .128$. The late positive parietal ERP (i.e., late positive component; LPC) was larger for correct ($M = 4.56$, $SD = 3.6$) than incorrect ($M = 4.21$, $SD = 3.4$) trials, suggesting memory access and updating during category learning. A topographic map of the ERP difference between correct and incorrect trials is illustrated in Figure 7.

Feedback Processing. ERPs recorded during feedback were examined for the presence of a P300 effect. Correct/incorrect P300 subtractions were performed. The dependent measure in our analyses was the 50% fractional area latency of the epoch from 300 ms to 600 ms after feedback onset at frontocentral, central, and parietal scalp sites, FC1, FC2, C3, C4, Cz, CP1, CP2, and Pz. We averaged the eight electrodes for the analysis of variance. As expected, a feedback-locked P300 ERP effect was found, $F(1, 34) = 59.86$, $p < .001$, $\eta^2_{partial} = .638$, with a pronounced P300 for incorrect ($M = 403.55$, $SD = 30.09$) compared to correct ($M = 373.28$, SD = 30.07) trials (see Figure 8). Note that although the timing of the P300s differed appreciably for correct and incorrect trials, we were primarily interested in the amplitude of this waveform, and not its timing. Thus the 50% fractional area latency allowed us to calculate their amplitudes in a way that did not require us to set different a priori time analysis windows for the two conditions. The difference in amplitudes suggests that the error feedback response was unexpected for learners. One could speculate that unidimensional learners would be surprised when they made an error, because using a unidimensional rule could still result in good performance if applied consistently. Two-dimensional rule learners would also be surprised when they made an error, because their rule tended to work for the majority of trials.

To confirm that ERP findings related to stimulus processing and feedback processing were not influenced by the inclusion of single-dimensional rule learners, the same analyses mentioned above were conducted excluding the nine single-dimensional rule learners. Results remained significant after removing these participants, indicating that single-dimensional rule learners did not drive the ERP effects found.

Stimulus difficulty. While on the surface it may appear that strong conjunctive rule-based learners were performing very well, they may have still been processing different types of stimuli differently. That is, the optimal rule required that participants respond Category B when the spatial frequency was high and the orientation was steep and to respond Category A otherwise (which included 3 different subsets of Category A), since there were 40 stimuli belonging to Category B, and 40 stimuli belonging to Category A (8 stimuli belonging to Category

A1, 16 belonging to Category A2, and 16 belonging to Category A3). In this regard, Category B could be interpreted as being the easier category to learn, because there is only one subset of this category and 40 exemplars per learning block (i.e., Category B has a stronger family resemblance structure compared to Category A). On the contrary, participants would have to learn that exemplars A1 (low frequency & shallow orientation), A2 (low frequency and steep orientation), and A3 (high frequency & shallow orientation) all belong to Category A, even though they look different. Among the three subtypes of Category A, A2 and A3 were the most difficult to learn because they shared one dimension in common with Category B. That being said, ERP analyses were used to examine how participants differed in processing easier (Category B) versus harder (Category A2 and A3) stimuli.

We were particularly interested in strong conjunctive rule-based learners, because these included individuals who learned a conjunctive rule and consistently applied it across a large number of trials. By analyzing data from strong conjunctive learners, this enabled us to have enough data points to compare correct responses on Category A2/A3 to Category B. A strong learner was classified as any participant who was best fit by a two-dimensional rule-based strategy during at least the last three learning blocks. There were 20 participants included in this analysis. The dependent measure in our analyses was the mean amplitude of the epoch from 500 ms to 800 ms post-stimulus onset at parietal and centro-parietal scalp sites, CP1, CP2, P3, Pz and P4. We averaged the parietal and centro-parietal electrodes for the analysis of variance, for correct responses only. As shown in Figure 9, a late positive slow wave emerged, $F(1, 19) = 22.46$, $p < .001$, $\eta^2_{partial} = .542$, reflecting a larger late positive slow wave for difficult ($M = 4.83$, $SD = 3.84$) compared to easy ($M = 3.49$, $SD = 2.89$) categories. These findings are in line with past research showing that positive slow waves with centro-parietal distributions were elicited by target stimuli in difficult perceptual discrimination tasks (Ruchkin, Johnson, Mahaffey, & Sutton, 1988; Ruchkin & Sutton, 1983). Furthermore, research by Gunter, Jackson, and Mulder (1995) suggests that positive slow waves reflect difficulty of perceptual operations and memory storage. Processing Category A2 and A3 stimuli require greater effort and memory requirements than processing exemplars from Category B. To further illustrate this finding, topographic maps of the ERP difference between hard versus easy stimuli, for correct responses only, is illustrated in Figure 10.

Discussion

The primary focus of the present study was to examine if the cognitive processes involved in categorization can be distinguished physiologically. Gabor patch stimuli were used, similar to those used in many behavioural studies involving rule-based category learning. Categorization accuracy was assessed, and stimulus-locked and feedback-locked ERPs were examined. Unlike past studies involving more simple, single-dimensional rule-based categorization tasks, the current study set out to examine how stimuli and feedback are processed when categorization is governed by a more complex rule. Behavioural data indicated that most participants were able to learn the conjunctive rule-based category set, albeit performance was low during the first few learning blocks as participants tested various rules and changed strategies. Strategy analysis of the behavioural data revealed that participants varied in the strategy they adopted when completed the conjunctive rule-based task, with the majority of participants adopting a two-dimensional rule-based strategy.

ERP findings demonstrated that various key components are involved in the categorization process. To begin, when processing the categorization stimuli, an LPC was found, which was larger for correct compared to incorrect trials. This finding is in line with previous research by Morrison and colleagues (2015), suggesting that an individual's working memory is continually updated when viewing categorization stimuli. As new rules are tested and dismissed, this information is updated in memory. Unlike Morrison et al., who used a single-dimensional rule-based category set, the current study used a conjunctive rule-based category set. One potential difference between these category sets is that in the single-dimensional category set, inhibitory control is required to inhibit one of the stimulus dimensions (e.g., categorize based on the frequency of the lines in the Gabor patch, while ignoring the orientation of the lines), whereas in the conjunctive category set we used, less inhibitory control may be recruited because participants have to integrate information from both dimensions to arrive at the correct categorization rule (e.g., categorize based on the frequency and orientation of the lines in the Gabor patch). It may also be plausible that conjunctive category learning require substantial inhibition at the rule level, with irrelevant rules being inhibited. However, this does not undermine our argument that this category is likely to require less inhibition than single-dimensional category set. In addition, the conjunctive category set places heavier demands on working memory capacity to solve, because more hypothesis testing is required to arrive at the correct, more complicated rule. The fact that an LPC was found in both a single-dimensional and a conjunctive rule-based category set suggests that this component indexes working memory processes.

Aside from Morrison and colleagues (2015), who examined the LPC in the context of a category learning task, other studies have also supported the idea that the LPC is involved in memory updating and decision accuracy. The LPC has been important in studies of explicit recognition memory (Rugg et al., 1998), and is generally found to be largest over parietal scalp sites. Additionally, the LPC has also been shown to be sensitive to decision accuracy. For example, Finnegan, Humphreys, Dennis, and Geffen (2002) found that a larger LPC amplitude was elicited in response to accurately categorized word stimuli. Participants were presented with new unstudied words and old words, which had been presented at an earlier time. Results revealed that LPC amplitude was larger in ERPs evoked by words, which were correctly recognized, compared to incorrect recognition decisions.

The second cognitive process of interest in the current study was the manner in which participants processed feedback regarding their categorization decisions. More specifically, we were interested in how learners would process negative feedback compared to positive feedback. We found that incorrect trials elicited a more pronounced P300 compared to correct trials, which is in line with research by Morrison and colleagues (2015). This suggests that learners developed confidence in their categorization strategy and were surprised upon receiving negative feedback. The current finding is in line with prior research demonstrating that P300s occur more based on deviant stimuli or stimuli that have lower probabilities (Duncan-Johnson & Donchin, 1982; Johnson, 1984). Similarly, when presented with a gambling task to complete, Hajcak, Holroyd, Moser, and Simmons (2005) found that P300 amplitude was largest for the unexpected outcomes, confirming that participants indeed formed expectations regarding feedback. Furthermore, it appears that learners in our study developed confidence in their categorization strategy over time, and as a result, reacted unexpectedly upon receiving negative feedback on some trials.

The examination of stimulus-locked and feedback-locked ERPs, included data from participants classified as learners. In the current study, learners were classified as a participant adopting either a single-dimensional rule-based strategy or a two-dimensional strategy. While the optimal conjunctive rule would yield the highest accuracy (100% possible); single-dimensional rules based on either dimension could provide an accuracy of up to 80%, and two-dimensional strategies could also be successful as well, due to the relatively high separability of the four underlying distributions (Zeithamova & Maddox, 2006). That being said, participants classified as random guessers based on computational modeling of strategy use, were removed from data analysis. To alleviate any concerns that single-dimensional rule learners may be processing the stimuli in a distinctly different manner from two-dimensional rule learners, additional analyses were conducted. Single-dimensional rule-learners were removed and the data was re-analyzed only including two-dimensional rule-learners. Results remained significant when excluding single-dimensional rule-learners from the analysis, suggesting that single-dimensional and two-dimensional strategy users processed the categorization stimuli in a similar manner. That is, both types of strategy users would need to take part in the same cognitive processes (i.e., hypothesis testing, rule switching, and memory updating). Future research would benefit from examining complex rule-based category learning using a category set where only a complicated rule can result in good performance. Adopting a more simplistic rule-based strategy will result in much lower categorization performance. If this were done, good learners and poor learners could be compared to determine whether any differences in processing existed based on the strategy the participant adopted.

In addition to examining stimulus-locked and feedback-locked ERPs for correct versus incorrect categorization responses, also of interest in the current study was if and how participants would differentially process easy and hard categorization stimuli. Since only correct trials were analyzed, only the strongest learners were included in the analysis to ensure that a sufficient number of trials (i.e., at least 30 trials per learner) in each of the category type bins (i.e., A1, A2, A3, and B). Strong learners were classified as any participant who was best fit by a two-dimensional rule-based strategy during at least the last three learning blocks. Additionally, we were interested in comparing hard versus easy stimuli in strong learners, in particular, because these participants were scoring the highest on the categorization task. Among these high-performers, we were interested in determining whether they would process stimuli differently (easy versus hard) even though they were categorizing all stimulus types with high optimal rule accuracy. Results revealed a larger late positive slow wave for difficult compared to easy categorization stimuli, confirming that processing differences exist based on stimulus difficulty and taking us a step past what behavioural research can tell us. Behavioural data can show that participants are performing well on the conjunctive category set, but this type of data can tell us very little about stimulus processing demands, highlighting the importance of ERP.

Numerous studies have reported the emergence of positive slow waves (PSW) in more difficult task conditions, unrelated to category learning (Cremer, Kok, Zeef, & Keuss, 1996; Kok, 1986/8; Kok, Vijver, & Rooijakkers, 1985). For example, research by Ruchkin (Ruchkin et al., 1988; Ruchkin & Sutton, 1983) showed that PSWs with a centroparietal distribution were elicited by target stimuli in difficult perceptual discrimination tasks. These findings are in line with results from the current study, which found a PSW for more difficult task stimuli in a perceptual-based category learning task across centroparietal electrodes. Various explanations have been given to explain PSWs. Ruchkin et al. (1988) suggested that PSWs reflect the difficulty of perceptual operations and memory storage. A second suggestion has been that PSWs are functionally closely related to P3 and represent continued processing of perceptually difficult stimuli (Kok & Looren de Jong, 1980; Ruchkin, Sutton, Kietzman, & Silver, 1980). These prior findings support the PSW found in the present study, because our more difficult categorization stimuli (i.e., A2 & A3) were perceptually similar to the opposing Category type (i.e., B), and furthermore required heavier processing efforts to retrieve information about these categorization stimuli from memory. Future research may benefit from examining whether PSWs are a functionally distinct ERP component related to task difficulty or whether PSWs represent a delayed P3. Furthermore, the emergence of PSWs for difficult categorization stimuli has important implications for our understanding of complex, rule-based category learning. In everyday life we often encounter various members of a category, some more difficult than others to categorize. Given the fact that not all members of a category set are treated/processed in the same way, ERP research on category learning can shed light on how stimulus difficulty impacts the categorization process.

In summary, the current study took a novel approach to understanding the cognitive mechanisms involved in making categorization decisions by measuring event-related potentials during a complex rule-based task. Overall, findings demonstrated the effectiveness of real-time neural monitoring during category learning and provide evidence highlighting the cognitive mechanisms involved in rule-based category learning. Results suggested that learning complex categories engenders qualitative changes in brain activity that are marked by a late positive component during stimulus processing, indexing the updating process of working memory. Additionally, a P300 component was present during feedback processing, indexing confidence in categorization decisions and showing the different neural responses to correct and incorrect trials. An analysis of the differential responding to difficult and less difficult stimuli revealed a larger late positive slow wave for more difficult stimuli. This finding suggests that all categorization stimuli are not treated/processed the same way, even if performance is quite high. The ERP analyses clearly allowed us to understand the nature of how these different kinds of stimuli were processed in a way that traditional behavioral analyses could not. This suggests that ERP analyses can be used to better understand the cognitive processes involved in different categorization tasks.

References

Ashby, F. G. (1992). *Multidimensional models of categorization*. Lawrence Erlbaum Associates, Inc.

Ashby, F. G., & Ennis, J. M. (2006). The Role of the Basal Ganglia in Category Learning. In *Psychology of Learning and Motivation Volume 46* (Vol. 46, pp. 1–36). Elsevier.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences*, *9*(2), 83–89.

Cincotta, C. M., & Seger, C. A. (2007). Dissociation between striatal regions while learning to categorize via feedback and via observation. *Journal of Cognitive Neuroscience*, *19*(2), 249–265.

Cremer, R., Kok, A., Zeef, E., & Keuss, P. (1996). Age-related effects of different types of noise and stimulus quality: An event-related potential (ERP) study. *Journal of Psychophysiology*. Retrieved from http://psycnet.apa.org/psycinfo/1996-06847-005

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.

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biological Psychology*, *14*(1-2), 1–52.

Filoteo, J. V., Maddox, W. T., Simmons, A. N., Ing, A. D., Cagigas, X. E., Matthews, S., & Paulus, M. P. (2005). Cortical and subcortical brain regions involved in rule-based category learning. *Neuroreport*, *16*(2), 111–115.

Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP "old/new" effects: memory strength and decisional factor (s). *Neuropsychologia*, *40*(13), 2288–2304.

Grossman, M., Koenig, P., DeVita, C., Glosser, G., Alsop, D., Detre, J., & Gee, J. (2002). The neural basis for category-specific knowledge: an fMRI study. *NeuroImage*, *15*(4), 936–948.

Gunter, T. C., Jackson, J. L., & Mulder, G. (1995). Language, memory, and aging: An electrophysiological exploration of the N400 during reading of memory-demanding sentences. *Psychophysiology*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.1995.tb02951.x/full

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology*, *42*(2), 161–170.

Haruno, M., & Kawato, M. (2006). Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Networks: The Official Journal of the International Neural Network Society*, *19*(8), 1242–1254.

Huang-Pollock, C. L., Maddox, W. T., & Karalunas, S. L. (2011). Development of implicit and explicit category learning. *Journal of Experimental Child Psychology*, *109*(3), 321–335.

Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, *53*(6), 891–903.

Johnson, R., Jr. (1984). P300: a model of the variables controlling its amplitude. *Annals of the New York Academy of Sciences*, *425*, 223–229.

Kok, A. (1986/8). Effects of degradation of visual stimuli on components of the event-related potential (ERP) in go/nogo reaction tasks. *Biological Psychology*, *23*(1), 21–38.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577.

Kok, A., & Looren de Jong, H. (1980). Components of the event-related potential following degraded and undegraded visual stimuli. *Biological Psychology*, *11*(2), 117–133.

Kok, A., Vijver, F. R., & Rooijakkers, J. (1985). Effects of Visual Field, Stimulus Degradation, and Level of Practice on Event-Related Potentials of the Brain. *Psychophysiology*, *22*(6), 707–717.

Li, S., Mayhew, S. D., & Kourtzi, Z. (2009). Learning shapes the representation of behavioral choice in the human brain. *Neuron*, *62*(3), 441–452.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. *Psychological Review*, *111*(2), 309–332.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70.

Maddox, W. T., Pacheco, J., Reeves, M., Zhu, B., & Schnyer, D. M. (2010). Rule-based and information-integration category learning in normal aging. *Neuropsychologia*, *48*(10), 2998–3008.

Merchant, H., Zainos, A., Hernández, A., Salinas, E., & Romo, R. (1997). Functional properties of primate putamen neurons during the categorization of tactile stimuli. *Journal of Neurophysiology*, *77*(3), 1132–1154.

Miles, S. J., & Minda, J. P. (2011). The effects of concurrent verbal and visual tasks on category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(3), 588–607.

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: a comparison of children and adults. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(6), 1518–1533.

Minda, J. P., & Miles, S. J. (2010). The Influence of Verbal and Nonverbal Processing on Category Learning. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 52, pp. 117–162). Academic Press.

Minda, J. P., & Rabi, R. (2015). Ego depletion interferes with rule-defined category learning but not non-rule-defined category learning. *Frontiers in Psychology*, *6*, 35.

Monchi, O., Petrides, M., Petre, V., Worsley, K., & Dagher, A. (2001). Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *21*(19), 7733–7741.

Morrison, R. G., Reber, P. J., Bharani, K. L., & Paller, K. A. (2015). Dissociation of category-learning systems via brain potentials. *Frontiers in Human Neuroscience*, *9*, 389.

Nomura, E. M., & Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience and Biobehavioral Reviews*, *32*(2), 279–291.

Poldrack, R. A., & Foerde, K. (2008). Category learning and the memory systems debate. *Neuroscience and Biobehavioral Reviews*, *32*(2), 197–205.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology:*

Official Journal of the International Federation of Clinical Neurophysiology, *118*(10), 2128–2148.

Psychology Software Tools, I. (2012). E-Prime 2.0 (Version 2). Pittsburgh, PA. Retrieved from http://www.pstnet.com.

Rabi, R., Miles, S. J., & Minda, J. P. (2015). Learning categories via rules and similarity: comparing adults and children. *Journal of Experimental Child Psychology*, *131*, 149–169.

Rabi, R., & Minda, J. P. (2014). Rule-based category learning in children: the role of age and executive functioning. *PloS One*, *9*(1), e85316.

Rabi, R., & Minda, J. P. (2016). Category learning in older adulthood: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *Psychology and Aging*, *31*(2), 185–197.

Ruchkin, D. S., Johnson, R., Jr, Mahaffey, D., & Sutton, S. (1988). Toward a functional categorization of slow waves. *Psychophysiology*, *25*(3), 339–353.

Ruchkin, D. S., & Sutton, S. (1983). 11 Positive Slow Wave and P300: Association and Disassociation. *Advances in Psychology*, *10*, 233–250.

Ruchkin, D. S., Sutton, S., Kietzman, M. L., & Silver, K. (1980). Slow wave and P300 in signal detection. *Electroencephalography and Clinical Neurophysiology*, *50*(1-2), 35–47.

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature*, *392*(6676), 595–598.

Schnyer, D. M., Maddox, W. T., Ell, S., Davis, S., Pacheco, J., & Verfaellie, M. (2009). Prefrontal contributions to rule-based and information-integration category learning. *Neuropsychologia*, *47*(13), 2995–3006.

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience and Biobehavioral Reviews*, *32*(2), 265–278.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203–219.

Smith, E. E., & Grossman, M. (2008). Multiple systems of category learning. *Neuroscience and Biobehavioral Reviews*, *32*(2), 249–264.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*(1), 168–176.

Williams, Z. M., & Eskandar, E. N. (2006). Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nature Neuroscience*, *9*(4), 562–568.

Zeithamova, D., & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition*, *34*(2), 387–398.

Author Notes

Rahel Rabi is currently a postdoctoral fellow at the The Rotman Research Institute of Baycrest Centre, Baycrest Hospital, Toronto, ON

Marc F. Joanisse, Tianshu Zhu, and John Paul Minda, are at the Department of Psychology and the Brain and Mind Institute at the The University of Western Ontario, London, ON

This research was supported by research grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada to MFJ and JPM and a NSERC Canada Graduate Scholarship to RR. Infrastructure support was provided by an NSERC RTI grant. We thank Kazunaga Matsuki for his development and assistance of the strategy analysis program. Address correspondence to John Paul Minda at jpminda@uwo.ca

Table 1: Distribution Parameters for the Conjunctive Rule-Based Category Set

Category Structure	μ_f	μ_o	σ^2	$cov_{f,o}$	N
Category A_1	283	98	75	0	8
Category A_2	317	98	75	0	16
Category A_3	283	152	75	0	16
Category B	317	152	75	0	40

Note. Stimuli from the A_1, $A_{2,}$ and A_3 distributions were all members of Category A.

Table 2: Number of subjects fit by each class of decision bound models

Model	Two-Dimensional	Single-Dimensional	Guessing
Block 1	1	21	13
Block 2	10	19	6
Block 3	17	12	6
Block 4	22	9	4
Block 5	28	7	0
Block 6	26	9	0

Note. The optimal model is shown in bold. There were 35 participants included in the study.

Figure Captions

Figure 1. Conjunctive category structure used. Open circles represent Category A and filled circles represent Category B. The dashed line represents the optimal decision bound.

Figure 2. Stimulus presentation during each trial.

Figure 3. Average proportion of correct categorization responses as a function of learning block for all participants compared to just the strong learners. Error bars denote the standard error of the mean.

Figure 4. An example of the modeling results across blocks for one participant's data. Triangles indicate actual Category A items, and circles indicate actual Category B items; Filled symbols indicate a participant's Category A responses, and open symbols indicate a participant's Category B responses. The red lines show the best fitting decision-boundary model for each block. In this example, the participant started off using a single dimensional rule (i.e., orientation rule indicated by the horizontal line), but switched to a conjunctive rule for the remainder of the categorization task.

Figure 5*.* Electrode montage. Filled circles indicate electrodes included in the analyses. Red outlined circles represent electrodes examined during stimulus processing (i.e., category

learning, stimulus difficulty, and stimulus frequency). Blue outlined circles represent electrodes examined during feedback processing.

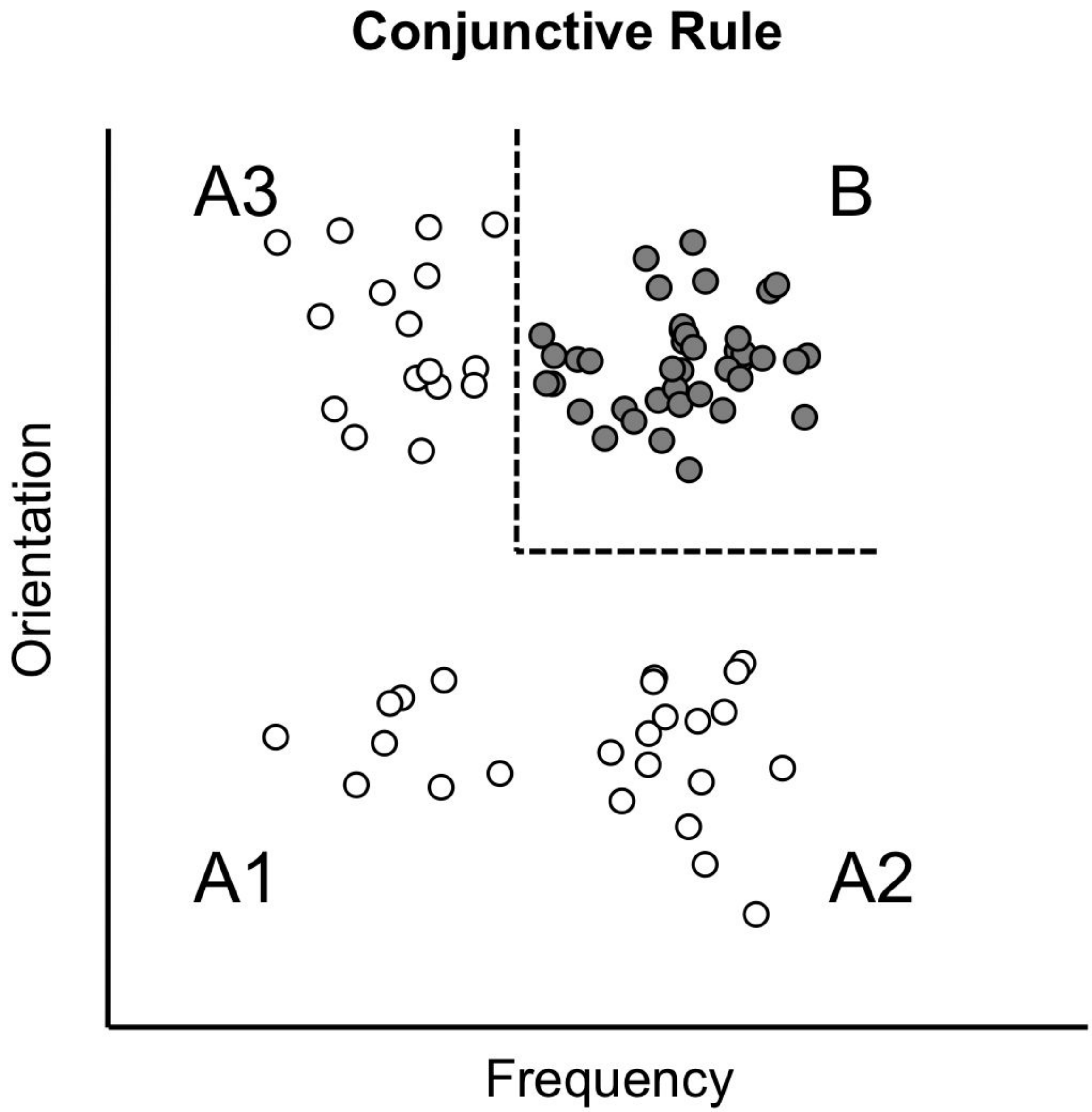
Figure 6. Stimulus-locked ERPs for correct vs. incorrect trials across parietal electrodes.

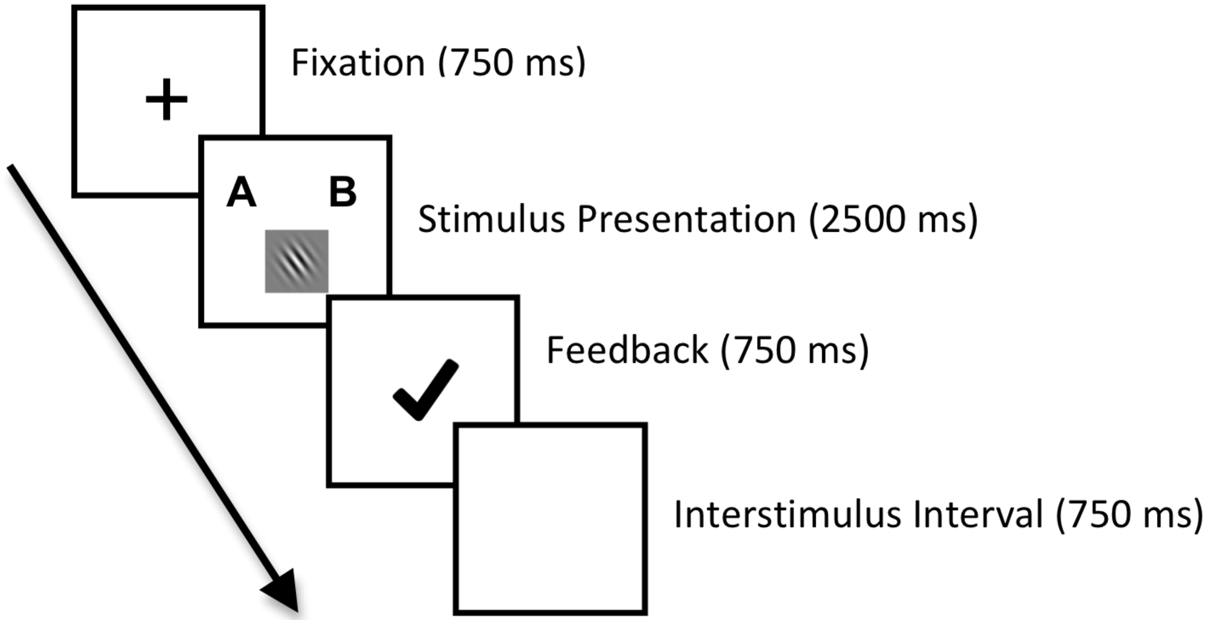
Figure 7. A topographic map of the correct minus incorrect subtraction of mean amplitude between 300-600 ms for stimulus-locked ERPs.

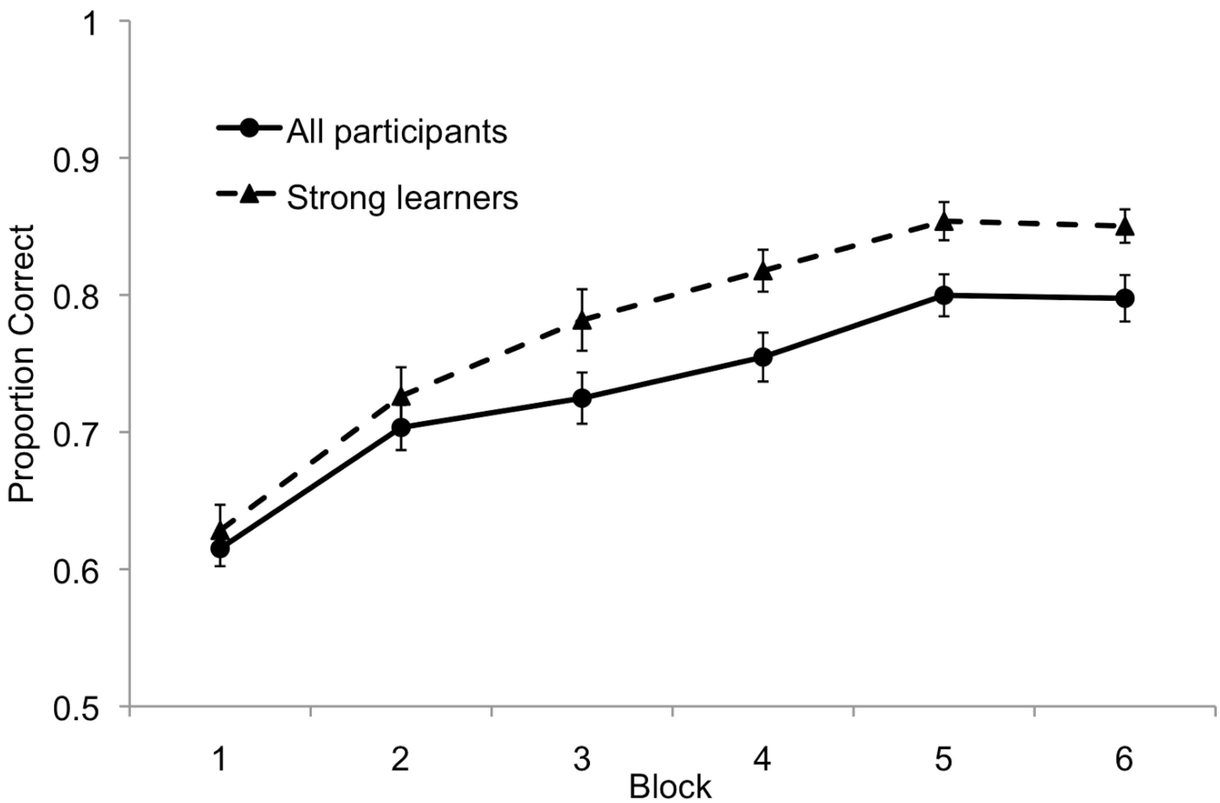
Figure 8. Feedback-locked ERPs for correct vs incorrect trials.

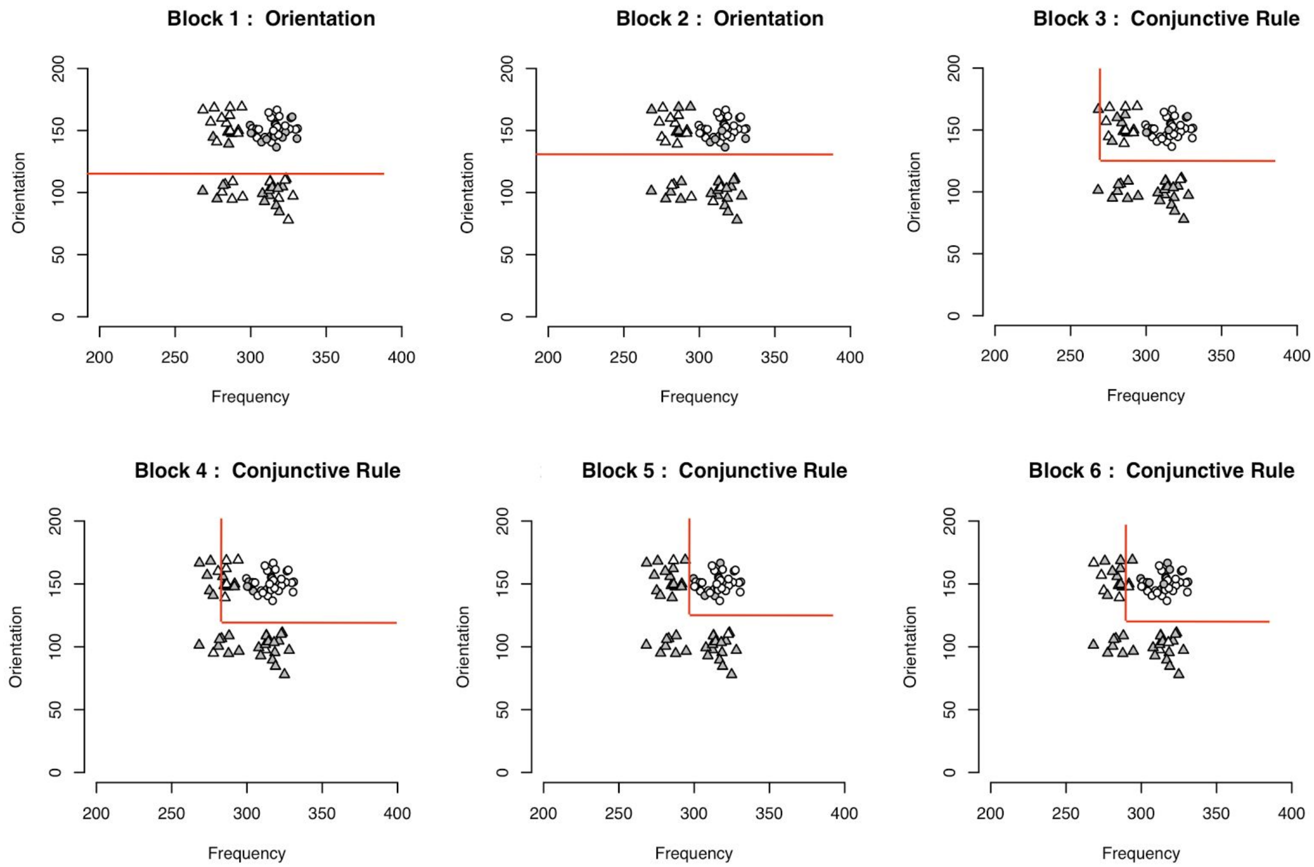
Figure 9. Stimulus-locked ERPs for easy versus difficult stimulus items.

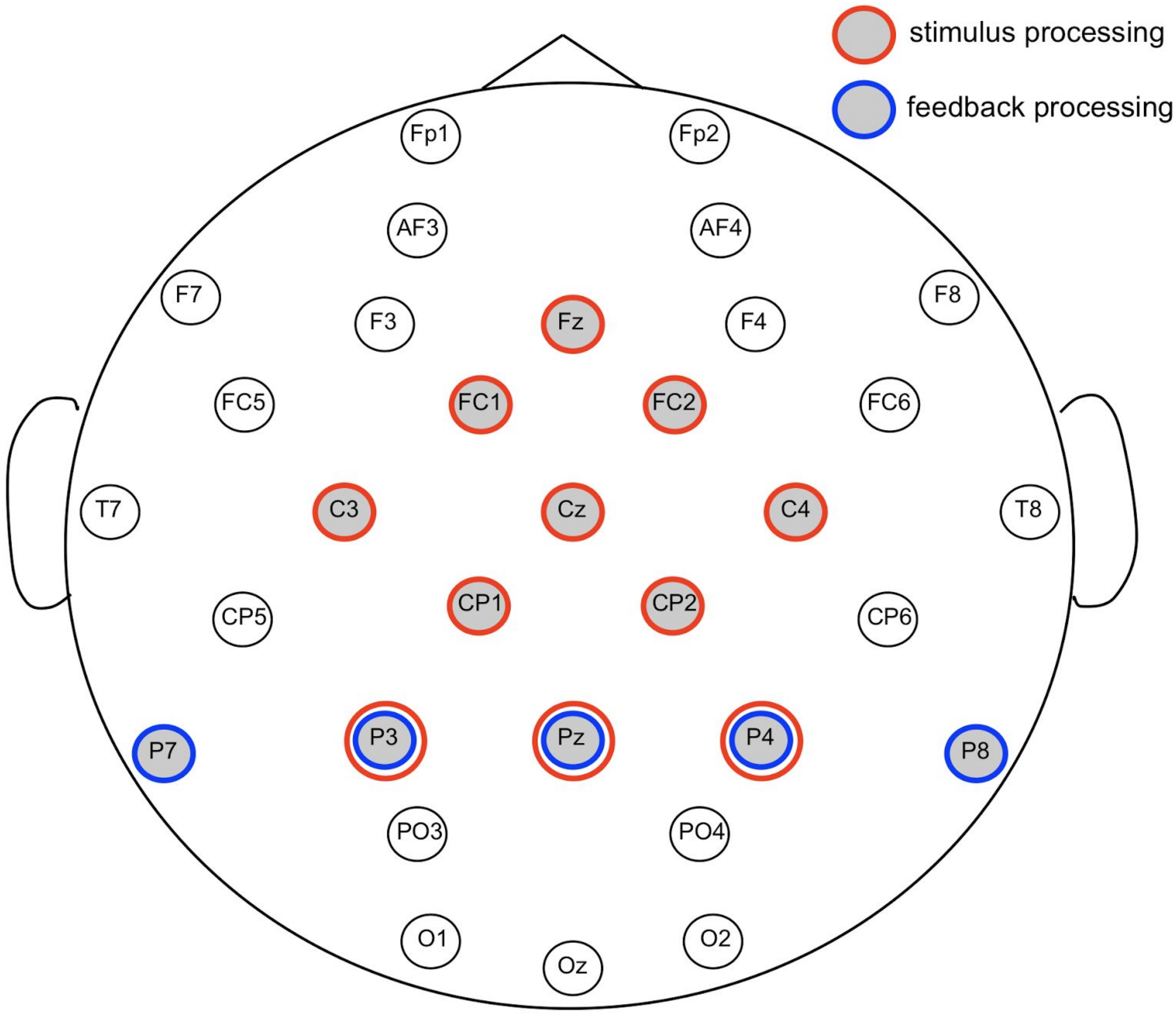
Figure 10. Topographic maps of hard/easy subtraction (correct responses only) of mean amplitude between 500-800 ms for stimulus-locked ERPs.

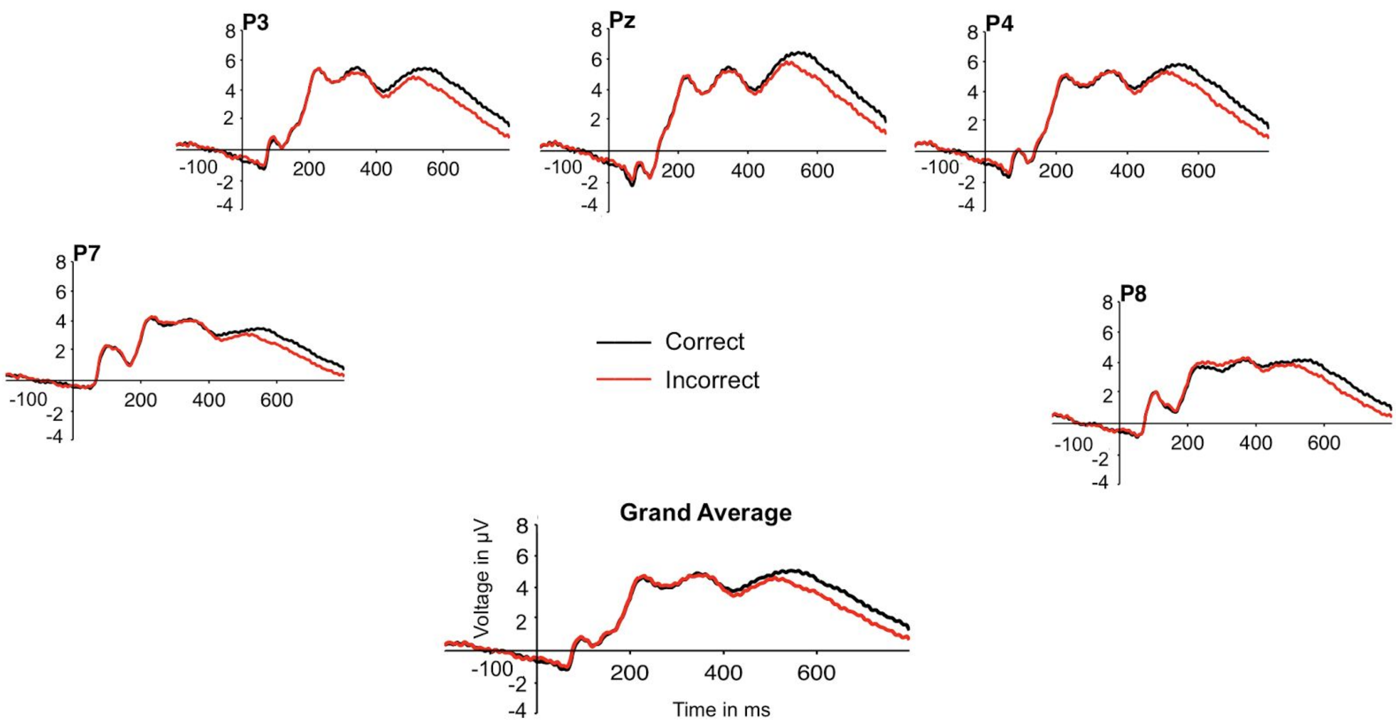

Conjunctive Rule
A3
B
A1
A2
Orientation
Frequency

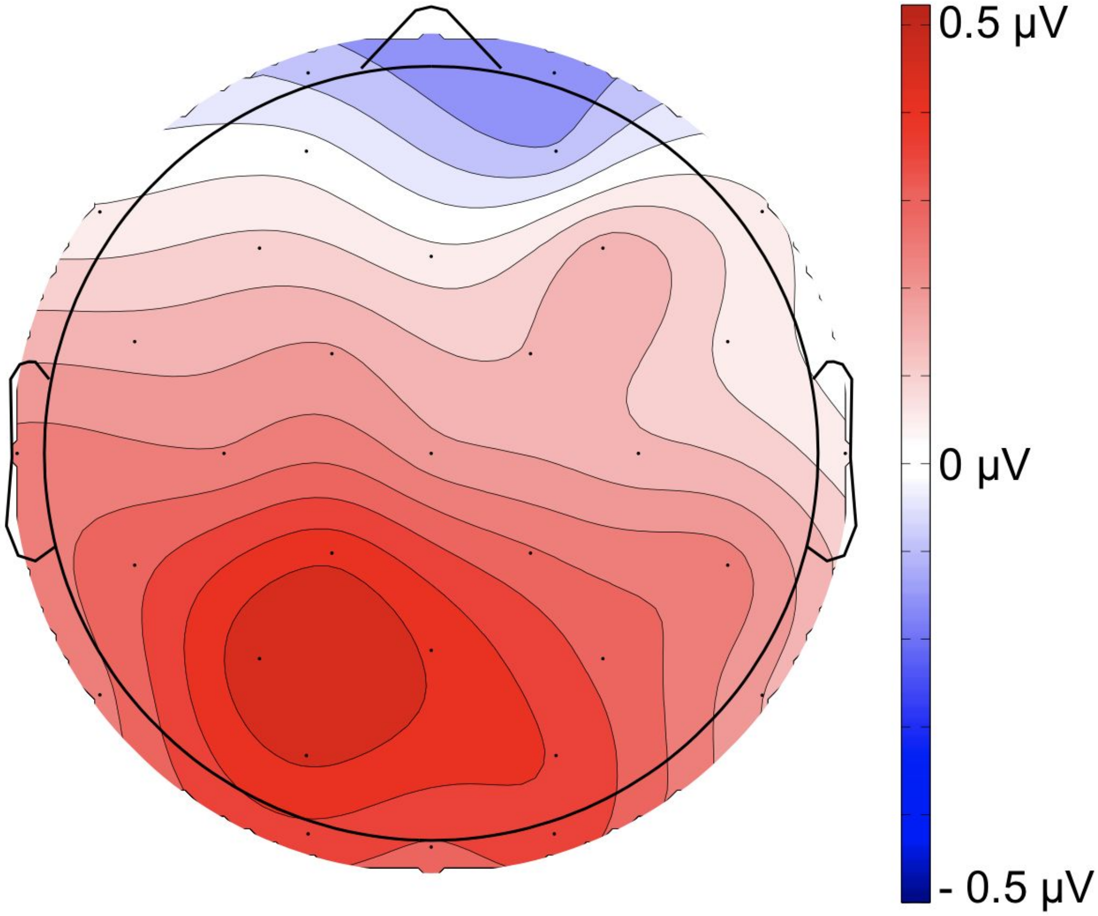

Fixation (750 ms)
A B
Stimulus Presentation (2500 ms)
Feedback (750 ms)
Interstimulus Interval (750 ms)

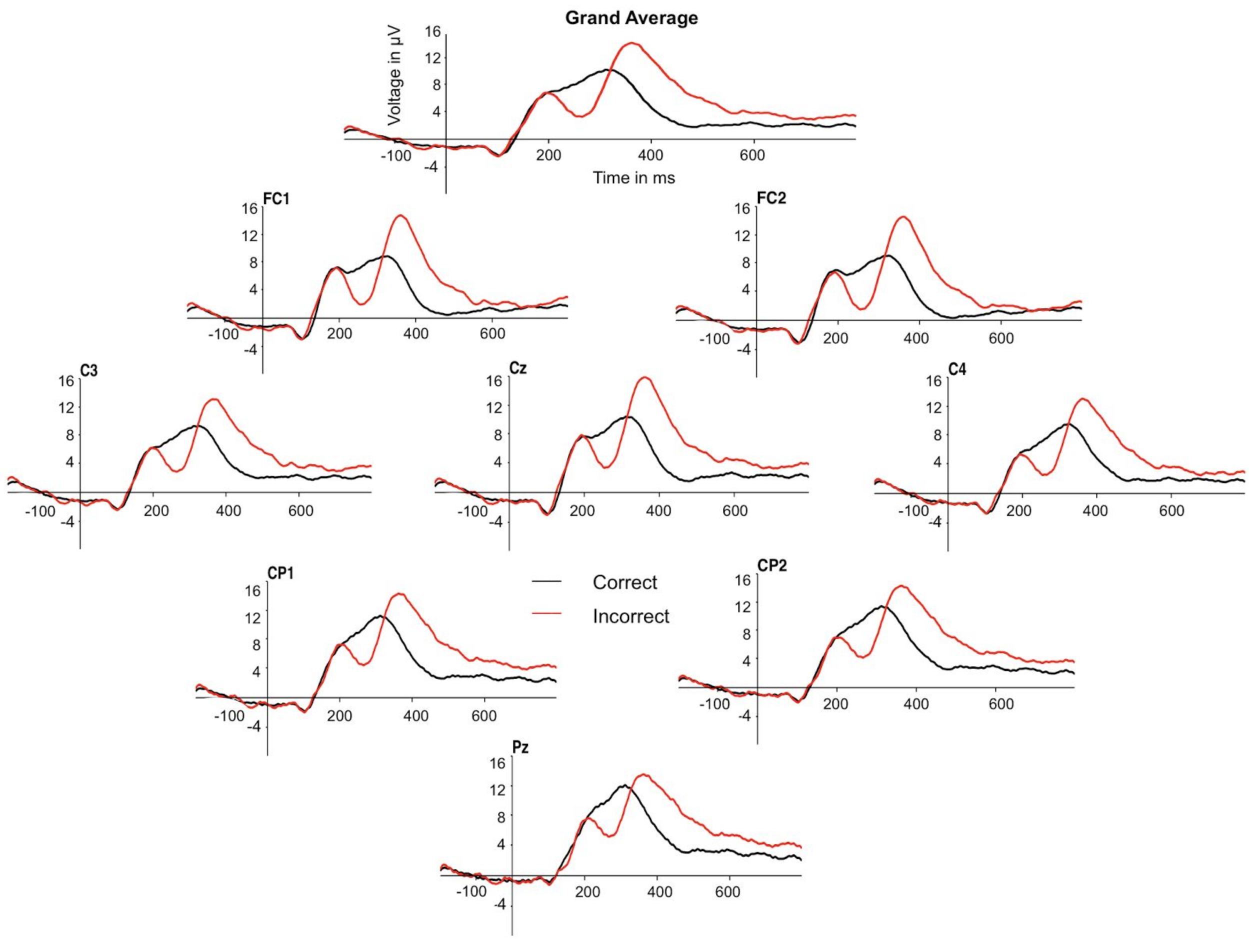

1
0.9
0.8
0.7
0.6
0.5
Proportion Correct
All participants
Strong learners
1
2
3
4
5
6
Block

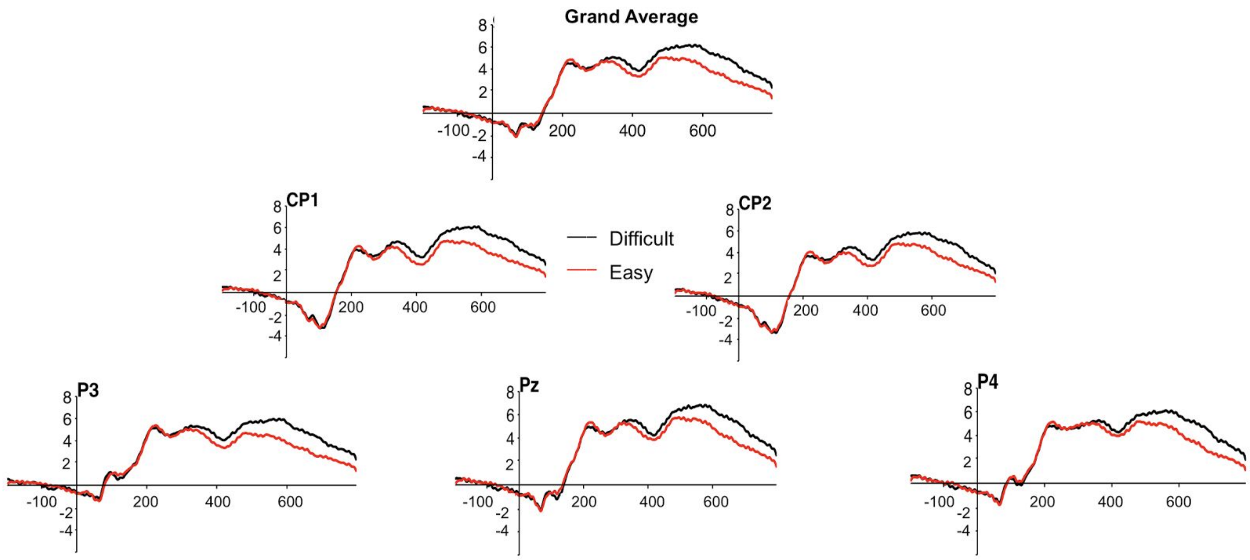

Block 1 : Orientation
Block 2 : Orientation
Block 3 : Conjunctive Rule
Block 4 : Conjunctive Rule
Block 5 : Conjunctive Rule
Block 6 : Conjunctive Rule
Orientation
Frequency
0
50
100
150
200
250
300
350
400

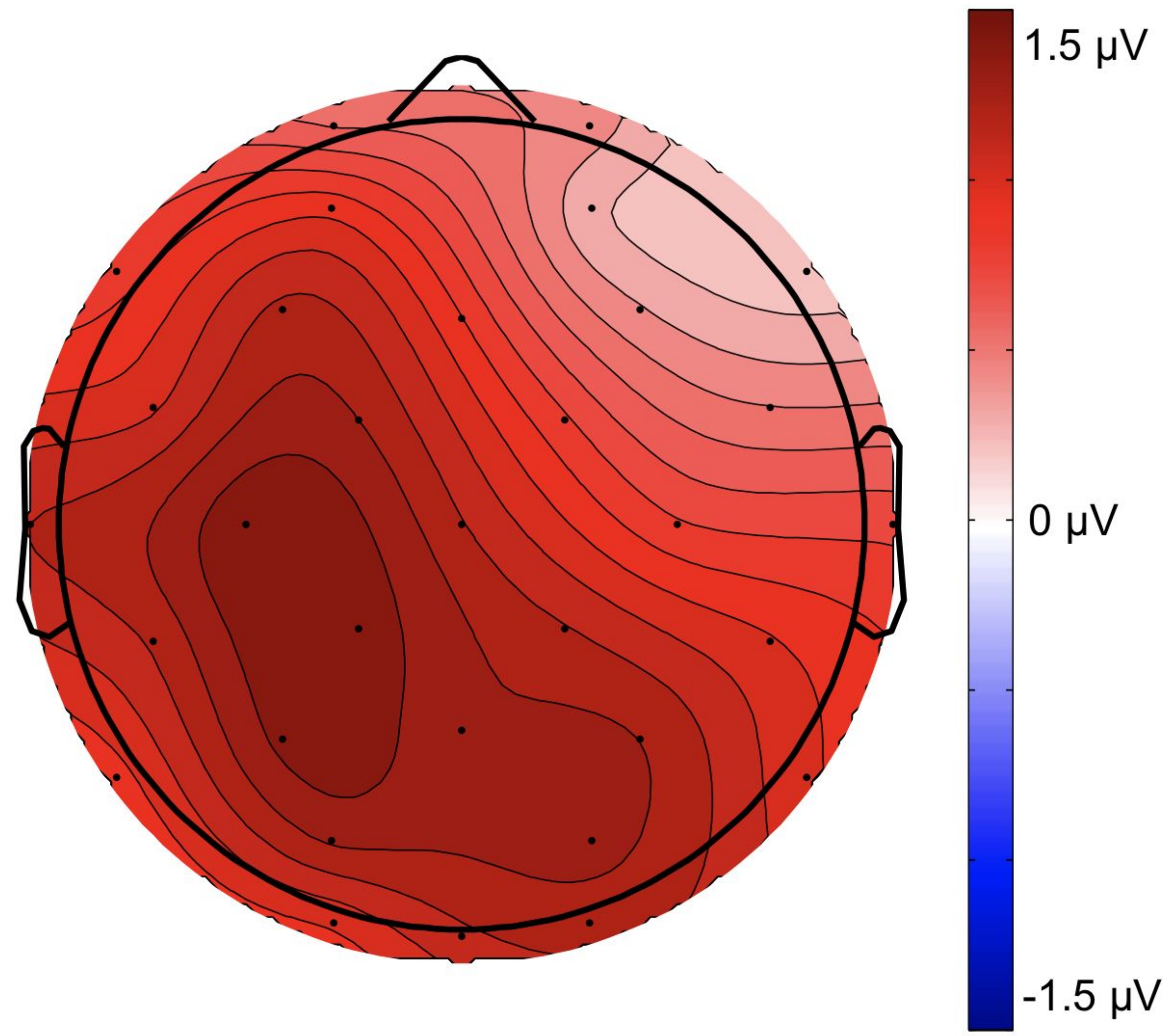

stimulus processing
feedback processing
Fp1
Fp2
AF3
AF4
F7
F3
Fz
F4
F8
FC5
FC1
FC2
FC6
T7
C3
Cz
C4
T8
CP5
CP1
CP2
CP6
P7
P3
Pz
P4
P8
PO3
PO4
O1
Oz
O2


P3
Pz
P4
P7
P8
Correct
Incorrect
Grand Average
Voltage in µV
Time in ms


0.5 µV
0 µV
- 0.5 µV


Grand Average
Voltage in µV
Time in ms
FC1
FC2
C3
Cz
C4
CP1
CP2
Pz
Correct
Incorrect


Grand Average
CP1
CP2
P3
Pz
P4
Difficult
Easy
8
6
4
2
-2
-4
-100
200
400
600


1.5 μV
0 μV
-1.5 μV