Midfrontal Theta and Cognitive Effort: Real World Applications in Medical Decision-Making

by

Jordan Middleton
Bachelor of Science, University of Victoria, 2017

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of

MASTER OF SCIENCE

in the School of Exercise Science, Physical & Health Education
Faculty of Graduate Studies

© Jordan Middleton, 2019
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.
Midfrontal Theta and Cognitive Effort: Real World Applications in Medical Decision-Making

by

Jordan Middleton
Bachelor of Science, University of Victoria, 2017

Supervisory Committee

Dr. Olave E. Krigolson, Supervisor
School of Exercise Science, Physical, & Health Education
Faculty of Education

Dr. Bruce Wright, Outside Member
Division of Medical Sciences
University of Victoria
Faculty of Medicine
University of British Columbia
ABSTRACT

Medical choices can be life or death, and thus improving the accuracy of diagnostic decisions within a time constrained environment has a large potential for positive change. To that end, an adaptation of Dual Process Theory was developed to create a theoretical framework for medical decision making. In order to effectively measure this framework, a possible electroencephalographical link was investigated. During a complex medical diagnostic task, 52 participants were asked to diagnose what liver condition simulated patients had based on procedurally generated biometric data. Feedback was provided during a learning phase until the pattern was learned. During the experimental phase, possible ranges for the biometric data were extended, allowing for increased diagnostic difficulty in some trials, thereby producing conflict for the participants. This difference between the control (Type 1) trials and the high conflict (Type 2) trials was measured using electroencephalography. It was predicted that an elevation in midfrontal theta power would be observed in high-conflict trials, which would provide a neurological correlate for Type 2 processing. This hypothesis was not verified, although several modifications to the experimental design were provided to inform future investigations. It is likely that an improved paradigm would be able to distinguish between the two processes, providing vital neurofeedback that could inform future medical students and emphasize effective learning to improve diagnostic outcomes.
TABLE OF CONTENTS

SUPERVISORY COMMITTEE ........................................................................................................ ii
ABSTRACT .................................................................................................................................. iii
TABLE OF CONTENTS ............................................................................................................ iv
LIST OF TABLES ....................................................................................................................... vi
LIST OF FIGURES .................................................................................................................... vii
ACKNOWLEDGEMENTS ...................................................................................................... viii

CHAPTER 1: LITERATURE REVIEW ................................................................................... 1

1.1 Introduction ........................................................................................................................... 1
1.2 Dual Process Theory ............................................................................................................. 1
  1.2.1 Default Interventionist Model. ....................................................................................... 4
  1.2.2 Parallel-Competitive Model. ......................................................................................... 4
  1.2.3 Emulation-Based Framework for Cognition. ................................................................. 4
  1.2.4 Unified Theory of Judgement. ....................................................................................... 5
1.3 Medical Decision-Making ..................................................................................................... 5
1.4 Behavioural Measures on Decision-making ......................................................................... 8
  1.4.1 Speed-Accuracy Trade-off. ............................................................................................ 8
  1.4.2 Self-Reported Confidence. ............................................................................................. 9
1.5 Neuroimaging of Decision-making ..................................................................................... 10
  1.5.1 Time Domain. .............................................................................................................. 12
  1.5.2 Frequency Domain. ..................................................................................................... 14
1.6 Analytical Techniques ......................................................................................................... 18
  1.6.1 Fast-Fourier Transform. ............................................................................................. 18
  1.6.2 Wavelet Transforms. .................................................................................................. 19
  1.6.3 Principle Component Analysis. ................................................................................... 19
1.7 SUMMARY ............................................................................................................................ 20

CHAPTER 2: EXPERIMENT 1 ................................................................................................ 21

  2.1 Introduction ....................................................................................................................... 21
2.2 Method ................................................................................................................................22
  2.2.1 Participants. ..................................................................................................................22
  2.2.2 Apparatus and Procedure ...........................................................................................22
  2.2.3 Data Acquisition and Processing ...............................................................................25

2.3 Results ................................................................................................................................27
  2.3.1 Behavioural Data. .......................................................................................................27
  2.3.2 Reward Positivity ........................................................................................................28
  2.3.3 Fast Fourier Transform ..............................................................................................28
  2.3.4 Wavelet .......................................................................................................................35
  2.3.5 Principal Component Analysis ...................................................................................36

2.4 Discussion ...........................................................................................................................37

CHAPTER 3: DISCUSSION ......................................................................................................41

3.1 Findings ................................................................................................................................41
  3.1.1 Behavioural ..................................................................................................................41
  3.1.2 Time Domain ..............................................................................................................41
  3.1.3 Frequency Domain .....................................................................................................42
  3.1.4 Summary .....................................................................................................................43

3.2 Implications ..........................................................................................................................44

3.3 Limitations and Future Directions ......................................................................................45
  3.3.1 Methodological Considerations .................................................................................45
  3.3.2 Future Research Directions .......................................................................................47
  3.3.3 Future Applications ....................................................................................................49

3.4 Conclusion ..........................................................................................................................50

REFERENCES ..............................................................................................................................52
**LIST OF TABLES**

**Table 1.1.** System 1 & 2 characteristics .................................................................2

**Table 1.2.** Comparison of EEG and fMRI as neuroimaging tools .........................10

**Table 2.1.** Summary of behavioural data ...............................................................28

**Table 2.2.** Summary of repeated measures ANOVA for stimulus .........................30

**Table 2.3.** Post-hoc $t$-test results between bands during stimulus .....................30

**Table 2.4.** Post-hoc $t$-test results for channel-band interaction during stimulus ......31

**Table 2.5.** Post-hoc $t$-test results for condition-band interaction during stimulus ....31

**Table 2.6.** Post-hoc $t$-test results for condition-channel-band interaction during stimulus .....32

**Table 2.7.** Summary of repeated measures ANOVA for response .........................33

**Table 2.8.** Post-hoc $t$-test results between bands during response ....................33

**Table 2.9.** Post-hoc $t$-test results for channel-band interaction during response ......34

**Table 2.10.** Post-hoc $t$-test results for condition-band interaction during response ....34
LIST OF FIGURES

Figure 1.1. Process Model of diagnostic reasoning ................................................................. 6
Figure 1.2. Sample of Stroop Task .......................................................................................... 8
Figure 1.3. Standard Reward Positivity signal .................................................................... 12
Figure 1.4. Standard P300 signal ......................................................................................... 13
Figure 1.5. Representation of neural synchronization and frequency signals ...................... 14
Figure 1.6. Representation of Time and Frequency domains .................................................. 15
Figure 2.1. Experimental paradigm ..................................................................................... 23
Figure 2.2. Summary plot of behavioural data .................................................................... 27
Figure 2.3. Reward Positivity comparison for learners and non-learners .............................. 29
Figure 2.4. FFT results at electrodes $F_z$ and $P_z$ for stimulus and response .................... 29
Figure 2.5. Wavelet results at electrodes $F_z$ and $P_z$ for stimulus and response ............... 35
Figure 2.6. PCA results at electrodes $F_z$ and $P_z$ for stimulus and response ..................... 36
Figure 3.1. Example images of portable EEG devices ......................................................... 50
ACKNOWLEDGEMENTS

First, I would like to thank Olav Krigolson for guiding me through my project, giving me the opportunity to excel, and for the occasional nudge to keep me on track. Second, I would like to thank Bruce Wright for his insight into the medical field and for his continuing support of the pursuit of knowledge. Third, I would like to thank Chad Williams for laying the foundation for my thesis, for his guidance, and his support throughout. Fourth, I would like to thank the members of the Krigolson Laboratory; each and every one of them acted as a sounding board at some point during this thesis, and their feedback was invaluable. Last, but certainly not least, I would like to thank my partner Greg Marlowe, who kept me grounded, focused, and helped me stay afloat during those times when I thought I was drowning. Without the support of these people, this project would have been a shadow of what it is, so again, thank you all.
CHAPTER 1: LITERATURE REVIEW

1.1 Introduction

Medical decision-making (DM) is a complex process, requiring a high level of expertise, experience, and support. Despite extensive training and systemic support, it is still estimated that practitioner diagnostic error is as high as 12% (Grabber, 2005); after all, doctors are still human. Although many of these errors do not lead to death, and some have no effect on outcome at all, many lead to increased morbidity, which are poorly represented in many statistical analyses. Surprisingly, the degree of training does not appear to be a factor. For instance, when both residents and attendings were asked to observe physical examinations of patients admitted to the emergency department, predicted diagnoses for both groups was correct in 52.2% and 53.9%, respectively (Cabrera et al., 2015).

Expecting outcomes to improve simply by raising awareness of these issues is unlikely and without a clear strategy for intervention, the status quo will probably be maintained. For current doctors, tools have been introduced to provide automated oversight, such as computerised diagnostic decision support systems, though their impact on diagnostic errors has been mixed (Nurek, Kostopoulou, Delaney, & Esmail, 2015). This is only one area to target; the way medical knowledge is imparted to students is another, and possibly far more effective opportunity for greater outcomes. If the approach to medical training can be targeted to work with cognitive processes as we know them, learning can be optimized for accuracy and hopefully performance after graduation. To do this, we must understand how clinicians think, how those processes can be measured, and what interventions will be effective.

1.2 Dual Process Theory

There have been many proposed theories on cognition; from the days of Aristotle, who first proposed dualism, to Freud, who introduced the ideas of *id*, *ego*, and *superego*. As our
understanding of cognition evolved, so too did our conceptual models for defining it. One prominent current model is known as Dual Process Theory (DPT), and it was first introduced in 1975 (Posner & Snyder). Over time, the model was developed into what we use today (Stanovich & West, 2000), and the characteristics of each system have been clearly, but broadly, defined (Table 1.1). The key differences are that System 1 is fast, easy, inaccurate, and implicit, while System 2 is slow, effortful, accurate, and explicit. For example, solving for the equation $2+2=4$ would be a System 1 process, whereas $x=\sqrt{169}$ would be a System 2 process. To clarify, it is believed that to fully qualify as a System 1 process, the neurological representation of the task must be linked to its solution, such that $2+2=4$, rather than quickly solving $2+2=x$ (Thompson et al., 2013). Conversely, System 2 processes approach a solution through prediction and

<table>
<thead>
<tr>
<th>Cognitive Style</th>
<th>System 1</th>
<th>System 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computational Style</td>
<td>Associative</td>
<td>Rule-Based</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Passive</td>
<td>Active</td>
</tr>
<tr>
<td>Capacity</td>
<td>High</td>
<td>Limited</td>
</tr>
<tr>
<td>Cognitive Awareness/Control</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Automaticity</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Rate</td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>Reliability</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Errors</td>
<td>Common</td>
<td>Rare</td>
</tr>
<tr>
<td>Effort</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Emotional Attachment</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Scientific Rigor</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

*Table 1.1 Summary of System 1 & 2 characteristic (Adapted from Stanovich & West, 2000).*
simulation, which is cognitively demanding, but extremely valuable in processing unknown outcomes. A key feature of System 2 is the ability to distinguish between these simulated worlds and the real world, a process called cognitive decoupling (Stanovich & Toplak, 2012), and it is essential to prevent reality and fantasy from becoming tangled. These characteristics may provide insights into how cognition works and how to improve it.

This model is not universally accepted, nor is it fully able to describe the varied degrees of human cognition. First, the lack of coherence in the models proposed by different authors, particularly in the clustering of attributes, suggests a lack of construct validity (Keren & Schul, 2009). An explanation for this inconsistency is the difference between defining and correlated features; although certain characteristics are often associated with a certain type of cognition, it is not necessarily universally tied to it (Evans & Stanovich, 2013). The second issue is that the spectrum of cognition extends to a larger scope—rather than binary categories as suggested by DPTs—and instead should be described as a continuum extending all the way to the level of rigor seen in scientific research (Cader, Campbell, & Watson, 2005). Stanovich responded by suggesting that System 2 need not be constrained to short term processing, but rather any effortful thought, even those extending for months or years, can be considered a System 2 process. While this interpretation of the model retains internal consistency and broad application, it results in reduced resolution. Another critique is that use of the term “Systems” implies wholly separate and distinct processing schema, which is why the more recent literature has adopted the terms “Type 1” and “Type 2” (Stanovich & Toplak, 2012), as they still allow for the differentiation between overall strategies without implying mutual exclusivity.

There is significant debate over when and how each type of decision-making is used. The four most prominent approaches are the Default Interventionist Model (DIM), the Parallel-
Competitive Model (PCM), Emulation-Based Framework for Cognition (EBFC), and the Unified Theory of Judgement (UTJ).

1.2.1 Default Interventionist Model. The DIM posits that Type 1 processes are the default response process of the brain, and Type 2 processes intervene when Type 1 is either insufficient or if the outcome would be undesirable (Kahneman, 2011). In order for Type 2 processes to be utilized, a degree of complexity, difficulty, motivation, and/or novelty must be present to initiate recruitment (Evans & Stanovich, 2013). While this model does package everything quite neatly, the suggestion that Type 2 processes are only called on when necessary was questioned. Based on this critique, the model was modified to state that Type 1 always computes a default response, while Type 2 processes determine whether or not to accept that response (Evans, 2011). This model requires Type 1 processes to complete prior to Type 2 assessment, which then predicts the outcome of the default as well as other possibilities, and either accepts or rejects the default. This modification blurred the distinctions between the DIM and the PCM.

1.2.2 Parallel-Competitive Model. The PCM suggests that both Type 1 and 2 processes occur simultaneously, in direct competition (Barbey & Sloman, 2007). While this does tie in many features seen in the DIM, there are subtle differences. First, before a decision can be made, both Type 1 and 2 processes must be completed, which is contrary to experimental evidence (Evans & Stanovich, 2013). Second, it suggests that Type 2 processes are always at play, which necessitates cognitive effort well beyond what is seen in studies (Evans & Stanovich, 2013).

1.2.3 Emulation-Based Framework for Cognition. This model seeks to differentiate Type 1 and 2 processes based on function. It suggests that Type 1 processes manage base-level tasks, such as immediate motor responses, short-term navigation of the environment, whereas Type 2 processes manage long-term goals and decisions (Colder, 2011). This framework is based on
research regarding forward models of motor planning (Desmurget & Grafton, 2000; Wolpert & Ghahramani, 2000). For example, Type 1 processes would manage the immediate task of navigating through a crowd, while Type 2 processes would manage ensuring you reach your intended destination. While this model does align with findings in both fields, it only accounts for the motor aspect of cognition, and fails to capture the full scope of decision-making.

1.2.4 Unified Theory of Judgement. The UTJ proposed that decision-making processes exist on a spectrum, rather than categorized into disparate bins. In particular, the distinction seen in classical dual process theories that Type 2 processes are rule based whereas Type 1 processes utilize heuristic associations is eliminated, and suggests that all decision-making is rule based (Kruglanski & Gigerenzer, 2011). Instead of categorizing processes, UTJ places approaches on a spectrum, ranging from heuristic based rules to those based on memory and hypothetical predictions. Selection of a rule is made by comparing the likely outcome of each rule and its difficulty of execution relative to available motivation and attentional capacity. Essentially, a cost/benefit analysis is determined for each available rule set, and the optimal response is selected. Should all presently available rules be deemed insufficient, additional resources are called upon to generate additional rules based on experiential memory and hypothetical models (Kruglanski & Gigerenzer, 2011). While this model does bring a novel approach to the debate, it does not align with models used to understand and explain medical decision-making (Croskerry, 2009b).

1.3 Medical Decision-Making

While DM in medicine is still DM, the unique demands, acuity, and stakes differentiate it from other forms of DM (Croskerry, 2014). As such, medical DM must be studied as its own field based on the findings of the more general field, then refined to simulate its own unique demands. This idea is often seen in medical simulation labs at medical schools across the country, using
reactive software that responds to student actions, as well as trained simulated patients that emulate real patient conditions in a clinical setting (UCalgary, 2018). While the DPT model proposed by Stanovich & West covers the subject in a general way, the field of medicine has unique challenges that requires a specific model. Such a model was proposed by (Figure 1.1), and effectively reduced the process to pattern recognition Croskerry (2009). In other words, recognized patterns would be addressed as previously successful, while unrecognized patterns would require additional cognitive effort to resolve. It also gave a mechanism by which explicit process could override the automatic process in the event of a changing external environment. One of the key components of the model is the training of Type 1 processes. In order for future Type 1 processes to be effective, repetitive and accurately executed Type 2 processes are required first. That is, without sufficiently accurate repetition, the conclusions at which Type 1 processes can arrive may be flawed, and potentially dangerously so. There are no shortcuts; slow, methodical exposure and repetition are

![Figure 1.1. The process model of diagnostic reasoning (Adapted from Croskerry, 2009a).](image-url)
necessary prerequisites to develop an accurate framework for Type 1 processes. To expand on the previously mentioned study by Cabrera et al. (2015), the predictions made were presumed to be primarily Type 1 processes and suffered from the speed/accuracy trade-off. After all, the fast-paced nature of the emergency department necessitates similarly fast decisions. In general, the observed clinicians had high specificity in all metrics, but were often quite poor in sensitivity, particularly when predicting ICU admission (29% for residents and 43% for attendings). Considering the nature of the ED and its propensity for reliance on fast-paced, type 1 decisions, the findings of this study are particularly concerning, given the risk to patient outcomes.

The implication of these findings is that insufficient Type 1 training is responsible for a significant proportion of Type 1 failures, and thereby diagnostic errors. While this may be true in some cases, it is not the whole picture. The topic has been debated by many, but the full scope of the literature is surprisingly scarce for such an important field. Norman et al. (2017) discussed the field extensively in a literature review and made two major assertions. First, if errors are based on Type 1 failures, then they are failures of trained heuristics, and this would be evidenced by vulnerability to cognitive biases like recency or premature conclusions. Second, if the failures are in Type 2 processes, then it would be due to insufficient reflection. Their conclusion was that neither explanation wholly describes the literature.

In Norman and colleague’s review, they stated that improved opportunity for review and reflection appeared to be the most effective way to improve error rates, but this relies on practitioners recognizing that they are in a “cognitive minefield”, as Kahneman put it. Unfortunately, evidence suggest that this is unreliable at best (Monteiro et al., 2015). As such, an objective tool to monitor cognitive strategies must be employed to provide feedback physicians are unable to effectively provide themselves.
This creates an intervention opportunity: by ensuring repetition is accurate and explicit, effective training of internal pattern recognition can be promoted. The next question is: how can it be determined which DM strategy is being used?

### 1.4 Behavioural Measures on Decision-making

The three primary behavioural areas being studied are reaction time, trial accuracy, and self-reported confidence (SRC). The two former categories are closely related, and better combined as the speed-accuracy trade-off (SAT).

#### 1.4.1 Speed-Accuracy Trade-off

The clearest and simplest task to illustrate the SAT is called the Stroop task (Stroop, 1935). In this task, there are a series of colours written in matching (congruent) or not matching (incongruent) ink (Figure 1.2; Pompon, McNeil, Spencer, & Kendall, 2015). The task is to name the ink colour, regardless of the text. Since reading is an automated process that is difficult to override, saying the ink colour instead of the word is challenging, and increases response time by approximately 74% (Macleod, 1991). In this task, accuracy remains relatively close to perfect in both conditions, but since incongruent trials are more difficult, reaction time increases to maintain high accuracy. Were time to be restricted, accuracy would fall instead, hence the trade-off nature of these two goals.

Additionally, there are variations of the Stroop task used to elicit a similar effect, such as number size incongruency. In this version, the participant is asked to select which one of two numbers has a larger numerical value when the numbers vary in physical size (Robertson, Hiebert, Seergobin, Owen, & MacDonald, 2015). This allows for more variability, as

---

**Figure 1.2.** A sample Stroop task. The top row is the congruent control, where the ink colour matches the word. The middle and bottom row are incongruent, where the ink colour and word do not match (Pompon et al., 2015).
size is a more continuous dimension than simple colour categories.

Complexity can be manipulated by changing the degree of difference between two variables, as discussed, and by altering the number of possible options. This effect is called Hick’s Law, and it states that the time to make a decision increases as a function of the number of available choices (Hick, 1952). The relationship is remarkably simple and can be approximated by the formula $T=b \cdot \log_2(n+1)$, where $T$ is the time to decide, $b$ is the base response time, and $n$ is the number of possible choices.

There is some debate around the neural basis of this effect. In particular, it remains unclear as to whether or not this is a direct trade, or simply the result of differing priorities ( Förster, Higgins, & Bianco, 2003). In other words, is it a zero-sum scenario as the SAT hypothesis suggests, or is it simply a strategic choice made to sacrifice one component if the other is considered more important? While the direct mechanisms behind the SAT remain elusive, the results themselves are robust, with over 80 years of research to back them ( Förster et al., 2003; Heitz & Schall, 2012; Hick, 1952; Macleod, 1991; Stroop, 1935).

1.4.2 Self-Reported Confidence. Self-assessment is always subjective, which makes the validity of any self-measure suspect. In particular, SRC is particularly biased, and has little to no correlation with actual ability (Liaw, Scherpber, Rethans, & Klainin-Yobas, 2012). It is thought that this effect is caused by global ability influencing our assessment of task-specific ability. That is, if you tend to do well at tasks in general, you are likely to rank yourself high on most tasks regardless of actual performance. This is especially true in medical students, as there is no correlation between SRC and clinical ability, or even medical knowledge (Barnsley et al., 2004).

Although the validity of SRC is questionable when applied to predicting ability, there is a correlation between SRC and certain strategies to complete a task. In particular, high SRC
assessments are correlated to higher error rates, suggesting that this high self-scoring is actually overconfidence (Vancouver, Thompson, Tischner, & Putka, 2002). However, this association is not well studied, making this assertion preliminary at best.

1.5 Neuroimaging of Decision-making

The two main forms of neuroimaging used to study DM are functional magnetic resonance imaging (fMRI) and electroencephalography (EEG). Each technique has its own strengths and weaknesses, which have been outlined in Table 1.2.

Extensive research has been done on DM using fMRI and even though there are some linguistic inconsistencies in the literature, the findings are congruent. Certain terms like “intuitive” and “perceptive” are used to describe processes designed to elicit Type 1 processes, while Type 2 processes are described by terms such as “effortful” or “deliberate”. These differences remain purely in word choice; at their heart, the findings closely relate to distinctions outlined by DPT.

It has been suggested that the difference between Type 1 and 2 processes is simply the degree of cognitive effort exerted to complete a task. In spatiomotor tasks, there is increased activity in

<table>
<thead>
<tr>
<th></th>
<th>fMRI</th>
<th>EEG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Startup Cost</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Operating Cost</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Portability</td>
<td>Immobile</td>
<td>Variable</td>
</tr>
<tr>
<td>Size</td>
<td>Large</td>
<td>Small</td>
</tr>
<tr>
<td>Spatial Resolution</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Temporal Resolution</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>
the right posterior lateral prefrontal cortex and bilateral presupplementary motor areas (preSMA; Ivanoff, Branning, & Marois, 2008). This suggests that the preSMA modulates the degree of sensory information needed to appropriately respond, while the posterior lateral prefrontal cortex plays a role in response inhibition.

Associative DM tasks elicit responses in entirely different areas. The orbitofrontal cortex (OFC) has been suggested to play a role in the early integration of incomplete information (Bar et al., 2006; Zander, Horr, Bolte, & Volz, 2016). It has also been suggested that the anterior insular cortex has a role in producing a subjective “gut feeling” (Hedne, Norman, & Metcalfe, 2016; Zander et al., 2016). But this finding may be misleading because the anterior insular cortex has direct connections to the anterior cingulate cortex (ACC), and thereby high activation in most tasks. Therefore, its activation in associative tasks may be merely coincidental (Chang, Yarkoni, Khaw, & Sanfey, 2013).

There are also clear differences between the neural structures used by novices and experts to solve a problem. In a medical DM task, residents and attendings were asked to make a series of diagnoses based on a brief case summary card, which contained patient history, presentation, and select lab test results (Hruska et al., 2016). Cases were classified as either easy or hard. While there were no significant differences between novices and experts in the easy condition—a Type 1 process—there were several key differences for hard trials. Novices showed greater activation in the ventrolateral prefrontal cortex, which is associated with accessing semantic information, whereas experts had elevated activity in the dorsolateral prefrontal cortex, which is associated with comparisons to a normative standard. In addition, the novices showed more activity in the left hemisphere, which suggest greater reliance on concrete information, while the experts showed more activity on the right, indicating more abstract, pattern-based processing. The overall finding
is that novices rely on concrete, semantic information to decide, while experts use patterns gleaned from experience, coupled with an understanding of context and scope.

The primary forms of data analysis for EEG originate in both the time and frequency domains. For the time domain, time-locked event-related potentials are the primary component of interest. For the frequency domain, both fast-Fourier and wavelet transforms will be used.

1.5.1 Time Domain. An event-related potential (ERP) is any continuous recording of EEG data time locked to the moment of an event. Typically, data are segmented to a window including, not bordered by, the zero line, and many trials are aligned for averaging. The final average waveform is an amplification of the real effect.

The major benefit of averaging ERP data is to remove the random noise. Live EEG recordings are contaminated by many additional factors that are not relevant to the actual event: muscle contractions, blinks, incidental neural activity, or recording artifacts. If the voltage values of all these influences is random, then averaging many trials should drive their influence to zero, unmasking the real effect.

The scale of random fluctuations, compared to the actual effect, is called the signal to noise ratio. The validity of the above depends on the noise being random and unrelated to the event. It is extremely easy to violate this assumption. As such, diligence in experimental design and execution must be taken to avoid introducing regular artifacts that may influence the final waveform. Some examples of such interventions include minimizing signal noise during electrode application by using conductive gel, reminding the participant to minimise eye and body movement during the experiment, and using carefully selected post-processing parameters to remove residual noise.
When using EEG, there are many signals that have been well defined and carry meaning to them. Of these signals, the reward positivity (RewP) and P300 have the most direct relevance to the study.

1.5.1.1 Reward Positivity. The RewP is a positive deflection seen in the difference wave when “Win” conditions are compared to “Loss” conditions in tasks with an uncertain outcome (Figure 1.3). The peak of the difference wave occurs about 250ms after feedback is delivered (Holroyd & Coles, 2002). Over time, the amplitude of the difference wave decreases, which reflects the decreasing novelty of correct responses.

The signal itself is believed to originate in the anterior cingulate cortex (ACC), which has been verified with fMRI (Philiastides & Sajda, 2007). It functions as a form of reward signal to refine behaviour based on expected outcome. When an outcome is uncertain, positive feedback stimulates the dopamine reward pathway, reinforcing the behaviour that caused the outcome. The response becomes smaller over time as the outcome becomes more certain; after all, one does not need to reinforce behaviours with a known outcome (Holroyd & Coles, 2002). As such, the relative amplitude of the difference wave can be used as evidence of learning effects.

1.5.1.2 P300. The P300, or P3 is very functionally named: it is a positive deflection with a mean peak at approximately 300ms after stimulus. The

*Figure 1.3*. A standard reward positivity signal, with win and loss conditions compared (Krigolson, Williams, Norton, Hassall, & Colino, 2017).
signal is most easily elicited using a simple oddball task, in which one of two coloured circles is
presented to a participant in rapid succession, with one colour being much more common than the
other. The less common colour, called the oddball, leads to the P300 signal when compared to the
other colour, or the control.

There are two main measures of P300: the amplitude and the latency. The amplitude is the
maximum deflection from pre-stimulus baseline within a standard window of 200–500ms, and the
latency is the time between the stimulus and that peak. While there is no unified understanding by
what mechanism these factors are altered, their association is clear: increased P300 amplitude
correlates with increased attentional engagement (Curran & Cleary, 2003), while latency is
inversely correlated with general cognitive performance (Pelosi et al., 1992) and age (Polich,
1996). The P300 is composed of two main sub-components: P3a and P3b (Figure 1.4). P3a is the
initial, high amplitude deflection at the beginning of the signal.

It is thought to indicate attentional allocation for novel stimuli
(Spencer & Polich, 1999), and has a strong signal between Cz
and Fz. The P3b is a longer duration, lower amplitude signal
with diffuse strength over the posterior scalp. It is thought to be
involved in context updating and short-term memory storage
(Polich, 2007).

1.5.2 Frequency Domain. The time and frequency
domain are interchangeable. That is, data in either form can be
converted to the other without loss; it is merely another way of
representing data (Figure 1.5). This conversion is commonly
done using a tool called a fast Fourier transform, which will be
discussed later. The frequency domain is divided up into several ranges called “bands”, and each band is associated with certain types of activity, in conjunction with a location. There are 5 commonly defined frequency bands: delta (1–4 Hz), theta (4–6 Hz), alpha (8–1 Hz), beta (14–30 Hz), and gamma (30+ Hz). We will be focusing on the theta and alpha bands.

1.5.2.1 Physiological Correlates. While detection of these signals provides a valuable tool for studying patterns of activity in the brain, it is important to understand what these signals represent physiologically. On a cellular level, all neural activity is simply action potentials traveling down axons towards synapses. Like a computer, neurons are either firing or they are not; there isn’t really a sinusoidal cycle of excitation and depolarization, although their activation can be rhythmic. Instead, what is observed as a rhythmic pattern in EEG is more representative of synchronicity of neuron activation (Jensen, Kaiser, & Lachaux, 2007). The more aligned the neurons are with their activity, the low the frequency band—and higher amplitude—observed in the EEG recording, as illustrated in Figure 1.6. This implies that lower frequency bands like delta indicate diffuse, but highly coordinated activity, which is expected for distributed processing (Başar, Başar-Eroğlu, Karakaş, & Schürmann, 2000). Higher frequencies like gamma indicate an information-dense signal, which is seen during attention switching and other processes that require a significant redirecting of neural resources (Başar et al., 2000).
In some cases, this synchronization is seen as an isolated burst of activity, or a spike. This is called a tonic event or signal, and indicates a singular alignment in a certain region, such as the P300—which is primarily a theta band signal—responding to an oddball stimulus (Polich, 2007). Alternatively, a consistent elevation in a certain band can occur when sustained coordination is required, and this is called a phasic signal, this type of activity is observed when a sustained behavioural change—such as switching from a default behaviour to an alternative—occurs (Hassall, Holland, & Krigolson, 2013).

1.5.2.2 Theta Waves. The theta frequency band is defined as ranging from 4–8 Hz. It is associated with several high level cognitive processes, namely novelty, conflict, punishment, and error processing (Cavanagh, Zambrano-Vazquez, & Allen, 2012). These relationships are strongly supported in the literature, though the degree of association varies. The relationship is the strongest in correct/incorrect feedback for reinforcement learning paradigms, as well as easy/hard stimuli for response conflict studies, though the effect size is smaller in the latter condition.

Based on the widespread influence of theta oscillations in all parts of the brain, it has been proposed that theta acts as a substrate by which other signals can be carried. Some have even gone so far as to dub theta a kind of lingua franca, or common

![Image](image-url)

Figure 1.5. Illustration of how neural synchronicity produces oscillatory signals recorded by EEG, in spite of neurons only carrying binary information (Izhikevich, 2003).
language, that links many brain regions (Cavanagh et al., 2012). In any case, the ubiquity of theta makes it very clear that it plays a major role in the connectivity and coordination of distributed processing in the brain. The functional role of theta is thereby dictated by the region it is observed; if elevated theta is seen in certain structures and contexts, it stands to reason that those areas are more heavily relied upon for those specific tasks.

The exact source of midfrontal theta (MFθ) is moderately debated, though it is most commonly demonstrated to originate somewhere in the mid-cingulate cortex, as indicated by EEG source estimation (Cohen & Ranganath, 2007), magnetoencephalography (Doñamayor, Marco-Pallarés, Heldmann, Schoenfeld, & Münte, 2011) and fMRI (Becker, Nitsch, Miltner, & Straube, 2014) studies. This region is closely associated with the ACC and posterior cingulate cortex, which explains the relationship between MFθ, reward learning, conflict, and attention. Additionally, the ACC is highly interconnected with many brain regions, forming a kind of processing hub (Cohen, 2011). It stands to reason that a region with such high connectivity would play a significant role in the regulation of distributed cognitive processes.

The most important association of theta waves is cognitive control, particularly at midfrontal locations. The relationship is suggested by the Adaptive Control Hypothesis (TACH). This hypothesis suggests that the same systems that give rise to anxiety are used for cognitive control, to address the same problems (Grupe & Nitschke, 2013). The reasoning is that cognitive control is utilized when the situation is too complicated or dangerous for habitual, automated responses to manage. It has been suggested that this is the primary role of the mid-cingulate cortex, a region closely associated with MFθ and cognitive control (Cavanagh & Shackman, 2015).

Alongside its regulatory role, theta oscillations have been found to facilitate the transmission of information dense, high-frequency, low power gamma activity (Canolty et al., 2006). The ability
for theta to facilitate the transmission of higher-frequency signals is critical for learning and action selection, as it synchronizes activity across multiple neural regions (Benchenane et al., 2010). As such, theta band activity can be seen as a way to focus cognitive resources under high demand, especially in conflict processing (Kok, Rahnev, Jehee, Lau, & de Lange, 2012).

1.5.2.3 Alpha Waves. The frequency range associated with alpha waves is 8–15 Hz. The presence of alpha waves is generally associated with reduced attention (Klimesch, 2012). Because of this association, the amplitude of alpha oscillations can be used as a direct measure of attention in many tasks. What is not clear is whether alpha waves are the cause or result of reduced attention (Mathewson et al., 2011). That is, are alpha waves the result of decreased attention in a certain region, or are alpha waves used as a mechanism to direct attention selectively? In either case, the association is strong, though understanding the mechanism of their formation may prove valuable in establishing the underlying systems governing resource management in the brain.

There is also a growing body of literature suggesting that it is more than just amplitude that matters; alpha phase is also an important characteristic when it comes to modulating attention. There is always some basal level of alpha present, even during heightened attention. It has been shown that the phase of this alpha at the time of stimulus onset can mean the difference between seeing or not seeing a stimulus (Mathewson, Gratton, Fabiani, Beck, & Ro, 2009).

1.6 Analytical Techniques

Several mathematic tools will be used to alter the data, allowing for additional analytical approaches. These techniques are the fast-Fourier transform (FFT), wavelet transform, and principle component analysis (PCA).

1.6.1 Fast-Fourier Transform. An FFT is used to convert time domain data into the frequency domain. The formula used for this conversion is $X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i kn / N}$, where
\( k=0, \cdots, N-1 \), \( x_n \) is the time-domain data, and \( N \) is the number of data-points in the dataset (Cooley & Tukey, 1965). The Fourier inversion theorem demonstrates that a Fourier transform is interchangeable with the source data; that is, they are merely different forms of the same data; nothing is lost by converting between forms (Folland, 1992).

An FFT is typically performed for each trial and channel individually, then averaged across trials for each participant. The individual participant FFTs can then be combined into a grand average for each electrode, which is a mean of means for the data from that location.

**1.6.2 Wavelet Transforms.** A wavelet is a tool for extracting frequency information from a dataset without sacrificing temporal resolution. A wavelet is defined as \( w(t,f) = A e^{2\pi tf} \cdot e^{-\frac{t^2}{2\sigma_t^2}} \) (Yordanova, Kolev, & Rothenberger, 2013), where \( \sigma_t \) is the wavelet duration, and \( A = (\sigma_t \sqrt{\pi})^{-\frac{1}{2}} \).

The analysis will use a ratio of \( \frac{f_0}{\sigma_f} = 6 \) to optimize the balance between temporal and frequency resolution. The frequency range analyzed was 1 to 30 Hz with 0.5 Hz steps (59 steps total). For each data point, the peak of the wavelet is lined up with the data point, and a dot product between the wavelet and the surrounding data is produced. This dot product is summed, and that value is reported for that data point. The wavelet is then moved to the next data point and the process is repeated. This whole process is repeated for all 59 frequency steps, producing a three-dimensional plot of time (s), frequency (Hz), and spectral power (\( \mu V^2 \)).

For final analysis, a wavelet transform is typically produced for each electrode and trial, then the trials are averaged for each participant. Finally, the participant averages are combined into a grand average, making the final wavelet transform a mean of means.

**1.6.3 Principle Component Analysis.** PCA is a tool that allows for the spectral decomposition of a signal (Dien, Spencer, & Donchin, 2003). It can be done in either the time or
frequency domain, depending on the source dataset. An additional variant of PCA can be done spatially as well which allows for one to approximate the location of primary signal generation on the scalp, if a spatial coordinate reference is provided to the algorithm.

The output of the PCA is the ten algorithmically generated components responsible for the largest amount of variance. These components can then be plotted to visualize where in time or frequency the respective variance occurs. This is useful to inform later analysis during exploratory research.

1.7 Summary

Like any human endeavour, the medical field is vulnerable to errors, but when lives are on the line, the stakes are much higher. Targeted interventions in medical education are one such domain by which this error rate may be reduced. Using Dual Process Theory to identify when less accurate Type 1 processes are being prematurely utilized may be a novel neurofeedback approach in encouraging more effective training of future physicians.

This raises the question: can Type 1 processes be differentiated from Type 2 reliably using EEG? Based on the literature, it seems possible. Since elevated MFΘ is positively correlated with conflict processing, and posterior alpha is negatively correlated with attention, it is hypothesized that these two measures can be used to distinguish between Type 1 and 2 processes in a medical decision-making task. I predict that Type 2 decisions will evoke elevated MFΘ, allowing for the determination of decision-making processes based on frequency data.
CHAPTER 2: EXPERIMENT 1

2.1 Introduction

Although medical doctors go through lengthy and intensive training they are still human, and still prone to error. Some estimates place the error rate as high as 12% (Graber, 2005), which has severe implications for patient outcomes. As such, understanding the decision making pathway and how each process along the way can influence decision making accuracy is paramount to improving education and patient outcomes.

The current model for medical decision-making is heavily based on Dual Process Theory (Croskerry, 2009a). This model divides decision-making into two general categories: Type 1, which are fast and easy, but have lower accuracy, and Type 2, which are slow and effortful, but tend to be more accurate (Kahneman, 2011). Type 1 processes are generally based on heuristics, and function on an unconscious level; they are responsible for the ‘gut’ feeling. Type 2 processes are being utilized whenever conscious, effortful deliberation takes place (Evans & Stanovich, 2013).

Utilizing electroencephalography, different components of decision-making can be observed. The most commonly studied signal is theta oscillations, which are a frequency signal defined as 4 Hz to 6 Hz (Cavanagh et al., 2012). Elevated MFΘ has been associated with increased conflict processing, which is a feature of Type 2 processing (Gärtner, Grimm, & Bajbouj, 2015). This signal is thought to originate from the anterior cingulate cortex, which has been identified as a structure responsible for conflict processing by fMRI studies, as well as EEG source analysis (Cavanagh & Shackman, 2015).

Another signal of interest is posterior alpha. This signal has been found to have an inverse relationship with attention (Mathewson et al., 2011). Although this is a more general concept than
conflict processing, posterior alpha can provide a secondary verification, as increased attention is necessary for processing higher conflict tasks (Mathewson et al., 2014)

Using a modified medical cards paradigm (Williams et al., 2017), this study aims to differentiate between Type 1 and Type 2 processes by utilizing MFΘ power in a complex medical simulation task. More specifically, MFΘ power is expected to be higher during high conflict trials requiring Type 2 processing when compared to no conflict trials requiring Type 1 processes.

2.2 Method

2.2.1 Participants. Forty-six participants aged 17 to 46 ($M = 21$, $SD = 4.7$) with no medical training were recruited from the University of Victoria. All participants had normal or corrected-to-normal vision, no neurological impairments, and volunteered for extra course credit in a psychology course. Fourteen participants were removed for failing to learn the paradigm, and two were removed for having abnormal results in a standard lab screening task. All participants provided informed consent approved by the Human Research Ethics Board at the University of Victoria, and the study followed ethical standards as prescribed in the 1964 Declaration of Helsinki.

2.2.2 Apparatus and Procedure. Participants were seated in a sound dampened room in front of a 19” LCD computer monitor and used a handheld 5-button RESPONSEPixx (VPixx, Vision Science Solutions, Quebec, Canada), controller to complete an adaptation of a Cards reinforcement learning paradigm (Williams et al., 2017) written in MATLAB (MathWorks, 2017) using the Psychophysics Toolbox extension (Kleiner et al., 2017).

Cards teaches participants through the application of reinforcement learning principles. Participants were presented with physiological data (e.g., liver enzyme values) which they then used to make clinical decisions. Specifically, participants learned to classify two types of liver and
biliary diseases: cholestatic extrahepatic and general hepatocellular. This classification mimics the first step of cognitive organization structures called “schemes”, a process particularly ascribed to expertise (Coderre, Mandin, Harasym, & Fick, 2003). During each clinical case (i.e., trial) of the experiment, participants were shown a patient case-study card followed by a multiple-choice presentation of the diagnostic classification options. Following a participant's diagnosis, a feedback screen indicated whether the diagnostic classification was correct or incorrect.

The patient card presented included a photo of the simulated patient and 10 physiological readings (Figure 2.1). The patient's photo was randomly determined, without replacement, by a pool of 357 profiles (Minear & Park, 2004). All ‘patients’ were 50 to 93 years old with no outward manifestations of any liver or biliary diseases. To extend the length of the task, the photos were repeated twice in a newly randomized order proceeding the first presentation of the entire set. This resulted in a total of 1071 possible trials. The physiological data were displayed in five rows and

---

*Figure 2.1. Summary of experimental paradigm and timing.*
two columns where the text was displayed in a green, purple, blue, yellow and white Arial font from top to bottom, respectively. To ensure participants were learning to discern which variables were necessary to classify clinical cases (rather than spatial locations), the physiological data were randomly placed across the card on each trial. Eight of the ten physiological readings (heart rate, blood oxygen level, blood pressure, respiratory rate, temperature, alkaline phosphatase, gamma-glutamyl transferase, ultrasound reading) were distractor variables and were not useful for diagnosing clinical cases. The remaining two variables (alanine aminotransferase and aspartate aminotransferase) were pertinent to the clinical cases and varied as a function of the patient's disease. For each variable presentation, a random number was generated within a respective and appropriate range thus ensuring that no two cards were the same. All cards were generated and verified by a medical expert in the clinical area as to their accuracy and validity. Importantly, participant did not receive any of the above information, nor were they trained on any of the variables or diseases in the experiment. Participants were able to view the patient card as long as needed and once ready to decide, they pressed a button select their diagnosis. For each trial, we measured the reaction time—the time participants took to make a diagnosis after the patient card was presented.

The diagnostic classification options were presented in the same array as the question box in a white Arial font. The options were bordered by a colour (left: green, right: red) to match the coloured buttons of the response box. Participants selected a specific diagnostic classification option by pushing the appropriate response button, after which the border of the selection raised in brightness. To ensure that participants were learning the diagnostic classification (rather than spatial locations), we randomized which diagnostic classifications appeared in a given location. Once participants submitted their response, a white fixation cross appeared for 400ms to 600ms
so that the upcoming feedback stimuli could be analyzed independently of motor activity. Feedback of the decision was then presented as either a ‘✓’ (correct) or an ‘X’ (incorrect) in a white Arial font. This feedback stimulus was presented for 1000ms. Importantly, this was the only feedback provided to participants – they were never given information on why they had made the correct or incorrect decision, or on how to reach the correct decision. At the offset of feedback, the participant would record how confident they were in their selection from one to ten before moving on to the next trial.

After verifying that the pattern had been learned by achieving 90% or higher in three consecutive blocks of 20, participants moved on to the second phase of the experiment. Those that were able to achieve this requirement were placed in the “learners” group, whereas those who were unable to were considered “non-learners”. In this phase, the data ranges were extended such that they met in the middle, creating a zone of previously unseen values called the “conflict zone”. The second phase consisted of 10 blocks of 20 trials. The nature of each trial was random, but probabilities were set to create approximately 100 trials with all values in previously learned ranged (no conflict), 50 trials with one of the two important variables in the conflict range (low conflict), and 50 trials with both variables in the conflict range (high conflict). Additionally, feedback was withheld in this phase to limit learning effects as much as possible. Otherwise, the paradigm was unchanged; the appearance was identical, and trial confidence was recorded before proceeding.

2.2.3 Data Acquisition and Processing. Accuracy rates, response time, and self reported confidence were recorded using the RESPONSEPixx controller (VPixx, Vision Science Solutions, Quebec, Canada). For each participant, behavioural analyses examined block accuracy rates,
response times, and self reported confidence. Grand average behavioural data were created by averaging the results of all corresponding participants.

EEG data were recorded from 64 electrodes which were mounted in a fitted cap with a standard 10-10 layout (ActiCAP, Brainproducts GmbH, Munich, Germany). Electrodes on the cap were initially referenced to a common ground. On average, electrode impedances were kept below 20 kΩ. The EEG data were sampled at 500 Hz, amplified (ActiCHamp, Revision 2, Brainproducts GmbH, Munich, Germany), and filtered through an antialiasing low-pass filter of 8 kHz. To ensure temporal accuracy of stimuli and data a DATAPixx processing box (VPixx, Vision Science Solutions, Quebec, Canada) was used.

EEG data were processed as follows using Brain Vision Analyzer software (Version 7.6, Brainproducts, GmbH, Munich, Germany). First, excessively noisy or faulty electrodes were removed. The ongoing EEG data were down sampled to 250 Hz, re-referenced to an average mastoid, and then filtered using a dual pass Butterworth filter with a passband of 0.1 Hz to 60 Hz in addition to a 60 Hz notch filter. Segments spanning -200ms to 2200ms locked to the card display, and -2200ms to 200ms locked to the response, were created to complement an independent component analysis which was used to correct ocular artifacts (Luck, 2014). Channels that were initially removed were then interpolated using spherical splines. A re-segmentation of the data was then conducted to yield 2000ms epochs for each time of interest. All waveforms were then processed through an artifact rejection algorithm with a $15 \mu V/ms$ gradient and a $150 \mu V$ absolute difference criteria. For each participant, condition and time point, ERP waveforms were created by averaging the epoched data for each electrode. The ERP component of interest in the learning phase was the reward positivity, simply as a replication of the previous paradigm (Williams et al., 2017). The peak value at FCz for the difference wave (correct–incorrect) was computed, and a
mean window of ±25ms was compared between conditions via \( t \)-test (Williams et al., 2017). This was repeated for both the learner and non-learner groups. In the conflict phase, wavelet plots were generated to compare the frequency effects of the two conditions (no conflict and high conflict).

The wavelets were generated using a modified Morlet wavelet with a Morlet parameter of 6, for 59 frequency steps (1–30 Hz in 0.5 Hz steps). The two conditions were compared at \( F_z \) for theta effects (Cavanagh & Shackman, 2015) and \( P_z \) for alpha effects (Williams, Kappen, Hassall, Wright, & Krigolson, 2019). Significance was tested using k-cluster analysis to isolate clusters which were then averaged to a single value for each condition and participant. ANOVA and post-hoc \( t \)-tests were used to test significance between conditions, and principle component analysis was used to investigate the variability in both frequency and time.

2.3 Results

2.3.1 Behavioural Data. There were three behavioural measures: accuracy, reaction time, and self-reported confidence. The participants were 96% accurate in no conflict trials and 67% accurate in high conflict trials (\( p < .001 \); Figure 2.2; Table 2.1). The mean time to decision for no conflict trials was 4.9s and 5.6s in high conflict trials (\( p < .001 \)). Participants reported a mean

![Figure 2.2. Summary plot of behavioural data (accuracy, reaction time, and self-reported confidence).](image)
Table 2.1.

Summary of behavioural data (accuracy, reaction time, and self-reported confidence).

<table>
<thead>
<tr>
<th></th>
<th>No Conflict</th>
<th></th>
<th>High Conflict</th>
<th></th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95% CI</td>
<td>95% CI</td>
<td></td>
<td>95% CI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M  SD  LL</td>
<td>UL</td>
<td>M  SD  LL</td>
<td>UL</td>
<td></td>
</tr>
<tr>
<td>Accuracy (%)</td>
<td>96  11  92.6  99.4</td>
<td></td>
<td>67  19  61.3  73.7</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Reaction Time (s)</td>
<td>4.9 1.6 1.12 2.08</td>
<td></td>
<td>5.6 2.2 4.93 6.21</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Self-reported Confidence</td>
<td>8.5 2.2 7.83 9.17</td>
<td></td>
<td>7.7 2.2 7.02 8.38</td>
<td>&lt;.001</td>
<td></td>
</tr>
</tbody>
</table>

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

A confidence score of 8.5 in no conflict trials and 7.7 in high conflict trials (p < .001).

2.3.2 Reward Positivity. To ensure learning did occur, the presence of the reward positivity was verified post-hoc for the participants that learned the task. The win and loss waveforms, time-locked to the display of the feedback, was plotted for both the win and loss conditions, and the difference wave was computed (Figure 2.3). The peak value of the difference waveform was identified at 280ms, with a mean value of 2.52 µV (p = .16). For the non-learners, the peak occurred at 300ms, with a mean value of 6.85 µV (p = .03).

2.3.3 Fast Fourier Transform. An FFT was calculated for the two electrodes of interest (Fz and Pz) at both time windows and both conditions (Figure 2.4). A difference plot was generated from this data as well. While there are elevated levels of theta at Fz and alpha at Pz for both time points, there was no visual difference between the no and high conflict conditions.

To further investigate the differences between the different conditions, a repeated measures ANOVA was calculated, comparing the interaction between frequency band (delta, theta, alpha, and beta), channel (Fz and Pz), and conflict level (control and conflict) within each time window.
Figure 2.3. The reward positivity for the participants that successfully learned the task (left) and those that did not (right). Plotted are the waveforms for the win and loss condition, as well as the difference wave. At the bottom there are black squares for every datapoint that showed a significant difference between the two conditions. Within-subject confidence intervals are displayed as shaded regions. Both plots used FCz as the electrode of interest.

Figure 2.4. The FFT plots above show that the frequency spectra for both channels of interest (Fz on left, Pz on right), and time windows (stimulus on top, response on bottom) are similar. The difference plot (control–conflict) visualizes this similarity.
(stimulus [Table 2.2], and response [Table 2.7]). Significant results were found during stimulus for band ($p < .001$) and band–channel ($p = .003$), and during response for band ($p < .001$), channel ($p = .019$), conflict ($p = .019$), band–channel ($p = .019$), and band–conflict ($p < .001$). Post-hoc testing revealed significant effects for all interactions between bands, but not between channel or condition within the same band for either stimulus (Tables 2.3-2.6) or response (Tables 2.8-2.10).

Table 2.2
A summary of repeated measures ANOVA results exploring the interaction between band, electrode, and conflict level during the stimulus time window.

<table>
<thead>
<tr>
<th>df</th>
<th>SS</th>
<th>F</th>
<th>p</th>
<th>Post-hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel</td>
<td>1</td>
<td>10.574</td>
<td>1.222</td>
<td>.278</td>
</tr>
<tr>
<td>Condition</td>
<td>1</td>
<td>3.331</td>
<td>5.599</td>
<td>.025</td>
</tr>
<tr>
<td>Band</td>
<td>3</td>
<td>7315.153</td>
<td>123.225</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Channel–Condition</td>
<td>1</td>
<td>3.273</td>
<td>4.011</td>
<td>.055</td>
</tr>
<tr>
<td>Channel–Band</td>
<td>3</td>
<td>90.257</td>
<td>4.551</td>
<td>.005</td>
</tr>
<tr>
<td>Condition–Band</td>
<td>3</td>
<td>6.848</td>
<td>4.975</td>
<td>.003</td>
</tr>
<tr>
<td>Channel–Condition–Band</td>
<td>3</td>
<td>6.570</td>
<td>4.226</td>
<td>.008</td>
</tr>
</tbody>
</table>

Table 2.3
Post-hoc t-test results for band during stimulus. Values are Bonferroni corrected for multiple comparisons.

<table>
<thead>
<tr>
<th>Delta</th>
<th>Theta</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theta</td>
<td>&lt; .001</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>1</td>
</tr>
<tr>
<td>Beta</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>
Table 2.4

*Post-hoc t-test results for channel–band interaction during stimulus. Values are Bonferroni corrected for multiple comparisons.*

<table>
<thead>
<tr>
<th></th>
<th>Delta</th>
<th>Theta</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_z$</td>
<td>$P_z$</td>
<td>$F_z$</td>
<td>$P_z$</td>
</tr>
<tr>
<td><strong>Delta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F_z$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_z$</td>
<td>.001</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Theta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F_z$</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$P_z$</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Alpha</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F_z$</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>$P_z$</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
<tr>
<td><strong>Beta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F_z$</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>$P_z$</td>
<td>.001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Table 2.4

*Post-hoc t-test results for channel–band interaction during stimulus. Values are Bonferroni corrected for multiple comparisons.*

<table>
<thead>
<tr>
<th></th>
<th>Delta</th>
<th>Theta</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CL</td>
<td>CT</td>
<td>CL</td>
<td>CT</td>
</tr>
<tr>
<td><strong>Delta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td>.45</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Theta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Alpha</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>CT</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
<tr>
<td><strong>Beta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>CT</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

*Note.* Conditions are control (CL) and conflict (CT), respectively.
Table 2.6

*Post-hoc* t-test results for channel–conflict–band interaction during response. Values are Bonferroni corrected for multiple comparisons.

<table>
<thead>
<tr>
<th>Channel</th>
<th>CL</th>
<th>CT</th>
<th>Delta</th>
<th>Fz</th>
<th>Pz</th>
<th>Fz</th>
<th>Pz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Channel</th>
<th>CL</th>
<th>CT</th>
<th>Theta</th>
<th>Fz</th>
<th>Pz</th>
<th>Fz</th>
<th>Pz</th>
<th>Fz</th>
<th>Pz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Channel</th>
<th>CL</th>
<th>CT</th>
<th>Alpha</th>
<th>Fz</th>
<th>Pz</th>
<th>Fz</th>
<th>Pz</th>
<th>Fz</th>
<th>Pz</th>
<th>Fz</th>
<th>Pz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Channel</th>
<th>CL</th>
<th>CT</th>
<th>Beta</th>
<th>Fz</th>
<th>Pz</th>
<th>Fz</th>
<th>Pz</th>
<th>Fz</th>
<th>Pz</th>
<th>Fz</th>
<th>Pz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Conditions are control (CL) and conflict (CT), respectively.
Table 2.7

A summary of repeated measures ANOVA results exploring the interaction between band, electrode, and conflict level during the stimulus time window.

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>F</th>
<th>p</th>
<th>Post-hoc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel</td>
<td>1</td>
<td>110.568</td>
<td>4.326</td>
<td>.05</td>
<td>0.0014</td>
</tr>
<tr>
<td>Condition</td>
<td>1</td>
<td>1.700</td>
<td>3.932</td>
<td>.06</td>
<td>N.S.</td>
</tr>
<tr>
<td>Band</td>
<td>3</td>
<td>10376.993</td>
<td>125.738</td>
<td>&lt; .001</td>
<td>Table 2.8</td>
</tr>
<tr>
<td>Channel–Condition</td>
<td>1</td>
<td>0.191</td>
<td>0.407</td>
<td>.53</td>
<td>N.S.</td>
</tr>
<tr>
<td>Channel–Band</td>
<td>3</td>
<td>265.656</td>
<td>5.242</td>
<td>.002</td>
<td>Table 2.9</td>
</tr>
<tr>
<td>Condition–Band</td>
<td>3</td>
<td>4.312</td>
<td>3.198</td>
<td>.03</td>
<td>Table 2.10</td>
</tr>
<tr>
<td>Channel–Condition–Band</td>
<td>3</td>
<td>0.432</td>
<td>0.399</td>
<td>.75</td>
<td>N.S.</td>
</tr>
</tbody>
</table>

Table 2.8

Post-hoc t-test results for band during response. Values are Bonferroni corrected for multiple comparisons.

<table>
<thead>
<tr>
<th></th>
<th>Delta</th>
<th>Theta</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Theta</td>
<td>&lt; .001</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Beta</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 2.9

*Post-hoc t-test results for channel–band interaction during response. Values are Bonferroni corrected for multiple comparisons.*

<table>
<thead>
<tr>
<th></th>
<th>Delta</th>
<th>Theta</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>( F_z )</td>
<td>( P_z )</td>
<td>( F_z )</td>
<td>( P_z )</td>
</tr>
<tr>
<td>( F_z )</td>
<td>( P_z )</td>
<td>( F_z )</td>
<td>( P_z )</td>
<td>( F_z )</td>
</tr>
<tr>
<td>Delta</td>
<td>( F_z )</td>
<td>( P_z )</td>
<td>( F_z )</td>
<td>( P_z )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( &lt; .001 )</td>
<td>( .001 )</td>
<td>( &lt; .001 )</td>
<td>( .001 )</td>
</tr>
<tr>
<td></td>
<td>( &lt; .001 )</td>
<td>( .001 )</td>
<td>( &lt; .001 )</td>
<td>( .001 )</td>
</tr>
</tbody>
</table>

Table 2.10

*Post-hoc t-test results for channel–band interaction during stimulus. Values are Bonferroni corrected for multiple comparisons.*

<table>
<thead>
<tr>
<th></th>
<th>Delta</th>
<th>Theta</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>( F_z )</td>
<td>( P_z )</td>
<td>( F_z )</td>
<td>( P_z )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Conditions are control (CL) and conflict (CT), respectively.
2.3.4 Wavelet. The difference wavelets were normalized to a decibel scale and plotted (Figure 2.5). A difference wavelet for the two seconds following the presentation of the simulated patient card and the two seconds preceding the selection of the diagnosis were generated at both $F_z$ and $P_z$, and the difference of the high conflict condition subtracted from the no conflict condition was used.

A first pass assessment consisting of a $t$-test for each data point was performed to identify regions of interest between conditions. For theta, there were four regions identified as significant ($p > .05$) at $F_z$ in the stimulus time window, and two in response. For alpha, both time windows

![Figure 2.5](image-url)

*Figure 2.5. The difference wavelets following the onset of the simulated patient cart (left) and preceding the selection of their diagnosis (right) at electrodes $F_z$ (top) and $P_z$ (bottom). All four plots are the high conflict condition subtracted from the low conflict condition. The frequencies of interest were theta (4-6 Hz) at $F_z$ and alpha (8-15 Hz) at $P_z$. 
had four regions at $P_z$ flagged as significant. These were plotted as contours over existing wavelet plots to inform further investigation.

### 2.3.5 Principal Component Analysis.

A temporal PCA (Figure 2.6) showed similar

![Temporal and frequency PCA](image)

*Figure 2.6.* Temporal and frequency PCA for $F_z$ (top row) and $P_z$ (bottom row) at the template (left column) and decision (right column) epochs. The temporal PCAs did not suggest any timepoints that had elevated variability for any electrode or epoch. The frequency PCAs did reveal that most of the variability can be attributed to beta frequencies as opposed to theta.
variability for all component loadings at both Fz and Pz, in both the template and decision epochs. The frequency PCA revealed much more variability in components associated with beta frequencies (15 Hz – 30 Hz) than theta (~50% beta, ~7% theta) at both electrodes and both epochs. At Pz, there was also increased variability in the component associated with alpha frequencies (15.9% at template, and 15.1% at decision).

2.4 Discussion

The behavioural results demonstrated that an increase in conflict does reduce accuracy while increasing reaction time, though the difference is much larger for accuracy. This is a replication of prior research (Williams et al., 2017), and suggests that accuracy was sacrificed to maintain a relatively low response time in both conflict levels. Additionally, self-reported confidence did not effectively reflect accuracy, verifying the assertion that self-reported confidence is not an effective measure of performance (Liaw et al., 2012). This suggests that effective external feedback could be a powerful tool to inform and improve decision making.

In the time domain, a significant difference was observed in the RewP for the non-learners, but not the learners. While counter-intuitive, this is in line with prior findings (Krigolson, Pierce, Holroyd, & Tanaka, 2009). If learning occurs very quickly, as it did for many participants in the present experiment, the RewP will dissipate quickly, since “correct” feedback becomes expected, whereas the reward effect remains for non-learners. Additionally, a large P300 effect is observed for learners that is not seen for non-learners, because “incorrect” feedback acts as an oddball stimulus for learners due to its rarity. This effect has been observed in other studies as well (Spencer & Polich, 1999).

In the frequency domain, the FFT results failed to distinguish between the control and conflict conditions in the second phase. While the repeated measures ANOVA did find a
significant difference, post-hoc testing revealed that significance only existed between bands, and this difference was so strong that it influenced the interactions (band–channel, band–condition, and band–channel–condition) as well. Although significant, this difference is not due to the experimental intervention, and is instead due to the Power Law. Defined as $P = \frac{1}{f}$, which says that an increase in frequency results in a decrease in power that is proportional and asymptotic. This pattern is clearly visible as a general trend in Figure 2.4.

There were two notable deviations from the asymptotic plot predicted by the Power Law: theta power at $F_z$, and alpha power at $P_z$. While the two conditions were not significantly different in these regions, the elevated levels do suggest that the paradigm did elicit a conflict effect for both conditions, which may be due to the control being unable to effectively force a Type 1 decision. This was one methodological error in the experimental design; Type 1 process have encode the question and answer closely due to repetition and familiarity, and it was unlikely that participants were able to learn the pattern firmly enough to utilize this approach.

The lack of within-band differences was also present in the wavelets, which did not replicate previous research suggesting that increased conflict should produce elevated MF$\Theta$ power (Cavanagh & Shackman, 2015; Williams et al., 2019). However, the effect in posterior alpha power seen for both time windows is consistent with prior findings (Mathewson et al., 2011). There is a reduction in alpha power at $P_z$ as conflict increases, which is likely due to the increased attentional resources required to address these trials and posterior alpha power is inversely proportional to attention.

Unexpectedly, beta power was the most variable in all conditions. The PCA shows that beta variability accounted for over 50% of all variability globally, suggesting that it may be a larger contributor to complex decision-making than predicted. Beta is prominent in the motor control
literature, primarily due to playing a major role in response inhibition (Kühn et al., 2004). Beta oscillations have also been observed to play a role in responding to negative visual stimuli (Güntekin & Başar, 2010), face recognition (Ö zgören, Başar-Eroğlu, & Başar, 2005), memory retrieval (Waldhauser, Johansson, & Hanslmayr, 2012), and feedback visual processing (Jensen, Bonnefond, Marshall, & Tiesinga, 2015). More controversially, beta power has been linked to cognitive control. Some have found beta power to be elevated during memory formation and categorical learning (Helfrich & Knight, 2016), while others have found that elevated beta indicates an adherence to the status quo—that is maintaining the current allocation of resources within the brain (Engel & Fries, 2010). The exact role if beta power in the resolution of complex decision-making tasks remains elusive, however it presents a possible area of investigation in future studies.

Moving forward, subsequent studies can address the shortcomings seen in the present paradigm. The first has already been discussed: Type 1 processes were not effectively isolated. The second was that a third conflict condition (low conflict) had to be dropped from the analysis because it could not be effectively resolved. Originally, the control was defined as having all variables within the previously learned range, low conflict only one, and high conflict with both variables outside the known range. Unfortunately, it became clear that most participants were only tracking one of the two relevant variables, making it impossible to use the low conflict condition: if only one variable was within the known range, was it the tracked or untracked variable? Should a third disease be added, it would require the use of two variables to distinguish, thereby providing further resolution to the analysis. The last oversight was the lack of temporal consistency between trials. EEG is extremely sensitive to variations in temporal alignment between trials, and this study allowed for too much variation. With so much information presented at the same time, it could not
be determined when a participant stopped reading, started thinking, or when they had made a
diagnosis. This led to an averaging out of any effect that may have been present in the data, and
could be addressed by a more constrained experimental design.

Although this study was not able to prove the hypothesis—that Type 2 processes elicit
elevated MFΩ power, and that Type 1 and 2 processes can be distinguished by that measure—it
did inform future studies that may be able to isolate this difference. It is a finding that could be
extremely valuable in providing effective feedback to medical students during training. Since self-
assessments of accuracy are inherently flawed, an objective external source of real-time feedback
would be invaluable for ensuring effective repetition of Type 2 processes to solidify accurate Type
1 processes in the future. This intervention has the potential to further improve the diagnostic
accuracy of future physicians and improve outcomes for patients.
CHAPTER 3: DISCUSSION

3.1 Findings

3.1.1 Behavioural. The behavioural results seen in this experiment were consistent with previous research; high conflict trials had reduced accuracy and increased reaction time (Williams et al., 2017). Additionally, the self-reported confidence did not align with actual performance in the second phase; although both trended down as conflict increased, accuracy dropped far more than confidence. This verifies the assertion that self-reported confidence is not an effective measure of performance, especially when performance is poor (Liaw et al., 2012). A potentially powerful opportunity for intervention is suggested here: where self evaluation flounders, an external input could provide valuable feedback, encouraging objective assessment and reflection on the cognitive effort used.

3.1.2 Time Domain. The reward positivity was significant for non-learners, but not for learners. Although counter-intuitive, this result is consistent with existing literature. As a pattern is learned more accurately, the amplitude of the reward positivity decreases (Krigolson et al., 2009). Since participants who learned the pattern learned it quickly, typically within the first block, there were many more trials in which the pattern was learned than not, particularly when compared to non-learners. Additionally, the large P300 effect seen for the learners can be attributed to a frequency effect; incorrect feedback was rare, and so it was perceived as an oddball stimulus (Spencer & Polich, 1999).

Between the behavioural findings and the reward positivity, it is clear that the participants who completed the training phase did learn the pattern effectively. As such, the distinction between no conflict (control; both variable in a learned range), and high conflict (both variables in the conflict range) can be established.
3.1.3 Frequency Domain. The FFT findings failed to clearly differentiate between conflict states during the stimulus window. Post-hoc testing revealed that the band effect, as well as band–electrode, was simply a power effect, where lower frequency bands have higher power relative to higher frequency (Demanuele, James, & Sonuga-Barke, 2007). This is due simply to the relationship between power and frequency, as seen in the formula $P = \frac{1}{f}$. Since frequency is in the denominator, as it becomes smaller, relative power increases proportionately.

Similarly, there was a power effect for both band and band–channel during the response window, although it was not the only effect observed. Electrode and conflict were both significantly different in this time window ($p = .02$ for both), and these $p$-values were replicated with post-hoc $t$-testing. The interaction effects of band–conflict and band–channel were explored post-hoc, and this revealed that there were no within-band effects. That is, within each band, there were no significant differences between either control and conflict, or $Fz$ and $Pz$.

The difference wavelets for $Fz$ did not reveal any significant theta regions on visual inspection, and global $t$-testing only found isolated points within the theta band. As such, this experiment failed to replicate the findings of previous research, in that increased conflict should result in increased MFO$\Theta$ power (Cavanagh & Shackman, 2015; Williams et al., 2019). However, the elevated alpha at $Pz$ for both conditions does align with expected results (Mathewson et al., 2011). This supports the current assertion that alpha power is inversely proportional to attention: easier trials require less attention to solve, and alpha power increases as a result.

The unexpected finding was the contribution of beta frequencies. Beta is often discussed in the motor control literature, particularly during response inhibition (Kühn et al., 2004). In a Go/No-Go task, elevated beta power is seen in No-Go trials, and is thought to originate from the subthalamic nuclei as a form of motor inhibition (Picazio et al., 2014). Beta oscillations have also
been observed to play a role in responding to negative visual stimuli (Güntekin & Başar, 2010), face recognition (Özgören et al., 2005), memory retrieval (Waldhauser et al., 2012), and feedback visual processing (Jensen et al., 2015). More importantly, beta power has been linked to cognitive control, though there is disagreement on its role. On the one hand, experiments have found that beta power is elevated during memory formation and categorical learning tasks (Helfrich & Knight, 2016). However, other studies have found that it enforces the “status quo”; that is, beta oscillations are used to regulate adherence to current resource allocation in the brain (Engel & Fries, 2010). In yet another study, beta power was linked to the level of conflict of the previous trial, rather than the current trial (Stoll et al., 2016).

Due to the lack of definitive findings in this study, it can not be directly applied to any of the previously discussed models of decision making. As such, the applicability of the various DPTs discussed previously can not be assessed for viability based on the evidence or conclusions found here.

3.1.4 Summary. The findings of the current study support the hypothesis that beta power is associated with conflict processing and cognitive control. The design of the task did result in variable cognitive requirements between trials, considering that the degree of conflict had a large range of possible values. Based on the findings of previous studies, it is of no surprise that beta power accounted for approximately half of the variability in the frequency domain. These findings align with those of Engel et al. (2010), as inter-trial variability was high. This would not allow for maintenance of the status quo in terms of cognitive control, resulting in an overall reduction in beta power, which is seen in the difference plots. While it would be interesting to determine the effect of the previous trial on a given trial’s beta power, as suggested by Stoll et al. (2016), that data was not retained during the analysis.
3.2 Implications

Behaviourally, this study was able to replicate the findings of previous research regarding the ability of non-experts to learn a convoluted pattern with limited feedback in a medical context (Williams et al., 2017). In spite of the multiple distractors, participants were able to identify the pattern extremely quickly, with many requiring only a single block. This effect was confirmed both behaviourally and through the reward positivity. Speed and accuracy are closely tied; as one rises, the other falls. Since accurate repetition is mandatory for effective training of Type 1 processes, emphasis on deliberate and correct reasoning should be emphasized over speed, especially in the foundational phase of training. For aspiring doctors, that means they must be encouraged to slow down and carefully reason through their answers, rather than responding quickly. It is unfortunate that speed is so heavily emphasized in medical testing, as allowing for additional time to provide a more carefully reasoned response would have benefits for improved training, in addition to the obvious benefits on test performance.

This study was unable to demonstrate that MFΘ power can be used to differentiate between Type 1 and Type 2 processing using a modified medical cards paradigm. The expected level of MFΘ for the no conflict condition, would according to the power law, result in an asymptotic curve on the FFT, but that was not the case. What was seen in this study was elevation in MFΘ in both conditions that did not follow the pattern expected for otherwise basal brain activity. Either the hypothesis was incorrect, or we are observing something other than what we thought. That is to say, instead of MFΘ not being able to differentiate between Type 1 and Type 2 processes, it is likely that our design was unable to force Type 1 decision making processes and that all questions functioned as though they were Type 2. This was only true of theta at Fz, emphasizing the idea that both categories of trials were recruiting Type 2 processes, and the lack of findings stems from
the inability of the paradigm to force a Type 1 process in the control condition. In addition, the increased variability seen in beta frequencies suggests that the picture maybe bigger than the hypothesis fully encompassed. The results suggest further inquiry into the role of beta oscillations in modulating Type 1 and 2 processes might be fruitful.

3.3 Limitations and Future Directions

3.3.1 Methodological Considerations. There were three methodological oversights that limit the generalizability of this study’s findings.

First, the original study intended to have a third condition called “low conflict”. This was intended to be achieved by having only one of the two variables in the conflict range, while “high conflict” trials have both variables in the conflict range. As the study progressed, it became increasingly apparent that most participants were only tracking a single variable to make their diagnoses. As such, it could not be known whether a “low conflict” trial had conflict for that participant; if they were only tracking one variable, was it that variable in the conflict range, or the other variable not being tracked? This would completely change how that trial would be processed, which resulted in all trials from that condition being removed from the final analysis. To address this issue, future research would need to have more important variables for the participant to track or design the paradigm for only two conditions from the start. The added complexity of additional variables may alter the difficulty of the task as well, thus requiring additional adjustments to the paradigm. Since \( n \) options can be distinguished with \( n-1 \) variables, an appropriate number of both must be chosen to ensure every trial is having the intended manipulation.

The second flaw was that this study likely did not establish learning to the degree required for truly Type 1 processes to be isolated. As previously stated, Type 1 processes are used for such things like \( 2+2=4 \), which are intertwined so closely that the answer is encoded alongside the
question; they become a single object, in memory (Stanovich & Toplak, 2012). While participants did identify the pattern quickly, it is unlikely that the paradigm was able to encode the diagnostic criteria to the degree required by true Type 1 processes. As such, the study was more likely trying to distinguish between easy Type 2 and hard Type 2 processes, which might explain the lack of clear differences in MFΘ power. To address this flaw, the task must consist of problems that are extremely well known to the participant prior to the initiation of the experiment. That would require either a participant pool of expert clinicians for a medical task, or a modification of the task context to be more generally approachable.

The third flaw was that the time course of the experimental paradigm was not consistent for each trial. That is, the time spent reading the patient values, making a diagnosis, incorporating outcomes into their diagnostic schema, and other processes, did not occur for the same time, at the same time, or even in the same order for each trial. Since EEG is so sensitive to changes in time when averaging over many trials, it is probable that any effect that does exist on a trial-basis was completely washed out in the grand average. This could be addressed in future experiments by simplifying the task. With less information on the screen at once, and by keeping the order of presentation consistent, cognitive processes can be more carefully managed to make subsequent analysis more effective. Alternatively, the same paradigm complexity could be used, but presented in such a way as to control when the participants are reading and when they are thinking, so different neurological processes can be clearly isolated for analytical clarity.

This study was unable to confirm that diagnostic reasoning follows the clear binary of DPT, or that DPT can be directly measured by MFΘ power. The former may prove a difficult task to verify, but the latter has potentially been verified by other studies. Using the Add-One task, which was previously used to link pupil dilation to conflict processing (Kahneman, Peavler, & Onuska,
1968), while recording EEG signals allowed for direct observation of cognitive processes (Williams et al., 2019). When in an “add one” trial, a clear elevation of MFΘ power was observed, alongside a reduction of posterior alpha power. This finding strongly supports the hypothesis that Type 2 processes are tied to conflict resolution, and provides a neurological correlate for the assessment of DPT.

3.3.2 Future Research Directions. This study suggests that in more complex tasks, there may be more to consider than simply that MFΘ power—although using such a straightforward metric to clearly distinguish between high and low conflict states would prove a convenient, and powerful finding. With increasing task complexity comes increased schema complexity, which may not allow for a clear delineation between Type 1 and Type 2 processes. The influence of beta power in the findings suggests that a medical paradigm may instead—or at least additionally—be influenced by the broader concept of cognitive control. This concept is described as the “active maintenance of task-relevant context and top-down biasing of local competitive interactions that occur during processing” (Braver, Cohen, & Barch, 2002), or more simply, the computational mechanisms in the prefrontal cortex that manage cognitive resources based on the needs of a given context. While the nexus of this control may be at the frontal end of the brain, the regulatory mechanism of that control, namely beta oscillations, can be seen globally, as it impacts each successive region (Stoll et al., 2016).

This leads to a hypothesis for future experimentation: complex tasks, like those seen in medical diagnostic reasoning, are not modulated by binary processes like DPT, and thus cannot be measured effectively by techniques like pupillometry or MFΘ. They instead rely on a process such as cognitive control and are more effectively measured by signals like global beta power.
If proven true, this hypothesis can provide insight into cognition and learning during physicians formative training years to provide neurofeedback that encourages more effortful learning early. This additional information can better inform the diagnostic schemas that a physician, and their patients, will rely on later. Doctors are still people, and any tool that can better support them in a field as nuanced as medicine can only be helpful.

Ultimately, I would like to demonstrate that neurofeedback is a viable mechanism for improving medical training. There are a few important steps that need to happen first: a definitive link between a neurological correlate, such as elevated MFΘ power, and Type 2 processes must be established. For any application to be possible, real-time feedback is a requirement, so any distinction must hold on a trial-by-trial basis; without averaging across numerous trials, any algorithm must be robust to artifacts and noise. Following these two requirements, the hardware must be portable, so the findings must be validated on a wearable EEG product, such as the MUSE™.

As with all applications, there is an implicit assumption that any patterns observed in a controlled laboratory setting still hold in the real world, an assumption that some disagree with (van Erp, Lotte, & Tangermann, 2012). It would also need to be demonstrated that an intervention requiring a change to the current standard in medical training would improve outcomes at all, even if the neuroscientific findings hold.

This is all a very long way off, and even if all goes exactly as I hope, it would take years, and even longer for the actual benefits to come to fruition. That said, good science takes time, and in fields with clearly tangible benefits like medicine, it is all the more imperative. If this research can save lives, it not only should be done, but it must be done.
3.3.3 Future Applications. In addition to the academic value this line of research holds, there is significant real-world value as well. Research into learning and decision-making can have widespread impacts, well beyond the intended field of study. As discussed previously, this research could reduce diagnostic errors, and save lives as a result. It could improve the efficacy of medical training in the same way an athlete trains for their sport: deliberate repetition of certain movements in training so that the same pattern can be executed quickly, and at a high level, when the pressure is on. In the end, our brain is a collection of neurons and synapses; why would training for a physical event be any different than training for a mental one, on a neurological level? Fundamentally, it is still forming and strengthening synapses, even if the region of the brain or goal of the task is different.

The opportunities also extend beyond the design goals of this thesis. Biofeedback is a rapidly growing consumer field that has yet to be touched by portable EEG technology. Activity monitors are commonplace, heart rate monitors are common, and Apple™ even recently had the latest generation of the Apple Watch™ approved by the Food and Drug Administration as an electrocardiograph. Electronic methods for tracking sleep, diet, bodily cycles, and so many other things are becoming ever more a part of daily living for many. With the recent partnership between MUSE™ and Smith™ to produce unobtrusive EEG glasses (Figure 3.1; Smith, 2018), the public has access to affordable hardware to track their brain patterns. Unfortunately, the software behind these tools is currently unimpressive, possibly hampering their spread. There are more accurate tools available, but they are currently used for specific applications, such as the EEG Patch developed by the American Epilepsy Society for seizure tracking (Lehmkuhle, Elwood, Wheeler, Fisher, & Dudek, 2015) (Figure 3.1).
Additionally, the biofeedback provided from wearable, portable EEG could have additional clinical benefits. Real-time recording of brain activity would no longer rely on specialized facilities with expensive equipment and highly trained staff. EEG data could inform psychiatrists of certain neurological responses during a session, and sleep studies could be done in one’s own home. More pertinently, course instructors could have access to real-time data to inform student engagement, the efficacy of their lessons, and retention, allowing for the design of more effective data-driven curricula.

Should the goals of this line of research be realized, there are endless applications that this new information could provide people in their day-to-day lives, alongside its academic and industrial uses. The possible applications of this budding field are exciting, and even though the possibility exists for this knowledge to lead to a nefarious dystopia of privacy invasion, I am hopeful for a human led renaissance of self realization and improvement.

3.4 Conclusion

This study was unable to clearly demonstrate that mid-frontal theta power can accurately distinguish between Type 1 and Type 2 processes, as defined by Dual Process Theory, although a redesign of this study’s paradigm could still do so. Additionally, the data suggested that medical

---

Figure 3.1. A small EEG patch (top) that was developed by the American Epilepsy Society for seizure tracking can be seen in images (Lehmkuhle et al., 2015). It was designed to be wearable and unobtrusive, for comfortable daily use by patients. Below is an image of the Lowdown Focus sunglasses produced by a partnership between Smith™ and MUSE™ (Smith, 2018). They are EEG enabled sunglasses that can stream live data wirelessly to a paired mobile device.
diagnostic reasoning relies on cognitive control, at least in part, which was evidenced by the overwhelming influence of beta power variability. This finding suggests a novel approach to providing neurofeedback to physicians during their early training. With a stronger foundation for diagnostic schema, error rates can be reduced and outcomes improved.

The behavioural findings in this study affirm that the effects of the speed-accuracy tradeoff hold relevance for medical training. In medicine, decisions are high stakes and it is paramount that aspiring doctors are given every chance possible to learn effectively. Part of this includes training that gives emphasis to thinking through complex situations in an environment where time is not restricted. Deliberate, accurate repetition leads to a strong foundation of faster paced decision-making in the future. This type of training is best done when lives to not hang in the balance, so that when time is of the essence, they can confidently and accurately have a positive impact on another’s life.
REFERENCES


