Generalising the Fringe-P3 Method for the Detection of Deception and Concealed Information Through Investigations into Stimuli and Analyses

A THESIS SUBMITTED TO

THE UNIVERSITY OF KENT AT CANTERBURY

IN THE SUBJECT OF COMPUTER SCIENCE

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Kathryn Louise Harris

March 2021

Acknowledgements

I would first like to thank my supervisor, Howard Bowman, for his invaluable guidance throughout my thesis. I would especially like to thank him for answering my email one day during my master's degree, enquiring if he would supervise my dissertation research, without which I may never have become involved with the Fringe-P3 method. The work I began for that master's dissertation has been continued in this thesis.

Next, I would like to thank all the staff I worked and taught with, especially Rogério de Lemos and Laura Bocchi for making teaching a wonderful experience, Sonnary Dearden and the administration team, whose support is vital to all in the School of Computing, and the marketing team and student ambassadors for providing me with so many opportunities to present my research to a wider audience. I would also like to thank everyone I shared an office with throughout my thesis, for making every day in the office an interesting one.

Finally, I would like to thank my parents, Pauline and John, and my brother, Andy, for their immeasurable kindness and support throughout not only my PhD, but my whole time at university, and my grandparents, Joan and Alf, who always supported and encouraged my love of education. I could not have done it without them.

Abstract

The Fringe-P3 method combined with a concealed information test is a counter-measure resistant method to detect (concealed) familiarity with stimuli (Alsufyani et al., 2019; Bowman et al., 2013, 2014). The Fringe-P3 method presents stimuli on the fringe of awareness through rapid serial visual presentation, where, typically, only familiar and salient stimuli can breakthrough into conscious awareness. When a salient stimulus breaks through into awareness, it generates a P3 brain response that can be detected through EEG. As such, if a familiar/salient probe stimulus (e.g., a famous name) breaks through into awareness and a P3 is detected, we can infer that the participant is familiar with that stimulus and, therefore, is concealing information about it. The Fringe-P3 concealed information test has been shown to detect familiarity with own-name and famous face stimuli at both the group and individual participants' level (Alsufyani et al., 2019; Bowman et al., 2013, 2014). Successful detection of familiarity at the individual participants' level is key, as the real world application of the method would be for individual suspects in forensic investigations to link them to crimes by detecting their familiarity with crime-related stimuli.

This thesis aimed to generalise the Fringe-P3 method by demonstrating that it can detect familiarity with a wider variety of stimuli beyond own-names and faces. Specifically, it aimed to demonstrate that the Fringe-P3 method can detect concealed information with familiar name, email address, and location image stimuli. In addition to using EEG to detect P3s in three experiments in this thesis, a fourth experiment used the attentional blink paradigm (where a participant's detection of a probe stimulus causes them to miss a target stimulus presented shortly after) as an alternative to EEG to detect concealed information. This thesis also proposed an alternative method of analysing EEG datasets to detect P3s using a template based on other participants' P3s, called the matched filter convolution analysis.

This thesis successfully provided proofs of concept that the Fringe-P3 method could be used to detect familiarity with famous name and email address stimuli at the group and individual participants' level. It also demonstrated potential to be used with location image stimuli and the attentional blink, following some methodological improvements suggested for future work. The matched filter convolution analysis successfully detected an equal number of P3s to the aggregated grand average of trials analysis that is currently used as the standard analysis for Fringe-P3 datasets.

Table of Contents

List of Figures

List of Tables

1. Introduction

It is vital to the police, counterterrorism, airport security, and other law enforcement groups that they can accurately detect if a suspect is lying or concealing information about their involvement in and/or knowledge of a crime. The most common methods of deception detection used by law enforcement groups are polygraphs and control question tests. The use of polygraphs and psychological methods employed by "lie detection experts" are also commonly seen in popular media such as TV shows and novels. However, these methods are unreliable, with inconsistent accuracy rates, and are susceptible to countermeasures used to cheat them. Therefore, researchers have been working on a variety of alternatives in the hopes of finding more reliable and accurate methods. These alternatives include different styles of questioning, concealed information tests, and the use of electroencephalography (EEG) and functional magnetic resonance imaging (fMRI).

The research in this thesis is focused on one particular method of detecting concealed information with EEG called the Fringe-P3 method. This promising method has been shown to have high accuracy rates and be resistant to countermeasures. This thesis will further explore the Fringe-P3 method's efficacy with different types of stimuli.

1.1. History of Deception Detection

The main methods of deception detection used by law enforcement can be broken down into two categories: psychological and physiological. Psychological methods of deception detection are based on verbal and non-verbal cues from suspects during interrogation. Studies on these methods have found that on average, people can detect lies with an accuracy around 50%. If people were to simply guess if someone was lying or telling the truth, they would also have a 50% chance of being correct, which means that any method of deception detection with an accuracy rate that is 50% or less is no better than guessing. A meta-analysis by Bond and DePaulo (2006) found that people could accurately detect lies 47% of the time and accurately detect truth 61% of the time. Another meta-analysis by Aamodt and Custer (2006) found that police officers, detectives, judges, and psychologists ("professional lie detectors") were not more accurate at detecting lies (55.51%) than students and other "non-professionals" (54.22%). A review of psychological lie detection studies by O'Sullivan et al (2009) found that lie detection accuracy varied across law enforcement groups, with some groups scoring mean accuracies of 73% while other groups watching the same video scored 51%. They also found that overall, law enforcement groups had higher mean accuracy in high stakes lie scenarios (67.15%) compared to low

stakes lie scenarios (55.17%). Vrij et al (2017) conducted a meta-analysis on three particular cognitive lie detection methods (imposing cognitive load, encouraging interviewees to say more, and asking unexpected questions) and found that participants could correctly detect lies and truth 67% of the time. While the 73% accuracy found by Aamodt and Custer (2006) is not a trivial score, this rate is not consistent across individuals and scenarios, and most meta-analyses and studies have found accuracy rates closer to chance, thus making psychological methods of lie detection unreliable.

The main form of physiological deception detection is the polygraph (also colloquially referred to as the "lie detector test"). This refers to a range of psychophysiological measures including heart rate, respiration rate, and skin conductance responses. Polygraphs are based on the idea that if the participant's responses for one of these measures changes when answering a particular question compared to their baseline (their measures while they are known to be responding truthfully), then they are being deceptive in their answer to that question. Common "guilty" responses include changes in heart rate (increase or decrease), decrease in respiration, and increased skin conductance (Meijer et al., 2014). Studies have found accuracy rates varying between 35% and 100% (Brett et al., 1986; Levenson, 2009; Saxe et al., 1985) for polygraph methods, thus making it unreliable across situations and investigators. Additionally, physiological measures have been shown to be very susceptible to countermeasures and participants can be trained to "cheat" the tests, typically by increasing their baseline measure or reducing their physiological responses while answering the critical questions (Brett et al., 1986; Honts et al., 1985, 1987, 1994; Saxe et al., 1985). Therefore, a more consistent and countermeasure resistant method is still needed. The Fringe-P3 method investigated throughout this thesis and discussed in detail in the next chapter aims to be that method.

While the polygraph's accuracy in detecting deception is unreliable, there is evidence that the use of polygraphs leads to participants admitting more embarrassing (E. E. Jones & Sigall, 1971; Tourangeau et al., 1997) or crime-related information (Gannon et al., 2014). A recent evaluation of polygraph testing by police with convicted or suspected sexual offenders (J. Wood et al., 2020) investigated the influence of polygraph use on participants making risk-relevant disclosures from three groups: supervisees (convicted sexual offenders undergoing police supervision), suspects (undergoing police investigation), and applicants (convicted sexual offenders who had applied for removal of notification requirements). Risk-relevant disclosures (RRDs) were defined as the participant providing new information that leads to a change in how they are supervised, investigated, or risk

assessed. They found that 71.2% of supervisees, 76% of suspects, and 43.6% of applicants who undertook mandatory or voluntary polygraph tests made at least one RRD compared to 25% of supervisees, 30.4% of suspects, and 0% of applicants who did not take a polygraph test. They also found that participants who took polygraph tests made more detailed RRDs and that most RRDs were made in pre-polygraph interviews, which suggests that just the knowledge of an impending polygraph test motivated participants to divulge more information. Additionally, in the post-polygraph interviews, more RRDs were made following a "failed" polygraph result, where a significant response was detected. This evaluation does not provide any specific results on the *accuracy* of the polygraph tests in detecting lies but does show that taking a polygraph test motivates people to divulge more information. It is very possible that this same increased motivation could apply if the police were to use other methods of detecting deception and concealed information, including neurological measures such as the Fringe-P3 method used in this thesis' research.

As well as different measures for detecting deception, there are also several different question types. There are three main question formats used by law enforcement across the world: the Control Question Test (CQT), the Differentiation of Deception (DoD), and the Concealed Information Test (CIT). In the CQT (Reid, 1947), there are relevant and control questions that are asked while recording the participant's physiological reactions. The relevant questions relate to the specific crime under investigation while the control questions are more general but include some undesirable behaviour such as "have you ever done something illegal?" It is thought that guilty participants will have stronger reactions to the relevant questions than the control questions, while innocent participants will have stronger reactions to the control questions. One of the main criticisms of this method is that the assumption that only guilty participants will have strong reactions to the relevant questions is flawed (Fiedler et al., 2002; Iacono, 2008; Lykken, 1998), as it is possible that innocent participants will realise the importance of the relevant questions and thus may show stronger reactions to them. The notion that innocent participants will have stronger responses to the control questions than the relevant questions is also debatable, as an innocent participant could have equal responses to both. It has also been shown that the CQT combined with physiological measures is susceptible to countermeasures with guilty participants being falsely classed as innocent 45-55% of the time when using countermeasures (Honts et al., 1985, 1987, 1994). The CQT is often seen by researchers as lacking scientific validity due to its flawed assumptions and unreliable

results, with some arguing that it should not be used as evidence in court (Gallai, 1999; Meijer et al., 2016; Saxe & Ben-Shakhar, 1999).

The DoD initially involved participants answering half of a series of questions truthfully and the other half deceptively (Furedy et al., 1988). Another, more recent, version of the DoD called the Sheffield Lie Test involves participants answering each question twice, once truthfully and the other deceptively (Spence et al., 2001). The aim is to isolate the reactions that occur while being deceptive. Spence et al (2001) found deceptive responses were associated with significantly longer reaction times and significantly more activity in bilateral ventrolateral prefrontal cortices (recorded with fMRI). However, there are many criticisms of the DoD. The questions used in most DoD research are simple autobiographical or semantic questions such as "Is your name X?" and "Is Rome the capital of Italy?" which are not representative of the sort of questions that would be asked in real criminal investigations. Therefore, the results may not be the same in the field. Most of the literature on the DoD has used physiological measures (Furedy et al., 1988; Gödert et al., 2001; Vincent & Furedy, 1992) or fMRI (Kozel et al., 2004; Spence et al., 2001). As mentioned previously, physiological measures of deception are susceptible to countermeasures, and fMRI equipment is expensive and requires more space and training than other measures, so is also less likely to be taken on by law enforcement. Another criticism is that the DoD is reliant on one condition being all truths and the other being all lies. If someone were to be honest in both conditions or lie in both conditions, the DoD would not be able to tell the difference and identify the lies. Finally, most DoD research is focused on the group level rather than the individual level. One study that did test the DoD at the individual level was Spence et al (2001), who used the Sheffield Lie Test with one real suspect who had been accused of harming a child but claimed she was innocent. The test showed that response times were longer and there was more activity in the relevant brain regions when she had to agree with the accusations than when she refuted them. However, there is no way to tell if these results are because she was lying when she agreed with the accusations or whether admitting guilt is more cognitively demanding. So far, the DoD has not proven to be a reliable method of detecting deception.

The final method used by law enforcement is the concealed information test. It is worth clarifying at this point that detecting *concealed information* is not exactly the same as detecting *deception*. "Deception detection" typically refers to detecting a *lie* - a false statement - whereas detecting concealed information involves detecting the presence of crime-related information in a participant's memory. Deception detection methods test for

physical, neurological, or behavioural responses that occur while people are lying (e.g., increased heart rate), which do not occur when the participant is telling the truth. The methods and studies discussed previously in this section are attempting to detect deception. Concealed information tests, on the other hand, measure responses associated with recognition of a stimulus. These recognition responses happen whenever someone recognises a salient stimulus, regardless of whether or not that person is admitting that they recognise it, and thus the tests are not technically detecting deception. This recognition can, however, suggest that the person is being deceptive if they have a significant recognition response but claim they do not recognise that stimulus.

Only one of the experiments (chapter 4) conducted for this thesis involved lying as a part of the experiment's task and so is directly detecting deception. The other experiments are, instead, focused on detecting concealed information via familiarity with salient probe stimuli. Concealed information tests are a vital part of the Fringe-P3 method and will be explained in full in the literature review in the next chapter.

1.2. Objectives

So far, the Fringe-P3 method has been shown to accurately detect familiarity with own-name (Bowman et al., 2013, 2014) and famous face stimuli (Alsufyani et al., 2019). In order for the Fringe-P3 method to be considered reliable enough to be used in real criminal justice situations, it needs to be tested with a much wider variety of stimuli. The main aim of this thesis is to generalise the Fringe-P3 method by testing it with famous name, email address, and location stimuli. In addition to this, there are three further aims. Firstly, it will test an alternative to the P3 measure – the attentional blink – which has the potential to be used instead of or in addition to the P3 measure when using the Fringe-P3 method. Secondly, the results and discussions from chapters 4 and 5 suggest that participants could be using strategies to help them search for the target that may be negatively impacting their detection of the probe. Chapter 6 will test two task types – categorisation and detection – to investigate the impact of these potential search strategies. Finally, it will test a possible alternative method of analysing EEG data that has the potential to be able to detect P3's that do not fit the typical shape or latency.

1.3. Thesis Structure

Chapter 1: This chapter contains an introduction to the context and history of deception detection and descriptions of the main objectives and structure of this thesis.

Chapter 2: A review of the literature covering concealed information tests, EEG, the Fringe-P3 method and other methods used throughout this thesis to detect concealed information. This chapter ends with the central hypotheses.

Chapter 3: An analysis of a dataset from an EEG experiment using the Fringe-P3 method with two-part famous name stimuli. Findings at the group and individual level and their implications for the Fringe-P3 method and detecting concealed information are discussed.

Chapter 4: An analysis of a dataset from an EEG experiment using the Fringe-P3 method with email address stimuli. Results at the group and individual level and their implications for the Fringe-P3 method and investigating cybercrime are discussed.

Chapter 5: An analysis of a dataset from an EEG experiment using the Fringe-P3 method with location stimuli. Results at the group and individual level and their implications for the Fringe-P3 method are discussed.

Chapter 6: An analysis of a dataset from a behavioural experiment using an attentional blink paradigm with location stimuli. In addition to testing the attentional blink as a measure for the Fringe-P3 method, this experiment also investigates potential search strategies participants may use by comparing two tasks: categorisation and detection. The results at the group and individual level and their and implications for the Fringe-P3 method are discussed.

Chapter 7: This chapter proposes and tests an alternative method of analysing EEG datasets in the context of detecting P3s: the matched filter convolution analysis. This analysis method is tested on the email addresses data set from chapter 4. The outcomes and suitability of the method are discussed.

Chapter 8: A summary of the findings of this thesis and their implications for wider research and forensic investigations are discussed. Overall conclusions are drawn and possibilities for future research are suggested.

Glossary: A glossary of key terms relating to the methods used in this thesis.

Appendices: Appendices relating to chapters 3 (A), 4 (B), and 6 (C & D).

References: A full list of all works cited in this thesis.

1.4. Collaborations and Contributions

This thesis used several data sets collected by other researchers at the University of Kent and University of Birmingham. All but one of these data sets was processed solely by myself. All analyses were conducted by myself. Howard Bowman had important input on all experiments. The following provides further information on the contributions of others to the datasets used in each research chapter.

Chapter 3: Famous Names Dataset

The EEG dataset using famous name stimuli was collected by Abdulmajeed Alsufyani at the University of Kent. Part of the initial data processing (i.e., filtering and epoching) was performed by Abdulmajeed Alsufyani. There was a problem with the archived raw EEG files which meant that new processing could not be conducted directly on the raw files, so Alsufyani's original partially processed files (epoched and filtered) were used for this chapter. Merging and baselining the ERPs and all analyses in this thesis were conducted by myself.

This experiment was designed by Abdulmajeed Alsufyani and Howard Bowman and the dataset was originally analysed by Alsufyani for his PhD thesis in 2015 but was not published anywhere. Alsufyani's analyses used an earlier version of the AGAT analysis, where the number of trials in each condition were equalised (i.e., if there were 60 probe trials and 50 irrelevant trials for a participant following artefact rejection, then only the first 50 probe and 50 irrelevant trials would be analysed). Since then, it has been shown that the AGAT is robust against trial count asymmetry between conditions (Brooks et al., 2017), so the current version of the AGAT does not need to equalise the number of trials and simply uses all trials from both conditions (following artefact rejection). I have used the current version of the AGAT to provide all new analyses at the group and individual level for this thesis.

For the recall and recognition tests, Alsufyani used Wilcoxon's signed-rank tests. However, new analyses by myself found that the recognition measures were normally distributed, so paired t-tests were used instead of Wilcoxon's signed-rank tests for the recognition data in this thesis. The recall measure was not found to be normally distributed, so new Wilcoxon's signed-rank and paired t-tests were used for the recall data in this thesis.

This dataset has now been published in Cortex using my new analyses, with myself and Alsufyani as co-first authors (Alsufyani et al., 2021).

Chapters 4 & 7: Email Addresses Dataset

The EEG experiment using email address stimuli was designed by Howard Bowman, Claire Miller, Anthony Beech, and Brendan Jose, and the dataset was collected by Claire Miller and Weiyan Hwang at the University of Birmingham in 2015. All data processing and analyses in this thesis were conducted by myself.

This dataset was originally analysed by Weiyan Hwang for her bachelor's dissertation, but the data and her analyses were not published anywhere. Hwang analysed the data at the group level but did not conduct any analyses at the individual participants' level. All new analyses were conducted for this thesis at both the group and individual participants' level using the AGAT analysis. The introduction of analyses at the individual level for this thesis is especially key, as the Fringe-P3 method would be used for individuals in the real world, as it is individual suspects that would be tested in real forensic investigations.

This dataset and my analyses have now been published in the European Journal of Neuroscience (Harris et al., 2020).

Chapter 5: EEG Locations Dataset

The EEG experiment using location stimuli was designed by Howard Bowman and myself and the dataset was collected by myself for my master's dissertation in 2015. The running of the experiment, stimuli curation (including photographing of the University of Kent campus), data collection and processing, and all analyses were done by myself, including all new data processing and analyses for this thesis.

This dataset was originally analysed by myself for my master's dissertation in 2015 using a predecessor to the AGAT, known as the aggregated grand average of grand averages (AGAGA). The AGAGA created the aggregated ERP from the probe and irrelevant *ERPs*, whereas the AGAT creates the aggregated ERP from the individual probe and irrelevant *trials*. The AGAGA also equalised the number of trials in each condition, whereas, as mentioned earlier for the famous names dataset, the AGAT is robust against between-condition trial asymmetry, so does not need to equalise the number of trials. The data was only analysed at the group level for my MSc dissertation, whereas this thesis has used the AGAT to run new analyses on the data at both the group and individual participant's level. As mentioned previously, the introduction of analyses at the individual level for this thesis is especially important, as the Fringe-P3 method would be used for individual suspects in the real world.

For my master's dissertation, the recognition data was only analysed using Pearson's correlation coefficient to compare the recognition and ERP data and found no significant correlations. This thesis did not include any correlation analyses, and instead conducted all new analyses on the recognition data using paired sample t-tests to compare the recognition of different stimulus types.

Chapter 6: Behavioural Locations Dataset

The behavioural experiment using location stimuli was designed by Howard Bowman, Hannah Bowman, and James Niblett and the dataset was collected by Hannah Bowman and James Niblett at the University of Birmingham. The stimuli used were the same as in chapter 5 and curated by myself. All data processing and analyses in this thesis were performed by myself.

This dataset was originally analysed by Hannah Bowman and James Niblett for their bachelor's dissertations in 2017. They conducted ANOVAs on the d' data at the group level only and did not conduct any analyses on the hits data or at the individual participants' level. New analyses were conducted by myself for this thesis using ANOVAs and binomial regression analyses on the hits data at the group level and binomial regression analyses on the hits data at the individual participants' level as the main analyses. Additionally, new ANOVAs were also conducted on the d' data at the group level for appendix C, and further ANOVAs and binomial regression analyses were conducted on hits data comparing two of the four lags for appendix D. The inclusion of analyses at the individual level is, as mentioned earlier, especially important as the real-world application of the Fringe-P3 method would be for individual suspects in forensic investigations.

2. Methods for Detecting Concealed Information

2.1. Concealed Information Tests

As explained in the introduction, there is an important distinction between detecting deception versus detecting the presence of concealed crime-related information in a participant's memory. The main method for detecting concealed information is the concealed information test (CIT), formerly known as the guilty knowledge test (Lykken, 1959, 1960; Munsterberg, 1908). While Japan is the only country to predominantly use the CIT over other methods of deception detection in law enforcement, the CIT has a huge amount of evidence behind it from researchers.

Concealed information tests are based on the simple idea that guilty participants have knowledge of a crime that innocent participants do not. It uses salient stimuli called probe stimuli (e.g., a piece of jewellery that was stolen) that are presented infrequently amongst frequent non-salient stimuli (e.g., other pieces of jewellery not related to the crime). The non-salient stimuli must be of the same or similar type as the probe stimuli, so that an innocent person would not be able to work out which is the probe. When a salient probe is seen by a participant that is familiar with it (guilty), it generates detectable changes in physiological (Ben-Shakhar & Elaad, 2003; Lykken, 1959, 1960), behavioural (Ganis & Patnaik, 2009), and neurological (Alsufyani et al., 2019; Bowman et al., 2013, 2014) measures. The responses to the probe stimuli are then compared to the irrelevant stimuli (key non-salient stimuli). A strong and significantly different reaction to the probe than to the irrelevants suggests that the probe is salient to the participant and they are concealing information related to that probe and the crime. Meanwhile, innocent participants would have no different reaction to the probe, since it would be no more salient to them than the irrelevant or any distractor stimuli.

The two main paradigm types most commonly used with the CIT are personal-item paradigms and mock crime paradigms. Personal-item paradigms use personal items and autobiographical information such as the participant's own name and date of birth as probe items. Mock crime paradigms have participants commit a mock crime (e.g., steal some jewellery) and then use mock-crime-related items and facts as probes (e.g., the "stolen" necklace).

Lykken's original CIT experiments (1959, 1960), that began to draw people's attention to the CIT, relied solely on skin conductance responses (SCR) as a measure of deception. His first experiment (1959) used a mock-crime paradigm, where "guilty" participants

committed a mock theft, mock murder, or both, while "innocent" participants did neither. He detected deception in 88% of guilty participants and in none of the innocent participants. Lykken's second experiment (1960) used a personal-item paradigm with biographical information and detected 100% of the guilty participants.

A meta-analysis by Meijer et al (2014) compared skin conductance responses, respiration line length (RLL), changes in heart rate (HR), and neurological P3 responses in concealed information test experiments using mock-crime and personal-item paradigms. They found that the P3 had the largest effect size overall, followed by SCR, RLL, then HR. The P3 was moderated only by the paradigm type, performing significantly better with the personal-item paradigm compared to the mock crime paradigm. When compared to the second best measure, SCR, the P3 had significantly larger effect sizes than the SCR when using the personal-item paradigm but not when using the mock-crime paradigm. While only the paradigm type moderated the P3 measure, the SCR measure was moderated by motivation level, the number of questions, and whether or not innocent participants were included. Meijer et al found that higher motivation levels, a larger number of questions, and including innocent participants led to larger mean effect sizes. Ben-Shakhar and Elaad's meta-analysis (2003) also found that higher motivation levels led to larger effect sizes when using SCR. The finding that the number of questions did not significantly moderate effect sizes in P3 studies is particularly important, as many P3 studies use one repeated probe question/stimulus or a very small number of probe stimuli, and this did not impact the effect sizes. This means that if law enforcement were to use the P3 measure, they would not have to find a large number of different probes related to the crime in order for it to be reliable.

The P3 measure relies on the saliency of critical items, so Meijer et al suggested that the difference in effect sizes between the personal-item and mock crime paradigms was because personal-item paradigms use more salient stimuli than mock-crime paradigms. While mock-crime paradigms try to simulate real life crimes, the participants still know it is fake, thus the critical items may not be as salient to them as they could be in a real crime. Personal-item paradigms, on the other hand, use much more salient items including real autobiographical information about the participant, which would lead to stronger P3s and more significant results. Therefore, personal-item paradigms may be the more reliable method for research using the P3 to detect concealed information. The concealed information experiments conducted for this thesis using the P3 measure are closer to

personal-item paradigms and all use real salient stimuli (e.g., the participant's email address or university campus), so the P3 measure should have an advantage.

Another benefit of the P3 measure over SCR and other physiological measures, is that ERPs occur much more rapidly than other physiological responses, meaning there can be much less time between trials and more trials can be performed. P3 experiments often present each probe stimulus 20 times or more, whereas it is common for physiological experiments to show each probe stimulus only a few times or even once. This has the further benefit of allowing for the larger number of ERP trials to be averaged together to remove random noise, thus resulting in a better signal-to-noise ratio than physiological measures.

Finally, concealed information tests overcome the main criticism of the CQT - that innocent participants may fail the test - as the neutral items in the CIT are specifically chosen to be indistinguishable from the probes to truly innocent people, meaning that only guilty participants who have prior knowledge of the probe will know its importance and have differential reactions to it compared to the neutral items.

2.2. Electroencephalography, Event-Related Potentials, and the P3

Electroencephalography (EEG) is a non-invasive method of detecting electrical activity in the brain via electrodes placed in a headcap on the scalp. Figure 2.1. presents the layout of a 32 channel headcap. The brain consists of neurons which communicate with each other via synapses. When a synapse is activated, either electrically or by chemical neurotransmitters, it causes ion channels to open or close, generating excitatory or inhibitory postsynaptic potentials (PSPs) by changing the potential across the cell membrane. EEG cannot detect a single PSP, but when thousands of synapses activate at the same time in response to an event, the electrical field their PSPs generate can reach the scalp and be detected by individual EEG electrodes as positive or negative voltages.

EEG's main strength lies in its ability to provide very high quality temporal information (e.g., with a temporal resolution of 1ms) that other methods of neuroimaging such as fMRI cannot. The main weakness of EEG is that it does not provide as detailed spatial information regarding the source of the activity as other methods such as fMRI can. EEG is best used for research that is interested in the timing of activity in the brain, while fMRI is better for research regarding the location of activity. This thesis research uses the Fringe-P3 method, discussed in detail later, which requires the recording of fast responses and high quality temporal information, so EEG is the best choice.

To record EEG data, a conductive gel is inserted between the electrodes and the skin to create a stable connection and reduce impedance. Scalp-electrode impedance is checked before and during each experiment to keep it within acceptable levels. All experiments in this thesis kept impedances below 10 kOhms. The electrodes then record the electrical potential created by synapses activating during the experiment. Each electrode will record activity, providing a separate waveform for each electrode. The number of electrodes recorded from or used in analyses varies depending on the type of experiment – some experiments use all available electrodes, some only use a few. The experiments in this thesis recorded from between three and six key electrodes.

The voltages recorded by electrodes are relative to activity recorded by the reference electrode(s). The experiments in this thesis used references formed from either two electrodes placed on the mastoids behind each ear or from a combination of a common mode sense (CMS) active electrode and a driven right leg (DRL) passive electrode.

Raw EEG data is recorded according to a sampling rate measured in Hz (the number of times the signal is sampled per second). The higher the sampling rate, the higher the temporal resolution of the data. All experiments in this thesis used the standard BioSemi sampling rate of 2048Hz for recording. However, recording at such a high sampling rate generates very large data files, so the data is typically resampled to a lower rate before analysis. All experiments in this thesis resampled the data to 512Hz (a commonly used rate).

Raw EEG data contains a huge amount of information, including genuine brain activity and non-neural activity (e.g., biological potentials created by muscle and eye movements or electrical activity from nearby electrical devices that can be picked up by the electrodes). Therefore, raw EEG data undergoes processing to remove as much noise (non-neural activity and neural activity not related to the events being studied) as possible, which could prevent us from seeing the signal (the response to the event being studied).

Firstly, the EEG data is filtered to reduce noise. Typical filters are low-pass filters that allow frequencies below a certain value (e.g., 45Hz) to pass and attenuate those above it, and high-pass filters that allow frequencies above a certain value (e.g., 0.5Hz) to pass and block those below it. If both low and high-pass filters are used, then it can be called a band-pass filter. Notch filters can also be used to filter out frequencies in a narrow range (e.g., 7-9Hz), such as filtering out the steady state visually evoked potentials (SSVEP) that are generated

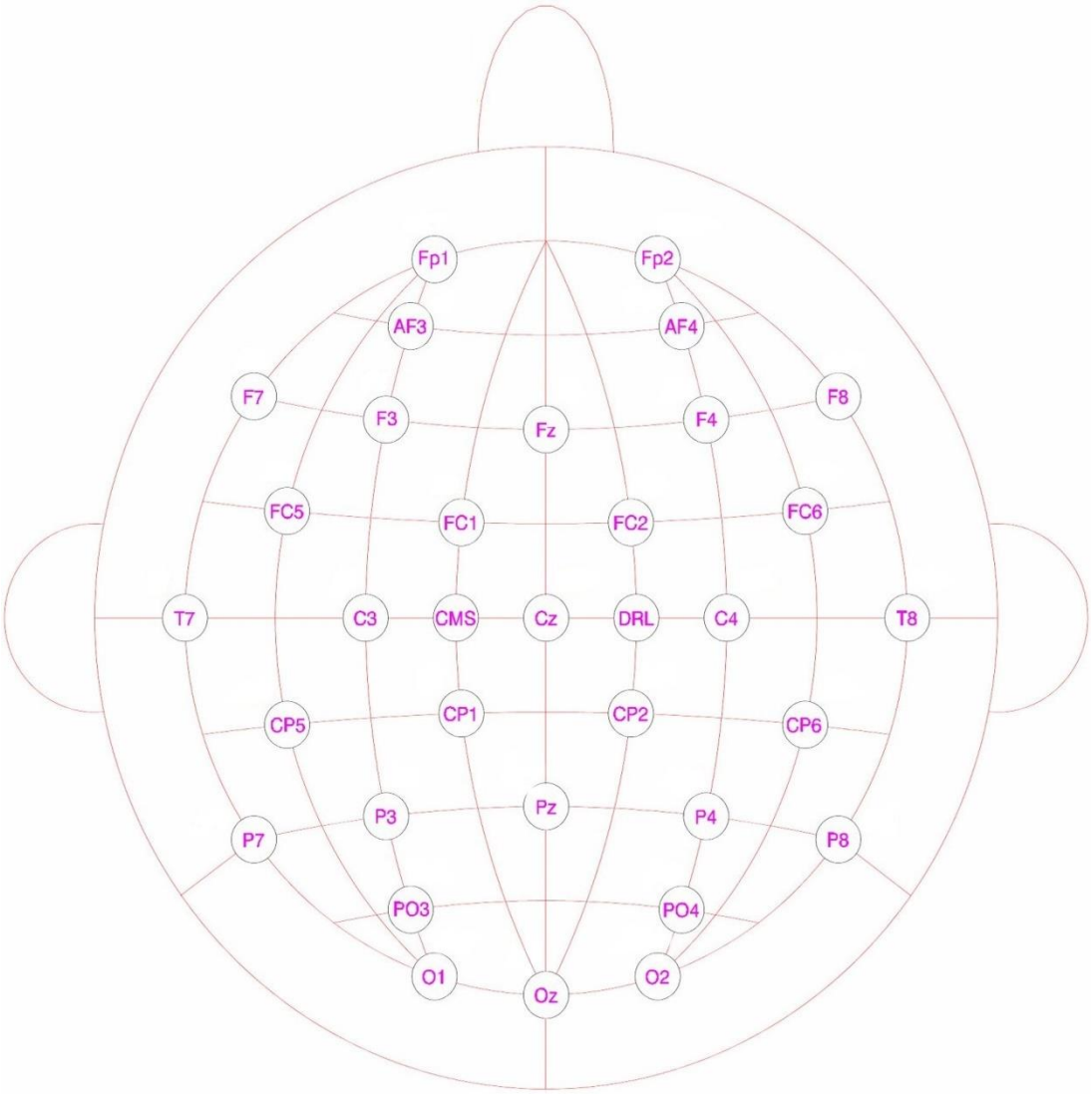
by visual stimuli being presented at regular rapid frequencies, where the waveforms oscillate at the stimulus presentation rate.

While recording EEG data, event codes are marked in the data whenever a specific event occurs, such as when a critical stimulus is presented. The EEG data is then split into epochs around these event codes (e.g., -100ms pre-stimulus onset to 1500ms post-stimulus), resulting in individual trial waveforms for specific events.

As mentioned earlier, EEG can also record artefacts caused by non-neural sources (e.g., blinking) as well as genuine electrical activity in the brain. Muscle and eye movements or blinks generate large potentials across the scalp (much larger than ERP signals) which can confound our ability to detect genuine brain activity. These artefacts can be detected manually, by looking at the EEG data for obvious too-large peaks in activity and/or automatically based on the voltage recorded from specific electrodes around the eyes as well as the key electrodes being studied. The eye electrodes used in the experiments in this thesis are two vertical electrooculogram (VEOG) electrodes placed above and below one eye and two horizontal electrooculogram (HEOG) placed on the outer canthus of each eye. Trials containing blinks detected by the VEOG and HEOG electrodes as well as the key electrodes can then be removed from further analysis. In this thesis, trials with voltages that went above 100μV and below -100μV at the eye electrodes or above 50μV and below -50μV at the key electrodes were rejected in order to reduce the noise and, ideally, leave only the signal in the EEG data.

Even after filtering and epoching, the single trial waveform at this point will still contain some background noise from non-response related neural activity. Therefore, the single trial data from particular electrodes and conditions (e.g., probe condition trials from the Pz electrode) are averaged together. This results in most of the unrelated background noise being averaged out, leaving (mostly) only the consistent responses to the events (signal) we are looking at. The resulting waveforms are called event-related potentials (ERPs) and are used in most statistical analyses.

Figure 2.1. The layout of a 32 channel BioSemi headcap for EEG.

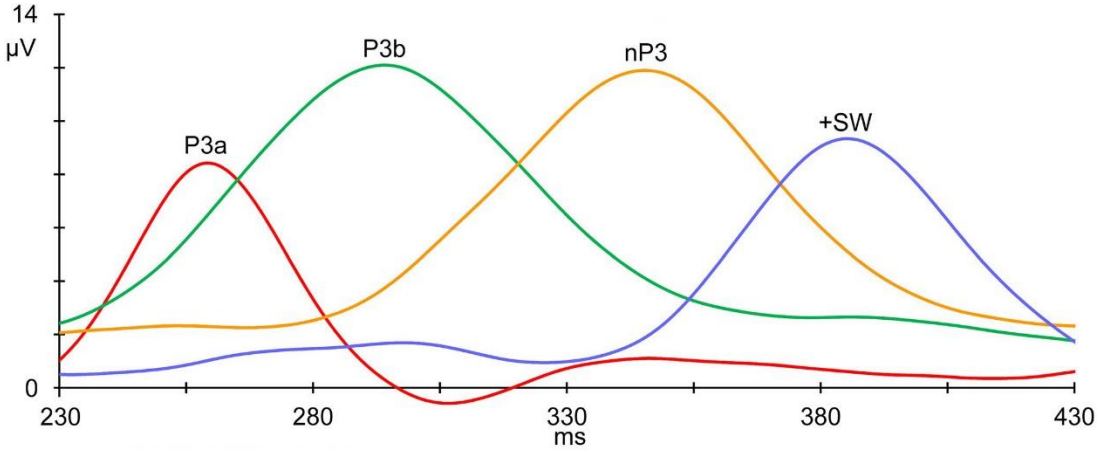


This figure was adapted from the BioSemi product website (BioSemi, Amsterdam, The Netherlands). It presents the layout of a 32 channel BioSemi headcap for EEG. The key electrodes of interest for this thesis are Fz, Cz, and Pz along the middle, with Pz being the main electrode of interest. In addition to these 32 EEG channels, two HEOG electrodes are usually placed on the outer canthus of each eye along with two VEOG electrodes placed above and below one eye to record electrooculogram data for artefact rejection. If using the linked mastoids as a reference, then two electrodes can also be placed on the mastoids behind each ear, or a combination of the CMS and DRL electrodes can be used as a reference.

The ERP component of interest to this thesis is the P3. The P3, also known as the P300, is traditionally generated during cognitive tasks where participants are discriminating between stimuli, especially when a stimulus is novel, unexpected, or salient to the participant. P3s typically appear as positive deflections in ERPs around 250-800ms post-stimulus (Meijer et al., 2014) and are usually recorded from the Pz, Fz, and Cz electrodes

(see figure 2.1 for their locations). The P3 is usually measured according to its amplitude (μV; the difference in voltage between the largest positive peak in the relevant time window and the mean pre-stimulus baseline) and its latency (ms; the time from stimulus presentation to the peak amplitude). The amplitude and latency of the P3 can be affected by stimulus and task types and by individual differences such as cognitive capability, age, and dementia (Emmerson et al., 1989; Fjell & Walhovd, 2001; Houlihan et al., 1998; Polich, 2007, 2012; Polich et al., 1986). The P3 can also be broken down into two subtypes: the P3a and P3b. P3as (or novelty P3s) are typically earlier, peaking around 250–280ms, and are maximal from the Fz and Cz electrodes. The P3b (also often referred to as simply the P3 or P300) is usually later, peaking around 300-500ms, and maximal from the Pz electrode (Comerchero & Polich, 1999; Polich & Kok, 1995). See figure 2.2. for examples of typical P3 component patterns. There is some debate about whether the P3 can be broken down into further subcomponents and whether the P3a is separate to the novelty P3 (nP3) or whether these are variations of the same potential (see Simons et al (2001), Polich (2007), and Barry et al (2020)) but this is beyond the scope of the current thesis research, which is most interested in differential general P3/P3b responses to probe and irrelevant.

Figure 2.2. Examples of P3 related components from Barry et al (2020).



This figure, adapted from Barry et al (2020) shows the latency of the P3a and P3b, as well as the novelty P3 (nP3) and positive slow wave (+SW). The P3a is typically earlier than the P3b, which peaks later and is broader. The nP3 and +SW are not key components for this thesis.

The P3 component was first discovered by Sutton et al (1965) when they found a large positive peak 300ms post-stimulus when participants could not predict whether the next stimulus would be auditory or visual. This peak was much smaller when participants were able to predict the stimulus. Since then, there have been thousands of studies into the P3.

One of the first studies to show that P3s could be generated by prior knowledge of a stimulus was by Karis et al (1984). They showed participants lists of words and later asked participants to discriminate between the previously seen words and new words. They found larger P3s for the previously seen words compared to the new words, which suggests that the P3 could be used to investigate if a person has prior familiarity with a stimulus. This finding could then be taken forward to research detecting deception and concealed information by comparing P3 responses to crime-related stimuli and irrelevant stimuli, where a larger P3 for crime-related stimuli would suggest prior familiarity with it and connect that participant to the crime.

Additionally, there is a wealth of research into the P3 using the oddball paradigm. The original 2-part oddball paradigm involves two types of stimuli: frequently presented irrelevant/distractor stimuli, and an infrequently presented target stimulus (the oddball). The infrequent stimulus will generate a large positive P3 response, while the frequent stimuli will generate either a smaller or no P3 response. In the 3-part version, there are salient probe stimuli presented in addition to an infrequently presented target and frequently presented irrelevant/distractor stimuli. P3s are generated by the target, as it is task-relevant, and the probe stimuli, as they are salient to the participant, but not by the irrelevant/distract stimuli as they are not salient. This, then, further shows that the P3 can be used to detect a participant's familiarity with a stimulus, as familiar/salient stimuli will generate P3s while the unfamiliar/non-salient stimuli will not. In the context of detecting concealed information using the P3, if there is a significant P3 response to the probe stimuli but not the irrelevant stimuli, then it suggests that the probe stimuli are familiar to the participant and they may be concealing information about them.

The Fringe-P3 method for detecting concealed information used in this thesis research uses a similar paradigm to the 3-part oddball paradigm, but with an additional stimulus type. The Fringe-P3 method uses four stimulus types: infrequent target stimuli that are task-relevant, infrequent salient probe stimuli, infrequent non-salient irrelevant stimuli, and (in)frequent non-salient distractor stimuli. There are many distractors in each stream, so as a class, distractors are presented frequently, but each individual distractor is presented extremely infrequently. The irrelevant stimuli are selected key distractor stimuli that are repeated as often as the probe and are used for comparison to the probe in the statistical analyses.

As the P3 has been shown to be generated by familiar stimuli when presented amongst unfamiliar stimuli (by Karis et al (1984) and the 3-part oddball paradigm), the P3 then began to be used with CITs, which aim to detect concealed knowledge of probe stimuli based on differential reactions to them compared to irrelevants. The first CIT experiments using the P3 were published by Rosenfeld et al and used mock-crime paradigms where participants pretended to steal an item (Rosenfeld, Cantwell, et al., 1988; Rosenfeld, Nasman, et al., 1987). Participants were then presented with streams consisting of the names of various items presented every 3 seconds, including the name of the stolen item (the probe). Participants were told to press a button for 'no' for every stimulus except the target when asked if they recognised each stimulus (thus being deceptive when saying 'no' after seeing the probe stimulus). They detected P3s for the name of the stolen item in 70-90% of participants, providing proof of concept that the P3 could be used with a CIT to detect concealed familiarity with a stimulus. Since then, there have been many experiments using the P3 with CITs to detect deception and concealed information. As mentioned in section 2.1, a meta-analysis by Meijer et al (2014) compared CIT studies using the P3 with CIT studies using other physiological measures (e.g., skin conductance responses) and found larger effect sizes for CITs using P3s than other measures, thus providing further evidence that the P3 is a suitable measure for detecting concealed information.

2.3. Countermeasures and Rapid Serial Visual Presentation

The CIT used alone is susceptible to countermeasures with physiological (Ben-Shakhar & Dolev, 1996), EEG (Rosenfeld et al., 2004), and fMRI (Ganis et al., 2011) measures with accuracy rates dropping as low as 18% (Rosenfeld et al., 2004). Countermeasures for physiological CITs include participants increasing their response to neutral/irrelevant items either physically (e.g., digging their nail into their skin to inflict pain) or mentally (e.g., thinking of an exciting memory). Countermeasures for neurological CITs include mental tactics such as focusing on not seeing the probe to decrease its saliency or focusing on or assigning a covert response to one or more of the irrelevants to increase their saliency. One attempt to overcome the impact of countermeasures on the P3 is the complex trial protocol (CTP) by Rosenfeld et al (2008). The CTP aims to prevent countermeasures by having participants respond to two stimuli within the same trial, the probe or irrelevant first, followed by a target or non-target after a delay (e.g., 1.5s), thus separating the responses to the probe and target stimuli. They suggest that by separating the two responses, they prevent resources being split between the probe and target. This would

then lead to a larger P3 for the probe (and a larger difference to the irrelevant), which would not be reduced enough by countermeasures to lose significance and allow deceptive participants to escape detection. However, research has shown that there are still countermeasures that can reduce probe P3s and the differential responses below significance, thus reducing accurate detection of concealed information while using the CTP (Hu et al., 2015; Lukács et al., 2016; Meixner & Rosenfeld, 2010). The Fringe-P3 method used by Bowman et al (2013, 2014) on the other hand, has been shown to have 100% accuracy rates with own-name stimuli (Bowman et al., 2013) and be resistant to mental countermeasures that attempt to manipulate salience (Bowman et al., 2014). This method combines a concealed information test with rapid serial visual presentation (RSVP) and the P3. The inclusion of RSVP is why the Fringe-P3 method is resistant to countermeasures.

RSVP is a method of stimulus presentation where items are shown at such short stimulus onset asynchronies (SOAs; the time between stimuli, e.g., 100ms) that they appear on the "fringe of awareness" rather than fully in conscious awareness. Studies have found that targets can be accurately detected in streams of letter, number, word, colour, and image stimuli at these fast speeds (Chun & Potter, 1995; Potter, 1976; Potter et al., 2008, 2010, 2014).

Research has also shown that emotional stimuli are better detected in RSVP streams and better remembered in post-stream memory tests (Barnard et al., 2004; Potter et al., 2014; Versace et al., 2010). Many crimes involve an emotional aspect, so this could mean that the crime-related probes are even more easily detected by guilty participants in RSVP streams. However, there is less clear evidence on how emotional stimuli affect the CIT. While there are many studies showing that emotional arousal and stress experienced by the participant during the CIT do not affect the efficiency of the test (Bradley et al., 2011; Kugelmass & Lieblich, 1966; Peth et al., 2012), there is very little specifically comparing emotional stimuli versus neutral stimuli with the CIT. One study (klein Selle et al., 2017) did compare the use of emotional and neutral stimuli with CIT and physiological measures and found that emotional stimuli increased detection efficiency for the SCR measure but not for heart rate or respiration measures. There is no evidence on how emotional stimuli may affect the P3 measure during CITs or RSVP combined with CITs. Therefore, we cannot be sure whether the increased detection of emotional stimuli in RSVP streams will be the same when combined with the CIT and P3 responses.

2.4. The Fringe-P3 Method

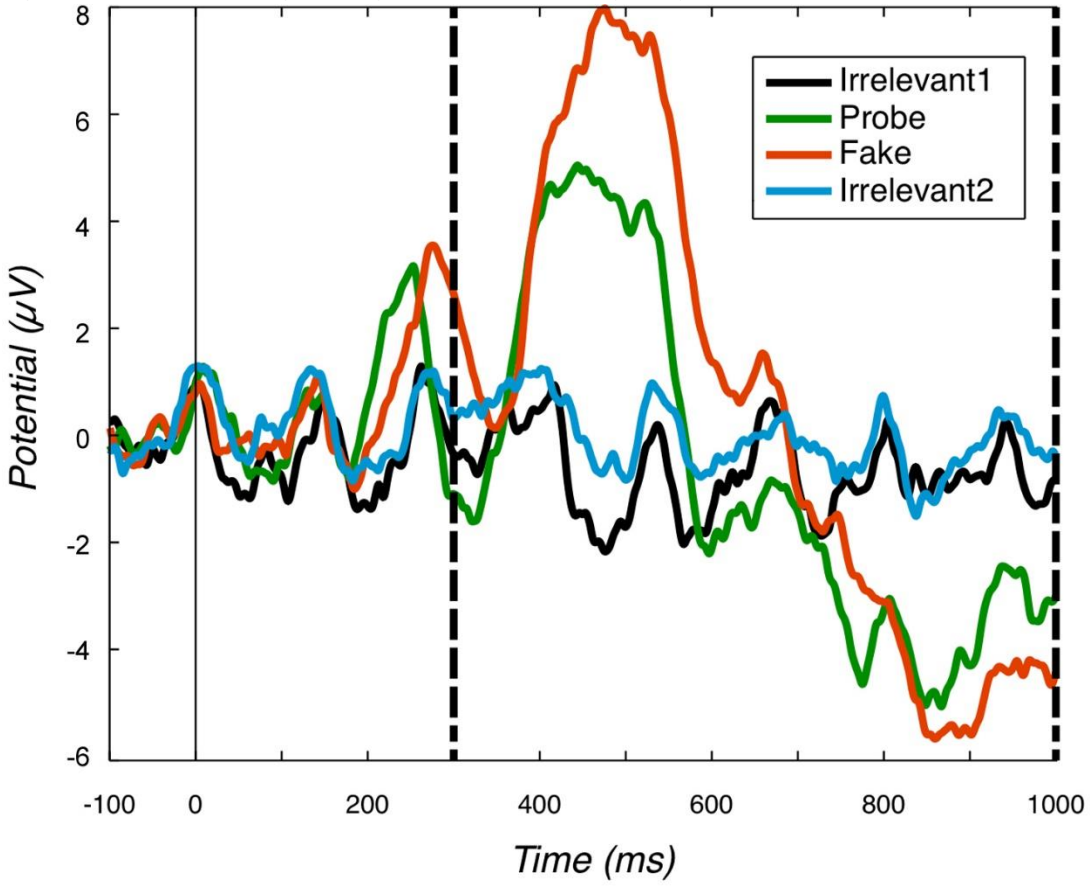
The Fringe-P3 method, proposed by Bowman et al (Alsufyani et al., 2019; Bowman et al., 2013, 2014), combines a CIT with RSVP and EEG. It uses four stimulus types: task-relevant targets, salient probes, non-salient irrelevants, and non-salient distractors. One critical stimulus (target, probe, or irrelevant) is presented in each RSVP stream and participants are tasked with detecting the target. Typically, the target's only role is to be the focus of the task and provides no information about the presence of concealed information. The probe stimuli are the stimuli we suspect the participant is concealing knowledge of. The irrelevant stimuli are non-salient key distractor stimuli that are used in comparison to the probe. The distractor stimuli are non-salient stimuli of the same type as the target, probe, and irrelevant and are used to fill the RSVP streams.

Bowman et al (2013) described how the perceptual system subliminally searches RSVP streams for salient stimuli (Subliminal Salience Search; SSS). They suggest that the brain (subliminally) attempts to match the items in the stream to templates of salient information stored in the brain, such as the person's real first name. When an RSVP stimulus matches a salience template, that stimulus breaks through into conscious awareness and is encoded into working memory. This breakthrough generates a P3 component that can be detected with EEG (Craston et al., 2009; Martens et al., 2006; Vogel et al., 1998). Only stimuli that are salient to the participant will match a template, break through, be encoded into working memory, and generate a P3. Non-salient stimuli will not match a template or generate a P3. As such, the Fringe-P3 method can detect concealed information by detecting which stimuli generate P3s and thus must be salient to the participant. If a participant was indeed concealing knowledge of the probes (guilty), the probes would be salient, generate a P3 response, and there would be a significantly different response to the probes compared to the irrelevants. If a participant truly had no knowledge of the probes (innocent), the probes would not be salient or generate a P3, and there would be no significantly different response to the probes compared to the irrelevants.

In Bowman et al's 2013 paper, they used the Fringe-P3 method with own-name stimuli. Participants chose a fake first name from a list of options and were tasked with pretending this was their real first name and looking for it in the streams. Each RSVP stream consisted of distractors and one of four critical stimuli: the target (their chosen fake name), the probe (their real name), and two key irrelevants (unfamiliar first names). At the end of each trial participants were asked "did you see your name?" and were to respond "yes" if they saw

their chosen fake name and "no" if they did not (including if they saw their real (probe) name). Bowman et al's analyses compared responses to the probe versus irrelevant, focusing on P3as at Fz and Cz, and P3bs at Pz. They found significant results at all three electrodes at the group level. At the individual level, they detected significantly different responses to the probe (P3as) for 13/15 participants at Fz and 12/15 participants at Cz, and significantly different responses to the probe (P3bs) for 12/15 participants at Pz. By combining the P3a and P3b analyses with a three-dimensional fisher combined probability procedure, they found significantly different responses to the probe for 100% of participants. They found no P3as or P3bs for the irrelevants. Figure 2.3 presents the grand average for the fake (the target), the probe, and both irrelevants. Clear P3s can be seen for the fake and probe but not for the irrelevants.

Figure 2.3. Grand Average at Pz from Bowman et al (2013).



This plot was taken from Bowman et al (2013) and flipped so that positive is plotted upwards on the y axis. The black vertical dashed lines represent the search boundaries for the window of interest. The "fake" is the fake name used as the target. There is a clear P3 positivity for the target and probe that is not present for the irrelevants. The target P3 is larger than the probe P3 because it is task relevant.

Bowman et al (2014) further investigated the robustness of the Fringe-P3 method by testing it against potential countermeasures. They tested two countermeasure types: the "probes as low-salient" method, where participants attempted to reduce their response to the probe, and the "irrelevant as high-salient" method, where participants attempted to increase their response to the irrelevants. Their 2014 experiments followed the same method as their 2013 paper, but with participants performing the additional task of trying to enact one of the two countermeasure methods. There were three countermeasure experiments and one without countermeasures. The first countermeasure experiment investigated the probes as low-salient method by informing participants that the experiment works by detecting their response to their real name and instructing participants to concentrate on "not seeing their name" in order to counter this. The second countermeasure experiment investigated the irrelevant as high-salient method by informing participants that there were two more names repeated in the experiment and instructing them to count how many times they appeared. The final countermeasure experiment also investigated the irrelevant as high-salient method by informing participants that there were two more names repeated in the streams and instructing them to identify one of them and then pretend that that was their real name. At the group level they found that, despite the countermeasures in three experiments, they still detected significant P3as at Fz and Cz, and P3bs at Pz for the probe and not the irrelevant in all four experiments. Using a fisher combined probability procedure at the individual level, they found significantly different responses to the probe for 12/12 participants in the no-countermeasures experiment, 8/10 in the probes as low-salient experiment (and the two misses were marginal), 9/10 for the first irrelevant as high-salient experiment, and 10/10 for the second irrelevant as high-salient experiment.

The use of RSVP in the Fringe-P3 method is thought to be the reason the method is so resistant to countermeasures, as participants cannot consciously control their responses to stimuli that are on the fringe of awareness (and thus they are not consciously aware of). By the time the stimulus has broken through into conscious awareness and participants realise they have seen the probe, a P3 has already been generated, making it too late for participants to apply any effective countermeasures. While two of the countermeasure experiments did result in one to two participants not having significant differences, Bowman et al estimated the area under the Receiver Operating Characteristic curve as 0.95 and 0.9938 respectively, and so concluded that the effect of interest was not reduced enough to suggest participants were able to consciously control their subliminal search

system effectively enough to confound the Fringe-P3 method. Therefore, they concluded that none of the countermeasures had a substantive effect on the accuracy of detecting concealed information. It is also worth noting that the irrelevants used in these experiments were taken from a list that was presented to all participants at the start, where they were asked to mark any names that they were familiar with so those names could be excluded from the streams. As a result, it is possible that the irrelevants may have been primed and thus easier to detect in the streams in order to enact a countermeasure. In a real forensic setting, the irrelevant names would not be presented beforehand and so would not be primed and would be even more difficult to detect.

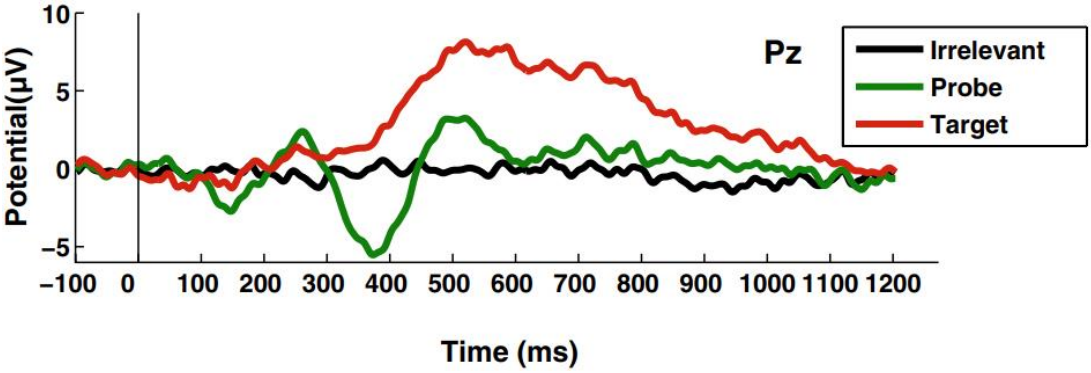
The failure of the irrelevant as high salient countermeasure to significantly impact detection accuracy also shows that participants found it very difficult to identify the repeated stimuli despite being presented up to 50 times. This fits with Avilés et al (2020), who investigated potential repetition effects by showing participants RSVP streams of words and tasking them with identifying the repeated stimulus (with no information on what the stimulus was, other than that it was repeated). They found that the probability of detecting an unknown repeated stimulus for the first time was consistent and did not increase with more repetitions. Avilés et al propose that stimuli that do not reach a threshold of awareness (and thus are not consciously perceived) create fleeting memory traces that do not last and dissipate so rapidly that repeated presentations across trials cannot accumulate, and therefore, repetition cannot increase the chances of a stimulus reaching the threshold of awareness and being consciously perceived. This means that presenting the irrelevant many times, as in the Fringe-P3 method, will not increase the likelihood of participants detecting it and using it to counter the test.

A fifth experiment in Bowman et al's 2014 paper was an innocents test and involved no countermeasures. Any "probe" used for an innocents test would actually be an irrelevant since it would be non-salient to the participant. Therefore, this innocents experiment did not use the participants' own-names as probes and instead used three irrelevants. They found no P3s for the irrelevants and no inflation of the false positive rate. This, combined with the irrelevant as high salient countermeasure experiment showing that participants found it difficult to "see" the irrelevants even when they were looking for them, and the research by Avilés et al (2020) providing evidence that repeating non-salient stimuli does not increase their chances of being detected, suggests that truly innocent participants will not generate P3s for repeated non-salient stimuli and inflate the false positive rate. This serves as a general proof that the method will not inflate the false positive rate, suggesting

that future Fringe-P3 research, such as that included in this thesis, does not need to include innocent groups unless there is a substantial change in the method.

Alsufyani et al (2019) took the Fringe-P3 method beyond the use of first name stimuli and tested it using face stimuli. This experiment followed the same Fringe-P3 method as Bowman et al (2013) except they used face stimuli in place of names and did not task the participants with pretending a fake name was their real name. Instead, participants were given an unfamiliar target face to look for in the streams and at the end of each stream were asked "did you see the target face?" to which they answered "yes" or "no". The probe stimuli were black and white photos of the faces of five famous people (Nelson Mandela, Barack Obama, Margaret Thatcher, David Beckham, and Angelina Jolie), and the irrelevants were unfamiliar faces. They investigated the effects based on two ERP components that had previously been shown to occur when viewing famous faces: the negative N400f (250-500ms) and the late positive P600 (500-800ms) (Curran & Hancock, 2007; Eimer, 2000; Touryan et al., 2011), rather than the P3. The windows selected for analysis were chosen using the aggregated grand average of trials analysis (AGAT) described in the next section. They found significant differential responses (both N400f and P600) to the probe compared to the irrelevant at Fz, Cz, and Pz at the group level. At the individual level for Pz only, there were significant differences in responses to the probe and irrelevant for 10/14 participants in the N400f window and for 7/14 participants showed significant differences within the P600 window. They also used the fisher procedure to combine the N400f and P600 p values into one combined p value, resulting in 14/14 participants having significant p values. This shows that the Fringe-P3 method can be used with famous face stimuli to differentially detect when participants see a familiar face with high accuracy. Figure 2.4 presents the grand average at Pz taken from Alsufyani et al (2019).

Figure 2.4. Grand Average at Pz from Alsufyani et al (2019).



This plot was taken from Alsufyani et al (2019). There is an N400f negativity followed by a P600 positivity for the probe that are not present for the irrelevant. There is also a P3 for the target.

A related method to the Fringe-P3 has also been used to present stimuli on the fringe of awareness by Rosenzweig and Bonneh (2019, 2020) using eye-tracking equipment to measure involuntary eye movements instead of EEG. This was based on the oculomotor inhibition (OMI) phenomenon where involuntary microsaccades are inhibited for a short time after perceiving a stimulus. The OMI is associated with attention shifts and stimulus saliency, and is prolonged by perceptual oddballs (Valsecchi & Turatto, 2009). Rosenzweig and Bonneh (2019) showed that the OMI is prolonged by familiarity with a face, including during passive viewing of face stimuli on the fringe of awareness in RSVP. Therefore, it can be used to detect concealed knowledge, as only stimuli that were familiar to the participant would be perceived and cause a prolonged OMI. Their second fringe of awareness experiment (Rosenzweig & Bonneh, 2020) was based on a "mock terror" design using the Fringe-P3 method. 13 participants in the study group covertly selected a "terror target" from 8 options and were given 20 minutes to review a set of video and text information on the target. The chosen target's face, name, and city were used as probes. The 12 participants in the control group did not pick a target or see any information on any of the options. All participants were then shown multiple RSVP streams of stimuli consisting of a critical stimulus, followed by a blank screen and two randomly selected colourful "relaxing" images to serve as masks. The critical stimulus was either a black and white image of a "universally familiar" face used as a reference, or one of the three probe stimuli. The study group showed prolonged microsaccade inhibitions for both the reference and probe stimuli, while the control group only showed prolonged inhibitions for the reference, and their reactions to the probe stimuli were indistinguishable to their reactions to the distractors. Based on their inhibition reactions to the stimuli, Rosenzweig and Bonneh were

able to accurately identify the chosen target for 100% of participants. This provides further evidence that fringe of awareness methods can be accurately used with face stimuli, and also shows that the Fringe-P3 method has applications beyond solely the use of EEG and the P3.

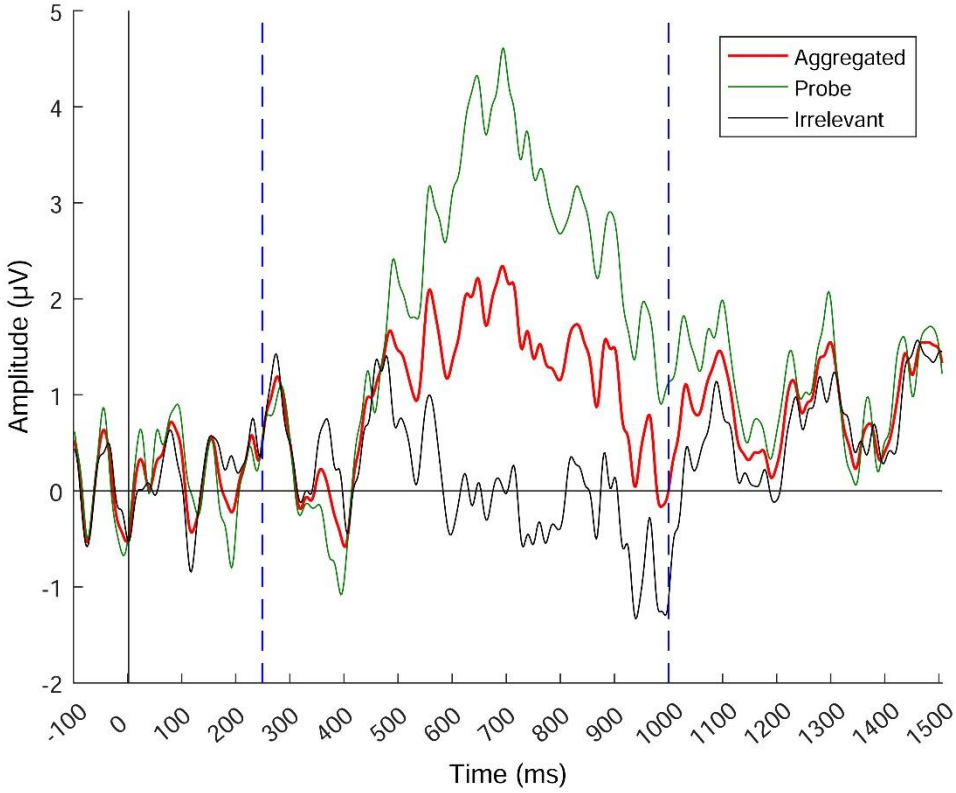
The Fringe-P3 method has shown very strong results, up to 100% accuracy, with face and own-name stimuli. This thesis will assess the effectiveness of the Fringe-P3 method with three more stimulus types: famous names, email addresses, and locations. It will also test the method while using the attentional blink (discussed later in this chapter) as a measure instead of EEG and the P3.

2.5. The Aggregated Grand Average of Trials Analysis (AGAT)

The AGAT (Bowman et al., 2020; Brooks et al., 2017) is a method of analysing ERP data and is the main analysis used in chapters 3, 4, and 5. In these chapters, it is used to compare the probe and irrelevant and is typically performed on data from the Pz electrode, but is sometimes also used on Fz and Cz data. The AGAT uses a data-driven orthogonal contrast between probe and irrelevant by merging these into one aggregated ERP. The window of interest is then selected based on this aggregated ERP, meaning that neither probe nor irrelevant is given biased treatment that could inflate the false positive (type I error) rate. Further justification for this approach can be found in Brooks et al (2017), Bowman et al (2020), and Friston et al (2006). There is some variation in parameters between the AGAT's implementation in each experiment in this thesis (e.g., window length and boundary start/end points), but the general procedure is the same for all.

Figure 2.5 shows an example of a 250ms – 1000ms window (blue dashed lines) within which the algorithm searched for the 100ms window of interest in the email addresses experiment in chapter 4. It shows the aggregated grand average as well as the probe and irrelevant grand averages.

Figure 2.5. Grand Average at Pz for the Aggregated, Probe, and Irrelevant ERPs.



This figure shows the aggregated grand average (red, bold) alongside the probe (green) and irrelevant (black) grand averages. The vertical blue dashed lines represent the 250-1000ms window within which the algorithm searched for the window of interest.

In this section, I will describe the AGAT method using common parameters for detecting a P3. First, the probe and irrelevant trials must be merged into one aggregated ERP for each participant. A window 100ms wide is then placed on each participant's aggregated ERP, starting 250-300ms after stimulus onset, and slid along the ERP until the upper boundary is reached at 1000ms. The starting boundary of the search window is 250-300ms post-stimulus. Previously, the AGAT always used a starting boundary of 300ms, as the P3 typically starts around 300ms, but following the research in chapter 7, this was expanded to 250ms to allow for variation in the onset of the P3. The upper boundary is usually 1000ms as the P3's positive deflection is expected to have returned to baseline by then (although, a negative rebound can continue on past 1000ms). These timings match ERPs and analyses from previous studies using the Fringe-P3 method (Bowman et al., 2013, 2014). The mean amplitude of the aggregated ERP within each window position is then calculated and the window of interest (i.e., the window position with the highest mean amplitude) is found for each participant. This window of interest is then applied to the

separate probe and irrelevant ERPs and the mean amplitude for each within that window is calculated and used in further analyses. Mean amplitude measures are used as they are more robust against high frequency noise (Luck, 2014).

For a group level analysis, the AGAT is typically followed by a two-tailed paired samples t-test on the probe and irrelevant mean amplitudes.

For an individual participants' level analysis, the AGAT is followed by a Monte-Carlo permutation test, permuting the individual trials for probe and irrelevant. Permutation works on the assumption of exchangeability, that if the null hypothesis is true, the data could be exchanged between groups thousands of times and the difference between probe and irrelevant would still be non-significant, subject to sampling error, and that any specific results would be just as likely to be found in one group as in the other. Permutation is used for the individual participants' analysis, as the generation of surrogate ERPs have a better signal-to-noise ratio than if t-tests are performed on the single trials, which contain more noise. Permutation tests are used instead of bootstrapping, since bootstrapping has been shown to be biased with peak measures but permutation is not (Zoumpoulaki et al., 2015).

To conduct the permutation test, the true irrelevant mean amplitude is subtracted from the true probe mean amplitude to provide the true observed difference. The probe and irrelevant individual trials are then permuted without replacement to create surrogate probe and irrelevant datasets. The same window of interest used with the true data is then applied to each of the permuted surrogate datasets and the surrogate probe and irrelevant mean amplitudes are calculated. The surrogate irrelevant mean amplitude is then subtracted from the surrogate probe mean amplitude and saved as the surrogate difference. This is typically repeated 1000-10,000 times. A p value is then calculated as the proportion of surrogate differences that are larger than the true observed difference.

For the individual participants' level analysis, the AGAT analysis is run on each individual participant as if they were a separate experiment. The individual participants are not compared with each other, therefore we are not doing multiple comparisons, so corrections for the type I error rate are not required. This applies to all individual participants' level analyses in this thesis.

Fringe-P3 research aims to show high detection rates of familiarity with the probe at both the group and individual participants' level. The individual participants' level analysis is especially important as the real world application of the Fringe-P3 method would be for

individuals rather than groups. For example, the Fringe-P3 method combined with a CIT could be used in forensic situations to investigate an individual suspect's familiarity with the face or name of another person of interest, such as an accomplice.

It should be noted that the inclusion of analyses at the individual participants' level is unusual. Most EEG research analyses effects at the group level only. Analysis at the individual participants' level is less common and more difficult, as it uses less data and is more susceptible to noise, so being able to find significant effects at the individual level is especially noteworthy.

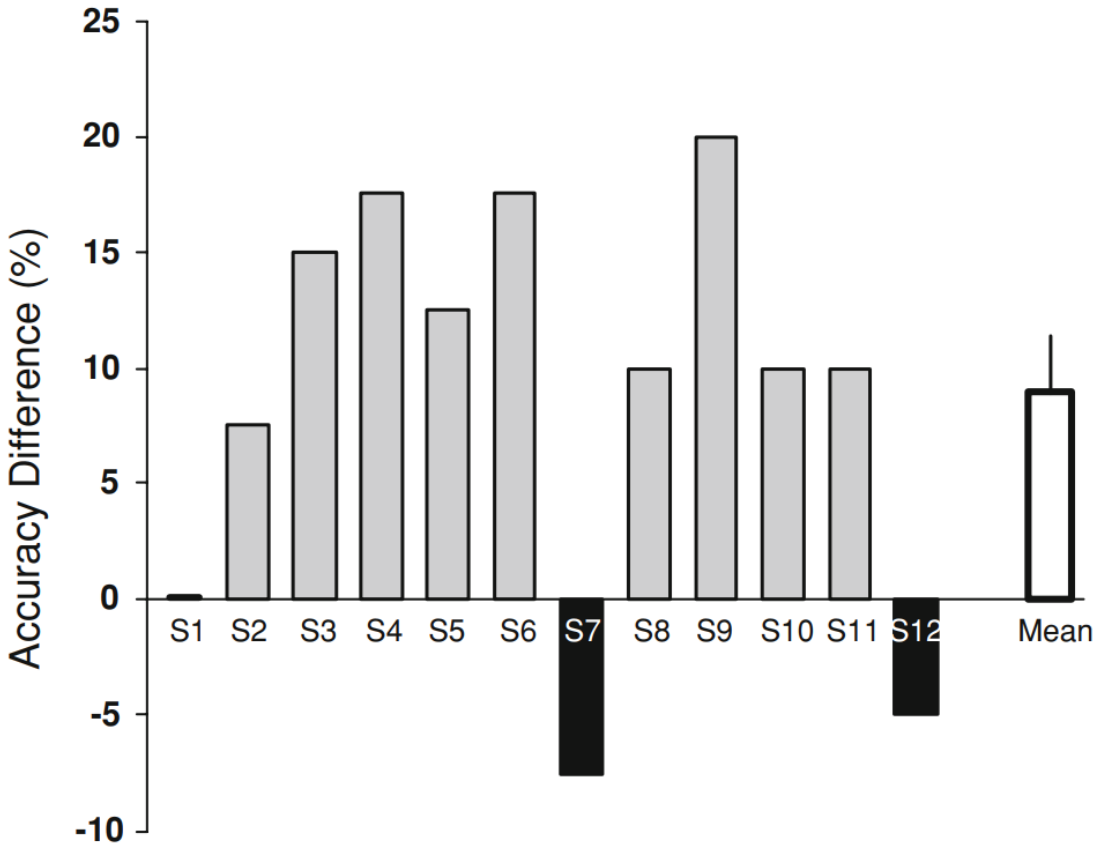
2.6. The Attentional Blink

The experiment in chapter 6 explores the use of the Fringe-P3 method with the attentional blink as a measure to detect concealed information instead of the P3. The attentional blink (Raymond et al., 1992) is a phenomenon where the detection rate of a second critical stimulus in an RSVP stream is significantly lower when it is presented 200-500ms after the first critical stimulus (Raymond et al., 1992; Shapiro, Arnell, et al., 1997). The classic attentional blink paradigm presents two targets (T1 and T2) in RSVP and varies the lag between them (e.g., lag 1 is T2 presented immediately after T1, lag 5 is T2 presented as the fifth item after T1). The phenomenon has been heavily researched and been shown to occur with letter, number, word, colour, face, and other picture stimuli (Evans & Treisman, 2005; Ganis & Patnaik, 2009; Joseph et al., 1997; Ross & Jolicœur, 1999; Shapiro, Caldwell, et al., 1997).

Not only does the attentional blink occur when two targets are presented, but it has also been shown to occur when the first critical stimulus is an emotional or salient probe stimulus (Barnard et al., 2005; Ganis & Patnaik, 2009). This can be combined with the Fringe-P3 method to detect concealed information by having the first critical stimulus be the probe we suspect the participant is familiar with/concealing knowledge of and the second critical stimulus be a target. The probe stimulus would only capture attention and cause an attentional blink if it were salient to the participant. Therefore, if a participant frequently misses the target when it is presented after the probe but not after an irrelevant, then it implies they are familiar with the probe stimulus and are concealing information about it. Whereas a truly innocent participant would not be familiar with the probe, and therefore would have no difference in detection accuracy when the target is presented after probe or irrelevant. Ganis and Patnaik (2009) investigated this using the attentional blink with a CIT and face stimuli. Rather than T1 and T2, they used one salient

famous face as a probe stimulus and one non-famous target face. They were able to detect concealed information in 9 out of 12 participants by looking for a decrease in target accuracy when it was presented shortly after the probe. Figure 2.6 presents a bar graph of the accuracy differences for individual participants.

Figure 2.6. Accuracy Difference Bar Graph from Ganis and Patnaik (2009).



This bar graph was taken from Ganis and Patnaik (2009). It shows the difference in accuracy between attentional blink trials (where the probe was presented before the target) and control trials (where the probe was not presented) for individual participants. Nine of the twelve participants had positive differences, meaning they had lower target accuracy in the attentional blink trials compared to the control trials.

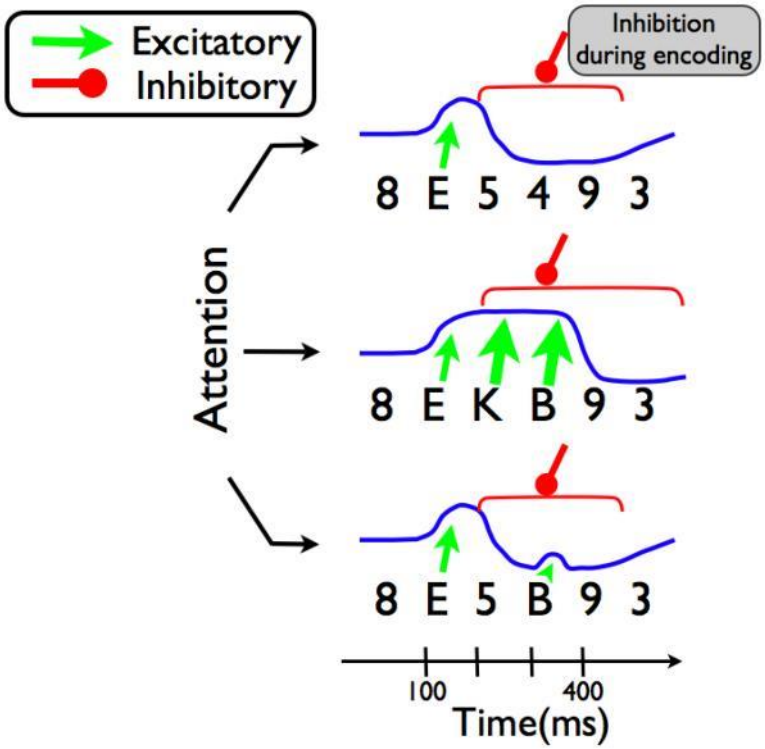
The Fringe-P3 method using EEG has already been shown to be resistant to countermeasures (Bowman et al., 2014), and the attentional blink has also been shown to be resistant to the effects of training to adapt to it (Braun et al., 1998). However, there are some circumstances in which the attentional blink may not appear.

Although the attentional blink is a very common phenomenon, there are some people who do not experience it, referred to as “nonblinkers” (Martens et al., 2006). Nonblinkers are thought to consolidate information to working memory faster than blinkers. Martens et al

(2006) compared blinkers and nonblinkers in a letter detection task while measuring their attentional blink and ERP for a P3 response. Martens et al (2006) found that nonblinkers had an earlier P3 peak in all lags, thus supporting the idea that they consolidate information to working memory faster than blinkers. This study also showed that the P3 measure can be combined with the attentional blink paradigm. This means the attentional blink could be combined with the Fringe-P3 method to provide an even more robust method of detecting the presence of concealed information.

Another exception to the attentional blink is lag 1 sparing. This occurs when T2 is presented up to 150ms after T1 and is still accurately detected, escaping the blink (Potter et al., 1998; Visser et al., 1999). One model of the attentional blink, which can also account for lag 1 sparing, is the episodic simultaneous type, serial token model (eSTST) by Wyble et al (2009). This model explains the attentional blink through competition between working memory encoding and attention allocation. This model is based on the idea that there is a window of attentional enhancement that is activated when T1 is detected and is kept open by consecutive targets, but if a new target is not presented within 200ms or a non-target stimulus is presented, then the window is deactivated and attention is briefly suppressed while the detected targets are encoded into working memory and consolidated. Any further targets that are presented once the window has closed and attention is supressed will receive no enhancement and are often missed. This means that multiple targets presented consecutively can be detected and encoded, thus accounting for lag 1 sparing, but if there is a non-target stimulus between two targets, then the attentional enhancement window closes and attention is suppressed, causing the attentional blink. Figure 2.7 presents a demonstration of this competition between attention and encoding. The theory also states that while multiple targets presented consecutively can be detected, this comes at the cost of sequential information as they compete to be encoded in parallel, resulting in them often being encoded and recalled out of order or with T2 being recalled more accurately than T1. Wyble et al suggest that the attentional blink is the result of the brain's attempt to keep targets episodically distinct, and that targets presented consecutively are encoded as one episode, but a target presented after a gap would be a separate episode. Such further episodes must wait until the first episode has finished encoding and the attentional enhancement window is ready to be opened again before it can be detected.

Figure 2.7. Competition Between Attention and Encoding from Wyble et al (2009).



This figure was taken from Wyble et al (2009) and is a schematic demonstration of the competition between attention and encoding using examples of target letters amongst distractor digits. In the first example, attention is excited by the target and then suppressed by encoding as the next stimulus is a distractor. In the second example, the first target excites attention and opens the attentional enhancement window which is then kept open by the following two consecutive targets and is only supressed by encoding when a non-target stimulus appears. This is how lag 1 sparing occurs. In the third example, attention is excited by the first target and is then suppressed by encoding as the following stimulus is a distractor. While attention is supressed, the appearance of a (single) second target following the distractor cannot boost attention enough to reopen the attentional enhancement window and be detected. This is when the attentional blink occurs.

Interestingly, while T2 following a distractor is often missed (T1, D1, T2), it has been shown that T2 can cue attention enough that a third target presented immediately after T2 (T1, D1, T2, T3) can reactivate the attentional enhancement window and be accurately recalled (Nieuwenstein et al., 2005; Olivers et al., 2007; Wyble et al., 2009). The eSTST explains this through competing inhibitory and excitatory inputs from working memory, where the encoding of T1 into working memory is trying to suppress attention, while T2 tries to excite attention but doesn't succeed enough to be detected itself, but this gives a boost to T3 which does excite attention enough to be detected and encoded.

The theory also suggests that the attentional blink is caused by targets only and not distractors, unlike other theories that claim that distractors influence the blink (Di Lollo et

al., 2005; Olivers et al., 2007; Raymond et al., 1992). This is backed up by studies that presented a blank screen/no stimulus between T1 and T2 and still found an attentional blink (Ouimet & Jolicœur, 2007; Visser, 2007), and from other research using complex and simple distractors that found that the complexity of the distractor stimuli and similarity to the target had no influence on the magnitude of the attentional blink (Grandison et al., 1997; McAuliffe & Knowlton, 2000).

2.7. Central Aims and Hypotheses

The overall aim of this thesis is to generalise the Fringe-P3 method by demonstrating its efficacy with more stimulus types, specifically famous names, email addresses, and locations. The benefits of testing famous name stimuli are two-fold. Firstly, it will show that the Fringe-P3 method can work with names other than own-name. Own-name is an especially highly salient stimulus, so it is important to show that the Fringe-P3 method is still accurate with name stimuli that are salient, but less so than own-name. Secondly, the famous name stimuli are two-part stimuli in the form of first and last-name pairs, and so will also demonstrate that the Fringe-P3 method can work with multi-part stimuli. This opens up the Fringe-P3 method to further multi-part stimuli such as street addresses and birthdays in future research. Demonstrating that the Fringe-P3 method can be used accurately with email addresses is vital for the detection of a person's online identity and is the first step towards showing that the Fringe-P3 method could be used to aid forensic investigations into cybercrime. The use of location stimuli will show that the Fringe-P3 method can detect familiarity with locations and can work with image stimuli that are more complex than greyscale famous faces.

A further aim of this thesis is to show that the Fringe-P3 method can also work using the attentional blink as a measure instead of the P3. The attentional blink is a cheaper method to perform than using EEG, and so, if successful, could be used in place of or, preferably, in combination with the P3 measure for an even more robust test. By combining the two measures, the test would still be able to detect concealed information even if one measure did not have a significant result (e.g., if the participant was a non-blinker, they should still have a P3, or if their P3 was not significant, they may still show a blink). In addition, the attentional blink experiments investigate two task types (categorisation and detection) that could affect the participant's ability to use search strategies when looking for a target in RSVP streams (e.g., looking for a turret in the left hand corner of the stimulus when looking for a specific castle image). Strategies that focus on one feature of the target might cause the participant to not see the probe if it does not also share that feature, and thus reduce

the accuracy of the test. These experiments will enable us to investigate whether task type affects the impact of search strategies on detection rates of probe stimuli using the Fringe-P3 method.

The final research chapter of this thesis covers a proposed new method of analysing ERP data to detect P3s. As the shape, amplitude, and latency of the P3 can be affected by differences in tasks and stimuli, the matched filter convolution analysis is designed to be able to better account for these differences than the AGAT method and should allow us to accurately detect P3s in experiments where the ERPs do not fit the typical P3 shape, amplitude, and latency.

The central hypotheses can be broken down into the following:

1) We hypothesise that the Fringe-P3 method will accurately detect differential ERP responses (in the form of a P3) to the probe compared to the irrelevant stimuli in chapters 3 to 5. This hypothesis can be further broken down into three further hypotheses for the three experiments using different stimulus types:
    a. For chapter 3, we expect to detect significantly different responses to the famous name probe stimuli compared to the non-famous unfamiliar name irrelevant stimuli, with only the probe stimuli generating a P3.
    b. For chapter 4, we expect to detect significantly different responses to the own email address probe stimuli compared to the unfamiliar email address irrelevant stimuli, with only the probe stimuli generating a P3.
    c. For chapter 5, we expect to detect significantly different responses to the familiar university campus photo probe stimuli compared to the unfamiliar university campus photo irrelevant stimuli, with only the probe stimuli generating a P3.
2) We predict that the Fringe-P3 method will accurately detect recognition of multi-item stimuli, in particular, the two-part first and last name pairs in chapter 3. We expect to detect differential responses to the probe famous name pairs compared to the irrelevant unfamiliar name pairs, with only the probe name pairs generating P3's.
3) We hypothesise that using the attentional blink as a behavioural measure of recognition in chapter 6 will allow us to accurately detect concealed information regarding familiarity with location stimuli (university campus photos), with only the familiar probe locations generating attentional blinks.

4) We predict that the categorisation task in chapter 6 will lead to a stronger attentional blink than the detection task.
5) We hypothesise that the matched filter convolution analysis in chapter 7 will better account for variability in P3 shape, amplitude, and latency than the AGAT, thus allowing us to detect P3s and differential responses to probe and irrelevant in participants whose ERPs do not fit the typical P3 shape and so may not have resulted in significant p values when the AGAT was used previously.

3. Detecting Concealed Knowledge of Famous Names using EEG

3.1. Introduction

This chapter consists of a re-analysis of a dataset using the Fringe-P3 method with famous name stimuli collected by Abdulmajeed Alsufyani. Part of the initial data processing (i.e., filtering and epoching) was performed by Alsufyani, the rest of the processing and all analyses in this chapter were conducted by myself (see section 1.4 for more details on previous and new uses of this data). This dataset has now been published in Cortex using my analyses, with myself and Alsufyani as co-first authors (Alsufyani et al., 2021). This chapter continues previous research into using the Fringe-P3 method with name stimuli, that began with Bowman et al's studies using own-name stimuli (2013, 2014), and will now use famous name stimuli.

As discussed in the literature review (section 2.4), the Fringe-P3 method has been shown to be 100% accurate when used with own-name stimuli and is resistant to key countermeasures (Bowman et al., 2013, 2014). Own-name stimuli, however, are a very special case of salient stimuli as they are one of the most frequently rehearsed stimuli we experience from birth. The most well-known demonstration of this is the Cocktail Party Effect, where own-names are rapidly and easily detected amongst other word stimuli (Cherry & Taylor, 1954; Moray, 1959; Wolford & Morrison, 1980; N. Wood & Cowan, 1995). Participants' responses to their own names have also been shown to be faster and stronger compared to other personally familiar or famous names (Cherry & Taylor, 1954; Mack & Rock, 1998; Yang et al., 2013). Furthermore, own-name stimuli have been shown to be resistant to the attentional blink, while other name and noun stimuli are not (Shapiro, Caldwell, et al., 1997). Therefore, it could be argued that the high detection rates and strong P3 responses to the own-name probes in Bowman et al's 2013 and 2014 papers are due to the stimuli being own-names and may not be applicable to other name stimuli. The current chapter aims to counter this argument by demonstrating that the Fringe-P3 method can accurately detect familiarity with "weaker" name stimuli that are still salient, but less so than own-names. This will be done using famous names: the names of famous people that the majority of the public would recognise, and thus would be familiar and salient, but to a lesser degree than the participant's own name.

In addition, previous Fringe-P3 research has only used single-item stimuli such as the participant's own first name. However, not all types of concealed information are single item; many, such as dates (e.g., First September), addresses (e.g., Twenty Two High Street),

and full names (e.g., Martin Jones) are multi-item. The lag 1 sparing phenomenon from attentional blink research shows that a second target presented shortly after a first target can be accurately reported, thereby showing that two stimuli presented in rapid succession can be detected, and suggests that the two may be processed together (Bowman & Wyble, 2007; Chun & Potter, 1995; Hommel & Akyürek, 2005; Simione et al., 2017; Wyble et al., 2009). This sparing effect has also been shown to occur when three or more targets are presented in rapid succession (Nieuwenstein & Potter, 2006; Olivers et al., 2007; Wyble et al., 2009). Therefore, the current chapter has the additional aim of showing that the Fringe-P3 method can detect familiarity with multi-item stimuli. In this study, the multi-item stimuli will be first and second (famous) name pairs presented sequentially in RSVP (e.g., "Barrack" followed immediately by "Obama").

In summary, this research has two main aims: 1) to show that the Fringe-P3 method can detect familiarity with weaker (famous) name stimuli than own-names, and 2) that it can detect familiarity with multi-item (first and second name pair) stimuli.

We hypothesise that there will be differential ERP responses (in the form of a P3) to the famous name probe stimuli compared to the non-famous/unfamiliar irrelevant name stimuli, and that the method will accurately detect recognition of familiar multi-item stimuli (first and last name pairs). For both of these hypotheses, we expect the probe stimuli to generate P3s while the irrelevant stimuli will not.

Recall and recognition tests will also be conducted at the end of each block of streams, to provide additional evidence on what stimuli breakthrough into conscious awareness. We hypothesise that only the salient stimuli, i.e., the famous names, will be reportable in these tests.

3.2. Method

3.2.1. Participants

Fifteen participants took part in the experiment, the same number as in the original own-names experiment (Bowman et al., 2013). Due to a technical error with the recording system, one participant was excluded, leaving fourteen participants for analysis (aged 19-24, 6 male, 8 female). All participants were students at the University of Kent, native English speakers, right-handed, had no neurological disorders, and had normal or corrected-to-normal vision. All participants gave their informed consent and were paid £10 for their participation. The University of Kent Sciences Research Ethics Advisory Group approved the experiment.

3.2.2. Stimuli and Presentation

The experiment was run in Matlab 2012 using Psychophysics Toolbox Version 3. The stimuli were 16-point, light grey, monospaced, sans-serif characters presented on a black background at the same location in the middle of the screen. All stimuli were pairs of first and second names, presented sequentially (i.e., first name followed immediately by second name), similar to Proverbio et al (2009). There were three categories of critical stimuli - probe, irrelevant, and target - and five name pairs in each category. Figure 3.1 shows an example section of an RSVP stream including a probe name pair.

The probe stimuli were the names of famous people from different fields: entertainment, sport, and politics. These were Justin Bieber, Leonardo DiCaprio, Barack Obama, David Beckham, and William Shakespeare. Due to a programming error, three participants (2, 3, and 4) were presented with Winston Churchill instead of William Shakespeare, and one participant (4) was also presented with Britney Spears instead of David Beckham. We do not believe that this error affected participants' results and further discussion of this can be found in appendix A.

The irrelevant stimuli were randomly generated names: Belia Labbe, Audrey Slater, Annie Rand, Blyth Tomayo, and Kylie Carr.

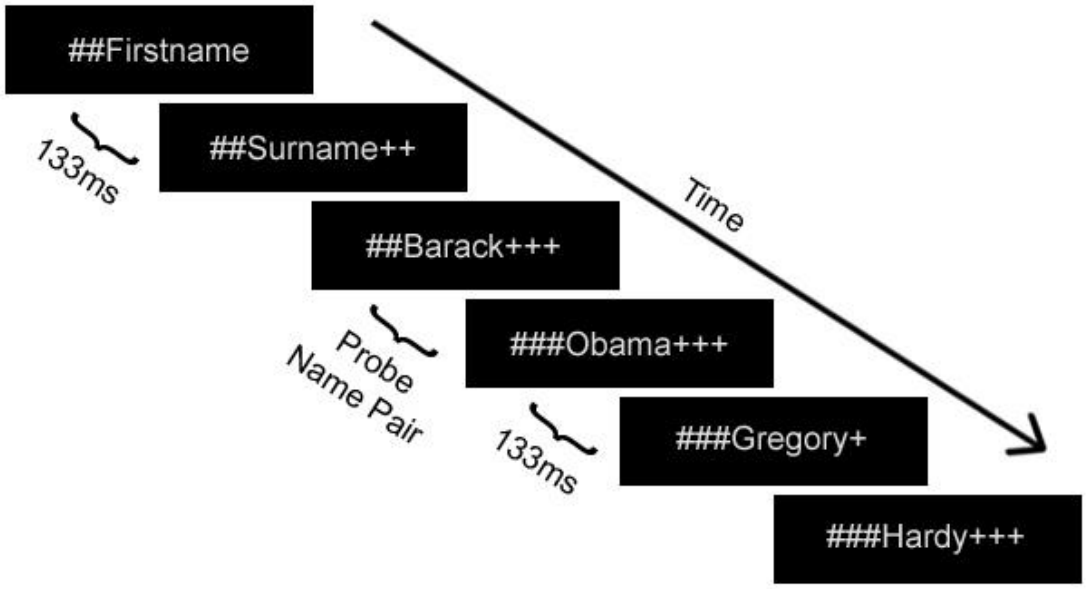
The target was a single pair of names that participants were instructed to respond to. At the start of the experiment, participants were presented with a list of five name pairs and were asked to choose their target name pair from the list, similar to Bowman et al (2014). Three participants (1, 5, and 6) chose Katherine Stevenson, two participants (9 and 12) chose Christopher Steffen, one participant (11) chose Waylon Travis, and the remaining eight chose April Harrington. No participant chose Waleed Finch.

The distractor names were randomly selected from a database of 10,000 pairs of common first and second names (50% male, 50% female). This database was created by using http://random-name-generator.info, which randomly generates first and last names based on the US Census database. The irrelevant and target names were randomly generated using the same name generator. Names that shared three or more letters in the same position as any of the critical stimuli were not used as distractors.

All names had a maximum length of 11 characters. The first letter of each name (first and second) was uppercase and all other letters were lowercase. In order to ensure that all stimuli were the same length, names that were shorter than 11 characters were padded by randomly adding "#" or "+" on either side of the name to make them 11 characters (e.g.,

"###Alice+++", "###Gregory+", "Danielle+++"), the same as in Bowman et al (2013, 2014). During the streams, distractor names were selected pseudo-randomly from the database so that no name could contain two or more letters in the same position as the name immediately before it, in order to prevent repetition.

Figure 3.1. Example Section of an RSVP Stream of Name Stimuli.



Each RSVP stream consisted of 15 pairs of first and second names, one of which was the critical pair (target, probe, or irrelevant) while the rest were distractors. The critical name pair in this example is the probe name pair Barack Obama.

All stimuli were presented in RSVP streams with an SOA of 133ms, the same as in Bowman et al (2013) and Alsufyani et al (2019). Each stream contained 15 pairs of first and second names, one of which was a critical name pair while the rest were distractors. The critical name pair appeared pseudo-randomly between the 5th and 10th positions (inclusive), in order to prevent any start or end of stream effects overlapping with the ERP response to the critical stimulus.

Each trial began with a fixation stimulus ("XXXXXXXXX") presented for 800ms to focus the participant's attention on the central stimulus presentation area. At the end of each trial, the final item, either "---------" or "=========" (randomly selected) was presented for 133ms. Participants were then asked to report what the final item was, in order to maintain participant's attention for the whole length of the stream.

The experiment was divided into 5 blocks. Each block used only one of the five probe name pairs and one of the five irrelevant name pairs, which were presented 15 times each. The same target name pair was used for every block and was presented 15 times in each block.

There were 45 trials total per block. The order in which the trials appeared within each block was randomised. Across the whole experiment, there were 75 probe trials, 75 irrelevant trials, and 75 target trials, resulting in 225 trials overall.

3.2.3. Tasks

Practice Trials

Before the main experiment, participants took part in a practice session consisting of 20 trials and a recall and recognition test. These practice trials contained only distractors and the target and did not contain any of the probe or irrelevant names nor the names of any other famous people. This practice session was simply to familiarise the participants with the task and RSVP streams and ensure that they could find the target.

Main Experiment

Participants were instructed to keep their eyes fixated on the centre of the screen and to avoid eye and body movements during each trial to reduce noise in the data. Participants were told to look for the target within the streams and to identify the final item in the streams. They were not told in advance that famous names would appear in the streams.

At the end of each stream, after being asked what the final item was, participants were asked “Did you see the target name?” Participants responded via a numeric keyboard, pressing 1 for “yes” and 2 for “no”. This task was included solely to keep participants attention on the streams.

At the end of each block, participants also completed recall and recognition tests. Recall is the more difficult test, since the recognition test will cue participants with a name pair to remember, while recall does not. The recall task must be performed before the recognition task, otherwise it would not be free recall as the recognition test would cue names to recall.

For the recall test, participants were asked on screen “What did you see?” and entered any names they saw during that block using the keyboard. There was no limit on the number of names that participants could enter and they were encouraged to enter as many names as they could.

For the recognition test, participants were presented with five name pairs and asked to rate how often they thought each name had appeared in that block. The confidence ratings were on a scale of 1 to 5, where 1 meant “Not appeared at all” and 5 meant “Appeared very often”. The five name pairs were taken from the following categories: target, probe, irrelevant, unpresented famous name (the name of a famous person that was not

presented in any streams during the experiment), and a distractor. There were five name pairs in each category – one for each block – excluding the target, of which there was only one for the whole experiment. The five distractor name pairs were randomly selected from the distractors database and had the same very small probability (0.14%) of being presented in a trial as any of the other potential distractor names in the database. The same five distractor name pairs were used in the recognition tests for every participant.

3.2.4. Recording Apparatus

We recorded EEG data using the BioSemi ActiveTwo system (BioSemi, Amsterdam, The Netherlands) and ActiView software. EEG data were recorded from the Fz, Cz, Pz, P3, P4, and Oz electrodes. Electro-oculogram data were recorded from both eyes using two HEOG electrodes on the outer canthus of each eye and two VEOG electrodes, one above and one below one eye. Data were referenced to a ground formed from a common mode sense (CMS) active electrode and driven right leg (DRL) passive electrode. Impedances were kept below 10 kOhms. The data were digitized at 2048Hz and filtered with a low-pass of 100Hz at the time of recording.

3.2.5. Analysis Procedure

The EEG data were analysed using Matlab 2016a and EEGLAB version 13.6.5b. The EEG data were re-referenced to the average of the combined mastoids and were resampled to 512Hz. The data were then filtered with a low-pass of 45Hz, a high-pass of 0.5 Hz, and a notch filter between 7 and 9Hz to filter out the SSVEP caused by the stimulus presentation rate. The data were then epoched into segments from -100ms to 1200ms, time-locked to the onset of the first name of a critical pair. Trial rejection was performed on the eye electrodes (with above 100µV and below -100µV as criteria), and on the Pz, Fz, and Cz channels (with above 50µV and below -50µV as criteria). Baseline correction was then performed on -100ms to 0ms.

After trial rejection, the maximum number of trials remaining per participant was 75 and the minimum remaining was 62 for both probe (*M* = 71.79, *SD* = 3.60) and irrelevant (*M* = 71.57, *SD* = 4.22).

3.2.6. ERP Analysis Procedure

The ERP data were analysed using the AGAT method (for a detailed description of the AGAT, see section 2.5). Data from the Pz electrode were the focus of the group and individual level analyses, as the P3 is typically greatest from that electrode (Comerchero & Polich, 1999; Polich & Kok, 1995), but data from the Fz and Cz electrodes were also analysed at the group level only. The probe and irrelevant ERPs and behavioural data were

the focus of all analyses. The target and distractor stimuli provide no information on the recognition of famous names, and so were not included in any analyses.

The window of interest for the AGAT was 100ms long and was searched for between 300ms and 1000ms, the same as in Alsufyani et al (2019) and Bowman et al (2013, 2014).

For the group level analysis, a two-tailed paired samples t-test was performed on the probe and irrelevant mean amplitudes from the windows of interest.

For the individual participants' level analysis, a Monte-Carlo permutation test was performed on the probe and irrelevant trials (as described in section 2.5) using 10,000 permutations.

3.3. Results

3.3.1. Group Level

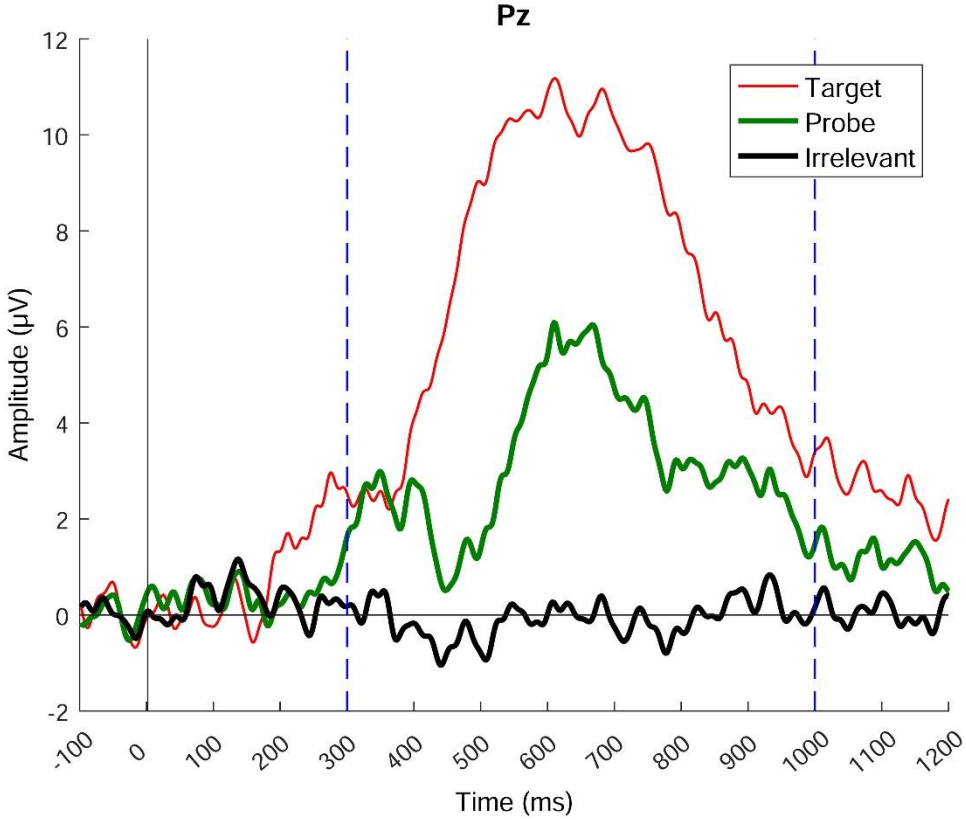
Table 3.1 shows the number of times participants answered "yes" when asked at the end of the trial if they saw the target in the target, probe, and irrelevant conditions as well as their d' scores. This shows that participants correctly said yes after most target trials and, with the exception of participant 2 (who may have misunderstood the instructions), rarely said yes (7 or less times) after probe or irrelevant trials (false alarms).

Table 3.1. The Number of Times Participants Answered "Yes" When Asked if They Saw the Target in Each Condition.

Participant	*Target*	*Probe*	*Irrelevant*	*d'*
1	62	1	2	2.995
2	69	30	19	1.854
3	52	2	3	2.339
4	70	6	6	2.906
5	65	4	5	2.666
6	56	7	5	2.069
7	73	1	1	4.149
8	60	2	3	2.676
9	48	6	1	2.037
10	74	1	4	4.050
11	48	7	2	1.913
12	65	5	2	2.789
13	66	5	4	2.730
14	64	7	3	2.552
Mean	*62.286*	*6.000*	*4.286*	*2.695*
Median	*64.500*	*5.000*	*3.000*	*2.671*
Std Dev	*8.534*	*7.296*	*4.497*	*0.701*

This table shows the number of times participants said "yes" after target trials (hits), and probe and irrelevant trials (false alarms). 10/14 participants scored at least 60 hits on target. The maximum possible score for each condition was 75.

Figure 3.2. Grand Average for All Participants from the Pz Electrode.

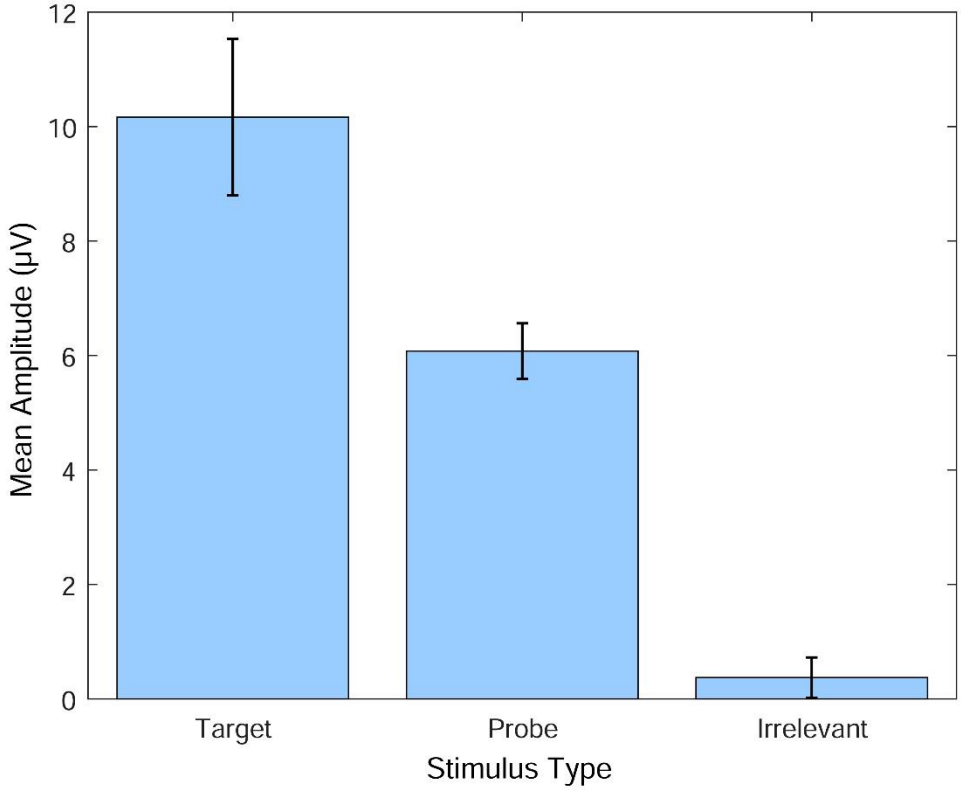


All ERPs were time locked to the onset of the first name in a pair. The vertical blue dashed lines represent the window within which the algorithm searched for the window of interest. Both the target and probe elicited a large P3, but the irrelevant did not. The target elicited the largest P3, which is expected since it was task-relevant, but provides no information on the recognition of famous names and so is of no further interest. The key comparison is between the probe and irrelevant.

Figure 3.2 presents the grand average from Pz and shows that the target and probe elicited large P3s at Pz within the 300ms to 1000ms time window, but the irrelevant did not. The mean amplitude values (presented in table 3.2) found in the AGAT analysis from Pz for probe were larger than those for irrelevant. The bar graph in figure 3.3 shows the grand mean amplitudes for target, probe, and irrelevant at Pz.

A paired samples t-test was conducted on the mean amplitude values for probe and irrelevant for all participants and found a highly significant difference between the probe (M = 6.077, SD = 1.816, Mdn = 6.497) and the irrelevant (M = 0.372, SD = 1.321, Mdn = 0.114), $t\ (13)$ = 9.668, p <0.0001, and a very large effect size, d = 3.593. This is consistent with there being a P3 response for probe but not for irrelevant, as expected.

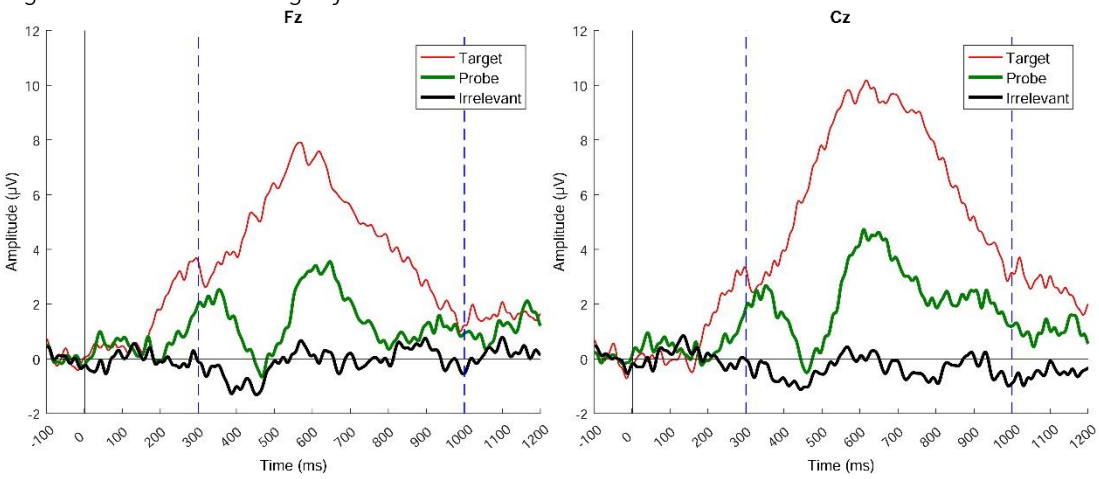
Figure 3.3. Grand Mean Amplitudes of Target, Probe, and Irrelevant from the Pz electrode.



This bar graph shows the grand average of the mean amplitudes and the standard error of the mean at Pz found within each participant's window of interest for each condition. The same window of interest used for each participant's probe and irrelevant ERPs in the AGAT analysis was used for their target ERP to generate this graph. There was a highly significant difference between probe and irrelevant at the group level, t (13) = 9.668, p <0.0001, and a very large effect size, Cohen's d = 3.593. The target was not statistically analysed.

Paired samples t-tests were also conducted on the highest mean amplitudes within the 300-1000ms window for Fz and Cz data. At Fz, there was a highly significant differences between the probe (*M* = 3.785, *SD* = 1.797, *Mdn* = 4.275) and the irrelevant (*M* = 0.469, *SD* = 0.815, *Mdn* = 0.538), *t (13)* = 7.218, *p* <0.0001, and a very large effect size, *d* = 2.376. At Cz, there was a highly significant difference between the probe (*M* = 4.874, *SD* = 1.621, *Mdn* = 5.080) and the irrelevant (*M* = 0.303, *SD* = 1.254, *Mdn* = 0.199), *t (13)* = 8.182, *p* <0.0001, and a very large effect size, *d* = 3.154. These results are consistent with there also being a P3 for the probe but not for the irrelevant at Fz and Cz, matching the result from Pz. The grand averages for Pz (figure 3.2) and for Fz and Cz (figure 3.4) show that the P3 amplitude was highest at the Pz electrode, as expected.

Figure 3.4. Grand Averages from the Fz and Cz Electrodes.



This figure shows the grand averages for Fz and Cz. The vertical blue dashed lines represent the window within which the algorithm searched for the window of interest.

Bowman et al's previous Fringe-P3 studies (2013, 2014) have found significant P3as for own-name stimuli at the Fz and Cz electrodes within a 150-300ms window. Other studies using own-name stimuli without the Fringe-P3 method have also found that own-name stimuli generated P3as (Fischler et al., 1987; Holeckova et al., 2006; Tateuchi et al., 2012). However, the current research found no significant difference between the probe and irrelevant conditions within a 150-300ms window at Fz, $t\ (13) = 2.014$, $p = 0.065$, $d = 0.766$, or at Cz, $t\ (13) = 1.537$, $p = 0.148$, $d = 0.608$, which is consistent with there not being P3as (although the p value at Fz is approaching significance). While these results do not fit with Bowman et al's studies showing P3as for own-name stimuli, it does fit with research using other familiar name stimuli that did not find a significant difference between familiar and unfamiliar names at frontal or central electrodes, and with Tacikowski and Nowicka (2010) who found that effects of familiarity were stronger at parietal electrodes. It may be that the particularly strong salience of own-name stimuli was behind the generation of P3as in the previous own-name studies, and that famous names, with weaker salience, do not generate P3as. However, the grand averages (figures 3.2 and 3.4) show a smaller positivity for the probe, and to a lesser extent the target, before the main P3 positivity, but this was later and wider than a typical P3a and peaked outside of the 150-300ms P3a analysis window. It is possible that these smaller positivities were P3as, but as most of the positivity occurred outside of the window being analysed, there was no significant difference in the analysis window. It is also worth noting that comparisons between the P3as and P3s in the present famous names study and those in Bowman et al's own-names studies (2013, 2014) are limited due to differences in the stimuli (this study used multi-item stimuli while

Bowman et al used single-item stimuli, which could explain why the current positivities were broader) and experimental tasks (this study had no explicit task relating to the famous names, while Bowman et al instructed participants to respond deceptively to their own names), which could have contributed to the differences in the waveforms. Additionally, individual differences can affect the latency of P3as and P3s (Polich, 2007, 2012), which could explain why the current positivities started later.

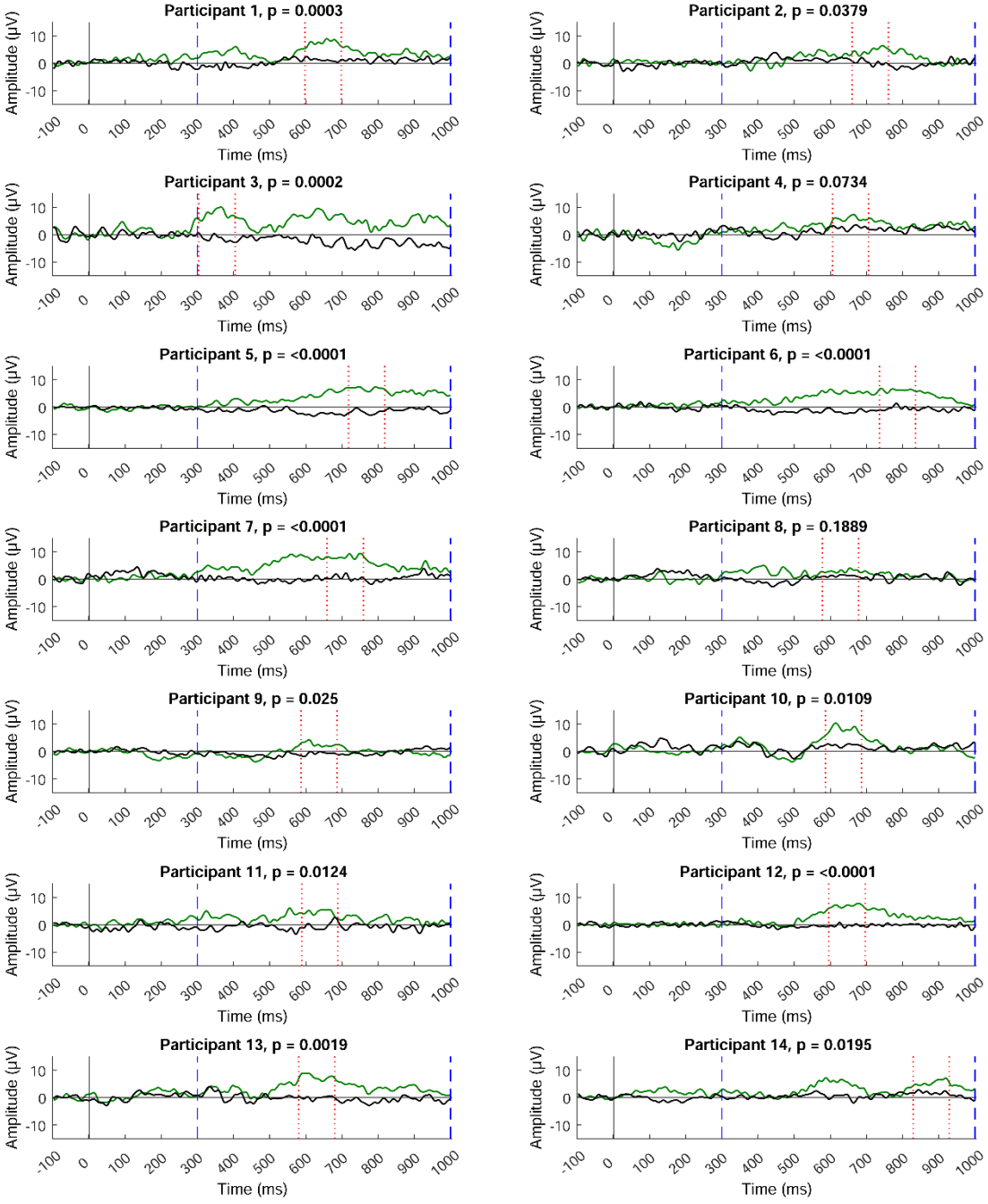
3.3.2. Individual Participants' Level

Table 3.2. Individual Participants' Mean Amplitudes and P Values from the Pz Electrode.

Participant	*Probe*	*Irrelevant*	*P Value*
1	7.578	1.229	0.0003*
2	4.638	0.270	0.0379*
3	7.628	-1.438	0.0002*
4	5.856	2.559	0.0734
5	6.801	-1.745	<0.0001*
6	6.194	-0.315	<0.0001*
7	8.184	0.641	<0.0001*
8	3.006	1.258	0.1889
9	2.613	-0.723	0.025*
10	8.083	2.271	0.0109*
11	4.518	-0.325	0.0124*
12	6.891	-0.278	<0.0001*
13	7.576	-0.042	0.0019*
14	5.515	1.846	0.0195*
Mean	*6.077*	*0.372*	*0.026*
Median	*6.497*	*0.114*	*0.006*
Std Dev	*1.816*	*1.321*	

This table contains the mean amplitudes from the AGAT analysis for probe and irrelevant as well as the p values from the individual participants level analysis. The mean and median amplitude values are larger for the probe than the irrelevant for all participants. 12/14 participants have p values below 0.05. 6/12 participants have p values below 0.001. The smallest p value that can be obtained from the 10,000 permutations is 0.0001. The asterisk marks p values that were significant at the 0.05 level.

Figure 3.5. Individual Participants' ERPs from the Pz Electrode.



This figure presents each participant's probe and irrelevant ERPs and p values from Pz. The green line represents the probe ERP and the black line represents the irrelevant ERP. The dashed vertical blue lines mark the start and end of the AGAT search window. The dotted vertical red lines represent the start and end of the window of interest with the highest mean amplitude used in the analysis for that participant.

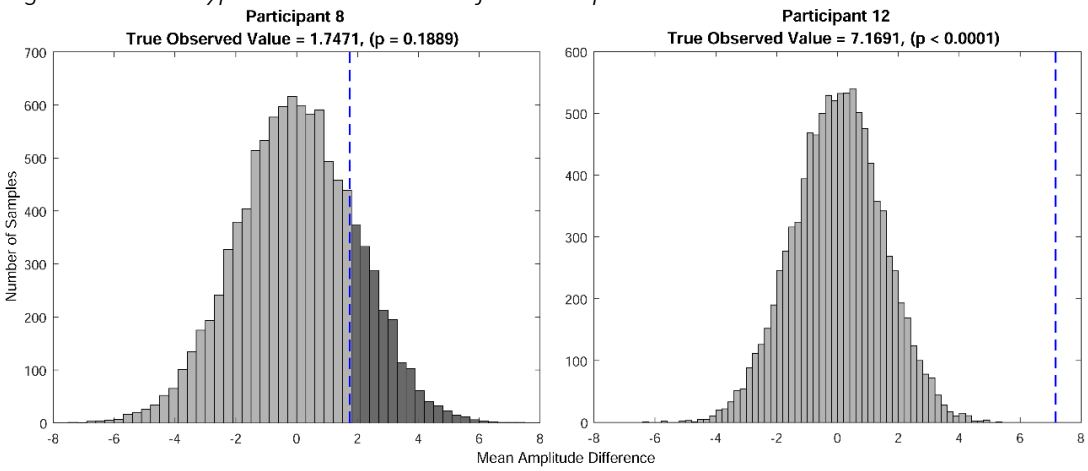
Table 3.2 presents the p values and mean amplitudes from the AGAT analysis for each participant from Pz. 12 out of 14 participants (86%) had a significant difference between the probe and irrelevant mean amplitudes at an alpha level of 0.05. Six participants (43%) had p values below 0.001 and six obtained p values between 0.001 and 0.05. The mean p value was 0.026 and the median was 0.006. The median is the most suitable measure of central tendency here, as the distribution of p values is skewed. Only Pz was analysed at the individual level as this was where we expected the P3 to be strongest.

Figure 3.5 presents the individual participants' probe and irrelevant ERPs along with their p values and windows of interest found in the AGAT analysis from the Pz electrode. Most participants have positive peaks in their probe ERP within the 300-1000ms window that are not present in their irrelevant ERP. Only two participants (4 and 8) did not have significant differences between their probe and irrelevant mean amplitudes in the analysis. Participant 4 had a positivity peaking around 650ms for the probe, and the AGAT correctly selected this as the window of interest, but there was not enough difference between the probe and irrelevant mean amplitudes to reach significance (p = 0.073). This is most likely due to the high noise level in the data causing the irrelevant to be positive within that window.

Participant 8 had a positivity peaking around 410ms for the probe, but as the irrelevant was negative at the same time, that window had a smaller amplitude in the aggregated ERP, which resulted in it not being selected as the window of interest. Instead, a window centred around 624ms was chosen, which had a smaller probe positivity, resulting in a smaller difference between the probe and irrelevant and a high p value (p = 0.189). The irrelevant being negative during the earlier probe positivity is likely to be due to noise in the data.

Participant 12, on the other hand, was one of the four participants with the largest significant differences, and showed a high positive peak for probe which resulted in a high probe – irrelevant mean amplitude difference and a highly significant p value (p < 0.0001). Figure 3.6 shows the null hypothesis distributions for participants 8 and 12. It can be seen that the true observed value (probe – irrelevant mean amplitude) could not reject the null hypothesis for participant 8, resulting in the high p value. In contrast, the true observed value for participant 12 is beyond the range of surrogate values, rejecting the null hypothesis and resulting in the highly significant p value.

Figure 3.6. Null Hypothesis Distributions for Participants 8 and 12.



This figure shows the probe - irrelevant mean amplitude differences from the permutation analyses for participants 8 and 12. The dashed vertical blue lines represent the true observed value.

3.3.3. Recall and Recognition Tests

Participants' recall scores are presented in table 3.3. There were five probes and five irrelevants used in the experiment, so total recall scores are out of five. All fourteen participants recalled two or more of the probes. Thirteen of the participants (93%) recalled at least three probes. Only two participants were able to recall any irrelevants, both recalling one each. A paired samples t-test found a significant difference between probe recall scores (M = 4.214, SD = 0.975, Mdn = 4.5) and irrelevant recall scores (M = 0.143, SD = 0.363, Mdn = 0), $t\ (13)$ = 13.350, p <0.0001, and a very large effect size, d = 5.534. The recall measure was found not to be normally distributed, so a Wilcoxon's signed-rank test was performed in addition and also found a highly significant difference between probe and irrelevant recall scores, Z = -3.347, p < 0.001 (approx.).

Participants' final confidence ratings from the recognition tests are presented in table 3.4. These final confidence ratings are the means of the confidence ratings given at the end of each of the five blocks. The main comparisons were probe (M = 4, SD = 0.532, Mdn = 4) against un-presented famous name (M = 1.286, SD = 0.321, Mdn = 1.2), and irrelevant (M = 1.586, SD = 0.454, Mdn = 1.5) against distractor name (M = 1.50, SD = 0.280, Mdn = 1.5). The paired samples t-tests found a highly significant difference between the probe and unpresented famous name confidence ratings, $t\ (13)$ = 16.480, p <0.0001, and a very large effect size, d = 6.179. There was no significant difference between the irrelevant and the distractor, $t\ (13)$ = 0.763, p = 0.459, and a small effect size, d = 0.227, suggesting that

participants were able to recognise the probe names but not the irrelevant names, despite the irrelevants being presented the same number of times in the streams.

Table 3.3. Participants' Recall Test Scores for Probe and Irrelevant.

Participant	*Probe*	*Irrelevant*
1	2	0
2	5	0
3	5	0
4	4	0
5	4	1
6	5	0
7	5	0
8	4	0
9	3	1
10	3	0
11	5	0
12	5	0
13	4	0
14	5	0
Mean	*4.214*	*0.143*
Median	*4.500*	*0*
Std Dev	*0.975*	*0.363*

This table presents the total number of probe and irrelevant names recalled by each participant during the recall tests at the end of each block. The highest possible score is 5. All participants recalled at least two probes and 13/14 recalled at least three probes. Only two participants recalled one irrelevant.

Table 3.4. Participants' Confidence Ratings from the Recognition Tests.

Participant	*Probe*	*Irrelevant*	*Unpresented Famous*	*Distractor*
1	3.8	1	1	1
2	3.8	2.4	2	1.8
3	4	1.6	1	1.6
4	4.4	1.8	1.4	1.2
5	4.2	2	1.6	1.6
6	4	1.2	1	1.8
7	4.8	1.4	1.4	1.4
8	3.4	1.2	1.2	1.4
9	3.4	2.2	1.8	1.8
10	4	2	1.2	1.8
11	4	1.4	1	1.2
12	5	1	1.2	1.4
13	4.2	1.8	1.2	1.2
14	3	1.2	1	1.8
Mean	*4.000*	*1.586*	*1.286*	*1.500*
Median	*4.000*	*1.500*	*1.200*	*1.500*
Std Dev	*0.532*	*0.454*	*0.321*	*0.280*

This table shows the final confidence ratings in each category for all participants. These ratings are the means of the ratings given at the end of the five blocks, where 1 meant the participant thought the name pair did not appear at all, and 5 meant the name pair appeared very often. Each participant's highest recognition score was for probe and only probes were given scores of 3 and above.

3.4. Discussion

The first aim of the current research was to show that the Fringe-P3 method could accurately detect familiarity with weaker (famous) name stimuli that are still salient, but less so than the own-name stimuli used in Bowman et al's previous research (2013, 2014). The experiment found a highly significant difference between probe and irrelevant and a very large effect size at the group level, as expected. However, it is the individual participants' level results that are key, as the practical application of the Fringe-P3 method would be for individual participants, such as in forensic investigations. There was a significant difference between probe and irrelevant for 12 of the 14 participants (86%), with six of these having highly significant p values below 0.001. This demonstrates that the Fringe-P3 method can indeed accurately detect familiarity using weaker (famous) name stimuli and that broad familiarity with a stimulus is enough for it to be salient and generate a detectable P3. These results fit with other research that has used famous names as stimuli without RSVP and found that famous names elicited a P3 response (Pickering & Schweinberger, 2003; Proverbio et al., 2009; Schweinberger et al., 2002; Stenberg et al., 2009; Tacikowski et al., 2011; Tacikowski & Nowicka, 2010).

The recall and recognition results further support these findings, with significant differences for recall between probe and irrelevant and significant differences for recognition between probe and un-presented famous names but not for irrelevant and distractor names. This suggests that the famous probe name pairs were salient to the participants and broke through into conscious awareness, allowing them to be recalled and recognised, while the irrelevant and distractor name pairs were not salient and so could not reliably be recalled or recognised.

The second aim of this research was to show that the Fringe-P3 method could accurately detect familiarity with multi-item stimuli, in this case: first and second name pairs presented consecutively. The ERP results show that famous name pairs did differentially breakthrough into consciousness and generate P3s for individual participants, while irrelevants did not, thus showing that multi-item stimuli can be effectively used with the Fringe-P3 method. This fits with theories of the attentional blink and lag 1 sparing that propose that two consecutive targets are processed together and can both be accurately reported (Bowman & Wyble, 2007; Hommel & Akyürek, 2005; Simione et al., 2017; Wyble et al., 2009). Research has also shown that three or more consecutive targets can be accurately reported and not generate an attentional blink (Nieuwenstein & Potter, 2006; Olivers et al., 2007; Wyble et al., 2009), meaning that it may be possible to use the Fringe-

P3 method with three or more item stimuli such as addresses (e.g., Twenty Two High Street).

Additionally, the P3 generated by famous name stimuli in the current research is broader (approx. 400-1000ms) than the P3 elicited by own-name stimuli (approx. 300-600ms in Bowman et al (2013) and 300-780ms in Bowman et al (2014)). This is likely to be caused by the famous name stimuli being two-part stimuli, and thus essentially being two salient names presented rapidly. It can be seen in figure 3.2 that there is only one P3, rather than two distinct P3s for the first and then second name stimuli, which further suggests that the name pairs are processed together, resulting in a single extended P3. This fits with research into the attentional blink that found a single long P3 for lag 1 (Craston et al (2009); see also Jones et al (2020) and Pincham et al (2016) for further discussion of extended P3s).

The extended P3 also provides further justification for the use of the mean amplitude measure with the Fringe-P3, instead of the peak-to-peak method used in Bowman et al's previous research (2013, 2014). The peak-to-peak method uses the difference between the maximum positive peak and the following negative deflection within a specified window. While this was suitable for Bowman et al's own-name research, where their ERPs had a positive peak followed by a negativity, such a negativity is not present in the ERPs of the current experiment. Therefore, the mean amplitude measure is more suitable than peak-to-peak for this data and is similar to the measures Alsufyani et al used in their Fringe-P3 study (2019) and that Craston et al (2009) used to analyse an extended P3 generated by T1 and T2 at lag 1.

In both the current research and the experiment by Alsufyani et al (2019), participants were not given a task related to the fame-related stimuli nor were they told in advance that fame-related stimuli would be present. As such, these experiments allow us to investigate whether the brain still selects salient stimuli even when there is no task related to them. In Bowman et al's own-name experiments (2013, 2014), participants were given explicit tasks relating to the own-name stimuli (e.g., look for your fake own-name), thereby turning it into a detection task. While the two fame-related experiments (the famous names reported here and the famous faces in Alsufyani et al (2019)) did not involve explicit tasks for the fame-related stimuli, participants are likely to realise at some point that fame-related stimuli are present and may start to (at least implicitly) look for fame-related stimuli, thus turning it into an implicit identification task. Therefore, the results of the fame-related experiments show that the Fringe-P3 method can accurately detect when

stimuli are both intrinsically salient (but not related to an explicit task, at least initially) and being searched for in an implicit identification task.

In summary, the results of this research showed that the Fringe-P3 method can accurately detect when participants perceived name stimuli that are salient, but less so than own-names, and that the method is effective with multi-item stimuli. Significant differences were found between the famous probe names and unfamiliar irrelevant names at both the group and individual participant levels. These results further demonstrate the applicability of the Fringe-P3 method to forensic situations beyond use with only own-name and famous face stimuli to now include other familiar name stimuli. The results also open the method up to use with multi-item stimuli such as full names, dates, and street addresses. These results, especially at the individual participants' level, are particularly important in regard to practical applications to forensic situations, where it could be used to detect whether or not a suspect is concealing knowledge of a specific person. In this case, if the suspect was familiar with a particular person (guilty), their name would be salient and would differentially breakthrough into awareness and generate a detectable P3 component. If the suspect was truly not familiar with the person (innocent), their name would not be salient and would not break into conscious awareness, generate a P3 or have a significant difference to the irrelevant. This method, therefore, could be used to identify criminals based on their familiarity with certain people, such as victims and accomplices.

4. Detecting Concealed Knowledge of Email Addresses using EEG

4.1. Introduction

This chapter is an analysis of a dataset using the Fringe-P3 method with email address stimuli collected by Claire Miller and Weiyan Hwang. All data processing and analyses in this chapter were conducted by myself (see section 1.4 for more details on previous and new uses of this data). This dataset and my analyses have now been published in the European Journal of Neuroscience (Harris et al., 2020). This chapter continues the use of the Fringe-P3 method with identity related information. While previous research used real world identity information in the form of names, the current chapter is a proof of concept experiment investigating the use of the Fringe-P3 method with online identities in the form of email addresses. This chapter contains the only experiment in this thesis that involves direct deception detection, as it asks participants to pretend a fake email address is their own email address.

Forensic investigations into cybercrime often know the online identity associated with a crime (e.g., email address, username), but are unable to link that online identity to a real world suspect. The current experiment attempts to show that the Fringe-P3 method can be used to link online identities to real world suspects by detecting their familiarity with that online identity.

We envision the Fringe-P3 method being used in situations where the police already know the online identity associated with a crime and suspect that a particular real world person is the user of that online identity. The police would then use the Fringe-P3 method with the real world individual, using the online identity as the probe to see if they are familiar with it. If the real world individual is indeed familiar with the probe online identity, then it will breakthrough into conscious awareness and generate a detectable P3. The police can then infer that the real world individual has a link with the online identity. Meanwhile, if an individual was truly not familiar with the probe online identity, then it would not breakthrough into conscious awareness and generate a P3.

So far, the Fringe-P3 method has only been used with name and face stimuli, but one of its core components, RSVP, has been used successfully with a much wider variety of stimuli including letter and number stimuli (Chun & Potter, 1995; Potter et al., 2008). RSVP streams of letters have even been used with EEG as a P3-speller (Acqualagna & Blankertz,

2013; Chennu et al., 2013), where participants search for a specific letter in RSVP streams, which generates a detectable P3 when it breaks through into conscious awareness, allowing participants to spell words. The email addresses used in the current Fringe-P3 experiment consist of strings of letters and numbers, and so the target and salient probe stimuli should be detectable, breakthrough into conscious awareness, and generate P3s.

The main aim of the current experiment is to provide a proof of concept that the Fringe-P3 method can be used with online identities, beginning with email addresses. We will investigate this by testing university students with their own university email address as the online identity probe.

We hypothesise that there will be significantly different responses (in the form of a P3) to the participants' own email address probe stimuli compared to the unfamiliar email address irrelevant stimuli, with only the probe stimuli breaking through into conscious awareness and generating a P3.

4.2. Method

4.2.1. Participants

Fifteen students from the University of Birmingham took part in the experiment and gave their informed consent. Three participants were excluded due to a technical error with the recording system. One further participant was excluded for having less than half the trials remaining after trial rejection, leaving eleven participants for analysis. The eleven participants were aged 19 - 28 (M = 23.27, SD = 2.9), seven female and four male. All participants had normal or corrected-to-normal vision and no known neurological disorders. The STEM Research Ethics Committee of the University of Birmingham granted ethical approval for the experiment.

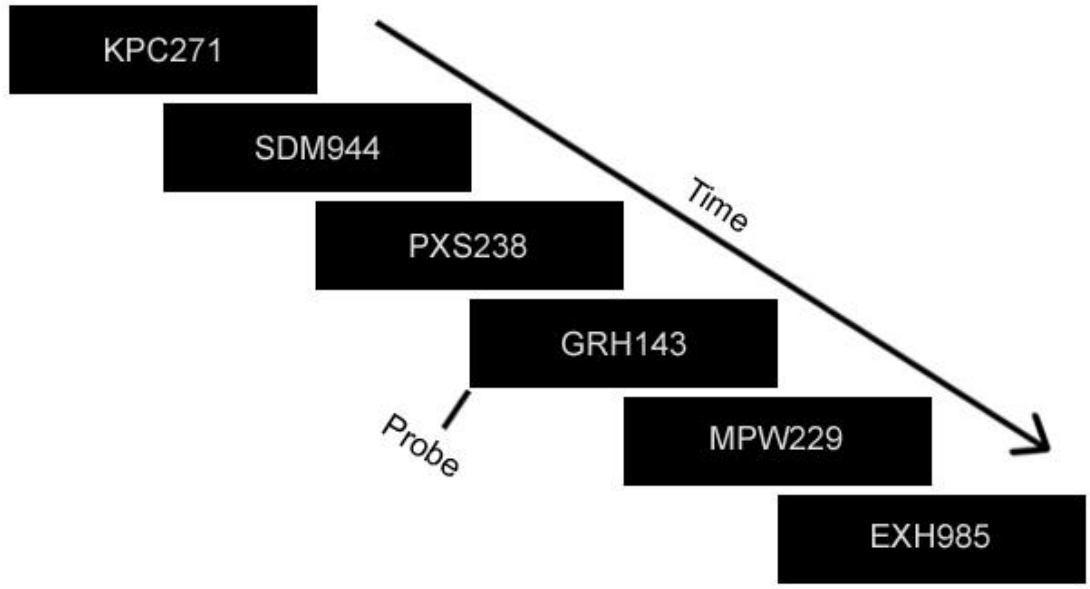
4.2.2. Stimuli and Presentation

All stimuli were email addresses using the University of Birmingham's email format. University of Birmingham email addresses are generated according to a set of rules, using three letters from the student's initials followed by three numbers (e.g., ABC123). If the student does not have three initials from their names, an X is added between their two initials (e.g., CXD456). All stimuli generated for this experiment matched this format.

There were three critical stimuli: probe, target, and irrelevant. The target stimulus was an email address generated at random that participants were to look for within the RSVP streams. Participants were tasked with pretending this target email address was their own and to look for it in the streams. The probe stimulus was the participant's real email

address, which they were not warned would appear in the streams. The irrelevant stimulus was a randomly generated unfamiliar email address that was not shown to the participants before the RSVP streams. These critical stimuli were presented amongst unfamiliar distractor stimuli. A bank of 3667 distractors was randomly generated, with the condition that they could not have more than two characters the same as any of the critical stimuli. Each distractor had the same small chance (0.38%) of appearing in a stream. Distractors and irrelevant stimuli are the same in terms of generation procedure and unfamiliarity, but each individual distractor is presented very infrequently while the irrelevant stimuli are presented as frequently as the probe and target stimuli.

Figure 4.1. Example Section of an RSVP Stream of Email Address Stimuli.



Each RSVP stream consisted of 15 email addresses, one of which was the critical stimulus (target, probe, or irrelevant) while the rest were distractors.

The stimuli were presented in streams of 15 items. One critical stimulus appeared in each stream, placed randomly between the fifth and eighth items (inclusive) of each stream. Distractors were selected pseudorandomly from the bank to fill the rest of the stream while ensuring that each distractor could not contain more than one character in the same position as the stimulus before it. Figure 4.1 shows an example section of a stream. At the start of each stream there was a fixation stimulus (“XXXXXX”) in the centre of the screen to focus the participant’s attention on the presentation area, and at the end there was a final item (“------” or “======”), which was chosen randomly. After the final item, participants were asked if they saw their email address (responding “yes” if they saw the target email

address) and which of the two possibilities the final item was. This final item question was to ensure that participants paid attention to the full length of the stream.

Streams were presented in five blocks of 36 trials. Participants were given the opportunity to take a break after each block. In each block, the critical stimuli were randomly presented 12 times each with one critical stimulus per stream. Over the five blocks, the target, probe, and irrelevant were presented 60 times each, with 180 trials in the whole experiment.

The Stimulus Onset Asynchrony (SOA) for each participant was chosen using a staircase procedure during practice trials to find the speed where the hit rate for the target was 75% and correct rejection rate was 80% for each participant. The SOAs used in the experiment were within a range of 100 – 250ms.

4.2.3. Recording Apparatus

The EEG data were recorded using the BioSemi ActiveTwo system (BioSemi, Amsterdam, The Netherlands) and ActiView software. Electroencephalographic data were recorded from Fz, Cz, and Pz. Electro-oculogram data were recorded from both eyes using two HEOG electrodes on the outer canthus of each eye and two VEOG electrodes, one above and one below the right eye. Linked mastoids were used as a reference. Impedances were kept below 10kOhm. The data were digitised as 2048Hz.

4.2.4. Analysis Procedure

The EEG data were analysed using Matlab 2016a and EEGLAB version 13.6.5b. The data were resampled to 512Hz and underwent a high-pass filter of 0.5Hz and low-pass filter of 45Hz, before being epoched into segments from -100ms to 1500ms, time-locked to the onset of the critical stimulus. Trial rejection was performed on the eye electrodes (with above 100µV and below -100µV as criteria), and on the Pz, Fz, and Cz channels (with above 50µV and below -50µV as criteria). Baseline correction was then performed from -100ms to 0ms.

After trial rejection, the maximum number of trials remaining per participant for probe and irrelevant was 60 and the minimum remaining was 46 for probe (M = 54.91, SD = 5.09) and 40 for irrelevant (M = 53.91, SD = 6.49).

4.2.5. ERP Analysis Procedure

The main analysis was the AGAT (for a detailed description, see section 2.5). Data from the Pz electrode were the focus of the group and individual level analyses, as we expect the P3 to be strongest there (Comerchero & Polich, 1999; Polich & Kok, 1995), but data from the Fz and Cz electrodes were also analysed at the group level only. No behavioural data was

provided, therefore all analyses used the ERP data only. The probe and irrelevant ERPs were the focus of all analyses. The target and distractor stimuli provide no information on the recognition of a participant's own email address, and so were not used in any analyses.

The search algorithm used a window of 100ms and searched between 250ms and 1000ms to find the window of interest. The 250ms start point is 50ms earlier than in the famous names chapter and was chosen based on the conclusions of chapter 7, which suggested 250ms as the start point for the AGAT to allow for individual differences around the start of the P3 response (typically around 300ms). The search algorithm selects the window of interest from the aggregated ERP, so this earlier start point will not be biased towards probe or irrelevant.

For the group level analysis, two-tailed paired samples t-tests were performed on the probe and irrelevant mean amplitudes from the windows of interest.

For the individual level analysis, a Monte Carlo permutation test was performed on the probe and irrelevant trials (as described in section 2.5) using 10,000 permutations.

4.3. Results

4.3.1. Group Level

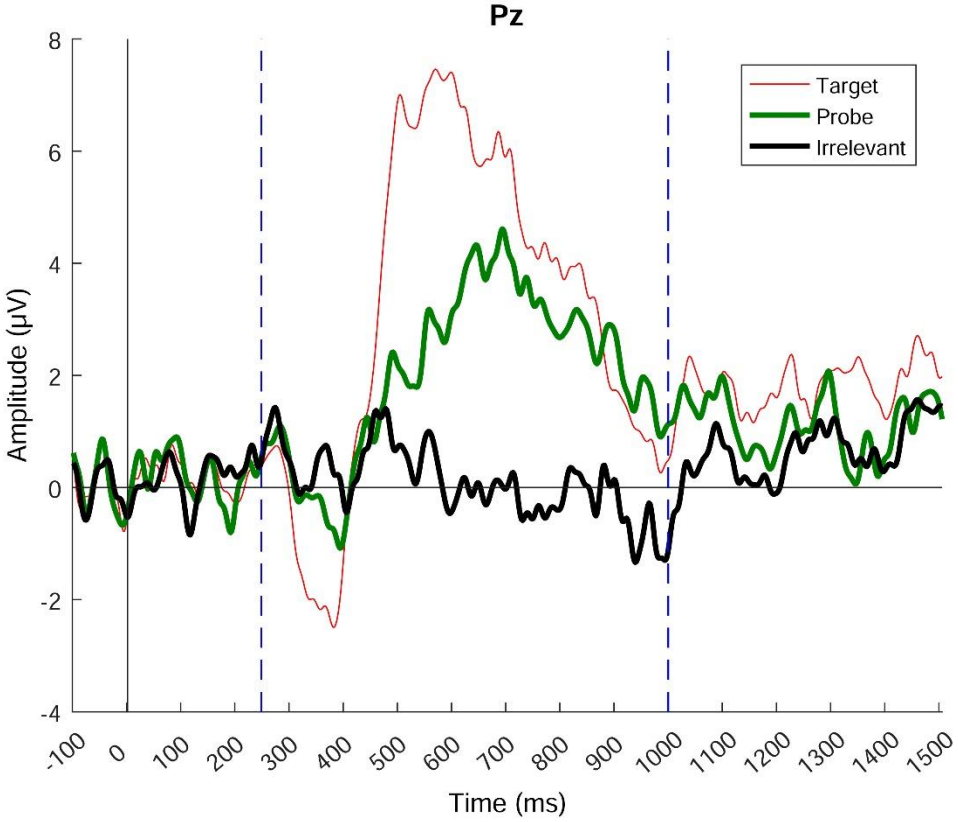
A grand average of target, probe, and irrelevant from the Pz electrode is presented in figure 4.2, where it can be seen that there is a clear P3 response for the probe condition that is not present for the irrelevant condition. Table 4.1 presents the individual mean amplitudes from Pz for probe and irrelevant as well as the overall mean, median, and standard deviation.

A paired samples t-test was performed on the mean amplitude data for all participants from Pz. There was a significant difference between probe (M = 5.615, SD = 3.293, Mdn = 4.792) and irrelevant (M = 0.703, SD = 1.744, Mdn = 0.509) at Pz, $t(10) = 4.246$, $p = 0.0017$, and a very large effect size, $d = 1.864$. This is consistent with there being a P3 response for probe but not for irrelevant, as expected.

Paired samples t-tests were also performed on the mean amplitude data from Cz and Fz and found a significant difference between probe (M = 5.440, SD = 3.372, Mdn = 4.120) and irrelevant (M = 1.119, SD = 1.424, Mdn = 1.204) at Cz, $t(10) = 3.835$, $p = 0.0033$, $d = 1.669$, and between probe (M = 4.092, SD = 2.041, Mdn = 3.710) and irrelevant (M = 1.208, SD = 1.165, Mdn = 1.058) at Fz, $t(10) = 3.504$, $p = 0.0057$, $d = 1.736$. These results are consistent

with there also being a P3 response for the probe but not the irrelevant at Cz and Fz, matching the result from Pz.

Figure 4.2. Grand Average for All Participants from the Pz Electrode.

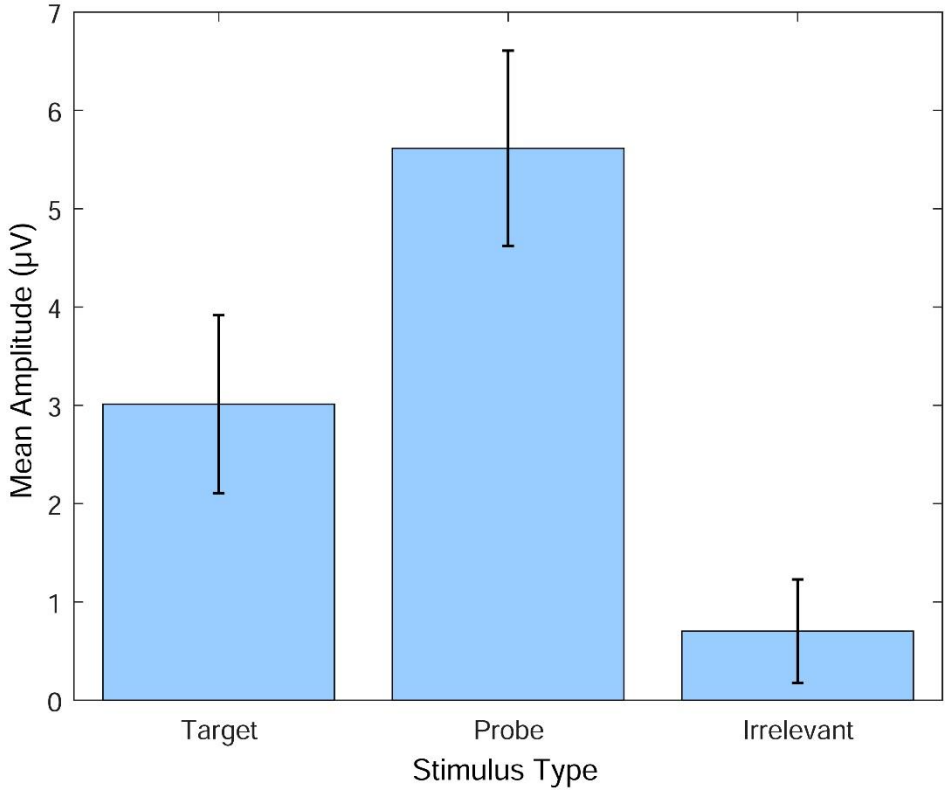


The vertical blue dashed lines represent the window within which the algorithm searched for the window of interest. A clear P3 positivity can be seen for the target and probe that is not present for irrelevant. The target elicited the largest P3, which is expected since it was task-relevant, but provides no information on the recognition of a participant's own email address and so is of no further interest. The key comparison is between the probe and irrelevant.

Figure 4.3 presents a bar graph of the grand mean amplitudes for the target, probe, and irrelevant from the Pz electrode. It can be seen that the probe grand mean amplitude was significantly larger than the irrelevant grand mean amplitude within the window of interest. The target grand mean amplitude is smaller than the probe grand mean amplitude despite having a larger peak in the grand average ERP (figure 4.2). This is because, to create this bar graph, the target mean amplitude for each participant was calculated using the same window of interest that was selected from the aggregated ERP in the AGAT analysis for the probe and irrelevant. For most participants, the target ERP peaked outside of this window of interest. See appendix B for further explanation. As the target provides no

information on the recognition of a participant's own email address, it was not involved in the main analyses and the target grand mean amplitude does not affect our conclusions.

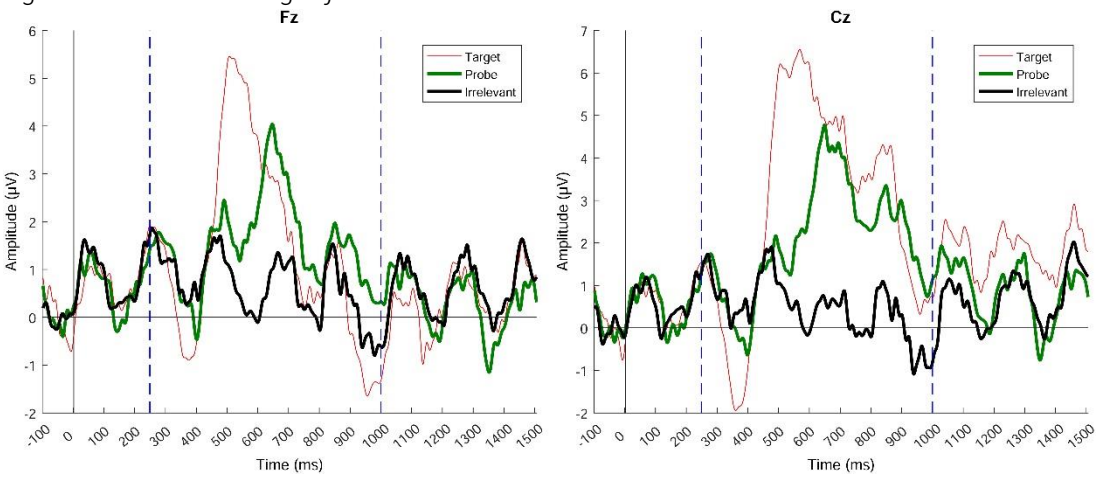
Figure 4.3. Grand Mean Amplitudes of Target, Probe, and Irrelevant from the Pz electrode.



This bar graph shows the grand average of the mean amplitudes and the standard error of the mean at Pz found within each participant's window of interest for each condition. The same window of interest used for each participant's probe and irrelevant ERPs in the AGAT analysis was used for their target ERP to generate this graph. There was a significant difference between probe and irrelevant at the group level, t (10) = 4.246, p = 0.0017, and a very large effect size, Cohen's d = 1.864. The target was not statistically analysed.

Previous Fringe-P3 studies using own-name stimuli (Bowman et al., 2013, 2014) have found significant P3a's at Fz and Cz within a 150-300ms window. No significant difference was found between the probe and irrelevant during this window for the current study at Fz, *t (10)* = -0.213, *p* = 0.836, *d* = -0.068, or Cz, *t (10)* = -0.207, *p* = 0.840, *d* = -0.069. As P3as were also not found when using famous name stimuli, it seems likely that the uniquely strong salience of own-name stimuli resulted in the P3as found in the previous own-name studies and that other stimuli that are salient, but less so than own-names, such as email addresses and, perhaps, famous names, do not generate P3as. Figure 4.4 presents grand average ERPs for Fz and Cz.

Figure 4.4. Grand Averages from the Fz and Cz Electrodes.



The vertical blue dashed lines represent the window within which the algorithm searched for the window of interest. Similar to the Pz electrode, a clear P3 positivity can be seen for the target and probe at both Fz and Cz that is not present for irrelevant.

4.3.2. Individual Level

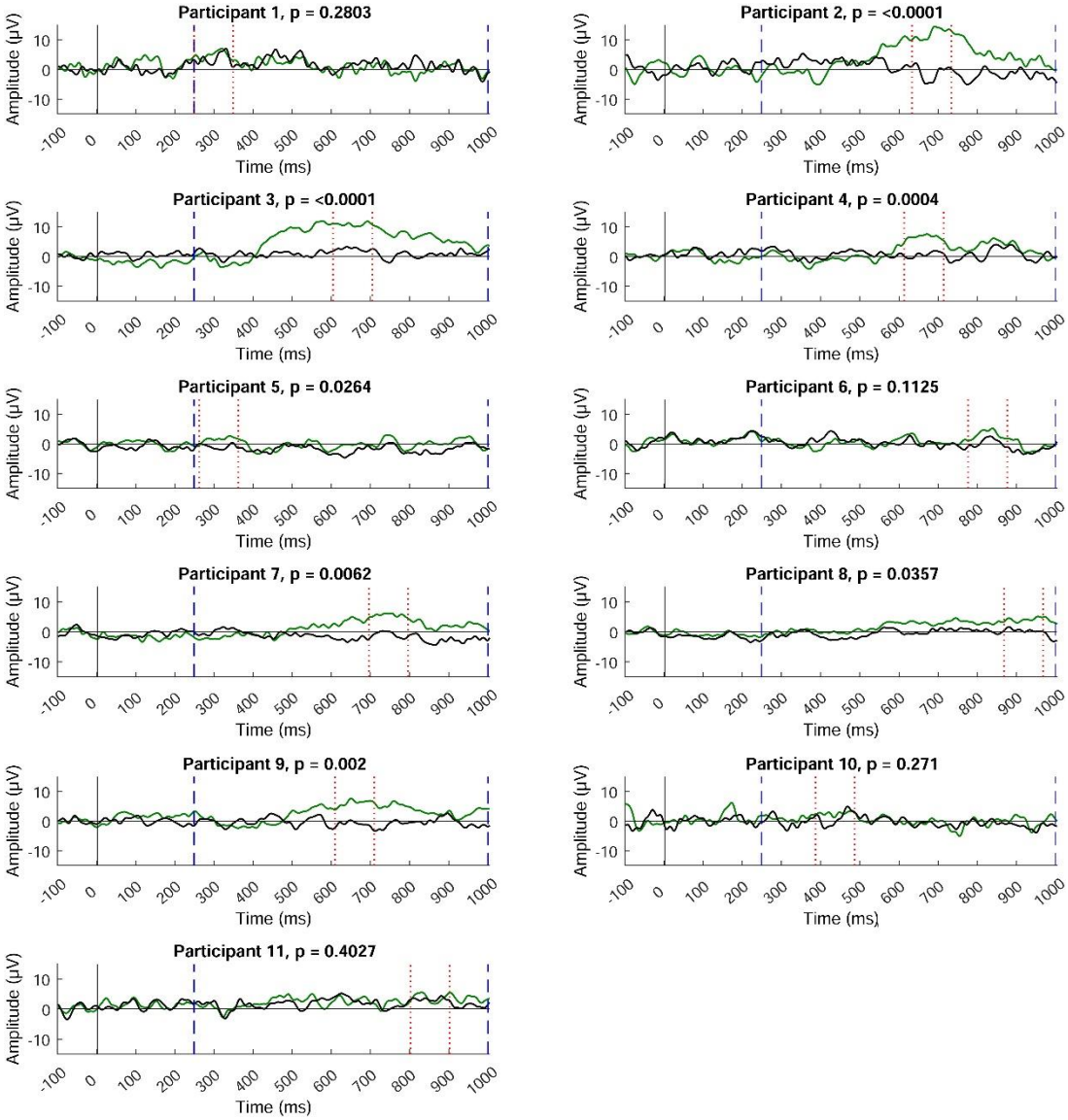
Figure 4.5 shows the ERPs for each individual participant and their window of interest from the AGAT analysis. Table 4.1 shows the mean amplitudes and p values for each participant obtained from the permutation tests on the probe and irrelevant data from Pz. There were seven out of eleven participants (64%) with a significant difference between the probe and irrelevant at an alpha of 0.05, with three (27%) of these having highly significant p values of less than 0.001. The mean p value was 0.103 and the median p value was 0.026. The median is the more appropriate measure of central tendency, rather than the mean, due to the skewed distribution of p values. Only Pz was analysed at the individual level as this was where we expected the P3 to be strongest.

Table 4.1. Individual Participants' Mean Amplitudes and P Values from the Pz Electrode.

Participant	*Probe*	*Irrelevant*	*P Value*
1	4.792	3.537	0.2803
2	12.197	-1.259	< 0.0001*
3	10.842	2.551	< 0.0001*
4	6.657	0.722	0.0004*
5	1.681	-0.999	0.0264*
6	3.199	0.509	0.1125
7	5.492	-0.919	0.0062*
8	4.141	0.481	0.0357*
9	6.287	-1.045	0.002*
10	2.623	0.926	0.271
11	3.853	3.234	0.4027
Mean	*5.615*	*0.703*	*0.103*
Median	*4.792*	*0.509*	*0.026*
Std Dev	*3.293*	*1.744*	*0.145*

This table shows the probe and irrelevant mean amplitudes from the window of interest and the p values for each participant. The mean amplitude values are larger for the probe than the irrelevant for all participants. 7 of the 11 participants (64%) had significant p values, with three of these participants having p values smaller than 0.001. The smallest p value that can be obtained from the 10,000 permutations is 0.0001. The asterisk marks p values that were significant at the 0.05 level.

Figure 4.5. Individual Participants' ERPs from the Pz Electrode.



The green ERPs are the probe and the black ERPs are the irrelevant. The vertical blue dashed lines represent the window within which the algorithm searched for the window of interest. The vertical red dotted lines represent the window of interest selected for that participant. For participants that had significant p values, a positivity in the probe ERP can be seen that is not present in the irrelevant ERP.

4.4. Discussion

This research presented an initial investigation into the use of the Fringe-P3 concealed information test for online identities, to link real world identities to online identities (e.g., email addresses, Skype addresses, usernames for online forums) connected to online crimes. The main aim of this experiment was to show that the Fringe-P3 method could be used to successfully detect familiarity with a participant's real own email address. The group level analysis found a significant difference between responses to the probe and irrelevant stimuli at the group level and a clear P3 response for the participants' own (probe) email addresses but not for the unfamiliar irrelevant email addresses, as expected. These results show that the method was successful in detecting when participants saw their own email address amongst non-familiar email addresses, even when they were trying to pretend the target email address was their own.

In addition to the group level analysis, a significant difference was found between the probe and irrelevant for seven out of eleven participants (64%) in the individual level analysis. Three of these participants had highly significant differences with p values smaller than 0.001. These promising results show that the method can work for an individual participant as well as at the group level, which is vital to forensic investigations where the method would most commonly be used with individuals.

While these results are generally significant at the individual level, as well as at group level, previous Fringe-P3 research using own names (Bowman et al., 2013, 2014) and faces (Alsufyani et al., 2019) have found even stronger results at the individual level. There are a few possible reasons why four participants did not have significant differences between their probe and irrelevant in this individual level analysis. The following explanations warrant further research that may result in improvements to make the Fringe-P3 concealed information test more successful.

One potential explanation is that participants' university email addresses may not be as salient to them as expected, especially compared to own names (Bowman et al., 2013, 2014). Our own names have especially significant meaning to us, making them highly salient and stand out more easily among other stimuli (Cherry & Taylor, 1954; Mack & Rock, 1998; Moray, 1959; Wolford & Morrison, 1980; N. Wood & Cowan, 1995; Yang et al., 2013), to the point of even escaping an attentional blink when presented shortly after another critical stimulus (Shapiro et al., 1997). In contrast, University email addresses are generated according to university rules, and not chosen by the participant, so they may not

have the same level of meaning and salience to all participants. This may have contributed to why the email addresses did not stand out as significantly in the streams compared to a participant's own name.

There is also a potential perceptual issue with the email address stimuli, since they are generated following the same rules and are the same length, making them very homogenous. Additionally, due to the generation rules, several stimuli have an X in the middle, further adding to their homogeneity. When distractors are too similar to the target (e.g., same category or colour), the distractors can capture attention (Folk et al., 1992; Su et al., 2011). There needs to be heterogeneity in RSVP stimuli in order for the critical salient stimuli to stand out from the distractors. It is possible the homogeneity of the email address stimuli may have contributed to some participants not demonstrating a significant effect, compared to previous research that used more heterogeneous stimuli, such as own-names (Bowman et al., 2013, 2014), famous names (Chapter 3), and famous faces (Alsufyani et al., 2019).

Finally, there is the possibility that some participants were (whether consciously or not) using a search strategy when looking for the target, such as looking for the first letter of the target email address and only processing the rest of the stimuli if the first letter matched. This could mean they did not read the probe if it did not start with the same letter as the target, which would prevent a P3 response. This fits with the glance-look model of cognitive control (Su et al., 2011), which argues that stimuli in RSVP streams are first "glanced" at to process the broad category or meaning of the stimulus (in our example, if it begins with a specific letter). If a stimulus matches the target category, then it receives a deeper "look", where specific meaning and detail is analysed (and the rest of that email address would be read). Additionally, the glance-look theory posits that distractor stimuli that are in the same or a semantically related category as the target can capture attention, receive a deeper "look", and initiate an attentional blink, causing participants to miss a target or probe stimulus presented shortly after (Folk et al., 1992; Su et al., 2011). In the case of email addresses, this could mean that distractors that began with the same letter as the target could have captured attention, reducing the chance of the participant seeing the probe and invoking a P3 if it appeared shortly after. A future experiment should attempt to prevent this search strategy by finding a task that forces participants to read the whole email address. Chapter 6 will investigate two task types that could affect the participant's ability to use search strategies when looking for image stimuli in RSVP streams.

Research into these potential issues outlined above could lead to improvements in both the stimuli and procedures used for future online identity deception experiments, and hence could potentially result in even stronger individual participant level responses. Despite these possible issues, the current study found significant results at both the group and individual level, thus providing the intended proof of principle demonstration that this method can successfully detect recognition of a participant's own email address. This is a vital step towards linking online identities to real world identities and aiding those fighting cybercrime.

5. Detecting Concealed Knowledge of Locations using EEG

5.1. Introduction

This chapter contains an analysis of a dataset collected by myself at the University of Kent for my master's dissertation. New data processing and all new analyses were conducted for this chapter (see section 1.4 for details on previous and new uses of this data). This data set moves away from identity related information and instead tests the Fringe-P3 method with location information in the form of photographs.

The Fringe-P3 method combined with a concealed information test has the potential to show that a participant is familiar with a location, such as one where a crime occurred, and may be concealing information regarding it. Specifically, the method could be used in forensic investigations where the police already know that a specific location was involved in a crime and suspect that a particular person has been to that location and may be linked to the crime. The method would then be used to detect the suspect's familiarity with that location, thereby linking them to it. Such locations could be indoors or outdoors, but the current experiment will focus on photographs of the exteriors of buildings.

The Fringe-P3 method has already been shown to accurately detect familiarity with images of famous faces by Alsufyani et al (2019), showing that the method can work with some image stimuli. A related method to the Fringe-P3 has also been used by Rosenzweig and Bonneh (2019, 2020) to detect familiarity with face image stimuli on the fringe of awareness using involuntary eye movements. However, these studies have only used images of faces and not other types of images. Our brains are especially good at detecting faces, to the point where there are specific areas of the brain that are thought to be dedicated to perceiving and recognising faces, such as the fusiform face area and occipital face area (Iidaka, 2014; Kanwisher et al., 1997; Pitcher et al., 2011). Therefore, it is important that we test the Fringe-P3 method with location image stimuli, not only to aid forensic investigations by linking suspects to locations, but also to show that the method can detect familiarity with picture stimuli that are more complex than the greyscale famous faces used by Alsufyani et al (2019).

While the Fringe-P3 method has only been tested with one type of image stimuli (faces), there have been many studies using one of its core components, RSVP, with a variety of image stimuli, especially while researching the attentional blink. These studies have shown that picture stimuli, including people, faces, objects, vehicles, plants, and animals can be accurately detected in RSVP streams (Balas & Momsen, 2014; Ganis & Patnaik, 2009;

Junghöfer et al., 2001; Potter, 1976; Potter et al., 2014; Trippe et al., 2007; Versace et al., 2010), including when the target picture is only briefly described (e.g., "a boat" or "a smiling couple") and not shown in advance (Potter, 1976; Potter et al., 2014). Image targets can even be detected in RSVP streams with SOAs as fast as 13ms (Potter et al., 2014). Additionally, some studies have shown that picture stimuli with emotional or threatening content are even better detected than neutral pictures in RSVP streams and are better remembered in post-stream memory tests (Junghöfer et al., 2001; Potter et al., 2014; Trippe et al., 2007; Versace et al., 2010). Many crimes involve an emotional or threatening aspect, so this could mean that some crime-related picture stimuli may be even more easily detected by guilty suspects in RSVP streams.

Another core component of the current research is the concealed information test. There are some studies that have successfully used CITs with location stimuli by presenting the *name* of a location, such as "London" (Meixner & Rosenfeld, 2011; Rosenzweig & Bonneh, 2020), but there do not appear to be any that have used *pictures* of locations. There are, however, several studies that have successfully used CITs to detect familiarity with pictures of objects (Ambach et al., 2010; Cutmore et al., 2009; Matsuda et al., 2013) as well as faces (Cutmore et al., 2009; Ganis & Patnaik, 2009; Lancry-Dayan et al., 2018; Rosenzweig & Bonneh, 2019, 2020), based on ERPs, eye movements, and the attentional blink.

The location picture stimuli used in this chapter were taken from a database used by Konkle et al (2010) to investigate memory for scenes using photographs of a variety of locations. They did not use RSVP or a CIT but did show participants (slower) streams of location image stimuli. In this experiment, participants were shown 2,912 pictures of locations from 128 scene categories (e.g., golf course, amusement park). The images were shown in 10 blocks of streams with each image shown for 3s, followed by a fixation stimulus and then the next image. At the end of all of the blocks there was a 20 minute break followed by forced choice recognition tests, where participants had to choose which of two images they had previously seen during the blocks. They found that recognition memory performance was high with accuracy varying between 76% and 96% depending on how many images participants were shown from that scene category during the streams (1-64). While this study presented the images at a much slower rate than in Fringe-P3 research, it does show that participants can remember and recognise location stimuli when they have only seen them for a short amount of time amongst a large number of other location stimuli. This, combined with the previously discussed RSVP and CIT research using other picture stimuli, suggests that participants should be able to perceive and detect

location image stimuli in the Fringe-P3 concealed information test used in the current study.

The main aim of this chapter is to show that the Fringe-P3 method can accurately detect familiarity with location picture stimuli, specifically photographs of the exterior of familiar buildings from the participant's university campus. Furthermore, this chapter aims to show that the Fringe-P3 method can work with picture stimuli that are more visually complex than the greyscale face stimuli used in previous Fringe-P3 research. We will test this with University of Kent students by comparing their ERP and behavioural responses to images of buildings on their own University of Kent campus (familiar probe) compared to buildings on the University of Birmingham campus (unfamiliar irrelevant).

We hypothesise that there will be significantly different responses to the probe campus image stimuli compared to the irrelevant campus stimuli, with only the probe stimuli generating a P3 response. We also predict that only the salient stimuli (i.e., probe locations) will be reliably recognised in the end-of-block and end-of-experiment recognition tests.

5.2. Method

5.2.1. Participants

Eight students from the School of Computing at the University of Kent took part in the experiment. All participants were male, aged between 23 and 36 (M = 27.75, SD = 4.77), had normal or corrected-to-normal vision and had no neurological disorders. All participants gave their informed consent and were paid for their participation. The University of Kent Sciences Research Ethics Advisory Group approved the experiment.

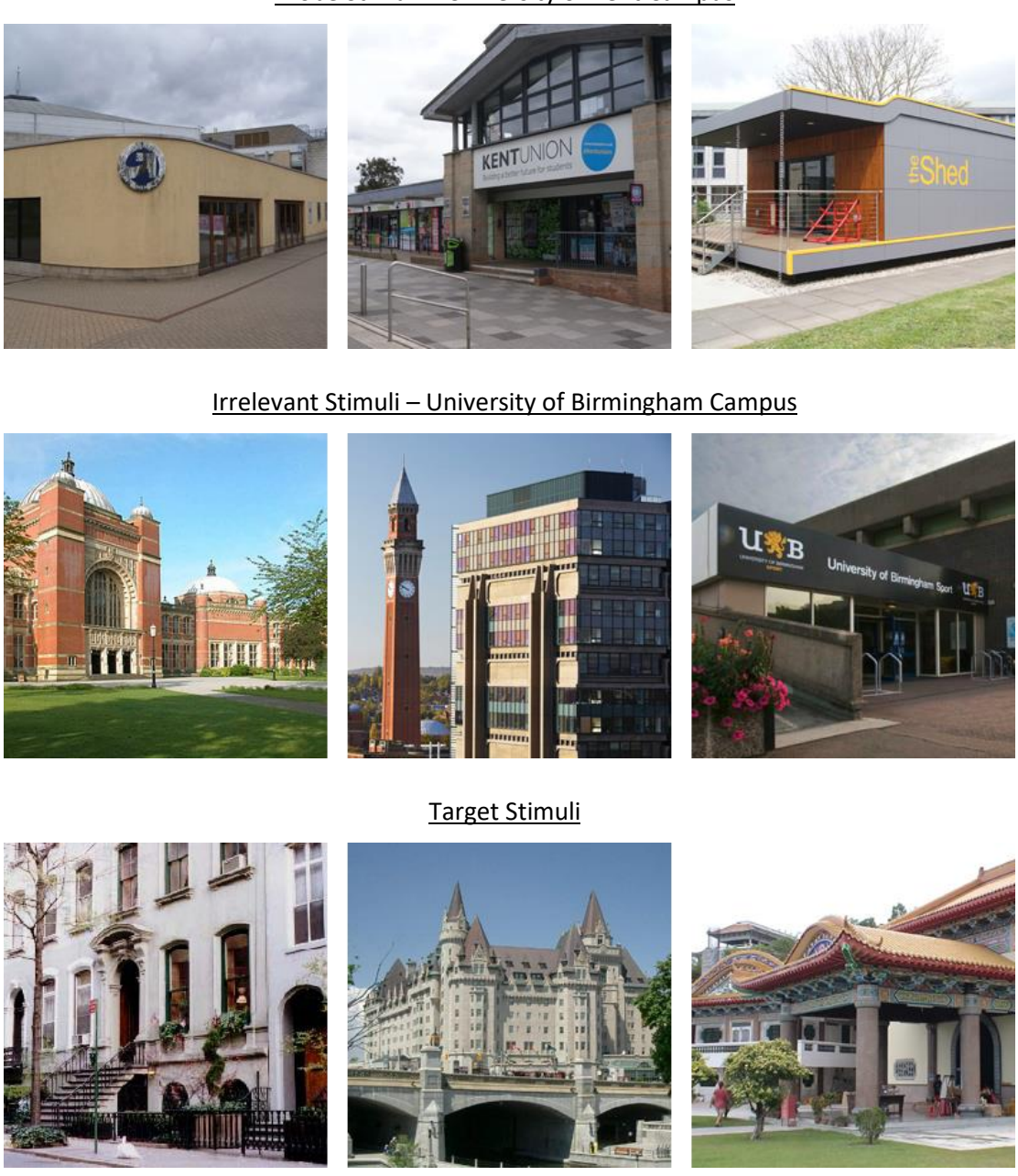
5.2.2. Stimuli and Presentation

The experiment was run in MATLAB R2014a using Psychophysics Toolbox version 3. All stimuli were 256px x 256px colour photographs of buildings presented on a black background at the same location in the middle of the screen. There were three categories of critical stimuli - probe, irrelevant, and target - and three images in each category. Figure 5.1 shows the critical stimuli used.

The three probe stimuli were photographs of buildings on the University of Kent's Canterbury Campus. A pilot study had been conducted previously to find which three University of Kent buildings were the most recognisable in RSVP streams to University of Kent students; these three most recognisable images were used as the probe stimuli in the current experiment. The three irrelevant stimuli were photographs of buildings on the

University of Birmingham campus. The target stimuli were three photographs of buildings chosen from Konkle et al's database (2010). A fourth target stimulus, also chosen from this database, was used only for the practice trials and not included in the main experiment. All of the distractors were taken from Konkle et al's database. As our probe and irrelevant stimuli were photographs of the outside of buildings, we only used distractors from Konkle et al's database that were also photographs of building exteriors. This resulted in the formation of a database of 411 distractors for the current experiment.

Figure 5.1. The Probe, Irrelevant, and Target Stimuli.



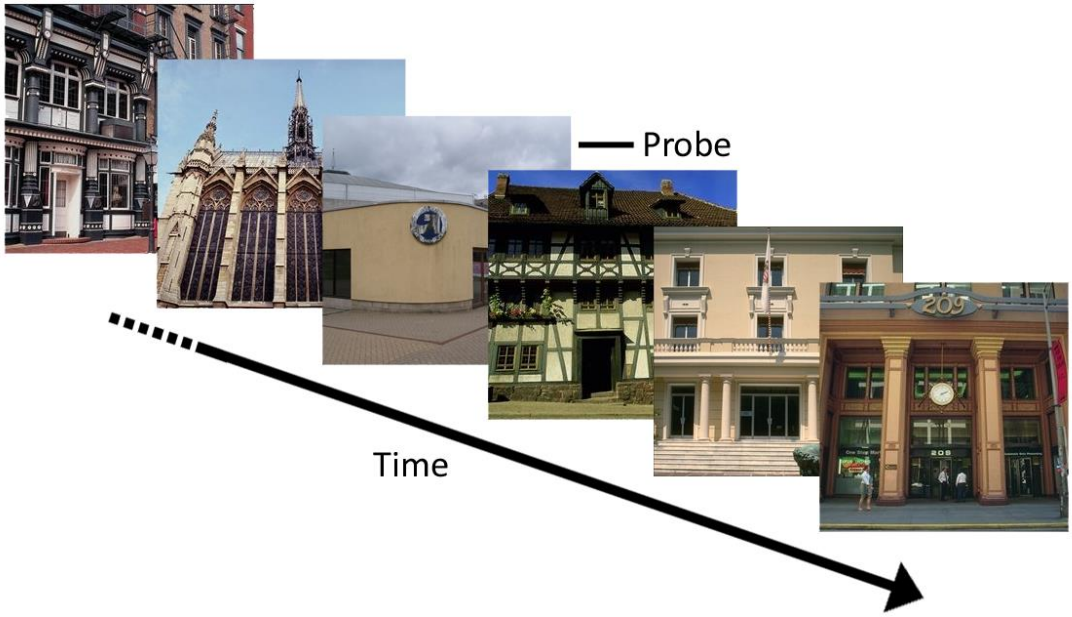
This figure presents the three probe, irrelevant, and target stimuli used in the main experiment.

All stimuli were presented in RSVP streams with an SOA of 167ms. This SOA was slower than previous Fringe-P3 research (133ms in chapter 3 and Bowman et al (2013), 100ms in Bowman et al (2014)) as it was thought that the location image stimuli would be more visually complex than name and face stimuli, so may need longer to be perceived and recognised within the RSVP streams. While picture targets have been shown to be detected in RSVP streams with faster SOAs, these targets were task relevant and would be expected to stand out more easily in the streams, whereas the current experiment's main focus is on the probe stimuli, which are not task relevant and, therefore, may not stand out as easily. Additionally, this SOA of 167ms was also used in two of Potter's RSVP experiments with picture stimuli (Potter, 1975, 1976) along with four other SOAs, one of which was faster in each experiment (113ms in 1975 and 125ms in 1976). While targets were detected by participants at the faster SOAs, their accuracy was lower than for 167ms, and the difference in accuracy between 125ms and 167ms was significant (Potter, 1976). Therefore, 167ms was chosen as a suitable SOA for the current experiment.

Each stream consisted of 18 location stimuli and started with a fixation stimulus ("XXXXXXX") presented for 800ms. Each stream contained one critical stimulus, which appeared randomly between the 5th and 9th positions in the stream (inclusive). The other positions in the streams were filled by distractors, pseudorandomly selected from our database so that each distractor could only appear once in each stream. Each stream ended with one of two final items ("-------" or "======="), which were included to ensure participants watched the entire stream. Figure 5.2 shows a section of an example stream.

There were three blocks of 75 trials. Each block was assigned one probe, one target, and one irrelevant stimulus, which were shown 25 times each within that block. The order in which the trials appeared within each block was randomised. Across the whole experiment there were 75 probe trials, 75 irrelevant trials, and 75 target trials, resulting in 225 trials overall.

Figure 5.2. Example Section of an RSVP Stream of Location Stimuli.



This is a section of an example stream, including one probe stimulus amongst distractors.

5.2.3. Tasks

Practice Trials

Before starting the main experiment, participants took part in ten practice trials to familiarise themselves with the task and viewing the streams. Six of these trials contained a target (that was not used in the main experiment) and the remaining four contained only distractors. None of the practice trials contained a probe or irrelevant stimulus or buildings from the University of Kent or University of Birmingham campuses.

Main Experiment

Participants were instructed to keep their eyes fixated on the centre of the screen and to avoid eye and body movements during each trial to reduce noise in the data. Participants were tasked with looking for the target and the final item in each stream. They were also told that "there will be images included of the university you are at" but were not given a specific task for these images.

At the end of each stream, participants were asked what the final item was and "did you see the target image?" Participants responded via a numeric keyboard, pressing 1 for "yes" and 2 for "no". Participants were shown the target stimulus when they were asked if they saw it at the end of each stream as a reminder. Both the target and final item tasks were included solely to keep participants' attention on the streams.

At the end of each block there was a recognition test on the screen consisting of five more questions. Participants were shown five images (separately) and asked “How many times did you see the following image?” next to each image. The five images were the target, probe, and irrelevant for that block, as well as a distractor and an “unpresented image” of a building on the University of Kent campus that was not shown in the streams. The same images were used for all participants. The response options were on a scale of 1 to 5, starting with 1 meaning “Never” and increasing through “Maybe once or twice”, “A few times”, “Many times”, and ending with 5 meaning “A lot”. The three distractor images were randomly selected from the distractors database and had the same small probability (4.14%) of being presented in a trial as any of the other potential distractor images in the database.

At the end of the final block, participants were also asked “Before the experiment, were you familiar with the University of Kent Campus?” to which all participants responded “yes”, and “Before the experiment, were you familiar with the University of Birmingham Campus?” to which all participants responded “no”. If there had been any participants that had responded “yes” to the University of Birmingham question, they would have been excluded from the analyses due to the likelihood of familiarity with the irrelevant stimuli. Participants were not asked about familiarity with the University of Birmingham before participating in the experiment as the question may have warned them that images of the University of Birmingham may be presented.

At the end of the experiment, participants were given a printed paper questionnaire, containing the three Kent Probe and three Birmingham Irrelevant stimuli presented in a random order amongst five more University of Kent and five more University of Birmingham images that were not presented in the experiment. Participants were asked to rate how well they felt they saw each image within the experiment on a scale of 1 (“Did not see”) to 5 (“Saw easily”).

5.2.4. Recording Apparatus

We recorded EEG data using the BioSemi ActiveTwo system and ActiView software. EEG data were recorded from the Pz, Fz, Cz, P3, P4, and Oz electrodes. Electro-oculogram data were recorded from both eyes using two HEOG electrodes on the outer canthus of each eye and two VEOG electrodes, one above and one below one eye. A common mode sense (CMS) active electrode and a driven right leg (DRL) passive electrode were used to form a ground and the linked mastoids were used as a reference. The data were digitized at 2048Hz and impedances were kept below 10 kOhms.

5.2.5. Analysis Procedure

The EEG data were analysed using MATLAB R2016a and EEGLAB v13.6.5b. The EEG data were resampled to 512Hz and then subjected to a high-pass filter of 0.5 Hz, a low-pass filter of 45 Hz, and a notch filter between 7 and 9Hz to filter out the SSVEP caused by the stimulus presentation rate. The data were then epoched into segments from -100ms to 1500ms, time-locked to the onset of the critical stimulus. Trial rejection was performed on the eye electrodes (with above 100µV and below -100µV as criteria), and on the Pz, Fz, and Cz channels (with above 50µV and below -50µV as criteria). Baseline correction was then performed on -100ms to 0ms.

After trial rejection, the maximum number of trials remaining was 75 for probe, irrelevant, and target, and the minimum was 51 for probe (M = 66.75, SD = 9.56), 49 for irrelevant (M = 67.38, SD = 10.42), and 43 for target (M = 65.88, SD = 11.05).

5.2.6. ERP Analysis Procedure

The ERP data were analysed using the AGAT method (for a detailed description of the AGAT, see section 2.5). Data from the Pz electrode were the focus of the group and individual level analyses, as the P3 is typically greatest from that electrode (Comerchero & Polich, 1999; Polich & Kok, 1995), but data from the Fz and Cz electrodes were also analysed at the group level only. The probe and irrelevant ERPs and behavioural data were the initial focus of all analyses. However, additional analyses were conducted using the target ERPs. The distractor stimuli provide no information on the recognition of familiar locations and so were not included in any analyses.

The AGAT search algorithm used a window with a width of 100ms and searched between 300ms and 1000ms to find the window of interest. For the main probe-irrelevant comparisons, the window of interest for each participant was found from their aggregated probe and irrelevant ERP. For the additional target-irrelevant comparisons, the window of interest for each participant was found from their aggregated target and irrelevant ERP.

For the group level analysis, two-tailed paired samples t-tests compared the probe and irrelevant mean amplitudes, and the target and irrelevant mean amplitudes.

For the individual level analysis, a Monte Carlo permutation test was performed on the probe and irrelevant trials (as described in section 2.5) and the target and irrelevant trials using 10,000 permutations.

5.3. Results

5.3.1. Group Level

Table 5.1 shows the number of times participants answered "yes" when asked at the end of the trial if they saw the target in the target, probe, and irrelevant conditions as well as their d' scores. This shows that participants correctly said yes (hit) after most target trials and rarely said yes (5 or less times) after probe or irrelevant trials (false alarms). Only participant 2 said yes on less than half of the target trials.

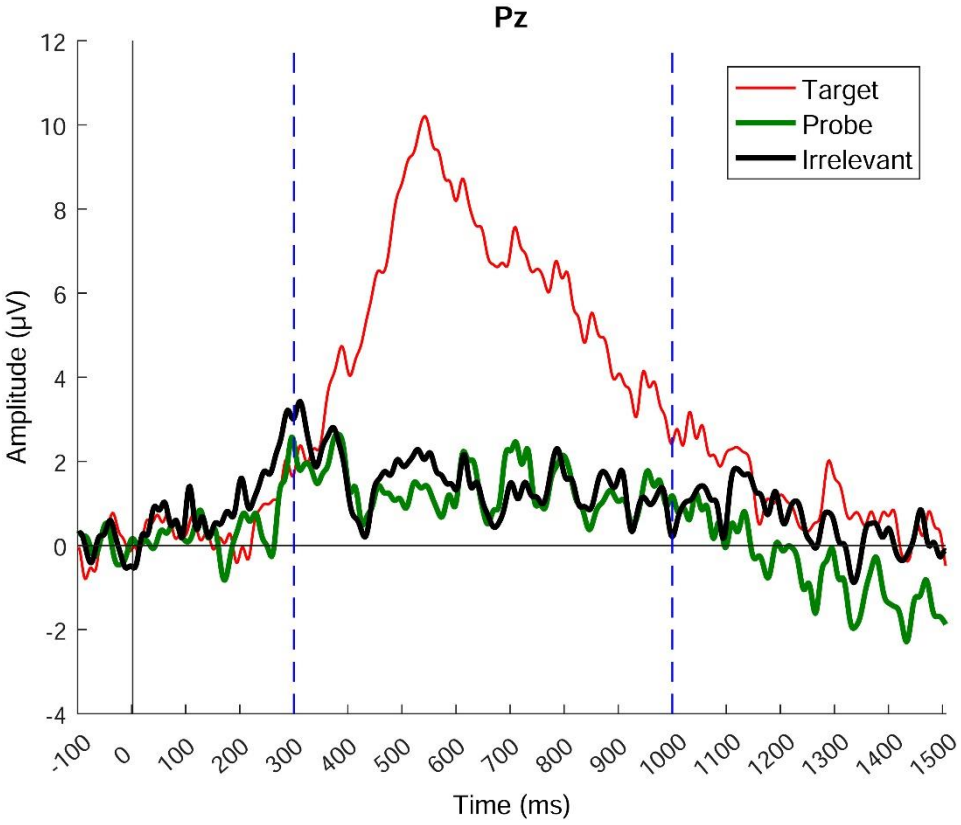
Table 5.1. The Number of Times Participants Answered "Yes" When Asked if They Saw the Target in Each Condition.

Participant	*Target*	*Probe*	*Irrelevant*	*d'*
1	72	2	1	3.804
2	32	1	0	2.290
3	62	4	2	2.692
4	48	3	2	2.192
5	73	0	1	4.407
6	52	5	3	2.119
7	70	1	0	3.976
8	49	0	5	2.228
Mean	*57.250*	*2.000*	*1.750*	*2.964*
Median	*57.000*	*1.500*	*1.500*	*2.491*
Std Dev	*14.489*	*1.852*	*1.669*	*0.941*

This table shows the number of times participants said "yes" after target trials (hits), and probe and irrelevant trials (false alarms). 4/8 participants scored hits on at least 60 target trials and 7/8 participants scored hits on over half the target trials. The maximum possible score for each condition was 75.

The grand averages show that there was a large positivity for the target at Pz (figure 5.3), Fz, and Cz (figure 5.5). There is also a smaller positivity for both probe and irrelevant at Pz, Fz, and Cz, that is clearest at Fz and Cz.

Figure 5.3. Grand Average for All Participants from the Pz Electrode.



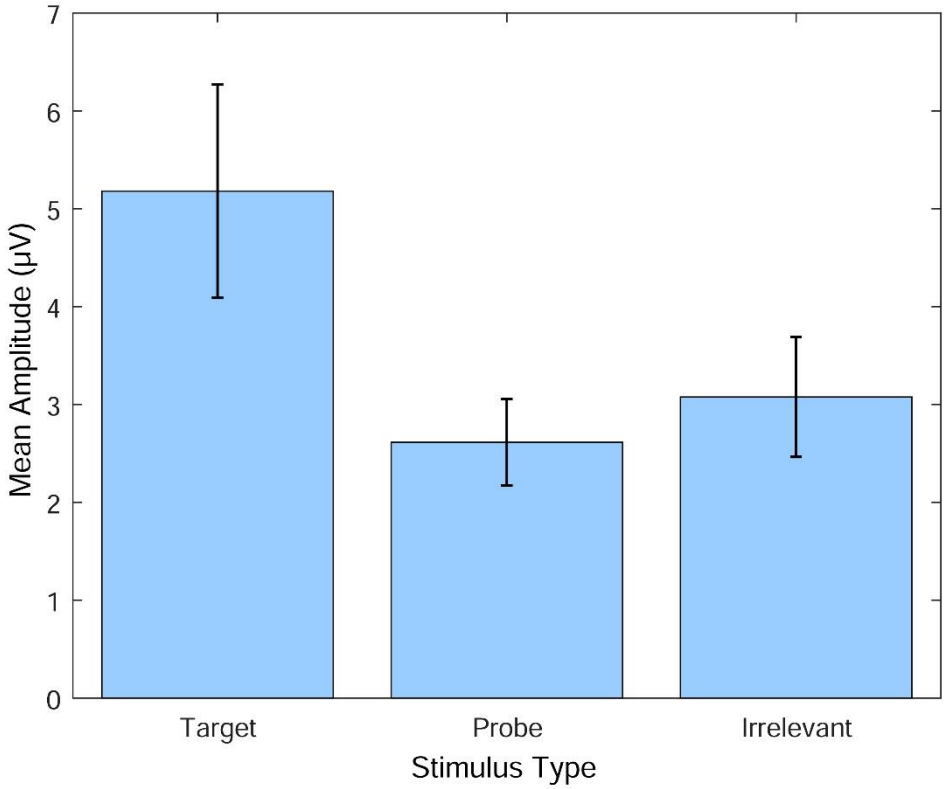
This figure presents the grand average for target, probe, and irrelevant at Pz. The blue dashed vertical lines represent the window within which the algorithm searched for the window of interest. There is a large positivity for the target and a small positivity for the probe and irrelevant.

Probe-Irrelevant Comparison

Table 5.2 presents each participant's probe and irrelevant mean amplitudes from Pz within their window of interest. A paired samples t-test was conducted on the probe and irrelevant mean amplitude data from Pz and found no significant difference between probe (M = 2.614, SD = 1.250, Mdn = 2.696) and irrelevant (M = 3.077, SD = 1.732, Mdn = 3.219), $t(7) = -1.473$, $p = 0.184$, $d = -0.307$. This result, combined with the grand average in figure 5.3, shows that there was not a significantly larger P3 positivity for the probe compared to the irrelevant at Pz.

The bar graph in figure 5.4 shows the grand mean amplitudes for target, probe, and irrelevant at Pz using the windows of interest found during the probe-irrelevant comparison AGAT analysis. The target grand mean amplitude is larger than the probe or irrelevant, which is expected since it is task relevant. The irrelevant grand mean amplitude is slightly larger than the probe grand mean amplitude, but this difference is not significant.

Figure 5.4. Grand Mean Amplitudes of Target, Probe, and Irrelevant from the Pz Electrode using the Window of Interest from the Probe-Irrelevant Comparison.

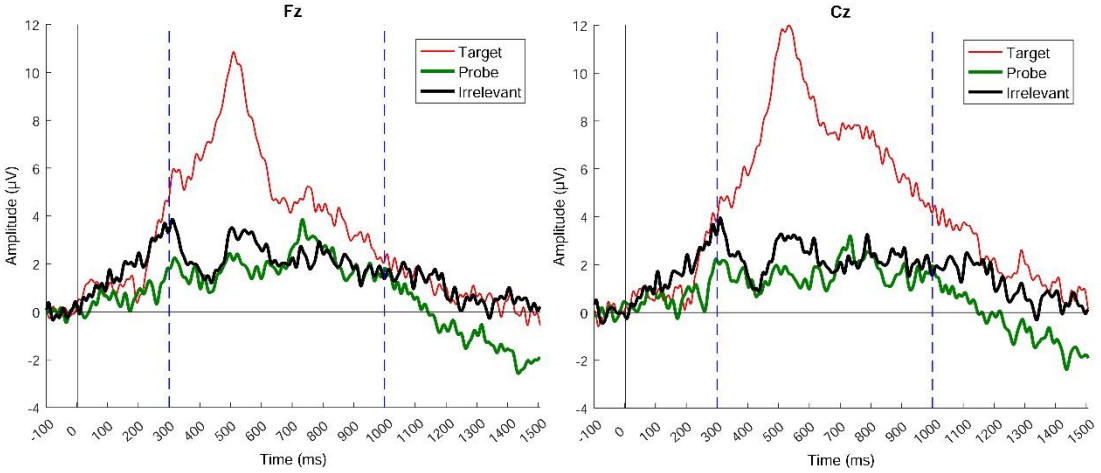


This bar graph shows the grand average of the mean amplitudes and the standard error of the mean at Pz found within each participant's window of interest for each condition. The windows of interest used for all three of these means were the same windows used in the probe-irrelevant AGAT analysis. There was no significant difference between probe and irrelevant at the group level, t(7) = -1.473, p = 0.184, d = -0.306.

The results from Pz were the main focus of the analysis, as the P3 is typically strongest from that electrode, but the mean amplitude data from Fz and Cz were also analysed. Paired samples t-tests found no significant difference between probe (*M* = 3.367, *SD* = 1.875, *Mdn* = 3.350) and irrelevant (*M* = 3.562, *SD* = 2.332, *Mdn* = 3.658) at Fz, *t*(7) = -0.273, *p* = 0.793, *d* = -0.092, and no significant difference between probe (*M* = 2.889, *SD* = 1.208, *Mdn* = 3.005) and irrelevant (*M* = 4.022, *SD* = 2.285, *Mdn* = 2.844) at Cz, *t*(7) = -1.776, *p* = 0.119, *d* = -0.620. These results are consistent with there not being a significantly larger P3 for probe compared to irrelevant at Fz or Cz, matching the result from Pz.

As there was not a significantly larger P3 for the probe within the 300-1000ms window at Pz, Fz, or Cz and the irrelevant is larger than the probe in the Fz and Cz grand averages within the 150-300ms window that is associated with P3as (figure 5.5), a P3a analysis was not conducted at Fz and Cz.

Figure 5.5. Grand Averages from the Fz and Cz Electrodes.

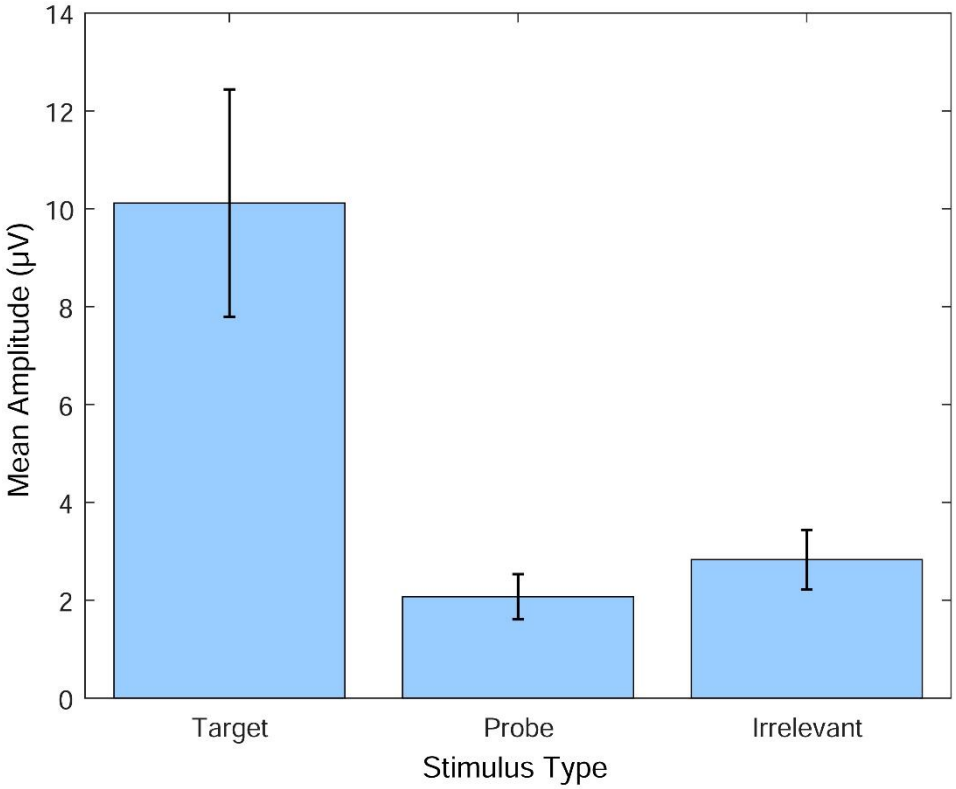


This figure presents the grand average for target, probe, and irrelevant at Fz and Cz. The blue dashed vertical lines represent the window within which the algorithm searched for the window of interest. There is a large positivity for the target and a small positivity for the probe and irrelevant.

Target-Irrelevant Comparison

The probe-irrelevant comparison results show that there was not a large positivity for the probe and there was no significant difference between probe and irrelevant at the group level. There was, however, a large positivity for the target in the Pz, Fz, and Cz grand averages (figures 5.3 and 5.5). This suggests that location stimuli can generate strong P3s and that there was an issue with our method or our specific probe or irrelevant stimuli that caused the probe ERPs to not be significantly different to the irrelevant ERPs. Therefore, the AGAT analysis was also used to compare the target and irrelevant data, using an aggregated target and irrelevant ERP (instead of the previous probe and irrelevant aggregated ERP) to find a new window of interest for each participant. Figure 5.6 contains a bar graph of the grand mean amplitudes for target, probe, and irrelevant from these windows. The target grand mean amplitude is over twice as large as the probe and irrelevant grand mean amplitudes. Table 5.3 presents the individual target and irrelevant mean amplitudes at Pz for each participant.

Figure 5.6. Grand Mean Amplitudes of Target, Probe, and Irrelevant from the Pz Electrode using the Window of Interest from the Target-Irrelevant Comparison.



This bar graph shows the grand average of the mean amplitudes and the standard error of the mean at Pz found within each participant's window of interest for each condition. The windows of interest used for all three of these means were the same windows used in the Target-Irrelevant AGAT analysis. There was a significant difference between target and irrelevant at the group level, t(7) = 2.801, p = 0.026, d = 1.517.

Paired samples t-tests were conducted on the target and irrelevant mean amplitudes and found a significant difference between target (*M* = 10.116, *SD* = 6.572, *Mdn* = 8.771) and irrelevant (*M* = 2.830, *SD* = 1.723, *Mdn* = 2.776) at Pz, *t*(7) = 2.801, *p* = 0.026, and a very large effect size, *d* = 1.517. There was also a significant difference between target (*M* = 12.086, *SD* = 7.558, *Mdn* = 10.208) and irrelevant (*M* = 3.575, *SD* = 2.045, *Mdn* = 3.971) at Cz, *t*(7) = 2.827, *p* = 0.026, and a very large effect size, *d* = 1.537. There was not, however, a significant difference between target (*M* = 10.070, *SD* = 7.868, *Mdn* = 7.387) and irrelevant (*M* = 3.739, *SD* = 2.501, *Mdn* = 3.901) at Fz, *t*(7) = 2.047, *p* = 0.080, *d* = 1.084. These results show that there was a significantly larger P3 for the target compared to the irrelevant at Pz and Cz, but not at Fz.

5.3.2. Individual Level

Probe-Irrelevant Comparison

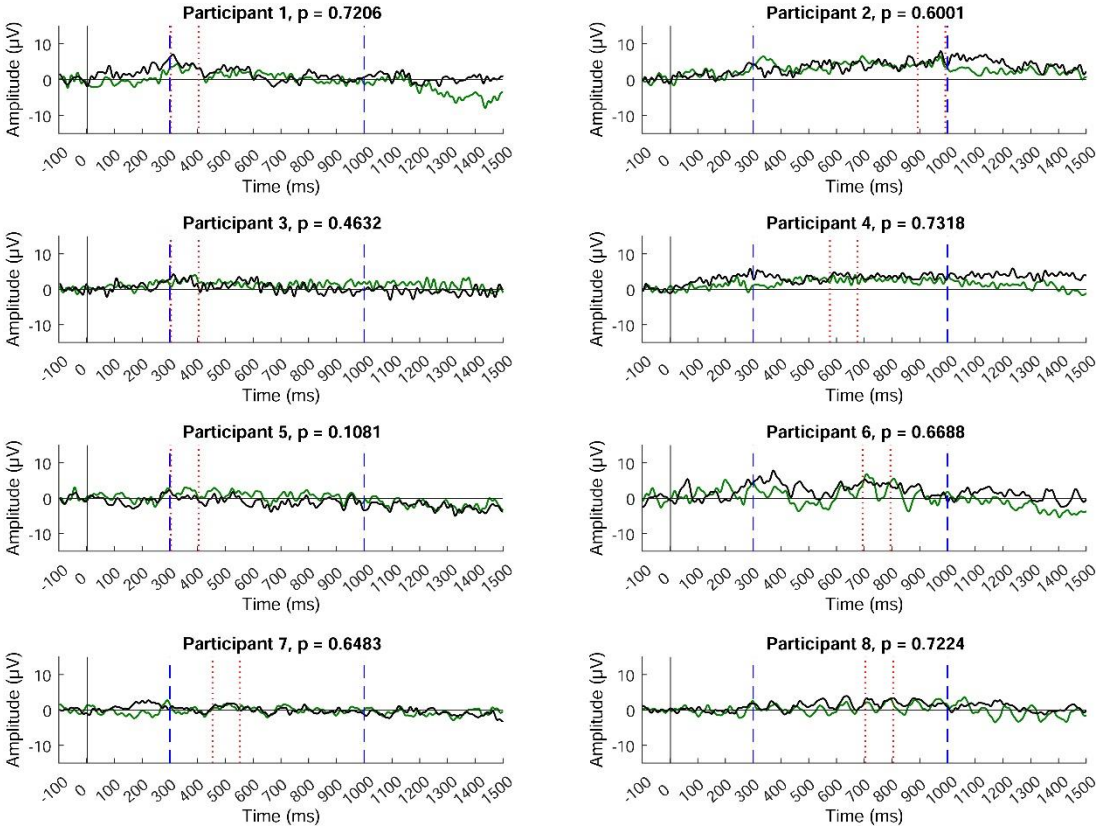
None of the participants had a significant difference between probe and irrelevant at Pz. Table 5.2 shows the mean amplitudes from the window of interest and the p value for each participant. The mean p value was 0.583 and the median p value was 0.659. Figure 5.7 presents the ERPs for each participant and the window of interest found for the probe and irrelevant comparison. Only Pz was analysed at the individual level as this was where we expected the P3 to be strongest.

Table 5.2. Individual Participants' Mean Amplitudes and P Values for Probe and Irrelevant from the Pz Electrode.

Participant	*Probe*	*Irrelevant*	*P Value*
1	3.375	4.531	0.7206
2	4.798	5.381	0.6001
3	2.671	2.517	0.4632
4	2.720	3.921	0.7318
5	1.777	0.334	0.1081
6	3.266	4.259	0.6688
7	0.849	1.308	0.6483
8	1.457	2.364	0.7224
Mean	*2.614*	*3.077*	*0.583*
Median	*2.696*	*3.219*	*0.659*
Std Dev	*1.250*	*1.732*	*0.211*

This table shows the probe and irrelevant mean amplitudes and the p values for each participant. None of the p values were significant.

Figure 5.7. Individual Participants' ERPs for Probe and Irrelevant from the Pz Electrode.



The green ERPs are the probes and the black ERPs are the irrelevants. The blue dashed vertical lines represent the window within which the algorithm searched for the window of interest. The red dotted vertical lines represent the window of interest found for each participant.

It can be seen in figure 5.7 that participant 2 had a positivity for probe between 300 and 400ms. This could potentially be a P3 but, as the irrelevant was low at this point but high later on, this 300-400ms window did not have the highest mean amplitude in the aggregated ERP, and so was not selected as the window of interest by the AGAT for analysis. Instead, a much later window was chosen where both probe and irrelevant were high and there was no significant difference. Participants 1 and 6 have positivities for probe within their selected windows of interest that also could have potentially been P3s, but the irrelevant was also high within these windows, so there was no significant difference.

Target-Irrelevant Comparison

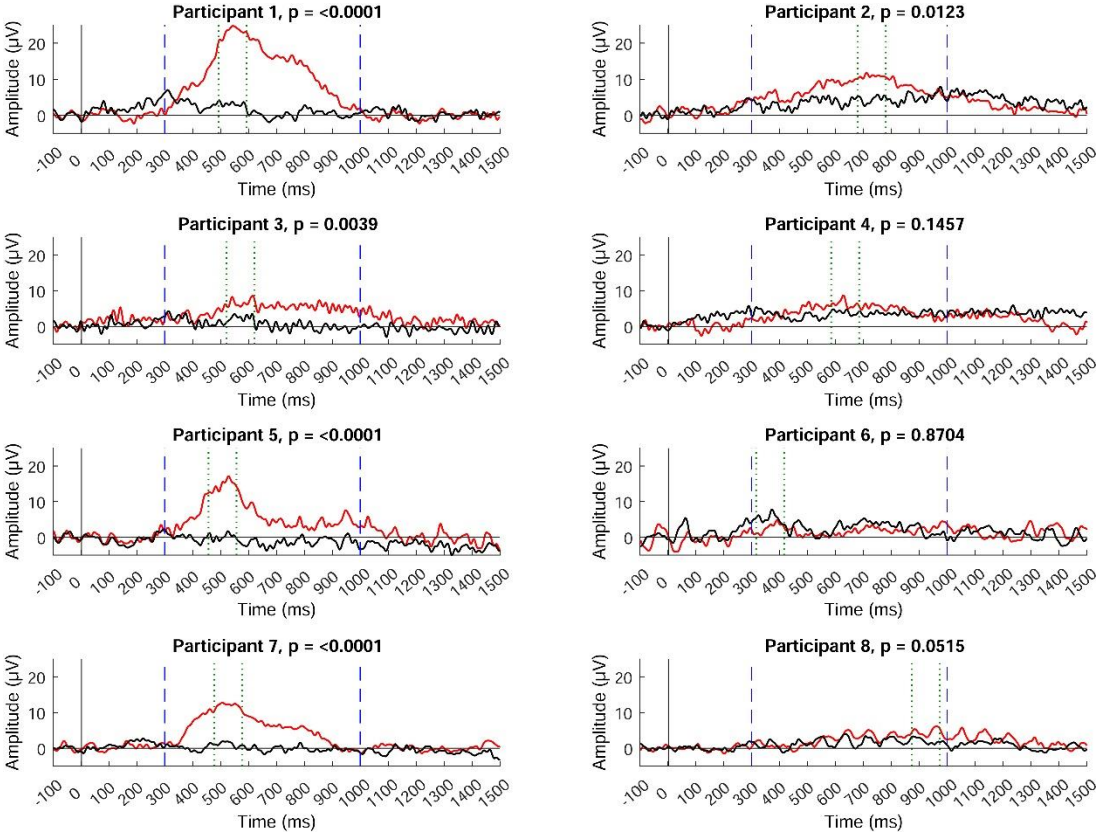
Five of the eight participants (63%) had a significant difference between target and irrelevant at Pz. Participant 8 was only just above the 0.05 significance threshold (p = 0.0515). Table 5.3 shows the mean amplitudes from the window of interest and the p value for each participant. The mean p value was 0.135 and the median p value was 0.008. The median is the more suitable measure of central tendency due to the skewed distribution of p values. Figure 5.8 shows the ERPs for each participant and the window of interest found for the target and irrelevant comparison. Only Pz was analysed at the individual level.

Table 5.3. Individual Participants' Mean Amplitudes and P Values for Target and Irrelevant from the Pz Electrode.

Participant	*Target*	*Irrelevant*	*P Value*
1	23.094	3.101	<0.0001*
2	10.839	4.390	0.0123*
3	6.702	2.450	0.0039*
4	6.347	3.944	0.1457
5	14.538	0.182	<0.0001*
6	2.742	5.363	0.8704
7	12.025	1.122	<0.0001*
8	4.640	2.086	0.0515
Mean	*10.116*	*2.830*	*0.135*
Median	*8.771*	*2.776*	*0.008*
Std Dev	*6.572*	*1.723*	*0.301*

This table shows the target and irrelevant mean amplitudes and the p values for each participant. 5/8 (63%) had significant p values. The target mean amplitude was larger than the irrelevant mean amplitude for all participants except participant 6. The asterisk indicates a significant p value.

Figure 5.8. Individual Participants' ERPs for Target and Irrelevant from the Pz Electrode.



The red ERPs are the targets and the black ERPs are the irrelevants. The blue dashed vertical lines represent the window within which the algorithm searched for the window of interest. The green dotted vertical lines represent the window of interest found for each participant. The target is visibly larger than the irrelevant within the search window for all of the five participants with significant p values, with participants 1, 5 and 7 having especially clear P3 positivities and highly significant p values.

5.3.3. Recognition Tests

Participants' confidence ratings from the end-of-block recognition tests are presented in table 5.4. These confidence ratings are the means of the ratings given in the recognition tests at the end of each of the three blocks.

The main comparisons were probe (*M* = 3.167, *SD* = 0.563, *Mdn* = 3.333) against an unpresented University of Kent image (*M* = 1.667*, SD* = 0.535, *Mdn* = 1.667), irrelevant (*M* = 2.250*, SD* = 0.496, *Mdn* = 2.167) against distractor (*M* = 3*, SD* = 0.797, *Mdn* = 3), probe against irrelevant, and irrelevant against an unpresented University of Kent image. The paired samples t-tests found a significant difference between the probe and unpresented Kent image scores, *t(7)* = 4.348, *p* = 0.003, and a very large effect size, *d* = 2.732, with the probe scores being significantly larger than the unpresented Kent image scores. There was a significant difference between irrelevant and distractor, *t(7)* = -2.908, *p* = 0.023, and a

very large effect size, *d* = -1.130, with the irrelevant scores being significantly smaller than the distractor scores. There was also a significant difference between probe and irrelevant scores, *t(7)* = 3.788, *p* = 0.007, and a very large effect size, *d* = 1.727, with the probe scores being significantly larger than the irrelevant scores. This shows that there was a significant difference in recognition even though the probe stimuli did not generate significantly larger P3s than the irrelevants. Finally, there was a significant difference between irrelevant and unpresented Kent image, *t(7)* = 2.498, p = 0.041, and a very large effect size, *d* = 1.132, with the irrelevant scores being significantly larger than the unpresented image scores.

Table 5.4. Participants' Confidence Ratings from the End-of-Block Recognition Tests.

Participant	*Probe*	*Irrelevant*	*Unpresented Kent*	*Distractor*
1	3.0	1.7	1.7	2.0
2	3.7	2.0	2.0	2.0
3	2.7	2.3	2.3	4.3
4	2.7	1.7	1.3	2.7
5	3.7	2.0	1.0	3.7
6	3.7	2.7	1.0	3.0
7	3.7	3.0	1.7	3.3
8	2.3	2.7	2.3	3.0
Mean	3.167	2.250	1.667	3.000
Median	3.333	2.167	1.667	3.000
Std Dev	0.563	0.496	0.535	0.797

This table shows the final confidence ratings in each category from the recognition tests given at the end of each block for all participants. These ratings are the means of the ratings given at the end of each block, where 1 meant the participant thought the location never appeared and 5 meant the location appeared a lot.

Table 5.5 presents participants' confidence ratings from the recognition test at the end of the experiment. This test presented all three probes and irrelevants along with five unpresented University of Kent and University of Birmingham images. The confidence ratings are the mean ratings from these categories.

The main comparisons were probes (*M = 4.167, SD = 0.797*, *Mdn = 4.167*) against unpresented University of Kent images (*M = 1.95, SD = 0.621*, *Mdn = 2.1*), irrelevants (*M = 3.083, SD = 1.035*, *Mdn = 3*) against unpresented University of Birmingham images (*M = 1.35, SD = 0.542*, *Mdn = 1.2*), and probes against irrelevants. The paired samples t-tests

found a highly significant difference between the probe and unpresented Kent image scores, $t(7)$ = 7.130, p <0.0001, and a very large effect size, d = 3.103, with the probe scores being significantly larger than the unpresented Kent image scores. There was a significant difference between the irrelevant and unpresented Birmingham image scores, $t(7)$ = 4.617, p = 0.002, and a very large effect size, d =2.098, with the irrelevant scores being significantly larger than the unpresented Birmingham image scores. There was also a highly significant difference between probe and irrelevant, $t(7)$ = 5.814, p = 0.001, and a very large effect size, d = 1.173, with the probe scores being significantly larger than the irrelevant scores, thus, again, showing that there was a significant difference in probe and irrelevant recognition even though there was not a significant difference in probe and irrelevant ERPs.

Table 5.5. Participants' Confidence Ratings from the End-of-Experiment Recognition Tests.

Participant	*Probe*	*Irrelevant*	*Unpresented Kent*	*Unpresented Birmingham*
1	3.7	3.0	1.8	1.4
2	5.0	3.7	2.4	1.4
3	4.3	3.0	2.6	2.6
4	3.3	2.3	2.4	1.0
5	5.0	3.7	1.2	1.0
6	5.0	5.0	2.6	1.4
7	3.0	1.7	1.4	1.0
8	4.0	2.3	1.2	1.0
Mean	*4.167*	*3.083*	*1.950*	*1.350*
Median	*4.167*	*3.000*	*2.100*	*1.200*
Std Dev	*0.797*	*1.035*	*0.621*	*0.542*

This table shows the final confidence ratings in each category given at the end of the experiment for all participants. These ratings are the means of the ratings given at the end of the experiment, where 1 meant the participant felt they did not see the image and 5 meant the participant felt they saw the image easily.

5.4. Discussion

There were two main aims for this experiment. Firstly, to show that the Fringe-P3 method could accurately detect familiarity with location image stimuli and secondly, as an extension to this, show that the Fringe-P3 method could work with image stimuli that are more complex than the previously used greyscale famous face stimuli (Alsufyani et al., 2019). It was predicted that there would be a large P3 for the probe stimuli but not for the irrelevant stimuli and there would be a significant difference between the probe and irrelevant mean amplitudes. The ERP data did not show a large P3 for the probe and the analyses did not find a significant difference between the probe and irrelevant at the group or individual participants' level. The ERP data did, however, show a large P3 for the target. To further investigate this, analyses were also performed on the target ERP data, comparing it to the irrelevant, to see if target location stimuli can breakthrough into awareness and generate significantly larger P3s than irrelevants. This would help us understand if it was the use of location stimuli in general that led to the lack of significant results for the probe-irrelevant comparison or whether there was an issue with our method or specific probe or irrelevant stimuli. The target analysis found a significant difference between target and irrelevant at Pz, where we expected the P3 to be strongest, and at Cz but not at Fz at the group level and significant differences for five of the eight participants at Pz at the individual level. These results show that location stimuli can breakthrough into conscious awareness, generate strong P3s, and be successfully detected by the Fringe-P3 method at both the group and, most importantly, at the individual participants' level. This also suggests that, since target locations did generate strong P3s, there was some aspect of the method or probe or irrelevant stimuli that caused the probes to not generate strong P3s that were significantly different to the irrelevants. While the probe-irrelevant results were not as expected, the target-irrelevant results are promising and mean that future experiments using the Fringe-P3 method with location stimuli, with some methodological and stimulus improvements, could result in accurate detection of familiarity with probe location image stimuli.

While there was not a significant difference between probe and irrelevant, it can be seen in the grand averages (figures 5.3 and 5.5), especially at Fz and Cz, that there is a small positivity for both probe and irrelevant. If there had been no brain response to the probe or irrelevant stimuli, then the ERPs should stay around zero, however, the presence of the positivity suggests that there may have been a P3-like response to the stimuli but this was weak, temporally-jittered (i.e., the latency varied considerably across trials, affecting the

breadth and amplitude of the positivity), and, as it occurred for both probe and irrelevant, there was no significant difference between the two. It is also worth noting that there was a negativity only for the probe following the positivity in the grand averages (figures 5.3 and 5.5) and in some individual participants' ERPs (it can be seen most clearly for participants 1 and 6 in figure 5.7). This suggests that, while the probe did not generate significantly stronger P3s than the irrelevant, there is something extra happening in the brain response for the probe that is not present for the irrelevant, which could be an effect of familiarity with the probe. This may relate to the MERMER, which is an ERP pattern consisting of a P3 positivity followed by a negativity that has been highlighted as a marker of concealed information (Farwell & Smith, 2001). It is possible that probe location image stimuli may generate a MERMER that irrelevants do not. A future experiment using the Fringe-P3 method could investigate if there is indeed a MERMER for probe location images and if this could be used to detect concealed knowledge of the probe in addition to or instead of the standard P3 AGAT analysis in Fringe-P3 experiments.

The most likely reason for why the probe stimuli did not generate large P3s may be due to participants using search strategies when looking for the target. One such search strategy could be to look at a particular feature of the target image in a specific location, for example, the turrets of the castle in the second target image or the tree in the bottom left of the third target image (figure 5.1). By narrowing focus to a specific feature in a specific location, the participant will not be viewing the entire image space, meaning that stimuli that do not share that feature in that location, such as the probe stimuli, will be less likely to capture attention, breakthrough into awareness, and generate a P3. It is also possible that the feature being searched for may be shared with an irrelevant or distractor, which could mean these might breakthrough into awareness while the probe does not. For example, if the participant did indeed look for turrets in the second target image, they may also detect the second irrelevant image, as this also contains a turret-like feature in a nearby location. One potential way to counter search strategies would be to have the target be a category of target images rather than one specific image (e.g., "castles" or "cafés"). This would force participants to view the whole image space rather than look for a specific feature or in a specific location, as they would not know exactly what the target looked like in advance. The attentional blink experiment in chapter 6 tests this idea using location stimuli by comparing the use of specific target images versus a category of target images.

The results of the end-of-block recognition tests from the current experiment found significantly higher recognition scores for the probe stimuli (M = 3.167) compared to the irrelevant stimuli (M = 2.250) and for the probes compared to the unpresented Kent buildings (M = 1.667). These comparisons were also significant for the end-of-experiment recognition test results. This fits with our expectation that the probe stimuli would be familiar to University of Kent students and would breakthrough into conscious awareness enough to be recognised in the streams. However, there was not a stronger positivity for the probe compared to the irrelevant despite these recognition results. The mean recognition score for the probe stimuli in the end-of-block tests was 3.167 out of 5, which is lower than anticipated based on previous experiments (i.e., 4 out of 5 in the famous names experiment in chapter 3). These results suggest that, while the probes were significantly more recognisable than the irrelevants and unpresented images, they were not as highly recognisable as expected or recognisable enough to generate stronger P3s. It is possible that a future experiment with different probe stimuli may result in increased recognition and differential P3s.

The end-of-block recognition results also found that the irrelevant stimuli (M = 2.250) had significantly higher recognition scores than the unpresented Kent images (M = 1.667), which is counter-intuitive as we expected neither to breakthrough into awareness nor be recognised. This result suggests that some irrelevants were consciously seen by some participants in the streams, although this was not enough to generate a strong P3. The end-of-experiment recognition results also found that the irrelevant stimuli (M = 3.083) had significantly higher recognition scores than the unpresented Birmingham images (M = 1.350), which further supports the possibility that the irrelevants were consciously seen. However, the irrelevant stimuli were shown on screen during the end-of-block recognition tests (where they received lower ratings) and this could, at least partially, explain their recognition in the later end-of-experiment test.

Interestingly, the end-of-block recognition results also found that the distractors (*M* = 3.00) had significantly higher recognition scores than the irrelevants (*M* = 2.250) and that over half the participants had scores between 3 and 4.3 out of 5 for the distractors. We expected there to be no significant difference between irrelevants and distractors, as these stimuli were not expected to breakthrough into conscious awareness and therefore would not be recognisable. However, these results suggest that some distractors did in fact breakthrough into awareness enough to later be recognised and given scores of 3 or more by some participants in the tests. This suggests a potential problem with the method or

stimuli that allowed these non-salient stimuli to be detected. More research would need to be done to be certain why this happened, but we suggest three possible explanations. Firstly, the critical stimuli, including the irrelevants, were limited to appearing between the 5th and 9th positions in the streams, whereas the distractors could appear at any point. Therefore, the distractors could have been more easily recognised if they were the first or last location stimulus of a stream. Secondly, we did not control for the visual salience of the stimuli (the innate perceptual qualities that make some stimuli stand out more compared to others), so it is possible that some of the distractors may have been more visually salient than the irrelevants, which would make them stand out and be more recognisable. Future research should attempt to control for the visual salience of the stimuli. Thirdly, RSVP is used so that stimuli appear on the fringe of awareness and only salient stimuli should be able to breakthrough into conscious awareness. However, the SOA used in the current experiment may have been too slow and allowed non-salient stimuli to enter conscious awareness and be recognisable, including the distractors used in the recognition tests.

The current experiment used an SOA of 167ms, which is slower than previous Fringe-P3 research which used 133ms in chapter 3 and Bowman et al (2013) and 100ms in Bowman et al (2014). The slower SOA was chosen as it was thought that location image stimuli are much more visually complex compared to the names and greyscale famous faces used in previous studies (Alsufyani et al., 2019; Bowman et al., 2013, 2014) and this would mean that none of the critical stimuli would be perceived and breakthrough into awareness if they were presented too fast. However, it is possible that the 167ms SOA was so slow that participants were able to consciously see non-salient stimuli, including the irrelevants and distractors, rather than just the salient target and probe stimuli. This possibility is supported by evidence from the grand average and individual participants' ERPs in figures 5.3 and 5.5. These show that both the probe and irrelevant were substantially above zero within the 300-1000ms AGAT search window. Is it possible that these positivities were P3-like responses for both the probe and irrelevant, but these were weak and temporally-jittered. The end-of-block recognition scores in table 5.4 also show that the irrelevants and distractors received higher ratings than the unpresented Kent images, and the analysis found that these irrelevant scores were significantly larger than the unpresented image scores, which, again, suggest that the irrelevants may have been consciously perceived. To prevent the non-salient stimuli from breaking through into awareness, future research using the Fringe-P3 method with location stimuli should consider using a staircase method (Cornsweet, 1987) to find the most suitable SOA for each participant. Using a staircase

procedure during practice trials, as used in the email addresses experiment in chapter 4, would allow us to account for individual differences and find an SOA specific to each participant where their target accuracy stays around predefined levels, for example, at a hit rate of 75% and a correct rejection rate of 80%. This would prevent the SOA from being so slow that it allows non-salient stimuli to be detected or so fast that it prevents the detection of salient stimuli.

Future research should also look into adding more distractor stimuli if possible. While Konkle et al's (2010) database consists of over 4000 stimuli, most of these were not photographs of building exteriors (e.g., golf courses, the inside of classrooms). As our probe and irrelevant stimuli were photographs of building exteriors, the distractors needed to be too, so most of Konkle et al's images were not suitable for this experiment and only 411 were used. Each distractor had a 4.14% chance of appearing in a stream, so it is likely that each distractor appeared more than once (but not as frequently as a critical stimulus) throughout the 225 streams. This should not have been a problem for the Fringe-P3 method as it has been shown that repeating non-salient stimuli does not increase their chance of breaking through into awareness from RSVP streams (Avilés et al., 2020), so they should not have been consciously seen. However, as we suspect that the SOA was slow enough to allow non-salient stimuli to be consciously seen (and our SOA was slower than the 84-133ms that Avilés et al analysed), it is possible that this repetition could have impacted the current study. A single distractor breaking through into awareness, even multiple times, would not have a substantial impact on the ERP results as the response would only occur in a small number of trials and would be averaged out when creating the ERPs. However, if multiple distractors were breaking through into awareness multiple times, then it could affect a larger number of trials and, thus, start impacting the ERPs. Additionally, the glance-look model of cognitive control suggests that distractors that are in the same or semantically related category as a target can capture attention and even generate an attentional blink, causing the participant to miss the following critical stimulus (Folk et al., 1992; Su et al., 2011). It is possible that some participants' ERPs, and therefore their P3s, could have been affected by this. Future research should use a larger database of distractors so that each distractor is very unlikely to appear more than once in order to prevent this possibility.

Similarly, as the distractors are all photographs of real buildings, it is possible that participants may recognise some of them. While the buildings were not known by the experimenters to be famous buildings, one participant did mention after the experiment

that one of the distractors was a castle from their home city and they had consciously seen it more than once in the streams. It would not be feasible to check that every participant is unfamiliar with every distractor, therefore, again, the most suitable way to overcome this problem in a future experiment would be to decrease the chances of a distractor being repeated by increasing the number of distractors in the database. Previous Fringe-P3 experiments have used much larger databases of 3667 email address distractors (chapter 4) and 1,000-10,000 name distractors (chapter 3; Bowman et al, 2013, 2014). However, location image distractors cannot be generated like the email address and name distractors, so it would be very difficult to find several thousand suitable location image distractors. Alsufyani et al's (2019) famous faces experiment used the same number of trials (225) as the current study but had a bank of 524 face image distractors and found significant results, therefore, it seems reasonable that future Fringe-P3 experiments using image stimuli should aim for (approximately) a minimum of 500 image distractors per 200 trials to reduce concerns about the distractors impacting the results.

It would also be beneficial to run the counterbalanced experiment at the University of Birmingham with the same stimuli but using the Birmingham images as the probe stimuli and the Kent images as the irrelevant stimuli. It is possible that the Kent images were simply not buildings the Kent participants were familiar with, or that the Kent stimuli had low visual salience and did not stand out more than the other stimuli, and this is why they did not generate large P3s. The University of Birmingham images, on the other hand, are brightly coloured, so may have higher visual salience and stand out more easily. If the counterbalanced study finds that the Birmingham students generate a strong P3 for the Birmingham probe stimuli, while the Kent students did not generate a strong P3 for the Kent probe stimuli, then it would suggest that the specific Kent stimuli were the reason they did not generate a large P3 in the current experiment. Future Fringe-P3 experiments using images should control for visual salience to prevent this possibility impacting the results. Similarly, if the Kent irrelevants did generate P3s for the Birmingham students, then it would suggest a further problem with the method.

Finally, this experiment only had 8 participants whereas previous Fringe-P3 studies have had 15. Preliminary analyses suggested that the experiment had potential problems with the method, so further participants were not collected for this experiment. Future experiments should include more participants as this would lead to more reliable results.

In summary, the results of this chapter provide a stepping stone towards using the Fringe-P3 method with location image stimuli. It has shown that location image stimuli can generate differential P3s, but that this was limited to the target stimuli. Methodological improvements, especially regarding the SOA and selection of stimuli, should increase the reliability of the experiment and may result in the probe stimuli generating strong P3s as well as the targets in future research.

6. Detecting Concealed Knowledge of Locations using the Attentional Blink

6.1. Introduction

This chapter contains an analysis of a dataset using the Fringe-P3 method with location image stimuli collected by Hannah Bowman and James Niblett at the University of Birmingham. All data processing and analyses in this chapter were performed by myself (see section 1.4 for details on previous and new uses of this data). This experiment used the attentional blink as a behavioural measure to detect concealed information instead of EEG and the P3. The location stimuli used for this experiment were the same as those used in chapter 5's EEG experiment with location stimuli, but, as this experiment was conducted at the University of Birmingham, the Birmingham images were used as probes and the University of Kent images were used as irrelevants, the reverse of in chapter 5.

The attentional blink is a behavioural phenomenon where a second critical stimulus is less accurately detected when presented 200-500ms after a first critical stimulus in RSVP (Raymond et al., 1992; Shapiro, Arnell, et al., 1997). The phenomenon has been shown to occur with letter, number, word, colour, face, and other picture stimuli (Evans & Treisman, 2005; Ganis & Patnaik, 2009; Joseph et al., 1997; Ross & Jolicœur, 1999; Shapiro, Caldwell, et al., 1997). This is not limited to when the target is a specific stimulus (e.g., the letter "A" or a specific image) in a detection task but has also been shown to occur when the target is a category (e.g., letters amongst numbers or pictures of animals) (Bowman & Wyble, 2007; Broadbent & Broadbent, 1987; Chun & Potter, 1995; Potter et al., 2010). The classic paradigm presents two targets (T1 & T2) as the two critical stimuli with T2 presented at specific lags following T1 (e.g., lag 1 is the stimulus immediately following T1, lag 7 is the seventh stimulus following T1). The detection of T1 generates an attentional blink and causes T2 to be missed if it is presented 200-500ms after T1. The eSTST model (Wyble et al., 2009) proposes that the attentional blink is caused by competition between working memory encoding and attention allocation. They suggest that T1 activates a window of attentional enhancement which is then deactivated if another target is not presented within 200ms, after which attention is supressed while the detected target(s) is encoded into working memory. Any targets presented while attention is suppressed will not be detected. See section 2.6 in chapter 2 for more on the attentional blink and the eSTST model.

The attentional blink has, importantly, also been shown to occur when the first critical stimulus is a salient probe rather than a target (Barnard et al., 2005; Ganis & Patnaik, 2009). This allows the attentional blink paradigm to be used to detect concealed information, as the probe stimulus would only breakthrough into conscious awareness and generate an attentional blink if it was salient. Therefore, if the participant's target accuracy is worse when the target is presented 200-500ms after the probe, then it suggests they are familiar with the probe and are concealing information about it.

There are two main exceptions to the attentional blink: lag 1 sparing and nonblinkers. Lag 1 sparing occurs when the second critical stimulus is presented up to 150ms after the first critical stimulus and is still accurately detected (Bowman et al., 2008; Bowman & Wyble, 2007). The eSTST model explains lag-1 sparing through its attentional enhancement window, which is activated by the first target and only closes if another target is not presented within 200ms of the previous target. Therefore, a second target presented at lag 1 that is within 150ms of the first target would be detected within this attentional enhancement window and be encoded into working memory, thus escaping the attentional blink (Wyble et al., 2009). Nonblinkers, on the other hand, are people who do not experience an attentional blink regardless of lag. Nonblinkers have been shown to consolidate information into working memory faster than blinkers, and this is thought to be the reason why they do not experience an attentional blink (Martens et al., 2006).

The attentional blink paradigm would be relatively simple to use while recording behavioural and EEG responses at the same time. The combination of these two measures would reduce the chance of random errors and lead to an even more robust and countermeasure resistant method of detecting concealed information. It will also be helpful for cases where individual differences may make one test fail to detect familiarity, for example if someone is a nonblinker but still generates a P3, or where someone's P3 is not strong enough to have a significant result in the AGAT analysis but they still show a strong attentional blink. Martens et al's study on nonblinkers (2006) has already shown that the attentional blink paradigm can be successfully combined with EEG to measure P3s concurrently. The combined use of EEG and the attentional blink to detect concealed information is a future goal of Fringe-P3 research.

In addition to testing the use of the attentional blink with the Fringe-P3 method, this chapter will also investigate the use of search strategies by participants to help them find the target. Such search strategies include looking for a specific distinctive feature of the

target, rather than looking for the whole target. For example, if the target was a word, then the search strategy could be to look for the first letter of that word and ignore all words that do not start with that letter. In the current experiment, which used images of castles as targets (figure 6.1), an example of a distinctive feature could be a turret in the top left corner. By looking for this specific feature, especially if they are only looking at a particular area (e.g., top left corner) of the images, the participant is far less likely to perceive images that do not share this feature. Therefore, if the probe stimulus does not share this feature, it will not be perceived, breakthrough into awareness, or generate a P3 or attentional blink. As a result, the fringe-P3 concealed information test would not detect the probe as being familiar and the participant would be seen as "innocent". This fits with the glance-look model of cognitive control (Su et al., 2011), which suggests that RSVP stimuli are first "glanced" at to process the broad features or meaning of that stimulus (e.g., turret-like shape in the top left corner). If the stimulus matches the broad features or meaning of the target, then it receives a deeper "look" to analyse the specific meaning and detail of the stimulus, at which point the participant would be able to detect if it was the specific target or not. If the stimulus does not match the broad features or meaning of the target during the "glance", then it will not receive the deeper "look" and will be analysed no further. Therefore, probe stimuli that do not match the same general features as the target will not receive the deeper "look" and therefore will not be perceived enough to breakthrough into conscious awareness and generate a P3 or attentional blink. It is vital that the task used in Fringe-P3 concealed information tests forces the participant to view the whole of the stimuli and not focus on one specific feature in order to counter the potential effects of such search strategies. It is also possible that location stimuli are more visually complex stimuli than the previously used words and greyscale faces, which may make detecting location targets a harder task, meaning participants may be even more likely to use a search strategy to aid them. The current experiment investigated search strategies by comparing two groups of participants with different search task types: categorisation and detection.

For the detection task, participants will be shown their specific target image before the blocks of streams and will look for that in the streams, as done in previous Fringe-P3 experiments (Alsufyani et al., 2019; Bowman et al., 2013, 2014) and in chapters 3, 4, and 5 of this thesis. For the categorisation task, participants will not be shown any of the target stimuli before the blocks and will instead be told to look for a category of targets: castles. Participants will not be informed about search strategies nor told to use or not use any in

the streams. It is expected that participants in the detection group will be more likely to use search strategies, as they will know what their target looks like and therefore could look for a specific feature such as a turret. On the other hand, participants in the categorisation group will not know what any of the specific targets will look like and, therefore, will not be able to look for a specific known feature. The categorisation participants will be forced to look at the streams more holistically and, as a result, are more likely to see the probe stimulus. As such, we expect that there will be an attentional blink in both groups, but that it will be more pronounced in the categorisation group.

There are two main aims of this chapter. Firstly, we seek to show that the attentional blink can be successfully used to detect concealed information using the Fringe-P3 method with location stimuli. Secondly, we seek to investigate two different search task types – categorisation and detection – to see how these impact the accuracy of the method in detecting concealed information and, through this, how they impact the effectiveness of search strategies.

We predict that the probe stimuli will generate attentional blinks (leading to fewer hits on target) while the irrelevants will not, and that these attentional blinks will occur at lags 1 and 3 (within the 200-500ms window following the first critical stimulus), but not at lags 5 and 7 (more than 500ms after the first critical stimulus). As such, we also predict that there will be an interaction between stimulus type (probe or irrelevant) and lag, with there being a bigger difference in the amount of hits inside than outside the blink for the probe condition compared to the irrelevant condition (i.e., a bigger attentional blink for probe than irrelevant). Finally, we expect that there will be a significant interaction between task type, stimulus type, and lag, with stronger attentional blinks occurring in the categorisation group than the detection group.

6.2. Method

6.2.1. Participants

All participants were students at the University of Birmingham and had normal or corrected-to-normal vision and no known neurological disorders. All participants gave their informed consent and were rewarded with Research Participation Scheme credit for their participation. The STEM Research Ethics Committee of the University of Birmingham granted ethical approval for the experiment.

30 participants took part in the categorisation group. Three were excluded due to familiarity with the University of Kent campus, as this would prevent the Kent stimuli being

useable as unfamiliar irrelevant stimuli, leaving 27 participants for the categorisation group. 22 participants took part in the detection group, none of which were familiar with the University of Kent campus.

The binomial regression analysis cannot complete if a participant's hit count is too close to ceiling, therefore additional exclusion criteria (as explained fully in section 6.2.4) resulted in 11 more participants being excluded from all categorisation analyses and six more participants being excluded from all detection analyses.

Following all exclusions, there were 16 participants in the categorisation group, all female, aged 18-20 (M = 19.19, SD = 0.75), and 16 in the detection group, 14 female and 2 male, aged 18-20 (M = 19, SD = 0.73).

6.2.2. Stimuli and Presentation

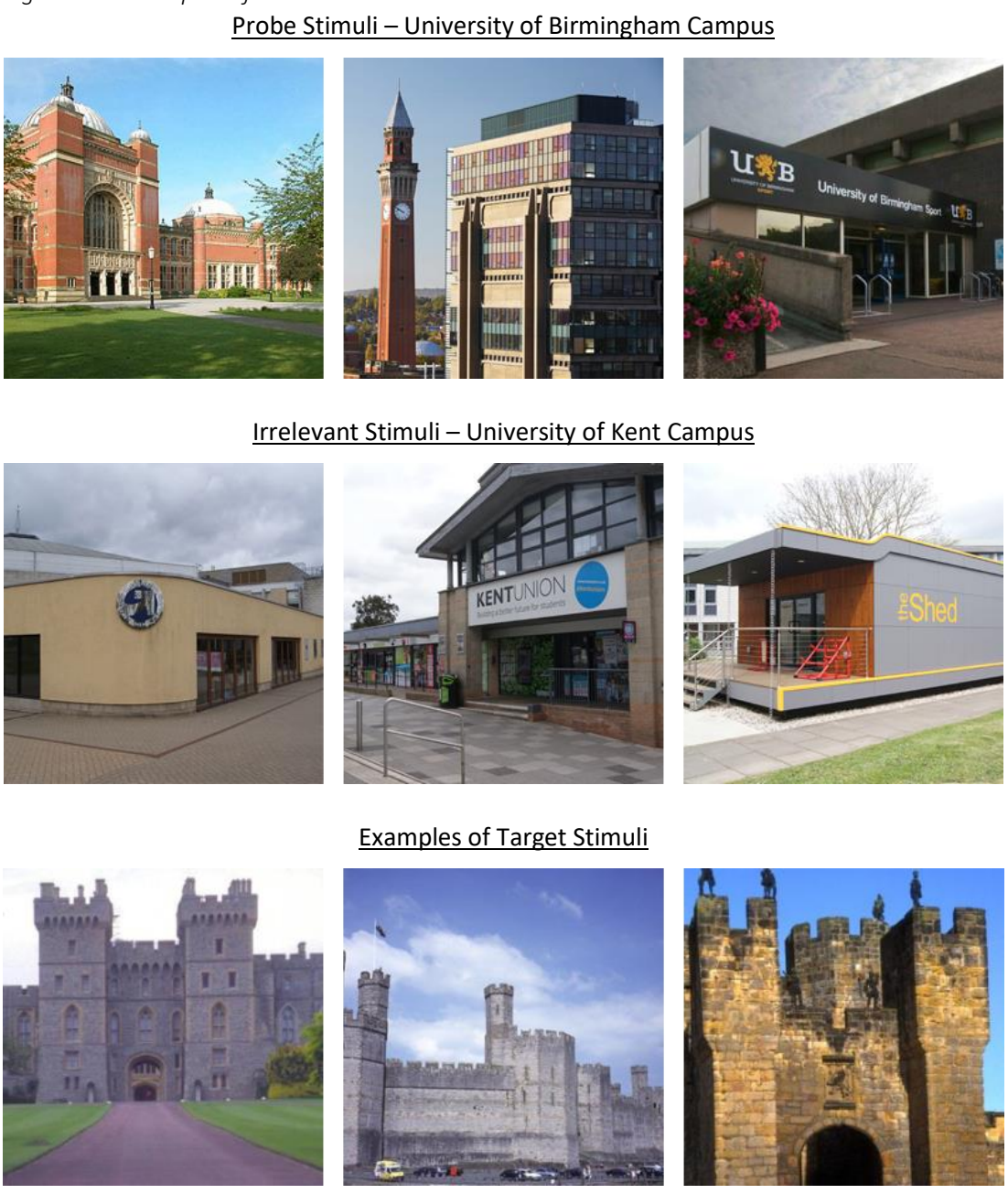
All stimuli were 256px x 256px colour photographs of buildings presented on a black background at the same location in the middle of the screen. The image stimuli were the same images as used in chapter 5 but were separated into conditions differently. For both the categorisation and detection tasks, the probe stimuli were three images of buildings from the University of Birmingham campus, and the irrelevant stimuli were three images of buildings from the University of Kent Canterbury campus. For the categorisation task, there were 10 target images of castles. For the detection task, there were three target images of castles. All target images were taken from Konkle et al's (2010) database of images of scenes. The bank of distractor stimuli was the same as used in chapter 5 and were all photographs of building exteriors taken from Konkle et al's database. The three target stimuli were removed from the bank of distractors for the detection task, and all images of castles were removed from the bank of distractors for the categorisation task. Figure 6.1 shows examples of the stimuli used.

All streams consisted of 16 images, one of which was either a probe or an irrelevant. Half of the streams also contained a target image, split equally between those containing probes and irrelevants. Randomly selected distractor stimuli filled the remaining spaces in the streams.

There were 240 streams total, separated into three blocks of 80 streams. Each block was allocated one probe and one irrelevant image, which were presented 40 times each per block and 120 times each across the whole experiment. Half of the trials also contained a target image, with 40 trials per block (120 trials total). In each block there were 20 trials with both a target and probe and 20 trials with both a target and irrelevant. Across the

whole experiment, there were 60 trials with both a target and a probe and 60 trials with both a target and an irrelevant. For the detection task, the three target images were separated into one for each block and shown 40 times each. For the categorisation task, each of the 10 castle images was presented an equal number of times (4) in every block.

Figure 6.1. Examples of the Critical Stimuli.



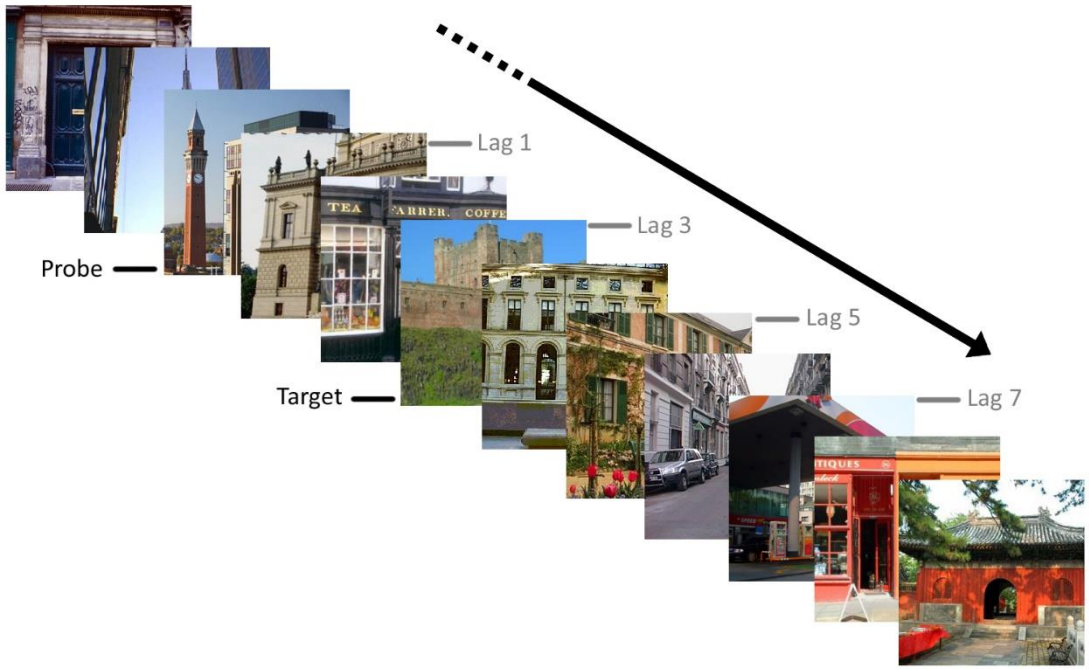
This figure presents the three probe stimuli, three irrelevant stimuli, and three examples of target stimuli.

The probe and irrelevant stimuli were still presented in trials where the target was not present. This repetition should not lead to either probe or irrelevant being more likely to be perceived, as repeated presentations of a non-salient stimulus has been shown not to increase the chances of it being perceived (Avilés et al., 2020). Therefore, the critical stimuli would still only breakthrough into awareness if they were salient and would not breakthrough if they were non-salient, despite the number of repetitions.

The SOA was 144ms. This was slower than the 133ms used in the famous faces and own-names Fringe-P3 studies (Alsufyani et al., 2019; Bowman et al., 2013) but faster than the 167ms used in chapter 5's EEG experiment with location stimuli. It was thought that the 167ms SOA used in chapter 5 was too slow and led to non-salient distractor and irrelevant stimuli breaking through into awareness. Therefore, the SOA needed to be faster for the current experiment. However, it was also thought that location stimuli are more complex than name and face stimuli and would need longer to be processed and breakthrough into awareness, therefore the SOA of 133ms used in previous Fringe-P3 experiments might be too fast. RSVP studies using picture stimuli by Potter (1975, 1976) also showed that target accuracy was lower at SOAs of 113ms and 125ms compared to 167ms and that the difference in accuracy between 125ms and 167ms was significant (Potter, 1976). As a result, the SOA of 144ms was chosen to be faster than the previous 167ms used in chapter 5 but slower than the fast SOAs in Potter's and Bowman et al's experiments.

The streams were divided into four separate lags: 1, 3, 5, 7. There was 144ms between the first critical stimulus (probe or irrelevant) and the target for lag 1, 432ms for lag 3, 720ms for lag 5, and 1008ms for lag 7. Each lag was used 10 times in each of the three blocks, meaning each lag was used 30 times for each participant across the whole experiment, with the target following the probe in half (15) of those trials and the target following the irrelevant in the other 15 trials. Figure 6.2 shows an example of a stream and the lag positions.

Figure 6.2. Section of an Example Stream.



This figure shows a section of an example stream, featuring a probe stimulus and a target stimulus (at lag 3). Marked in grey are the four lag positions.

6.2.3. Tasks

Participants were told to look for a target image in the streams and that the target would not always be present. They were shown a practice trial to adjust to the experiment before beginning. This practice trial contained a target that was not used in the rest of the experiment.

Participants in the detection group were shown the target image for that block before each stream, so they would know what their specific target was. Participants in the categorisation group were not shown any of the 10 castle targets before the streams and were only told to look for images of castles.

At the end of each stream, participants were asked if they saw a target image and responded yes or no via their keyboard with the 1 or 2 keys respectively. At the end of the experiment, participants were also asked if they were familiar with the University of Kent Canterbury campus so that any who said yes could be excluded. Participants were not asked if they were familiar with the University of Kent campus before participating in the experiment as the question may have warned them that images of the University of Kent may be presented. Breaks were offered between blocks.

6.2.4. Analysis Procedure

ANOVAs and Binomial Regression Analysis

All of the main analyses used the hits data only. While d' scores can usually provide more information on target accuracy than hits alone (since they consider the number of false alarms as well as hits), d' scores would not be appropriate for our analyses because the false alarm data cannot be sorted into the separate lags or collapsed lags. This is because the lags are based on when the target is presented and there is no target presented when there is a false alarm, so there is no lag information associated with false alarms. Therefore, d' scores cannot be used to compare target accuracy inside and outside of the attentional blink window, which is necessary for attentional blink research. The hit counts, on the other hand, can be separated into all lags and conditions, so are the more appropriate data for analysis. Additionally, the binomial regression analysis uses data at the individual trial level (hit or miss), so can only be performed using hits data. Therefore, all main analyses only used hits data. Although not used in the main analyses, the overall d' scores for categorisation and detection can be found in appendix C.

The main aim of the lag analyses was to compare the hits inside and outside the attentional blink window, therefore the four separate lags were combined into two "collapsed lags", with lags 1 and 3 collapsed to be "inside the blink" since they are within the 500ms attentional blink window, and lags 5 and 7 collapsed to be "outside the blink" since they are beyond the 500ms attentional blink window. The collapsed lags were used in the group level ANOVAs and the group and individual level binomial regression analyses. The separate lags could not be used in the binomial regression analysis, so were only used in additional ANOVAs at the group level.

ANOVAs were performed on the group level hits data in SPSS 24. There were four main effects: search task type (categorisation v detection), stimulus type (probe v irrelevant), collapsed lag (inside v outside the attentional blink window), and separate lag (1, 3, 5, 7). We were also interested in two-way interactions between stimulus type and collapsed lag, and between stimulus type and separate lag, as well as three-way interactions between search task type, stimulus type, and collapsed lag, and between search task type, stimulus type, and separate lag.

A binomial regression analysis was used to analyse the data at the individual participants' level in Matlab 2016a. This used the generalised linear model with a probit link function and a linear hypothesis test. There were two main effects - stimulus type and collapsed lag - and a two-way interaction between stimulus type and collapsed lag. The binomial

regression analysis was also performed on the group level data as an additional analysis. An alpha level of 0.05 was used for all statistical analyses.

Additional Exclusion Criteria

There were several participants whose hit counts were close to ceiling. When this happens, the binomial regression analysis, in particular fitting the generalised linear model, will reach an iteration limit and will not converge. The hits are divided into four bins for the analysis: ProbeOut (the target presented after the probe outside of the 500ms attentional blink window), ProbeIn (the target presented after the probe inside the attentional blink window), IrrelOut (the target presented after the irrelevant outside the attentional blink window), and IrrelIn (the target presented after the irrelevant inside the attentional blink window). Every participant who had scored a full amount of hits on at least one of the four bins threw an "iteration limit reached" warning and the analysis could not be completed. The capacity to observe a significant attentional blink effect is also impacted by proximity to ceiling, since the range of possible scores is restricted by the ceiling, limiting potential differences between conditions. Therefore, some participants had to be excluded. In order to be fair, participants' hits for the ProbeOut and IrrelOut bins were calculated and added together. The maximum possible score was 60. The distribution of these scores showed a clear bimodality, with a higher mode reflecting a clear compression to ceiling. We excluded all participants who scored above the intersection between these two modes. The score at the intersection was 55 for categorisation and 56 for detection, therefore participants with scores of 55 or above were excluded from both groups. 11 participants were excluded from all categorisation analyses and 6 participants were excluded from all detection analyses, leaving 16 participants in both groups.

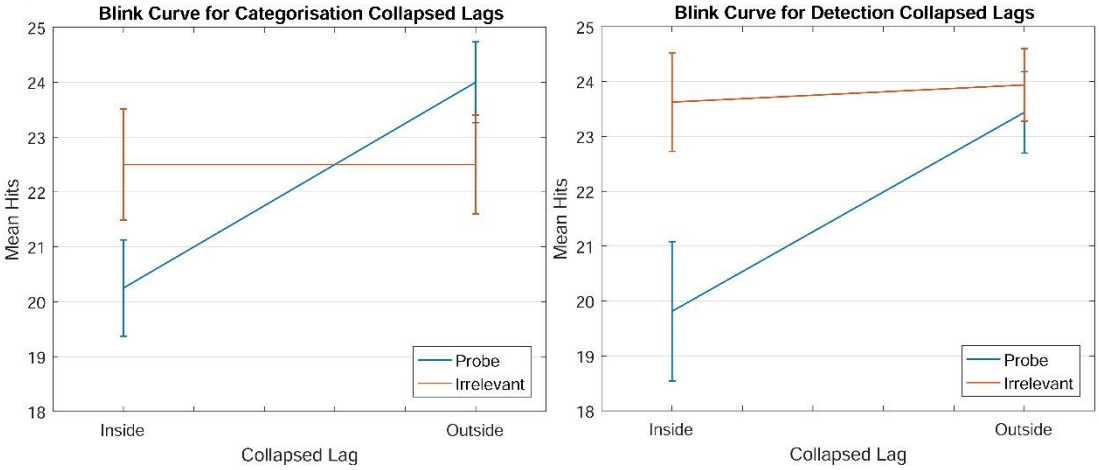
6.3. Results

6.3.1. Group Level

Blink Curves

Figure 6.3 contains the blink curves for the categorisation and detection search tasks with collapsed lag on the x-axis. In both search tasks, the mean hits for the irrelevant condition were very similar inside and outside the blink window, showing that the timing of the target did not affect detection accuracy in the irrelevant condition. The mean hits for the probe condition, on the other hand, were smaller inside the blink window than outside, thus demonstrating an attentional blink when the probe was presented within the 200-500ms attentional blink window but not outside that window.

Figure 6.3. Blink Curves for the Categorisation and Detection Collapsed Lags.

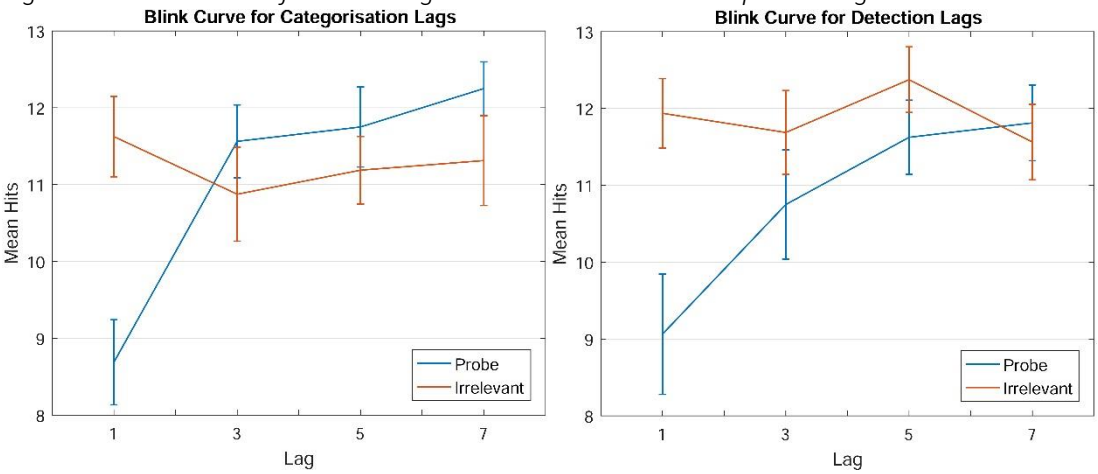


These blink curves show the mean hits and standard error of the mean for the collapsed lags and stimulus types for both the categorisation and detection tasks. The mean hits for the probe condition inside the blink window was smaller than the others, demonstrating an attentional blink.

Figure 6.4 contains the blink curves for the categorisation and detection search tasks with the four separate lags on the x-axis. In both search tasks, the mean hits for the probe was smallest at lag 1 and increased as lag length increased. This suggests an attentional blink when the target is presented at lag 1 after the probe. The irrelevant mean hits were stable and varied only slightly across the lags, showing no attentional blink.

Both figures 6.3 and 6.4 suggest that the presence of the probe caused an attentional blink inside the 200-500ms attentional blink window but not outside of the window, and that the irrelevant caused no attentional blink.

Figure 6.4. Blink Curves for the Categorisation and Detection Separate Lags.



These blink curves show the mean hits and standard error of the mean for the separate lags and two stimulus types for both the categorisation and detection tasks. The mean hits for the probe condition at lag 1 was smaller than the others, demonstrating an attentional blink.

Categorisation

Table 6.1 shows the total hits for each participant for each of the collapsed lags and stimulus types in the categorisation task group.

A 2x2 repeated measures ANOVA was performed on the stimulus type and collapsed lag data for the categorisation group. There was a significant main effect of collapsed lag ($F(1,15) = 7.979$, $p = 0.0128$), showing that there were significantly fewer hits inside the blink window than outside. There was also a significant interaction between stimulus type and collapsed lag ($F(1,15) = 9.507$, $p = 0.0076$), showing that there was a bigger difference in the amount of hits inside than outside the blink window for the probe condition compared to the irrelevant condition (i.e., a bigger attentional blink for probe than irrelevant). These results match our hypothesis that there would be an attentional blink in the probe condition but not in the irrelevant condition. There was not, however, a significant main effect of stimulus type ($F(1,15) = 0.352$, $p = 0.5616$), which does not match our expectation that there would be significantly fewer hits overall in the probe condition than the irrelevant condition.

A two-tailed binomial regression analysis was also performed on the stimulus type and collapsed lag data. There was a significant main effect of collapsed lag ($F(1,60) = 9.694$, $p = 0.0028$) and a significant two-way interaction between stimulus type and collapsed lag ($F(1,60) = 9.694$, $p = 0.0028$), but there was not a significant main effect of stimulus type ($F(1,60) = 0.185$, $p = 0.6686$). These results match the 2x2 ANOVA results.

Table 6.1. Collapsed Lag Hits for Categorisation.

Participant	*Probe Outside*	*Probe Inside*	*Irrel. Outside*	*Irrel. Inside*
1	27	25	25	18
2	26	22	18	25
3	24	21	25	23
4	25	20	21	23
5	24	18	20	21
6	27	23	27	28
7	18	16	19	13
8	25	18	24	23
9	23	26	26	25
10	27	21	27	26
11	20	20	17	23
12	28	24	23	27
13	22	12	20	16
14	22	21	17	20
15	20	19	26	24
16	26	18	25	25
Mean	*24.00*	*20.25*	*22.50*	*22.50*
Median	*24.5*	*20.5*	*23.5*	*23*
Std Dev	*2.94*	*3.51*	*3.60*	*4.05*

This table shows the number of hits per participant for the collapsed lags and stimulus types. The maximum number of hits per column for each participant was 30.

Table 6.2 shows the total hits for each participant for each of the four separate lags and two stimulus types in the categorisation task group.

A 2x4 repeated measures ANOVA was conducted on the separate lag and stimulus type data. There was a significant main effect of separate lag ($F(3,45) = 5.315$, $p = 0.0032$), with the fewest hits at lag 1, and a significant interaction between stimulus type and separate lag ($F(3,45) = 13.579$, $p < 0.0001$), with a bigger difference in hits inside than outside the attentional blink window in the probe condition. These results match our expectation that there would be an attentional blink when the target was presented at the short lags (1 & 3) after the probe. There was not a significant main effect of stimulus type ($F(1,15) = 0.352$, $p = 0.5616$). These results match the 2x2 ANOVA and binomial regression results.

Table 6.2. Separate Lag Hits for Categorisation.

Partici-pant	*Probe Lag 1*	*Probe Lag 3*	*Probe Lag 5*	*Probe Lag 7*	*Irrel. Lag 1*	*Irrel. Lag 3*	*Irrel. Lag 5*	*Irrel. Lag 7*
1	12	13	13	14	10	8	12	13
2	9	13	13	13	13	12	9	9
3	7	14	12	12	11	12	13	12
4	9	11	14	11	13	10	12	9
5	8	10	12	12	12	9	10	10
6	11	12	14	13	13	15	14	13
7	6	10	9	9	8	5	8	11
8	7	11	12	13	13	10	13	11
9	11	15	10	13	13	12	12	14
10	12	9	14	13	13	13	13	14
11	9	11	10	10	11	12	9	8
12	10	14	14	14	13	14	12	11
13	4	8	11	11	7	9	11	9
14	9	12	10	12	9	11	9	8
15	8	11	7	13	14	10	11	15
16	7	11	13	13	13	12	11	14
Mean	*8.69*	*11.56*	*11.75*	*12.25*	*11.63*	*10.88*	*11.19*	*11.31*
Median	*9*	*11*	*12*	*13*	*13*	*11.5*	*11.5*	*11*
Std Dev	*2.21*	*1.90*	*2.08*	*1.39*	*2.09*	*2.45*	*1.76*	*2.33*

This table shows the number of hits per participant for each of the separate lags and two stimulus types. The maximum number of hits per column for each participant was 15.

Detection

Table 6.3 shows the total hits for each participant for each of the collapsed lags and stimulus types in the detection task group.

A 2x2 repeated measures ANOVA was performed on the stimulus type and collapsed lag data for the detection group. There was a significant main effect of stimulus type ($F(1,15) = 27.310$, $p = 0.0001$), showing that there were significantly fewer hits in the probe condition, and a significant main effect of collapsed lag ($F(1,15) = 8.675$, $p = 0.0100$), showing that there were significantly fewer hits inside the blink window compared to outside. There was also a significant interaction between stimulus type and collapsed lag ($F(1,15) = 7.804$, $p = 0.0136$), showing that there was a bigger difference in the amount of hits inside than outside the blink window for the probe condition compared to the irrelevant condition (i.e., a bigger attentional blink for probe than irrelevant). These results match our hypotheses that there would be an attentional blink for the probe condition but not for the irrelevant condition.

A two-tailed binomial regression analysis was also performed on the stimulus type and collapsed lag data. There was a significant main effect of stimulus type ($F(1,60) = 12.251$, $p = 0.0009$), a significant main effect of collapsed lag ($F(1,60) = 9.998$, $p = 0.0025$), and a significant interaction between stimulus type and collapsed lag ($F(1,60) = 6.682$, $p = 0.0122$). These results match those of the 2x2 ANOVA.

Table 6.3. Collapsed Lag Hits for Detection.

Participant	*Probe Outside*	*Probe Inside*	*Irrel. Outside*	*Irrel. Inside*
1	21	23	23	24
2	25	13	22	22
3	25	17	25	23
4	26	25	26	26
5	23	16	24	18
6	27	17	26	27
7	26	29	26	25
8	20	14	19	20
9	25	23	27	27
10	22	19	20	24
11	22	16	23	23
12	22	22	25	27
13	18	13	20	18
14	19	19	23	18
15	27	23	27	27
16	27	28	27	29
Mean	*23.44*	*19.81*	*23.94*	*23.63*
Median	*24*	*19*	*24.5*	*24*
Std Dev	*2.97*	*5.08*	*2.64*	*3.59*

This table shows the number of hits per participant for the collapsed lags and stimulus types. The maximum number of hits per column for each participant was 30.

Table 6.4 shows the total hits for each participant for each of the four separate lags and two stimulus types in the detection task group.

A 2x4 repeated measures ANOVA was also performed on the stimulus type and separate lag data. There was a significant main effect of stimulus type ($F(1,15) = 27.310$, $p = 0.0001$), a significant main effect of separate lag ($F(3,45) = 4.918$, $p = 0.0049$), and a significant interaction between stimulus type and separate lag ($F(3,45) = 3.681$, $p = 0.0187$). This fits with our expectation that there would be fewer hits for the short lags (1 & 3) and a bigger attentional blink in the probe condition and matches the 2x2 ANOVA and binomial regression results.

Table 6.4. Separate Lag Hits for Detection.

Partici-pant	*Probe Lag 1*	*Probe Lag 3*	*Probe Lag 5*	*Probe Lag 7*	*Irrel. Lag 1*	*Irrel. Lag 3*	*Irrel. Lag 5*	*Irrel. Lag 7*
1	9	14	9	12	13	11	14	9
2	4	9	12	13	11	11	11	11
3	9	8	11	14	13	10	11	14
4	13	12	14	12	13	13	13	13
5	9	7	10	13	9	9	14	10
6	6	11	12	15	14	13	14	12
7	15	14	14	12	11	14	14	12
8	7	7	10	10	10	10	9	10
9	11	12	14	11	13	14	13	14
10	10	9	11	11	11	13	10	10
11	8	8	9	13	11	12	12	11
12	10	12	11	11	13	14	12	13
13	6	7	9	9	11	7	13	7
14	5	14	12	7	9	9	10	13
15	9	14	14	13	14	13	14	13
16	14	14	14	13	15	14	14	13
Mean	*9.06*	*10.75*	*11.63*	*11.81*	*11.94*	*11.69*	*12.38*	*11.56*
Median	*9*	*11.5*	*11.5*	*12*	*12*	*12.5*	*13*	*12*
Std Dev	*3.13*	*2.84*	*1.93*	*1.97*	*1.81*	*2.18*	*1.71*	*1.97*

This table shows the number of hits per participant for each of the separate lags and two stimulus types. The maximum number of hits per column for each participant was 15.

The categorisation and detection task groups had similar results. Both had significant main effects of collapsed lag and separate lag, showing that hits performance was worse in lags 1 and 3/inside the attentional blink window. Both also had significant interactions between collapsed lag and stimulus type and between separate lag and stimulus type, showing that there was more evidence for an attentional blink for probe than for irrelevant. However, only the detection group had a significant main effect of stimulus type, showing that there was a significant difference between hits performance in the probe and irrelevant conditions for the detection task but no significant difference for the categorisation task.

Three-way Analyses

A 2x2x2 mixed ANOVA was performed on the search task type, stimulus type, and collapsed lag data. There was a significant main effect of stimulus type ($F(1,30) = 11.256$, $p = 0.0022$), a significant main effect of collapsed lag ($F(1,30) = 16.648$, $p = 0.0003$), and a significant two-way interaction between stimulus type and collapsed lag ($F(1,30) = 17.288$, $p = 0.0002$). These results match the 2x2 ANOVA results and our predictions that there would be fewer hits in the probe condition than the irrelevant condition, fewer hits inside the attentional blink window than outside, and more evidence for an attentional blink for probe than for irrelevant. There was also a significant two-way interaction between search task type and stimulus type ($F(1,30) = 5.574$, $p = 0.0249$), which fits with the finding of a significant main effect of stimulus type in the 2x2 ANOVA for the detection task group but not for the categorisation task group and shows that hits performance was better in the irrelevant condition than the probe condition in the detection group but there was no difference in the categorisation group. There was not, however, a significant main effect of search task type ($F(1,30) = 0.140$, $p = 0.7113$), showing that there was not a significant difference between the two tasks overall. There was also not a significant two-way interaction between search task type and collapsed lag ($F(1,30) = 0.010$, $p = 0.9214$) or a significant three-way interaction between search task type, collapsed lag, and stimulus type ($F(1,30) = 0.066$, $p = 0.7985$). The three-way interaction is key, as it does not match our prediction that there would be a more pronounced attentional blink for the categorisation task compared to the detection task.

A 2x2x4 mixed ANOVA was also performed on the search task type, stimulus type, and the separate lags data. There was a significant main effect of stimulus type ($F(1,30) = 11.256$, $p = 0.0022$) and a significant main effect of separate lag ($F(3,90) = 9.774$, $p < 0.0001$), but there was not a significant main effect of search task type ($F(1,30) = 0.140$, $p = 0.7113$). There was a significant two-way interaction between stimulus type and separate lag

(F(3,90) = 13.505, p <0.0001), and between search task type and stimulus type (F(1,30) = 5.574, p = 0.0249). However, there was not a significant two-way interaction between search task type and separate lag (F(3,90) = 0.475, p = 0.7005) and there was not a significant three-way interaction between search task type, separate lag, and stimulus type (F(3,90) = 0. 0.779, p = 0.5085). These results match the 2x2x2 mixed ANOVA results.

6.3.2. Individual Participants' Level

Categorisation

Table 6.5 shows the individual participants' results from the two-tailed binomial regression analysis for the categorisation group. There were four significant results; one participant (15) had a significant main effect of stimulus type, one participant (13) had a significant main effect of collapsed lag, and two participants (2 and 12) had a significant two-way interaction between stimulus type and collapsed lag.

Table 6.5. Individual Binomial Regression Results for Categorisation.

	Stimulus Type Main Effect		*Collapsed Lag Main Effect*		*Stim. Type x Coll. Lag Interaction*	
Participant	*F*	*p*	*F*	*p*	*F*	*p*
1	3.532	0.0627	3.532	0.0627	0.534	0.4662
2	0.970	0.3267	0.189	0.6648	5.326	0.0228*
3	0.411	0.5226	1.175	0.2806	0.023	0.8800
4	0.084	0.7718	0.441	0.5080	2.176	0.1428
5	0.083	0.7738	1.038	0.3103	1.974	0.1627
6	1.563	0.2138	0.292	0.5901	1.563	0.2138
7	0.126	0.7230	2.154	0.1449	0.538	0.4649
8	0.469	0.4949	2.641	0.1068	1.389	0.2409
9	0.189	0.6648	0.189	0.6648	0.911	0.3419
10	1.024	0.3137	2.564	0.1120	1.024	0.3137
11	0.005	0.9428	1.372	0.2439	1.372	0.2439
12	0.314	0.5760	0.032	0.8590	4.167	0.0435*
13	0.094	0.7594	6.726	0.0107*	1.259	0.2642
14	1.325	0.2520	0.119	0.7309	0.575	0.4499
15	5.327	0.0228*	0.493	0.4841	0.123	0.7269
16	1.143	0.2873	2.579	0.1110	2.579	0.1110
Mean	*1.041*	*0.482*	*1.596*	*0.368*	*1.596*	*0.318*
Median	*0.440*	*0.509*	*1.107*	*0.295*	*1.315*	*0.254*
Std Dev	*1.451*	*0.278*	*1.766*	*0.274*	*1.435*	*0.235*

This table shows the F and p values for the individual participants in the categorisation group. The asterisk marks significant p values at the 0.05 level. The degrees of freedom for all participants were 1, 116.

Detection

Table 6.6 shows the individual participants' results from the two-tailed binomial regression analysis for the detection group. There were five significant results; three participants (2, 3, and 5) had a significant main effect of collapsed lag and two participants (2 and 6) had a significant two-way interaction between stimulus type and collapsed lag. No participants had a significant main effect of stimulus type.

Table 6.6. Individual Binomial Regression Results for Detection.

	Stimulus Type Main Effect		*Collapsed Lag Main Effect*		*Stim. Type x Coll. Lag Interaction*	
Participant	*F*	*p*	*F*	*p*	*F*	*p*
1	0.397	0.5297	0.397	0.5297	0.032	0.8587
2	0.804	0.3718	5.201	0.0244*	5.201	0.0244*
3	1.184	0.2788	4.076	0.0458*	1.184	0.2788
4	0.064	0.8015	0.064	0.8015	0.064	0.8015
5	0.338	0.5622	6.389	0.0128*	0.013	0.9088
6	2.687	0.1039	2.687	0.1039	4.987	0.0275*
7	1.725	0.1917	0.772	0.3813	1.725	0.1917
8	0.822	0.3666	0.822	0.3666	1.667	0.1993
9	2.264	0.1351	0.172	0.6787	0.172	0.6787
10	0.399	0.5290	0.069	0.7929	2.010	0.1590
11	2.338	0.1290	1.211	0.2734	1.211	0.2734
12	3.442	0.0661	0.338	0.5624	0.338	0.5624
13	1.658	0.2004	1.658	0.2004	0.275	0.6009
14	0.397	0.5299	0.994	0.3208	0.994	0.3208
15	0.861	0.3553	0.861	0.3553	0.861	0.3553
16	0.216	0.6433	1.160	0.2838	0.216	0.6433
Mean	*1.225*	*0.362*	*1.679*	*0.358*	*1.309*	*0.430*
Median	*0.841*	*0.361*	*0.928*	*0.338*	*0.928*	*0.338*
Std Dev	*1.016*	*0.219*	*1.923*	*0.257*	*1.616*	*0.292*

This table shows the F and p values for the individual participants in the detection group. The asterisk marks significant p values at the 0.05 level. The degrees of freedom for all participants were 1, 116.

One-Tailed Binomial Regression

The binomial regression analysis is two-tailed by default and the results in the tables 6.5 and 6.6 are from these two-tailed analyses. However, since we have directional hypotheses, a one-tailed binomial regression analysis would be more suitable. Therefore, the outcomes of the main effect and interaction equations were used to turn it into a one-tailed test. When the outcome of the equation matched our expectation, the p value was halved. When the outcome did not match our expectation, the p value was halved and then subtracted from one. In cases where the outcome of the equation was zero, this was treated as a positive result. The equations are explained below, where *ProbeOut* is the target presented after the probe outside of the 500ms attentional blink window, *ProbeIn* is the target presented after the probe inside the attentional blink window, *IrrelOut* is the target presented after the irrelevant outside the attentional blink window, and *IrrelIn* is the target presented after the irrelevant inside the attentional blink window.

Stimulus Type Main Effect = (ProbeOut + ProbeIn) - (IrrelOut + IrrelIn)

For the stimulus type main effect, we expected *ProbeIn* to have fewer hits than *IrrelOut* and *IrrelIn*, so the outcome of the equation would be negative. When the outcome of the equation was negative, as expected, we halved the p value. When the outcome of the equation was positive, we halved the p value and subtracted it from one.

Collapsed Lag Main Effect = (ProbeOut + IrrelOut) - (ProbeIn + IrrelIn)

For the collapsed lag main effect, we expected the collapsed lags outside the blink window (*ProbeOut* and *IrrelOut*) to have more hits than those inside the blink window (*ProbeIn* and *IrrelIn*), so the outcome of this equation would be positive. When the outcome of the equation was positive, as expected, we halved the p value. When the outcome of the equation was negative, we halved the p value and subtracted it from one.

*Stimulus Type * Collapsed Lag Interaction = (ProbeOut – ProbeIn) – (IrrelOut – IrrelIn)*

For the interaction, we expected there to be a similar number of hits for *ProbeOut*, *IrrelOut*, and *IrrelIn*, but there would be less hits for *ProbeIn*, so the outcome of this equation would be positive. When the outcome of the equation was positive, as expected, we halved the p value. When the outcome of the equation was negative, we halved the p value and subtracted it from one.

Categorisation

There were five significant results in the one-tailed binomial regression analysis for the categorisation group. One participant had a significant main effect of stimulus type (*M* = 0.464, *Med* = 0.454, *SD* = 0.299): participant 15 ($F(1,116) = 5.327$, $p = 0.0114$). Two participants had a significant main effect of collapsed lag (*M* = 0.290, *Med* = 0.199, *SD* = 0.280): participant 1 ($F(1,116) = 3.532$, $p = 0.0314$) and 13 ($F(1,116) = 6.726$, $p = 0.0054$). Two participants had a significant interaction between stimulus type and collapsed lag (*M* = 0.284, *Med* = 0.127, *SD* = 0.297): participant 2 ($F(1,116) = 5.326$, $p = 0.0114$), and 12 ($F(1,116) = 4.167$, $p = 0.0217$). There was one additional significant p value for the collapsed lag main effect in the one-tailed test compared to the two-tailed test.

Table 6.7 presents the number of p values below certain values (e.g., p <= 0.1) where the outcome of the equation was in the expected direction. It includes the results from both the one-tailed and two-tailed versions of the binomial regression analysis for the categorisation group. Overall, there were more p values below 0.05, 0.1, and 0.2 in the one-tailed test compared to the two-tailed test. The number of equation outcomes in the expected direction overall for the interaction is especially important as this shows that, while there were few significant p values, most of the equation outcomes (12/16) were in the expected direction. This provides evidence that there may have been an attentional blink effect for most participants, but this was only strong enough to reach significance for a few.

Table 6.7. One and Two-Tailed Results for Categorisation.

Number of results in the expected direction	*One-Tailed*			*Two-Tailed*		
	Stim. Type	*Coll. Lag*	*Interaction*	*Stim. Type*	*Coll. Lag*	*Interaction*
With significant p values (p <= 0.05)	1	2	2	1	1	2
With p values <= 0.1	1	6	5	1	2	2
With p values <= 0.2	4	8	10	1	6	5
Num. equation outcomes in the expected direction overall	8	12	12	8	12	12

This table presents the number of p values below certain values where the outcome of the equation was in the expected direction for the categorisation group. Each cell is out of 16.

Detection

There were six significant results in the one-tailed binomial regression analysis for the detection group. One participant had a significant main effect of stimulus type (*M* = 0.232, *Med* = 0.185, *SD* = 0.209): participant 12 (*F*(1,116) = 3.442, *p* = 0.0330). Three participants had a significant main effect of collapsed lag (*M* = 0.332, *Med* = 0.180, *SD* = 0.310): participant 2 (*F*(1,116) = 5.201, *p* = 0.0122), 3 (*F*(1,116) = 4.076, *p* = 0.0229) and 5 (*F*(1,116) = 6.389, *p* = 0.0064). Two participants had a significant interaction between stimulus type and collapsed lag (*M* = 0.317, *Med* = 0.291, *SD* = 0.269): participant 2 (*F*(1,116) = 5.201, *p* = 0.0122), and 6 (*F*(1,116) = 4.987, *p* = 0.0137). There was a significant p value for the stimulus type main effect in the one-tailed test when there were none in the two-tailed test.

Table 6.8 presents the number of p values below certain values where the outcome of the equation was in the expected direction. It includes the results from both the one-tailed and two-tailed versions of the binomial regression analysis for the detection group. Overall, there were more p values below 0.05, 0.1, and 0.2 in the one-tailed test compared to the two-tailed test. The number of equation outcomes in the expected direction overall for the interaction shows that most of the equation outcomes (13/16) were in the expected direction. This matches the categorisation group and, again, provides evidence that there may have been an attentional blink effect for most participants, but this was only strong enough to reach significance for a few.

Table 6.8. One and Two-Tailed Results for Detection.

Number of results in the expected direction	*One-Tailed*			*Two-Tailed*		
	Stim. Type	*Coll. Lag*	*Interaction*	*Stim. Type*	*Coll. Lag*	*Interaction*
With significant p values (p <= 0.05)	1	3	2	0	3	2
With p values <= 0.1	4	4	4	1	3	2
With p values <= 0.2	9	9	7	4	4	4
Num. equation outcomes in the expected direction overall	15	11	13	15	11	13

This table presents the number of p values below certain values where the outcome of the equation was in the expected direction for the detection group. Each cell is out of 16.

6.4. Discussion

There were two main aims for this study. Firstly, we wanted to show that the attentional blink can be successfully combined with the Fringe-P3 method to detect familiarity with location stimuli, and secondly, to see how two different search task types (categorisation and detection) might impact the attentional blink.

At the group level, the two-way analyses found significant main effects of collapsed lag and separate lag, and significant two-way interactions between stimulus type and collapsed lag, and between stimulus type and separate lag in both the detection and categorisation task groups. However, there was only a significant main effect of stimulus type in the detection group. The three-way analyses found significant main effects of stimulus type, collapsed lag, and separate lag, and significant two-way interactions between stimulus type and collapsed lag, stimulus type and separate lag, and search task type and stimulus type. There was not a significant main effect of search task type, nor were there significant two-way interactions between search task type and collapsed lag or search task type and separate lag. There were, most importantly, not significant three-way interactions between search task type, collapsed lag, and stimulus type or between search task type, separate lag, and stimulus type.

The significant main effects of separate lag and collapsed lag show that target accuracy was lower inside the 200-500ms attentional blink window than outside it. The significant interactions between stimulus type and separate lag and between stimulus type and collapsed lag show that the attentional blink effect was strongest for the probe. These findings match our hypotheses and demonstrate an attentional blink in both groups. However, there was not a significant main effect of search task type or a significant three-way interaction between search task type, stimulus type, and collapsed lag, showing that there was no significant difference between hits performance for the categorisation and detection tasks overall, and that the attentional blink was not stronger for the categorisation task. These findings are inconsistent with our hypotheses.

At the individual participants' level, there were four significant results for categorisation and five significant results for detection in the two-tailed analyses. There were four significant results for the key interaction between stimulus type and collapsed lag (two for categorisation and two for detection), showing that the attentional blink effect can be found individually. In the additional one-tailed analyses, there were five significant results (two for the interaction) for categorisation and six significant results (two for the

interaction) for detection. Although the majority of individual p values did not reach significance in the one- and two-tailed analyses, the mean and median p values for the interaction in both analyses for the categorisation and detection groups were the correct side of chance for our expectation. The outcome of the interaction equation was in the expected direction for most participants (12/16 for categorisation and 13/16 for detection) despite there being few significant p values. This suggests that there could be an attentional blink effect occurring for most individuals, but it was only strong enough to reach significance for a few participants in the current experiment. The "expected direction" analysis is somewhat analogous to the results that would be seen if a simple machine learning approach were applied, where individuals would simply be classified into the most likely of the two possible outcomes, with no assessment of the confidence in that outcome. Most importantly, though, these findings suggest that future experiments with methodological improvements may detect more significant results at the individual participant's level, which, as a result, could more strongly demonstrate that the Fringe-P3 method combined with the attentional blink paradigm could be used to detect concealed information at the individual participants' level.

The presence of a significant main effect of stimulus type at the group level for the detection task but not for the categorisation task, and a significant interaction between search task type and stimulus type in the three-way ANOVAs, could suggest that the detection task would be more suitable for using the attentional blink to detect concealed information as it resulted in a larger difference between the probe and irrelevant conditions. However, there was no significant main effect of search task type, so the different tasks did not have a significant effect on participants' overall hits performance, and there was no significant three-way interaction between search task type, stimulus type, and collapsed lag, so neither task led to a more pronounced attentional blink than the other. These results are surprising, as we expected there to be a significant main effect of search task type and a significant three-way interaction, with the probe generating a more pronounced attentional blink in the categorisation group than in the detection group, making the categorisation task better for detecting concealed information. This was expected as it was thought that having participants look for a category of target images, as opposed to detecting a specific target image, would force the participants to view the whole of the image space in the streams, thereby preventing them from using search strategies based on specific features or specific locations within the image space that could result in them missing the probe. However, the results did not support these hypotheses.

Potential methodological flaws that may have contributed to these non-significant results and future improvements to these methods are discussed later in this section.

It should also be noted that, although not part of the main analyses, the d' scores in appendix C show that, while there was not a significant main effect of search task type in the hits data, there was a significant main effect of search task type in the d' scores data. D' scores represent target accuracy, taking into account the number of false alarms as well as hits. Overall target accuracy was significantly higher in the detection task group than the categorisation task group. This was expected, as the detection task has specific targets to look for and therefore should be easier than the categorisation task. As the hits data did not have a significant difference, this difference in d' accuracy must be caused by there being significantly more false alarms for the categorisation task, as can be seen in the tables of hits and false alarms in appendix C (tables C.1 and C.2). The d' data was not included in the main analyses as the false alarms could not be separated into the four separate lags or collapsed lags, so could not be used to compare accuracy inside and outside of the attentional blink window, and the binomial regression analysis can only be run on hit and miss data, so d' scores could not be used in the individual participants' level analyses. Therefore, this significant difference in false alarms and accuracy could not be taken into account in the main analyses in the current study.

One of the main limitations encountered with both groups was the number of participants whose total hits scores were at ceiling. This prevented the binomial regression analyses from being run on their data and resulted in 11 participants being excluded from the categorisation analyses and 6 from the detection analyses (these exclusions are explained fully in section 6.2.4). It is thought that so many participants' hits scores reached ceiling due to the 144ms SOA used for both tasks. This SOA was chosen to be faster than the previous Fringe-P3 experiment using location stimuli from chapter 5, which used an SOA of 167ms that was thought to be too slow and allowed non-salient stimuli to be consciously perceived. However, the current SOA was still slower than other Fringe-P3 research that used SOAs of 133ms (Alsufyani et al., 2019; Bowman et al., 2013) and 100ms (Bowman et al., 2014) with famous face and own-name stimuli. The current SOA was chosen to be slower than 133ms because the location images were thought to be more visually complex than names and greyscale faces and, therefore, may need longer in order to be perceived enough to be (subliminally) recognised as salient or not. It is possible, however, that this 144ms SOA was still too slow and allowed participants to consciously see more items in the streams than just the critical stimuli, thus interfering with the attentional blink and leading

to high hit counts in all conditions. Future research could consider using an SOA of 133ms, which has been shown to be suitable for Fringe-P3 research in the famous names experiment in chapter 3 and in the Fringe-P3 experiments by Bowman et al (2013) and Alsufyani et al (2019). However, an even more appropriate suggestion would be to use a staircase procedure to select the best SOA for each participant. This would involve adjusting the SOA during practice trials to find the SOA where the participant's hit rate for the target is 75% and correction rejection rate is 80%. This would ensure that the correct SOA is chosen for each participant so that it is neither too fast nor too slow and reduces the chance of participants' scoring too close to ceiling, thus allowing the binomial regression analysis to work for all participants and reduce the need for exclusions. This staircase procedure was used successfully in chapter 4 for the Fringe-P3 experiment using email address stimuli.

Due to the slow SOA, when the target was presented at lag 3, it was 432ms after the probe or irrelevant, which is close to the end of the 200-500ms attentional blink window. It was thought that participants may have recovered from an attentional blink by the time the target was presented at lag 3, and that may have affected both the collapsed lag and separate lag analyses. Therefore, additional group and individual level analyses comparing lag 1 with lag 7 only were conducted. However, these analyses did not find a substantial improvement of the results at the group or individual participants' level. Therefore, it was concluded that the inclusion of lag 3 (and 5) did not have a substantial impact on the collapsed lag and separate lag analyses. The full results of the lag 1 v 7 analyses can be found in appendix D.

There were two main aims of this chapter: to show that the Fringe-P3 method combined with the attentional blink could be used to detect familiarity with a salient probe, and to investigate how two different search tasks could affect the attentional blink. The results from the two groups successfully act as a proof of concept that the attentional blink can indeed be used to detect familiarity with a probe stimulus in the context of detecting concealed information. The significant interactions between stimulus type and collapsed lag, and stimulus type and separate lag, demonstrate that target accuracy was lowest when the target was presented within 200-500ms of the probe, thus demonstrating an attentional blink for the probe as expected. There were also some significant results at the individual level, showing that there is potential for the method to be used for individual participants. Future research using the suggested methodological improvements may find stronger results, particularly at the individual participants' level. This method could also be

combined with EEG to test for a P3 alongside an attentional blink. This would provide an even more robust test where, for example, even if the participant does not experience an attentional blink after the probe (such as non-blinkers), they may still generate a P3 for the probe and thus suggest they are concealing information. The results of the categorisation versus detection comparisons were inconclusive as there was no significant main effect of search task type or significant three-way interaction between search task type, stimulus type, and collapsed lag using the hits data but there was a significant main effect of search task type using the d' prime data. More research needs to be done to compare the two task types, taking into account the suggested methodological improvements.

7. Matched Filter Convolution Analysis for using EEG to Detect Concealed Information

7.1. Introduction

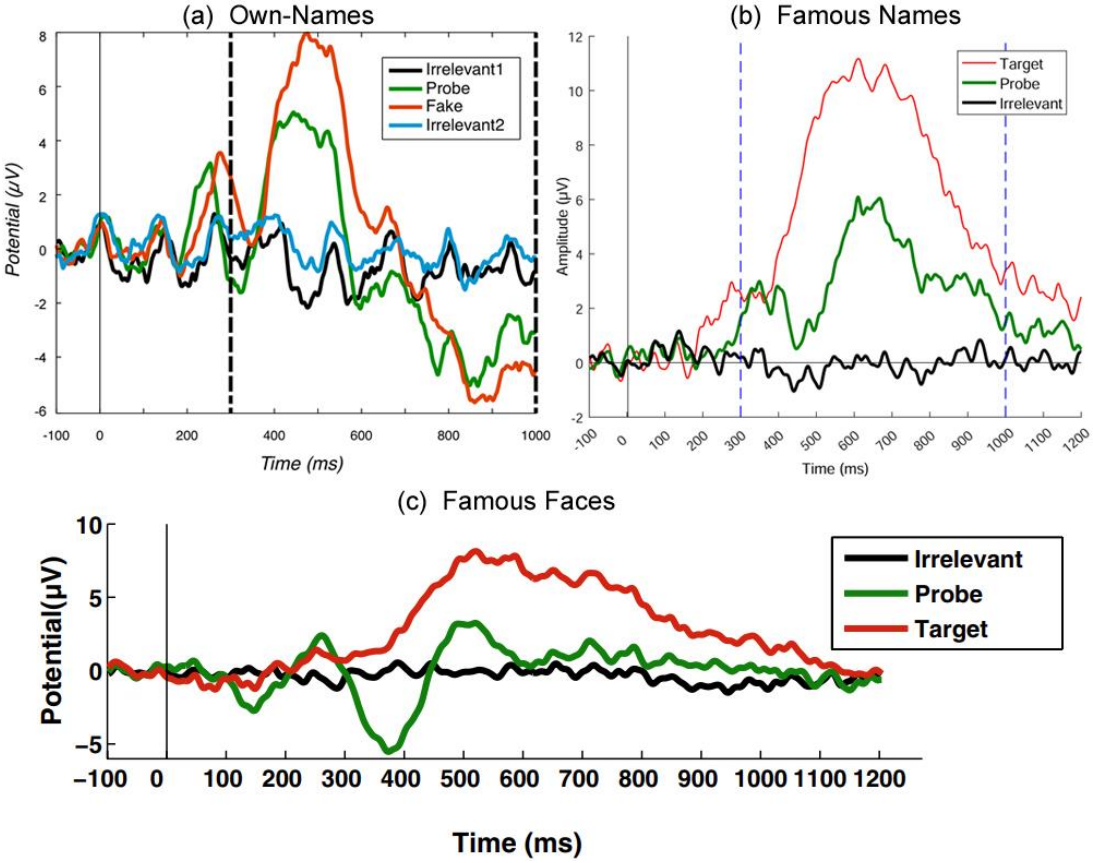
This chapter uses the same email addresses dataset used in chapter 4. For information on the experiment and collection of this data, please see chapter 4. The current chapter explores ways to more accurately classify and detect P3s at the individual participants' level through variations of the AGAT and a matched filter convolution analysis (MFC).

The Fringe-P3 concealed information test can be used to detect concealed information (through the presence of P3s) at both the group and individual participants' level. Real forensic situations are mostly focused on individual suspects, therefore, having a sensitive and reliable analysis for detecting P3s at the individual participants' level is key. The classic P3 (also called the P3b) appears as a positive deflection around 250-800ms post-stimulus, usually peaking around 300-500ms, and is maximal from the Pz electrode (Comerchero & Polich, 1999; Meijer et al., 2014; Polich & Kok, 1995). Sometimes a P3a is also present and appears before the P3b, peaking around 250–280ms, and is maximal from the Fz and Cz electrodes. While there is a classic P3 waveform, the latency, amplitude, breadth, and overall shape of the P3 can vary depending on the stimuli and tasks involved (Polich, 2007, 2012). Any analysis to detect P3s needs to be able take this variability into account in order to be reliable in both experimental and real forensic situations. The matched filter convolution analysis proposed in this chapter attempts to account for such variance.

Figure 7.1 presents the grand averages at Pz from three Fringe-P3 experiments: a) Bowman et al's (2013) first Fringe-P3 experiment using own-names, b) the experiment in chapter 3 using famous names, and c) Alsufyani et al's (2019) experiment using famous faces. These plots demonstrate some ways in which P3s can vary for different stimuli and slightly different methods. The probe and target waveforms in the own-names (a) and famous names (b) plots have large P3 positivities, but the own-names P3s peak between 450ms and 500ms, while the famous names P3s peak later, between 600ms and 650ms. Both also have smaller early positivities before the main P3, but these positivities in the own-names experiment peak within the 150-300ms P3a window while the positivities in the famous names experiment peak outside of the P3a window, so it is unclear whether these are late P3as or not. The probe and target in the own-names experiment also have a negativity following the P3 which is not present for the famous names. The famous names P3s, on the

other hand, are broader than the own-names P3s, which is most likely due to the famous names experiment using multi-item stimuli (first and second names presented consecutively) while the own-names experiment used single-item stimuli (first names only). These two grand averages show that even though the two experiments used very similar methods and both used name stimuli, even a small change to the method and stimuli (multi-item famous names instead of single-item own-names) can have a substantial effect on the ERPs. The famous faces experiment (c), which used pictures of faces as stimuli instead of names, has an even more different grand average. While the target has a large, broad P3, peaking between 500ms and 550ms, similar to the famous names experiment, the probe, on the other hand, does not have a P3 but instead has a negativity followed by a positivity that were interpreted as an N400f followed by a P600. This demonstrates that not only can ERPs vary between different stimulus types and experiments, but they can even vary within the same participant in one experiment with one stimulus type. These variations in P3 amplitude, shape, and latency demonstrate how vital it is that methods of analysing and classifying P3s take into account potential differences from the classic P3 waveform and from P3s from other experiments.

Figure 7.1. Grand Averages at Pz from Fringe-P3 Experiments Using Different Stimuli.



These plots are grand averages at Pz from Fringe-P3 experiments. Plot (a) was taken from Bowman et al's (2013) experiment using own-name stimuli. There is a clear P3a, P3b, and a post-P3 negativity for probe and target. Plot (b) was taken from the experiment using famous name stimuli in chapter 3. There is a clear P3b and a smaller positivity before the P3b that is outside of the 150-300ms P3a window. Plot (c) was taken from Alsufyani et al's (2019) experiment using famous face stimuli. There is a negativity (N400f) followed by a positivity (P600) for the probe while the target only has a positivity (P3) and no negativity.

In Bowman et al's two own-names Fringe-P3 experiments (2013, 2014), peak-to-peak analyses were used, as those datasets had clear negativities and positivities to compare. However, the three Fringe-P3 EEG datasets in this thesis (chapters 3, 4, and 5) did not have clear negativities, so a peak-to-peak analysis would not have been appropriate. Instead, the AGAT, using mean amplitude measures, was used as the main analysis (see section 2.5 for a full description of the AGAT). The AGAT does allow for some variation in the latency and amplitude of the P3, as it searches for a window of interest (on the aggregated ERP) with the highest mean amplitude within a 300-1000ms window. However, this analysis, as applied in the Fringe-P3 work in this thesis, is based purely on peak amplitude and does not take into account negativities or differences in the breadth and overall shape of the P3s. The AGAT could be modified to detect negativities, but only if there is a precedent in the literature for a negative effect. The MFC proposed in the current chapter can take into account all of these details by default.

The core feature of the MFC is that it allows us to search more accurately for a P3 based on the shape of other real P3s from the same experiment, rather than relying on the shape of the classic P3 or P3s from other experiments using different methods or stimuli. By using P3s from the same experiment, we can take into account any differences in latency, shape, or negativities caused by that experiment's method or stimuli.

The MFC is related to previous work on a weight template method by Alsufyani et al (2018) (see also Alsufyani (2018)). The weight template method is focused on inter-participant variability within the same experiment, where, for example, one participant's P3 may have more negativity than the others and so may not be detected and classified as a P3 by other analysis methods. Instead of basing the analysis on mean amplitude or peak-to-peak measures, the weight template analysis classifies P3s for the probe and irrelevant based on an individual-specific template made from that participant's target ERP. P3s are typically strongest for the target, so using the target ERP for the template should provide a clear example of that participant's personal P3 shape. The method begins by normalising the probe, target, and irrelevant ERPs. The weight template is then created from that participant's target-irrelevant difference time series within the P3 window. This template is then applied at a fixed position within the P3 window on the probe-irrelevant difference time series and the two time series are multiplied point-by-point. This increases points where the probe-irrelevant and target-irrelevant differences are both greater (i.e., where a P3 is). The resulting vector is then summed to create the true observed weighted difference and a permutation test is run to find the p value. They found that their weight

template had a higher hit rate at detecting P3s than the peak-to-peak method. However, this method will only work well when the target and probe P3s are similar. If the probe and target P3s have very different shapes or latencies, then the target P3 template may not be similar enough to the probe P3s to correctly detect them. This method will also not work well if the target and probe ERPs contain different components. For example, it can be seen in the grand average from the famous faces experiment (Alsufyani et al., 2019) in figure 7.1c, that there is an N400f negativity for the probe followed by a P600 positivity, whereas the target has no negativity but instead has a large P3 positivity. The weight template would not be appropriate for this data since the target and probe ERPs are so different. The MFC proposed in this chapter, on the other hand, would be suitable.

The MFC does not rely on the target and probe ERPs being similar as it does not use the target ERP as the template. Instead, it averages the probe ERPs from all other participants (leaving out the current participant being analysed) and uses this minus-one grand average as the filter template (see section 7.4.1 for the full procedure). This allows us to then detect P3s for individuals based on their similarity to probe P3s from other individuals in the same experiment. While this does mean that the MFC will not be as well tailored to the shape of each participant's personal P3 shape (as it uses other peoples' P3's as a template), it will be more tailored to the shape of the P3 generated by the specific stimuli and task. The MFC will also slide the filter template along the real ERPs within a broad search window and, therefore, will also take into account individual differences in the latency of the P3.

This chapter will compare the MFC with the AGAT using the EEG data from the email addresses experiment in chapter 4. We expect that the MFC will be able to correctly classify some P3s that the AGAT analysis missed. A significant difference between the probe and irrelevant in the analysis is consistent with there being a P3. Therefore, we will measure the success of the MFC variations by the number of significant p values compared to the AGAT.

7.2. The Dataset and Analysis Methods

7.2.1. The Dataset

The EEG data used for this chapter was the email addresses data used in chapter 4. Full details of the experimental method and data analysis procedures can be found in section 4.2 of chapter 4. The following is a summary of those procedures.

Data from eleven participants, aged 19 - 28 (M = 23.27, SD = 2.9), seven female and four male, were included in the analyses. The EEG data were recorded using the BioSemi ActiveTwo system and ActiView software. Electroencephalographic data were recorded from Fz, Cz, and Pz but only Pz was used in the current analyses as that is where the P3 is expected to be strongest. Electro-oculogram data were recorded from both eyes using two HEOG electrodes on the outer canthus of each eye and two VEOG electrodes, one above and one below the right eye. The linked mastoids were used as a reference and impedances were kept below 10kOhm. The data were digitised at 2048Hz.

The EEG data were then analysed using Matlab 2016a and EEGLAB version 13.6.5b. The data were first resampled to 512Hz, then underwent a high-pass filter of 0.5Hz and low-pass filter of 45Hz, before being epoched into segments from -100ms to 1500ms, time-locked to the onset of the critical stimulus. Trial rejection was performed on the eye electrodes (with above 100μV and below -100μV as criteria), and on the Pz, Fz, and Cz channels (with above 50μV and below -50μV as criteria). Baseline correction was then performed from -100ms to 0ms.

After trial rejection, the maximum number of trials remaining per participant for probe and irrelevant was 60 and the minimum remaining was 46 for probe (M = 54.91, SD = 5.09) and 40 for irrelevant (M = 53.91, SD = 6.49).

Due to the small number of participants (11) in this dataset, we will not be able to report robust statistical differences between the AGAT and MFC methods. Therefore, the inferences made in this chapter must be considered provisional at this stage.

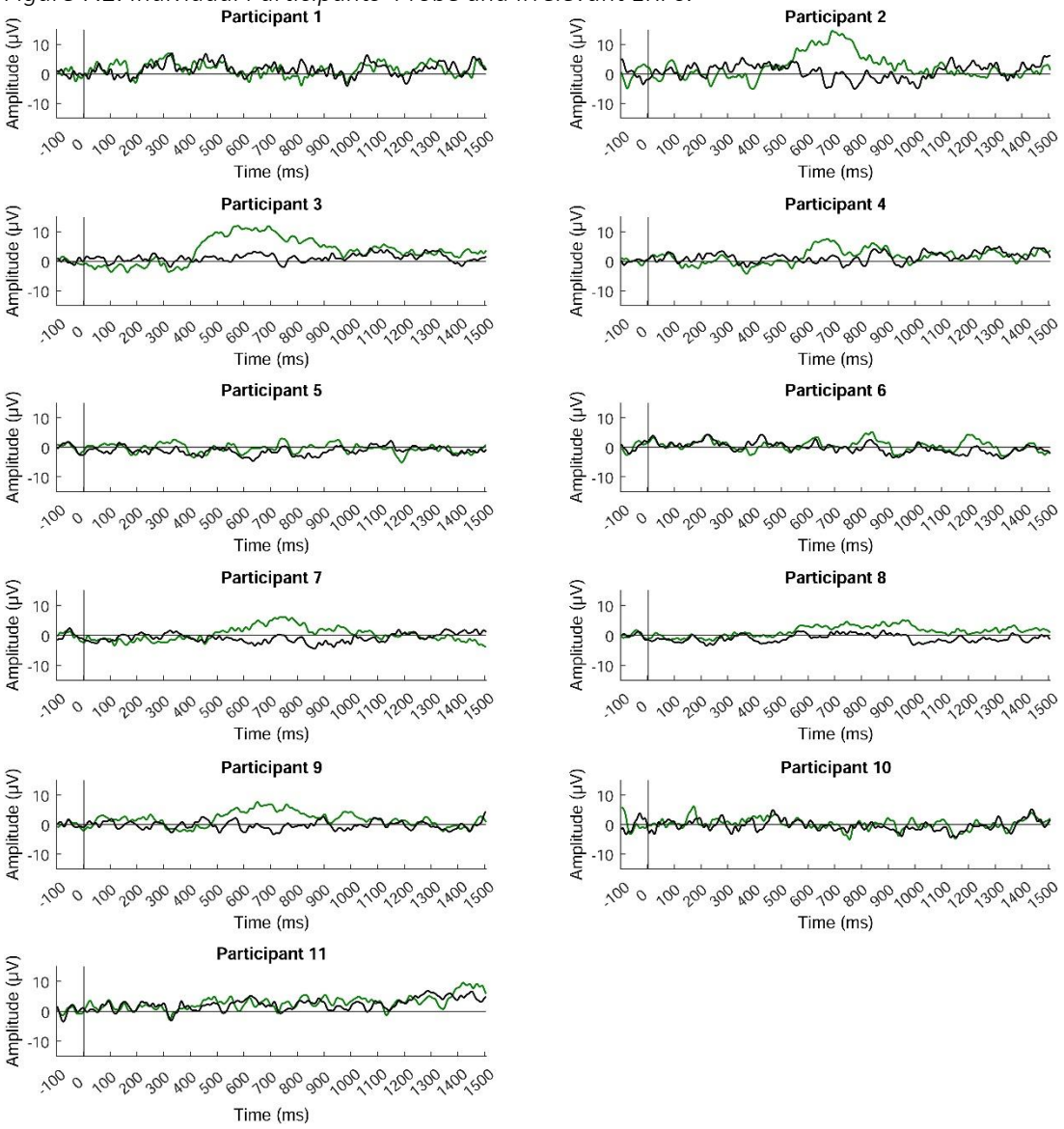
7.2.2. The ERPs and Analyses

All statistical analyses were performed at the individual participants' level, using data from the Pz electrode and the probe and irrelevant conditions only. The analyses used were the AGAT analysis (explained in full in section 2.5) and several versions of the MFC (each explained later in the relevant sections of this chapter). Every version of the AGAT and MFC used permutation tests with 10,000 permutations to calculate the p value.

As mentioned in section 2.5, the individual participants' level analyses treat each participant as a separate experiment and do not compare them with each other, so corrections for the type I error rate are not required. This applies to all versions of the individual AGAT and MFC analyses. Corrections for multiple comparisons are only required in section 7.8.1, when the four measures of the MFC are compared with each other, resulting in six comparisons.

Figure 7.2 shows the individual participants' probe and irrelevant ERPs from Pz and figure 7.3 shows the grand average for probe and irrelevant from Pz.

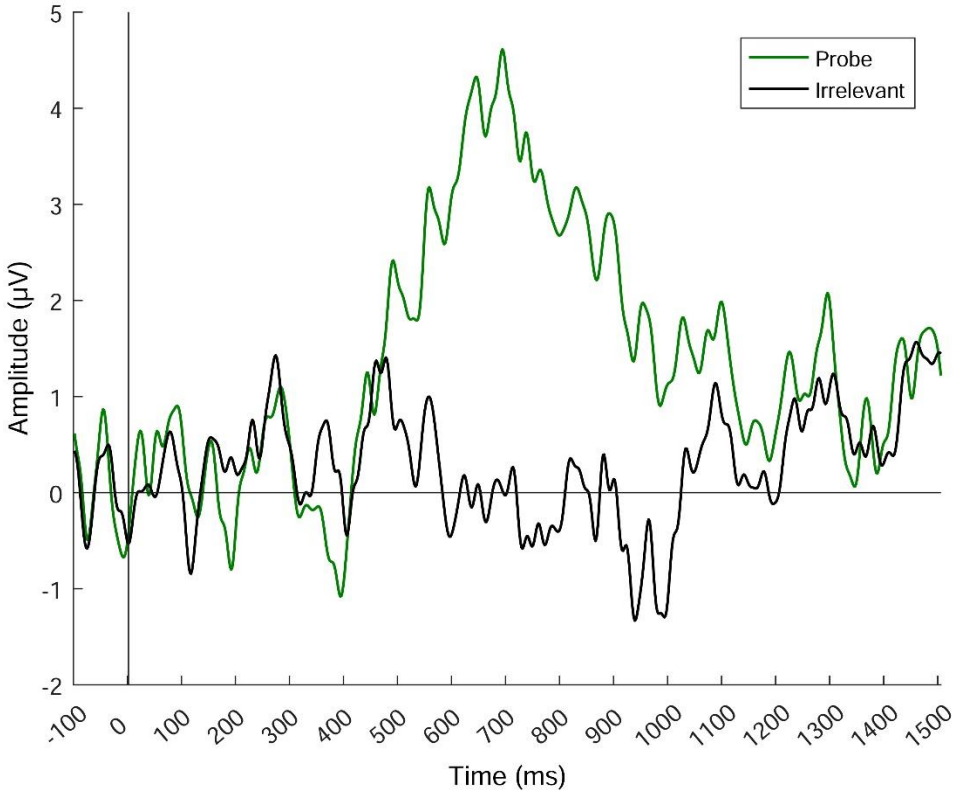
Figure 7.2. Individual Participants' Probe and Irrelevant ERPs.



This figure shows the individual participant's probe (green) and irrelevant (black) ERPs from Pz up to 1500ms post-stimulus.

It can be seen in figure 7.2 that there are clear P3 positivities for the probe but not the irrelevant for participants 2, 3, 4, 7, 8, and 9. Participant 5 has a positivity for the probe around 250-400ms, which could also be a P3, but this is less likely to be detected by the standard AGAT analysis as it starts before the AGAT's 300-1000ms search window. It may, however, be detected by the MFC as it uses a wider search window. Participant 6 has two positivities for the probe around 550-650ms and 750-900ms but the irrelevant also has positive deflections at the same time. Similarly, participant 1 has a positivity for the probe around 200-400ms but the irrelevant also has a positivity at this time. The AGAT is less likely to detect a significant difference for these participants due to similar amplitudes for the probe and irrelevant, however, as the MFC takes into account the shape of the waveform and not just the amplitude, it is possible that the MFC may be able to detect differential P3s for these participants. There is a clearer positivity for the probe around 1150-1300ms for participant 6, however the P3 positivity is typically over by this time, so this is less likely to be a P3. Overall, it is likely that both the AGAT and the MFC will detect P3s (and significant differences) for participants 2, 3, 4, 7, 8, and 9 as they have clear positivities for the probe that are not present for the irrelevant. However, the MFC may also be able to detect P3s for participants 1, 5, and 6, that the AGAT is less likely to detect. Due to the noise and drift in the data, it is less likely that the AGAT or MFC will detect P3s for participants 10 and 11.

Figure 7.3. Grand Average for Probe and Irrelevant.



This figure shows the grand average from Pz for all participants. A clear P3 positivity can be seen for the probe but not the irrelevant.

7.3. 300-1000ms AGAT

The first analysis was the AGAT analysis as described in section 2.5. This AGAT used the standard search window of 300-1000ms, matching the P3 bounding windows used in Bowman et al's own-names experiments (2013, 2014) and in chapters 3 and 5 of this thesis. This is the default analysis and will serve as a baseline against which the MFC variations will be compared.

7.3.1. 300-1000ms AGAT Results

Table 7.1 shows that six of eleven participants (55%) had significant p values.

Table 7.1. Mean Amplitude and P Value Results of the 300-1000ms AGAT.

Participant	*Probe*	*Irrelevant*	*P Value*
1	3.017	4.746	0.7640
2	12.197	-1.259	< 0.0001*
3	10.842	2.551	< 0.0001*
4	6.657	0.722	0.0003*
5	0.945	-0.771	0.1434
6	3.199	0.509	0.1080
7	5.492	-0.919	0.0065*
8	4.141	0.481	0.0457*
9	6.287	-1.045	0.0024*
10	2.623	0.926	0.2719
11	3.853	3.234	0.3971
Mean	*5.387*	*0.834*	*0.158*
Median	*4.141*	*0.509*	*0.046*
Std Dev	*3.468*	*1.945*	*0.240*
Num Sig			*6*

This table shows the probe and irrelevant mean amplitudes in the window of interest and the p values for each participant. 55% of participants (6/11) had significant p values. The asterisk marks p values that were significant at the 0.05 level.

The 300-1000ms window is the standard search window for the AGAT, but a 300-1500ms AGAT was also run, as some versions of the MFC in this chapter have search windows up to 1500ms. There was no substantive difference in results to the standard 300-1000ms version, with the same number of significant p values (6 total, *M* = 0.210, *Med* = 0.038, *SD* = 0.316). Only three participants (7, 13, and 14) had different (later) windows of interest

selected by the 300-1500ms version, none of which were significant in either version of the AGAT. As there was no substantive difference between the 300-1000ms and 300-1500ms AGAT, the standard 300-1000ms AGAT will still be used as the baseline against which the MFC will be compared.

7.4. Non-Weighted Matched Filter Convolution Analysis

The MFC uses a filter template made from other participants' probe ERPs as a template of what the probe P3 for that experiment looks like. The MFC slides this filter template along the individual participant's probe and irrelevant ERPs (separately) and calculates the covariance, correlation, dot product, and cosine at each point. Calculating these distance measures allows us to detect P3s by finding the window where that participant's ERP most closely resembles (e.g., correlates with) the filter template and, therefore, most closely resembles the expected P3 shape for that experiment. An additional objective is to assess the effectiveness of the four distance measures as part of the MFC. By using a leave-one-out procedure to generate the filter template, we will not bias the template towards the current participant being analysed, as their data is left out of their template. This first version of the MFC does not involve any weighting, so is referred to as the non-weighted MFC (a weighted MFC is introduced in section 7.5).

7.4.1. Non-Weighted MFC Procedure

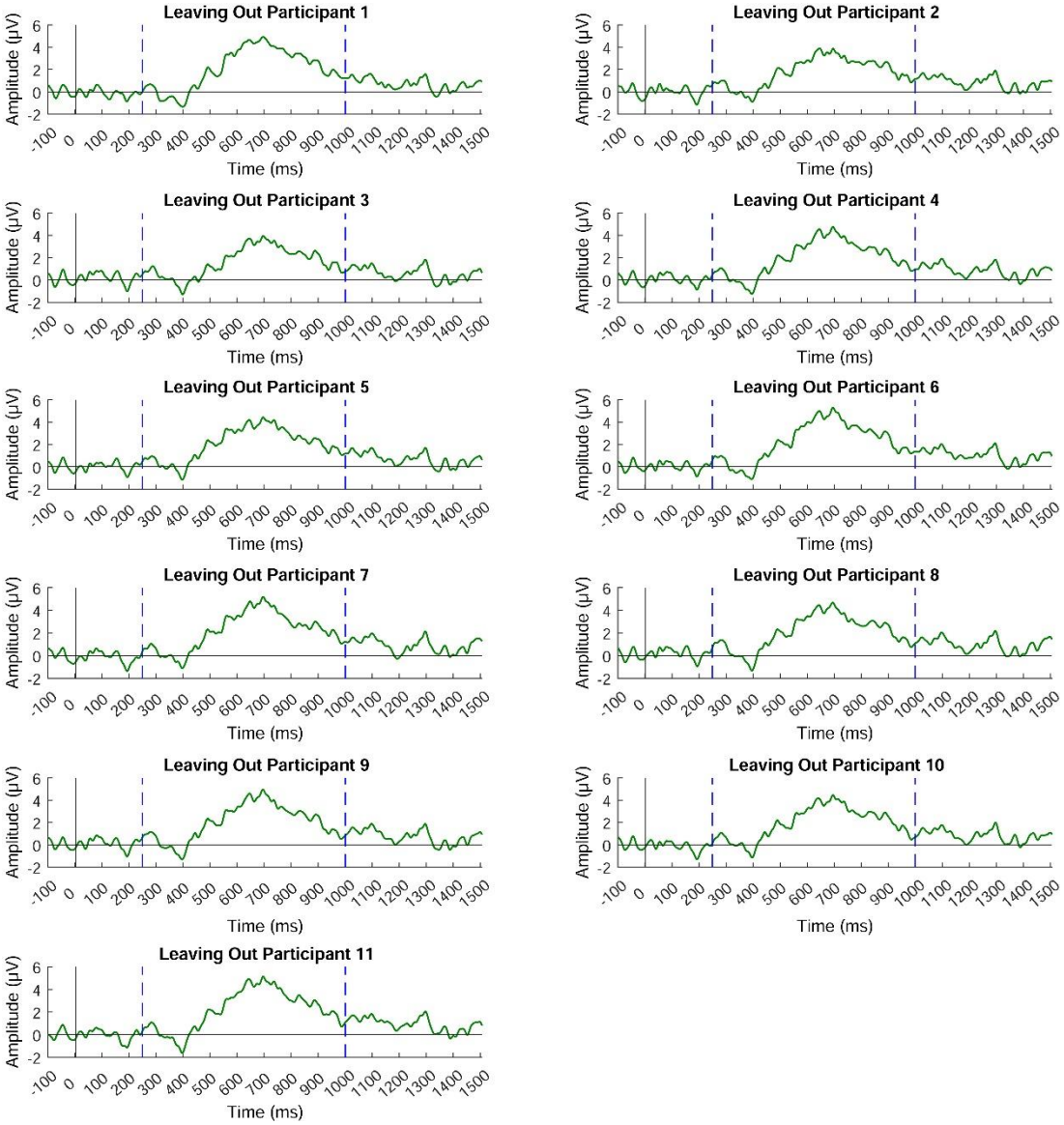
1) Generate a "minus-one" grand average, using a leave-one-out procedure by averaging together the probe ERPs from all participants, excluding the current participant being analysed (referred to from this point onwards as the "current participant"). For example, to analyse participant 1, the minus-one grand average would be the average of participants 2-11. Figure 7.4 presents the minus-one grand averages and the sections used as filter templates for each participant.

2) The filter template is cut from the minus-one grand average from 300ms to 1000ms, making it 700ms long. These filter template sections are marked with vertical blue dashed lines on figure 7.4. The 300-1000ms window for the filter template was chosen because the P3 positivity typically starts around 300ms and has returned to baseline by 1000ms. In experiments where the P3 in the grand average starts earlier than 300ms or ends later than 1000ms, this window will still contain the main body of the positivity so will still provide a reliable template for the P3.

3) The filter template is then placed at the starting point of the search window (e.g., 150ms) on the current participant's probe ERP and slid along the ERP until the upper edge of the template reaches the end of the search window (e.g., 1500ms). At each position, the covariance, correlation, dot product, and cosine are calculated between the filter template and the ERP, resulting in convolution time series. The largest covariance, correlation, dot product, and cosine results are saved. This is then repeated with the irrelevant ERP, using the same filter template. The difference between the largest covariance, correlation, dot product, and cosine results for probe and irrelevant are calculated as the true observed results.

4) The individual trial data then undergoes a permutation test, permuting the current participant's probe and irrelevant trials 10,000 times to create surrogate probe and irrelevant ERPs. The same filter template is then slide along these surrogate ERPs, as in step 3, and the difference between the largest covariance, correlation, dot product, and cosine results for probe and irrelevant are found and saved as the surrogate difference for each permutation. The p value is calculated as the proportion of surrogate differences that are greater than or equal to the true observed difference.

Figure 7.4. Minus-One Probe Grand Averages and Filter Templates.



This figure presents the minus-one probe grand averages from Pz for each participant. The blue vertical dashed lines represent the 300-1000ms window used as the filter template in the analyses. For example, the "leaving out participant 1" plot is the average of participants 2-11 and the section within the blue dashed lines would be used as the filter template to analyse participant 1.

7.4.2. 150-1500ms Non-Weighted MFC

The first version of the non-weighted MFC used a search window of 150-1500ms. The 150ms start point was chosen to allow for individual variability before the typical P3 onset at 300ms. The 1500ms end point was chosen to allow for variability in the latency of the P3 beyond the typical 1000ms end point.

Results

Table 7.2 shows that the covariance measure had the highest number of significant p values of the four measures, with six significant results (55%). This is the same number of significant results as the 300-1000ms AGAT (section 7.3). All other measures had less significant p values than the 300-1000ms AGAT. Interestingly, the 150-1500ms MFC covariance measure found a significant p value for participant 6 but not for participant 8, while the 300-1000ms AGAT found a significant p value for participant 8 but not for participant 6. This shows that, despite finding the same number of significant p values, the covariance measure did not detect P3s for exactly the same participants.

Table 7.2. P Value Results Summary of the 150-1500ms Non-Weighted MFC.

Participant	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
1	0.078	0.0548	0.5616	0.4206
2	0.0005*	0.0423*	0.0068*	0.0175*
3	< 0.0001*	0.0002*	< 0.0001*	0.0014*
4	< 0.0001*	0.0006*	0.1415†	0.2504†
5	0.945	0.9908	0.3537	0.3428
6	0.0486*†	0.0513	0.3625	0.3698
7	0.0227*	0.0536†	0.0482*	0.1386†
8	0.3468†	0.344†	0.0395*	0.0371*
9	0.0012*	0.0155*	0.0123*	0.0175*
10	0.3122	0.3349	0.4256	0.4297
11	0.489	0.5171	0.4615	0.4626
Mean	*0.204*	*0.219*	*0.219*	*0.226*
Median	*0.049*	*0.054*	*0.142*	*0.250*
Std Dev	*0.300*	*0.311*	*0.215*	*0.187*
Num Sig	*6*	*4*	*5*	*4*

This table show the results of the covariance, correlation, dot product, and cosine measures for the MFC with a 150-1500ms search window. The covariance measure had the highest number of significant p values. The asterisk marks p values that were significant at the 0.05 level. The dagger (†) marks p values that have changed significance compared to the 300-1000ms AGAT.

7.4.3. 200-1100ms Non-Weighted MFC

The large 150-1500ms search window from the previous version resulted in some windows of interest being chosen that were late in the ERP, past the point where the P3 has usually finished and were thus not likely to be P3s. For example, all of participant 1's windows of interest, except for the cosine measure, were around 740ms to 1440ms. While the MFC is intended to detect P3s that do not fit the typical shape and timing, it is very unlikely that P3s would appear so late after the stimulus when they typically peak far earlier, around 300-500ms post-stimulus. We cannot risk the MFC misclassifying other ERP components or drift as P3s, therefore, a shorter search window was tested in this second version of the MFC: 200-1100ms. These specific timings were chosen because they provide 100ms of flexibility on either side of the typical 300-1000ms P3 window to allow for variation in the start and end of the P3 for each participant, while still narrowing focus around the 300-1000ms window where P3s are most expected to occur.

Results

It can be seen in table 7.3 that the number of significant results for covariance decreased by one compared to the previous 150-1500ms version (section 7.4.2), while the number of significant results increased by one for correlation and dot product and increased by two for cosine. The cosine measure had the biggest difference in results between the 150-1500ms and 200-1100ms versions of the MFC, but none of the measures had a higher number of significant p values than the 300-1000ms AGAT (section 7.3).

Table 7.3. P Value Results Summary of the 200-1100ms Non-Weighted MFC.

Participant	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
1	0.7922	0.8220	0.7129	0.7393
2	< 0.0001*	< 0.0001*	0.0011*	0.0021*
3	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*
4	0.0001*	0.0002*	0.0244*†	0.0487*†
5	0.0771	0.0742	0.0681	0.0579
6	0.1251†	0.1388	0.3233	0.3295
7	0.0003*	0.0002*†	0.0114*	0.0009*†
8	0.3474	0.3450	0.0367*	0.0345*
9	0.0004*	0.0001*	0.0074*	0.0063*
10	0.8119	0.8453	0.4609	0.4571
11	0.4975	0.5217	0.4137	0.4746
Mean	*0.241*	*0.250*	*0.187*	*0.196*
Median	*0.077*	*0.074*	*0.037*	*0.049*
Std Dev	*0.322*	*0.335*	*0.248*	*0.260*
Num Sig	*5*	*5*	*6*	*6*

This table show the results of the covariance, correlation, dot product, and cosine measures for the MFC with a 200-1100ms search window. The dot product and cosine measures had the highest number of significant p values. The asterisk marks p values that were significant at the 0.05 level. The dagger (†) marks p values that have changed significance compared to the 150-1500ms non-weighted MFC.

7.4.4. 300-1000ms (Fixed Position) Non-Weighted MFC

The final version of the non-weighted MFC did not slide the filter template along the search window, but instead placed it in a fixed position at 300-1000ms. This version was tested to investigate how narrowing the window to the standard P3 window would affect the results and to test whether having a broader search window, as used in the previous versions, was indeed beneficial to detecting P3s.

Results

Table 7.4 shows that restricting the MFC to the 300-1000ms P3 window resulted in one less significant p value for cosine in this version compared to the 200-1100ms version (section 7.4.3). The other measures had the same number of significant results as the 200-1100ms version. The loss of one significant result for cosine suggests that reducing the search window to a fixed 300-1000ms window was too restrictive and having some flexibility around the start and end the typical P3 window is beneficial to detecting P3s.

Table 7.4. P Value Results Summary of the 300-1000ms Non-Weighted MFC.

Participant	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
1	0.7703	0.8108	0.6994	0.7008
2	< 0.0001*	< 0.0001*	0.0007*	0.0008*
3	< 0.0001*	0.0026*	< 0.0001*	0.0001*
4	< 0.0001*	< 0.0001*	0.0252*	0.0574†
5	0.2826	0.2644	0.0738	0.0628
6	0.0801	0.0825	0.2593	0.2601
7	0.0003*	0.0004*	0.0120*	0.0029*
8	0.2465	0.3334	0.0472*	0.0493*
9	0.0008*	0.0006*	0.0066*	0.0040*
10	0.8097	0.8447	0.4567	0.4434
11	0.5927	0.6224	0.4029	0.5196
Mean	*0.253*	*0.269*	*0.180*	*0.191*
Median	*0.080*	*0.083*	*0.047*	*0.057*
Std Dev	*0.323*	*0.339*	*0.240*	*0.252*
Num Sig	*5*	*5*	*6*	*5*

This table shows the results of the covariance, correlation, dot product, and cosine measures for the MFC with a 300-1000ms fixed window. The dot product measure had the highest number of significant p values. The asterisk marks p values that were significant at the 0.05 level. The dagger (†) marks p values that have changed significance compared to the 200-1100ms non-weighted MFC.

Of the three versions of the non-weighted MFC, the 200-1100ms version is the better choice as it has a smaller window than the full 150-1500ms version, and so reduces the chances of detecting a late component that is not a P3, but still allows for individual flexibility in the start and end of potential P3s. The dot product and cosine measures had the highest number of significant results (6 each, 55%) in the 200-1100ms version, which is the same amount as the 300-1000ms AGAT.

7.5. Weighted Matched Filter Convolution Analysis

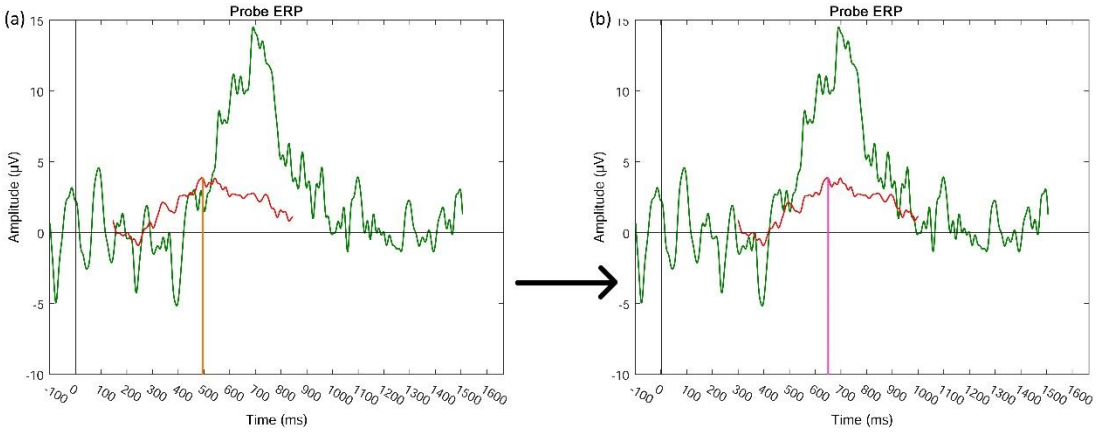
The weighted MFC takes the non-weighted MFC one step further by adding a weighting template. The weighted MFC begins the same as the non-weighted MFC by sliding the filter template along the ERP and creating covariance, correlation, dot product, and cosine time series. However, these time series then undergo additional steps where they are weighted by multiplying point-by-point with the weighting template. This weighting template is made from the full length of the minus-one grand average and does not slide but stays in its "original" position, so it will give greater weighting to points in the convolution time series that are closer (in time) to the peak of the weighting template (and therefore are closer to the peak of the minus-one grand average and are more likely to be P3s). The further in time a point of the convolution time series is from the peak of the weighting template, the less emphasis it will receive. High convolution results (e.g., a strong correlation) that occur close in time to the peak of the minus-one grand average are more likely to be P3s and will receive greater weighting, whereas high convolution results that occur very early before or very late after the peak are less likely to be P3s and will receive less weighting. This should lead to greater accuracy in the detection of P3s as it places additional emphasis on the results that most closely resemble the real probe P3 pattern for that stimulus type but still allows for variance in the latency of genuine P3s while reducing the chance of incorrectly classifying other early or late positivities as P3s when they are unlikely to be.

7.5.1. Weighted MFC Procedure

Figures 7.5, 7.6, and 7.7 provide visual aids for the following explanation using participant 2's data and the covariance measure as examples.

1) Follow steps 1-3 of the non-weighted MFC procedure in section 7.4.1. The weighted MFC starts the same as the non-weighted MFC, by sliding the filter template along the individual participant's ERP (as shown in figure 7.5) and calculating the covariance, correlation, dot product, and cosine at each point. This results in time series of covariance, correlation, dot product, and cosine results, as shown in figure 7.6a. The orange and pink vertical lines in figure 7.5a and b mark the peak of the filter template (red waveform) as it slides along the probe ERP (green waveform).

Figure 7.5. Example of the Filter Sliding and the Filter's Peak Position Timings.



This figure uses participant 2's data as an example of how the filter template slides along the participant's probe ERP. The orange and pink vertical lines show the position of the peak of the filter as it slides.

2) The weighting template uses the full length of the minus-one grand average (-100 to 1500ms). As the minus-one grand average can be negative at points, the absolute values are used for creating the weighting template. The absolute minus-one grand average ($absGA$) is then normalised by dividing each point by the sum of all points and multiplying by the total number of points, resulting in the weighting template ($Wtemp$; figure 7.6b).

$$Wtemp(t) = \frac{absGA(t).N}{\sum_{t=1}^{N} absGA(t)}$$

Where t is time in the time series and N is the length of the time series.

3) The weighting template is then applied to the convolution time series (e.g., time series of covariance results in figure 7.6a). Each point in the convolution time series is then multiplied point-by-point with the weighting template (figure 7.6b).

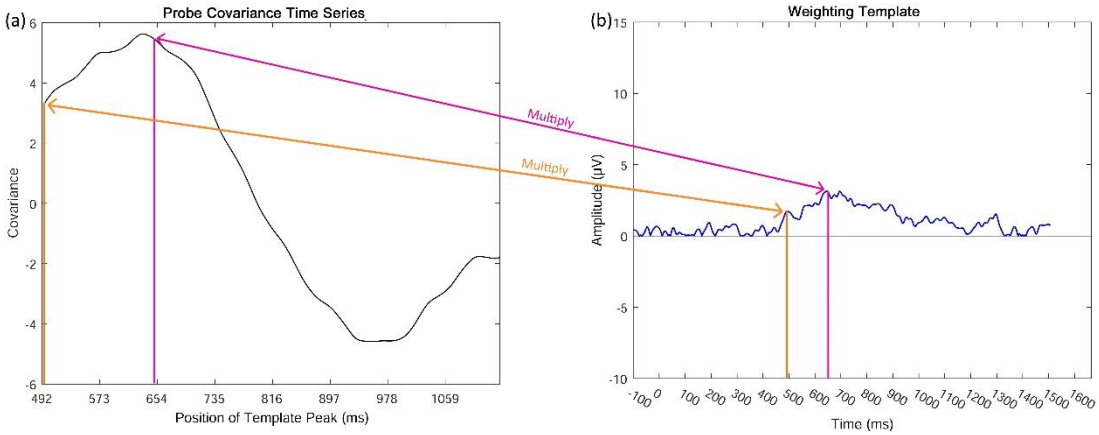
$$res(t) = Wtemp(t) \,.\, Conv(t)$$

Where Conv is the convolution time series.

Each point in the convolution time series represents a window of 700ms (not one millisecond). Therefore, the x-axis of the convolution time series (figure 7.6a) is labelled using the position of the peak of the filter template over the original ERP within the search window. The x-axis begins with the position of the peak when the template was first placed over the original ERP (the orange line in figure 7.5a) and increases by one time point (e.g., 1ms) each time the filter slides. The millisecond of the convolution time series x-axis is paired with the same millisecond from the weighting template for the point-by-point

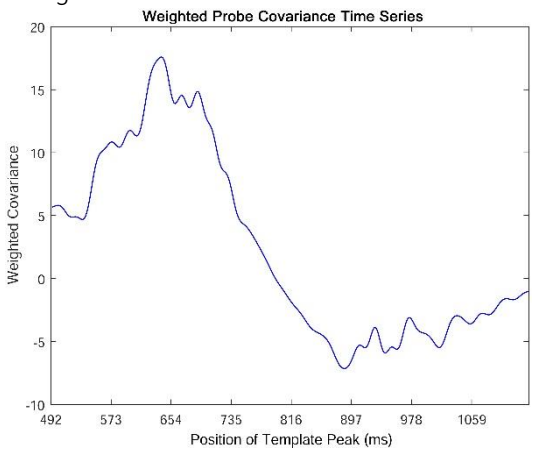
multiplication (as shown by the orange and pink lines in figure 7.6) to create the weighted convolution time series (figure 7.7). The highest result is then saved as the true observed value.

Figure 7.6. Example of a Probe Covariance Time Series and a Weighted Template.



This figure shows an example of a convolution time series (i.e., probe covariance time series) and a weighting template. The orange and pink lines demonstrate which milliseconds of the time series are multiplied together.

Figure 7.7. Example of a Weighted Covariance Time Series.



This figure shows the final weighted covariance time series.

4) The data then undergoes a permutation test, similar to step 4 of the non-weighted MFC but with the additional weighting step. The current participant's probe and irrelevant trials are permuted to create surrogate probe and irrelevant ERPs. The same filter template is then slide along these surrogate ERPs to create the surrogate covariance, correlation, dot product, and cosine convolution time series, which are then weighted using the same weighting template according to step 3 above and the highest value saved. The p value is calculated as the proportion of surrogate results that are above or equal to the true observed value.

7.5.2. 150-1500ms Weighted MFC

The first version of the weighted MFC used a search window of 150-1500ms. This was chosen to match the first version of the non-weighted MFC to compare how much of an affect the weighting had on the results.

Results

Table 7.5 shows that the 150-1500ms weighted MFC had a higher number of significant results than the 150-1500ms non-weighted MFC (section 7.4.2) for the correlation, dot product, and cosine measures, suggesting that the weighting did improve the ability to detect P3s for those measures. The covariance measure had one less significant result in the weighted MFC version.

Table 7.5. P Value Results Summary of the 150-1500ms Weighted MFC.

Participant	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
1	0.2454	0.2256	0.6823	0.7201
2	< 0.0001*	0.0037*	0.0030*	0.0069*
3	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*
4	< 0.0001*	< 0.0001*	0.0330*†	0.0594
5	0.2591	0.2970	0.2057	0.2056
6	0.0727†	0.0762	0.3089	0.3148
7	0.0002*	0.0003*†	0.0113*	0.0275*†
8	0.2776	0.3377	0.0425*	0.0377*
9	0.0004*	0.0011*	0.0061*	0.0042*
10	0.6207	0.6554	0.4381	0.4402
11	0.5110	0.5411	0.4029	0.5246
Mean	*0.181*	*0.194*	*0.194*	*0.213*
Median	*0.073*	*0.076*	*0.043*	*0.059*
Std Dev	*0.223*	*0.238*	*0.234*	*0.252*
Num Sig	*5*	*5*	*6*	*5*

This table show the results of the covariance, correlation, dot product, and cosine measures for the MFC with default weighting and a 150-1500ms search window. The dot product measure had the highest number of significant p values. The asterisk marks p values that were significant at the 0.05 level. The dagger (†) marks p values that have changed significance compared to the 150-1500ms non-weighted MFC.

7.5.3. 200-1100ms Weighted MFC

While the 150-1500ms weighted MFC had a higher number of significant results than the 150-1500ms non-weighted MFC, none of the measures had a higher number of significant results than the 200-1100ms non-weighted MFC. Therefore, this next version tested the weighted MFC with the shorter 200-1100ms window, to see if using a shorter window would further aid the weighted MFC in detecting genuine P3s and preventing the detection of late non-P3 positivities.

Results

Table 7.6 shows that the 200-1100ms weighted MFC had one additional significant p value for every measure compared to the 150-1500ms weighted MFC (section 7.5.2). This shows that the shorter window, again, prevented late peaks being selected that are unlikely to be P3s and instead selected earlier peaks that were more likely to be P3s. The current version also had one additional significant p value for the covariance, correlation, and dot product measures compared to the 200-1100ms non-weighted MFC (section 7.4.3), suggesting that the weighting did aid the classification of P3s. Most importantly, the current version resulted in the highest number of significant p values so far out of all analyses, including the AGAT (7/11 for the dot product measure, 64%). This shows that using a combination of the weighted MFC and the shorter search window led to an increased ability to detect P3s.

Table 7.6. P Value Results Summary of the 200-1100ms Weighted MFC.

Participant	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
1	0.9980	0.9953	0.7140	0.7575
2	< 0.0001*	< 0.0001*	0.0013*	0.0031*
3	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*
4	< 0.0001*	< 0.0001*	0.0269*	0.0593‡
5	0.0020*†‡	0.0005*†‡	0.0004*†‡	0.0015*†‡
6	0.0738	0.0800	0.3151	0.3211
7	< 0.0001*	< 0.0001*	0.0002*	< 0.0001*
8	0.2806	0.3441	0.0409*	0.0376*
9	< 0.0001*	< 0.0001*	0.0042*	< 0.0001*
10	0.9725	0.9823	0.4659	0.4598
11	0.5157	0.5430	0.4042	0.5231
Mean	*0.258*	*0.268*	*0.179*	*0.197*
Median	*0.002*	*0.001*	*0.027*	*0.038*
Std Dev	*0.395*	*0.398*	*0.253*	*0.272*
Num Sig	*6*	*6*	*7*	*6*

This table show the results of the covariance, correlation, dot product, and cosine measures for the MFC with default weighting and a 200-1100ms search window. The dot product measure had the highest number of significant p values. The asterisk marks p values that were significant at the 0.05 level. The dagger (†) marks p values that have changed significance compared to the 150-1500ms weighted MFC and the double dagger (‡) marks p values that have changed significance compared to the 200-1100ms non-weighted MFC.

7.6. α Weighted Matched Filter Convolution Analysis

The final version of the weighted MFC used an α parameter to vary the strength of the weighting. This was to investigate if there was a particular strength of weighting that led to more optimised classification of P3s than either the non-weighted or the default fully weighted versions. The α used were 0, 0.25, 0.5, 0.75, and 1. Where α = 0, the weighting is the same as the default full weighting. Where α = 1, it will result in a time series of 1s, making the weighting effect the same as using no weighting. If varying the α makes a substantial difference to the number of significant results, then the next version could involve an algorithm that finds the best α value for that participant, rather than using one of the five arbitrary α values chosen here. All α versions used a search window of 150-1500ms. This search window was chosen to allow full flexibility for the MFC and to see how titrating the α may impact potential incorrect detection of late peaks that are not P3s.

7.6.1. α Weighted MFC Procedure

1) Follow steps 1 and 2 of the default weighted MFC procedure in section 7.5.1.

2) After creating the weighting template, introduce an α parameter to affect the strength of the weighting, where 0 ≤ α ≤ 1.

$$\overline{W}temp(t) = (\alpha.1) + ((1-\alpha).Wtemp(t))$$

3) Step 3 here is the same as step 3 of the default weighted MFC procedure in section 7.5.1 but using the α weighting template ($\overline{W}temp$) in place of the default weighting template ($Wtemp$) for the point-by-point multiplication with the convolution time series.

$$res(t) = \overline{W}temp(t) \,.\, C(t)$$

4) Perform the permutation test the same as in step 4 of the default weighted MFC procedure in section 7.5.1 but continue using the α weighting template ($\overline{W}temp$) in place of the default weighting template ($Wtemp$) for the point-by-point multiplication with the surrogate convolution time series.

7.6.2. α Weighted MFC Results

The following five tables (7.7 through to 7.11) show the results of the different α weighted filters. The tables all show the results of the covariance, correlation, dot product, and cosine measures for their respective α weighted MFC.

Table 7.7 shows the results of the α weighted MFC where α = 0. An α of 0 is the same as using the default full weighting and the number of significant p values here matches the number found in the 150-1500ms default fully weighted MFC (section 7.5.2), as expected.

Table 7.7. P Value Results Summary Where α = 0.

Participant	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
1	0.2461	0.2244	0.6815	0.7206
2	< 0.0001*	0.0031*	0.0017*	0.0037*
3	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*
4	< 0.0001*	< 0.0001*	0.0292*	0.0581
5	0.2611	0.3008	0.2100	0.2101
6	0.0705	0.0777	0.3108	0.3156
7	0.0003*	0.0016*	0.0132*	0.0307*
8	0.2817	0.3437	0.0413*	0.0380*
9	0.0004*	0.0016*	0.0078*	0.0045*
10	0.6083	0.6444	0.4347	0.4360
11	0.5219	0.5553	0.4073	0.5282
Mean	*0.181*	*0.196*	*0.194*	*0.213*
Median	*0.071*	*0.078*	*0.041*	*0.058*
Std Dev	*0.223*	*0.238*	*0.234*	*0.253*
Num Sig	*5*	*5*	*6*	*5*

This table presents the p values for each measure. The dot product measure had the highest number of significant p values. The asterisk marks p values that were significant at the 0.05 level.

Table 7.8 shows the results of the α weighted MFC where α = 0.25. The cosine measured had one less significant result than the α = 0 version. The other measures had the same number of significant p values as the α = 0 version.

Table 7.8. P Value Results Summary Where α = 0.25.

Participant	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
1	0.1731	0.1499	0.6823	0.7085
2	0.0001*	0.0044*	0.0023*	0.0060*
3	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*
4	< 0.0001*	< 0.0001*	0.0331*	0.0575
5	0.3179	0.3689	0.2746	0.2588
6	0.0671	0.0727	0.3122	0.3190
7	0.0007*	0.0026*	0.0161*	0.0677†
8	0.2955	0.3717	0.0460*	0.0419*
9	0.0003*	0.0019*	0.0098*	0.0152*
10	0.5870	0.6258	0.4422	0.4442
11	0.5116	0.5395	0.3992	0.5126
Mean	*0.178*	*0.194*	*0.202*	*0.221*
Median	*0.067*	*0.073*	*0.046*	*0.068*
Std Dev	*0.220*	*0.239*	*0.234*	*0.246*
Num Sig	*5*	*5*	*6*	*4*

This table presents the p values for each measure. The dot product measure had the highest number of significant p values. The asterisk marks p values that were significant at the 0.05 level. The dagger (†) marks the p value that has changed significance compared to the α = 0 version.

Table 7.9 shows the results of the α weighted MFC where α = 0.5. All measures had the same number of significant results as the α = 0.25 version.

Table 7.9. P Value Results Summary Where α = 0.5.

Participant	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
1	0.1340	0.1074	0.6808	0.6757
2	< 0.0001*	0.0087*	0.0020*	0.0054*
3	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*
4	< 0.0001*	< 0.0001*	0.0423*	0.0587
5	0.4233	0.4830	0.3266	0.3078
6	0.0648	0.0684	0.3235	0.3321
7	0.0012*	0.0040*	0.0225*	0.1130
8	0.3332	0.3786	0.0400*	0.0357*
9	0.0007*	0.0029*	0.0110*	0.0189*
10	0.5596	0.5888	0.4431	0.4458
11	0.5184	0.5504	0.4100	0.5175
Mean	*0.185*	*0.199*	*0.209*	*0.228*
Median	*0.065*	*0.068*	*0.042*	*0.113*
Std Dev	*0.228*	*0.246*	*0.237*	*0.239*
Num Sig	*5*	*5*	*6*	*4*

This table presents the p values for each measure. The dot product measure had the highest number of significant p values. The asterisk marks p values that were significant at the 0.05 level.

Table 7.10 shows the results of the α weighted MFC where α = 0.75. The dot product measure had one less significant p value than α = 0.5 version. The other measures had the same number of significant results as the α = 0.5 version.

Table 7.10. P Value Results Summary Where α = 0.75.

Participant	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
1	0.0880	0.0638	0.6472	0.6260
2	0.0001*	0.0128*	0.0027*	0.0077*
3	< 0.0001*	< 0.0001*	< 0.0001*	< 0.0001*
4	< 0.0001*	< 0.0001*	0.0600†	0.0830
5	0.6674	0.7976	0.3538	0.3402
6	0.0567	0.0612	0.3256	0.3345
7	0.0009*	0.0072*	0.0321*	0.1270
8	0.3501	0.3621	0.0415*	0.0351*
9	0.0017*	0.0053*	0.0122*	0.0164*
10	0.4516	0.4758	0.4327	0.4339
11	0.5086	0.5367	0.4086	0.4974
Mean	*0.193*	*0.211*	*0.211*	*0.227*
Median	*0.057*	*0.061*	*0.060*	*0.127*
Std Dev	*0.251*	*0.283*	*0.229*	*0.226*
Num Sig	*5*	*5*	*5*	*4*

This table presents the p values for each measure. The covariance, correlation, and dot product measures were tied for the highest number of significant p values. The asterisk marks p values that were significant at the 0.05 level. The dagger (†) marks the p value that has changed significance compared to the α = 0.5 version.

Table 7.11 shows the results of the α weighted MFC where α = 1. An α of 1 is the same as using no weighting. The number of significant results for the covariance and correlation measures here are different to those found in the 150-1500ms non-weighted MFC (section 7.4.2), but these differences can be explained simply by the randomisation in the permutation procedure, as both of the p values that changed significance are close to 0.05.

All measures had the same number of significant results as the α = 0.75 version.

Table 7.11. P Value Results Summary Where α = 1.

Participant	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
1	0.0667	0.0461*†‡	0.5551	0.4238
2	0.0002*	0.0438*	0.0048*	0.0188*
3	< 0.0001*	0.0002*	< 0.0001*	0.0011*
4	< 0.0001*	0.0009*	0.1482	0.2551
5	0.9469	0.9930	0.3518	0.3433
6	0.0510‡	0.0510	0.3642	0.3700
7	0.0221*	0.0556†	0.0437*	0.1372
8	0.3410	0.3349	0.0356*	0.0315*
9	0.0022*	0.0131*	0.0132*	0.0193*
10	0.3168	0.3347	0.4285	0.4341
11	0.4988	0.5256	0.4658	0.4607
Mean	*0.204*	*0.218*	*0.219*	*0.227*
Median	*0.051*	*0.051*	*0.148*	*0.255*
Std Dev	*0.302*	*0.312*	*0.215*	*0.188*
Num Sig	*5*	*5*	*5*	*4*

This table presents the p values for each measure. The covariance, correlation, and dot product measures were tied for the highest number of significant p values. The asterisk marks p values that were significant at the 0.05 level. The dagger (†) marks p values that have changed significance compared to the α = 0.75 version and the double dagger (‡) marks p values that have changed significance compared to the 150-1500ms non-weighted MFC, most likely due to the randomisation in the permutation.

As can be seen from tables 7.7 through to 7.11, the number of significant p values slowly decreased for the dot product and cosine measures as α increased. The number of significant results for the covariance and correlation measures did not change. Varying the α did not greatly affect the number of significant results overall and did not lead to a higher

number of significant results than the 150-1500ms non-weighted or default weighted MFCs. Therefore, the α weighted MFC was tested no further.

The 200-1100ms default weighted MFC (section 7.5.3) remains the version with the highest number of significant results.

7.7. 250-1000ms AGAT

After seeing that the version of the MFC with the highest number of significant p values used a search window of 200-1100ms and had detected several P3s that started before 300ms, we needed to also test the AGAT with an earlier starting point for the search window for a fair comparison. Therefore, the AGAT was tested using a 250-1000ms search window instead of the typical 300-1000ms window. While using a 250-1000ms window would be too unnecessarily restrictive for the MFC - which was specifically intended to allow for variation - this is a suitable window for the AGAT as it still allows for some variation at the start while also remaining close to the 300-1000ms search window that has been tested and successfully used in the AGAT for previous Fringe-P3 research.

7.7.1. 250-1000ms AGAT Results

As can be seen in table 7.12, the 250-1000ms AGAT resulted in seven significant p values (64%). This is an increase of one additional significant result compared to the 300-1000ms AGAT (section 7.3) and matches the highest number of significant results found by the 200-1100ms weighted MFC (section 7.5.3). The additional significant result was for participant 5, whose window of interest (selected by the current AGAT) was 263-363ms. This is consistent with this participant having a P3 for the probe that was not detected by the previous AGAT because it began before the standard 300-1000ms search window.

Table 7.12. Mean Amplitude and P Value Results of the 250-1000ms AGAT.

Participant	*Probe*	*Irrelevant*	*P Value*
1	4.792	3.537	0.2803
2	12.197	-1.259	< 0.0001*
3	10.842	2.551	< 0.0001*
4	6.657	0.722	0.0004*
5	1.681	-0.999	0.0264*†
6	3.199	0.509	0.1125
7	5.492	-0.919	0.0062*
8	4.141	0.481	0.0357*
9	6.287	-1.045	0.002*
10	2.623	0.926	0.2710
11	3.853	3.234	0.4027
Mean	*5.615*	*0.703*	*0.103*
Median	*4.792*	*0.509*	*0.026*
Std Dev	*3.293*	*1.744*	*0.145*
Num Sig			*7*

This table shows the probe and irrelevant mean amplitudes in the window of interest and the p values for each participant. 64% of participants (7/11) had significant p values. The asterisk marks p values that were significant at the 0.05 level. The dagger (†) marks the p value that has changed significance compared to the 300-1000ms AGAT.

7.8. Comparing the AGAT and MFC

Table 7.13 presents the median p values for each analysis method and measure. The median is the more appropriate measure of central tendency than the mean, due to the skewed distribution of p values, so is more suitable for comparing the MFC and AGAT.

Table 7.13. Median P Values for Each Analysis Method and Measure.

Analysis Method	*AGAT*	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
AGAT 300-1000ms	0.046				
AGAT 250-1000ms	0.026*				
Non-Weighted MFC 150-1500ms		0.049	0.054	0.142	0.250
Non-Weighted MFC 200-1100ms		0.077	0.074	0.037	0.049
Non-Weighted MFC 300-1000ms		0.080	0.083	0.047	0.057
Weighted MFC 150-1500ms		0.073	0.076	0.043	0.059
Weighted MFC 200-1100ms		0.002	0.001*	0.027	0.038
Weighted MFC α = 0 150-1500ms		0.071	0.078	0.041	0.058
Weighted MFC α = 0.25 150-1500ms		0.067	0.073	0.046	0.068
Weighted MFC α = 0.5 150-1500ms		0.065	0.068	0.042	0.113
Weighted MFC α = 0.75 150-1500ms		0.057	0.061	0.060	0.127
Weighted MFC α = 1 150-1500ms		0.051	0.051	0.148	0.255
Min	*0.026*	*0.002*	*0.001*	*0.027*	*0.038*
Max	*0.046*	*0.080*	*0.083*	*0.148*	*0.255*

This table contains the medians of every analysis method and MFC measure in this chapter. The smallest median p values for the AGAT and the MFC are marked by an asterisk.

The 200-1100ms weighted MFC has the smallest median p values for each of the four measures across all versions of the MFC. The correlation measure in this version had the smallest median p value (p = 0.001) overall, but it was the dot product measure in this version that resulted in the highest number of significant p values across all measures and versions of the MFC. The median p value for the 250-1000ms AGAT was smaller (p = 0.026) than the 300-1000ms AGAT (p = 0.046). All four measures of the 200-1100ms weighted

MFC had smaller median p values than the 300-1000ms AGAT, but only the correlation and covariance measures had smaller median p values than the 250-1000ms AGAT (although the dot product median p value was only very slightly larger).

Table 7.14 presents the number of significant p values found for each analysis method and measure. The 250-1000ms AGAT and the dot product measure of the 200-1100ms weighted MFC had the highest number of significant p values (7).

Table 7.14. Number of Significant P Values for Each Analysis Method and Measure.

Analysis Method	*AGAT*	*Covariance*	*Correlation*	*Dot Product*	*Cosine*
AGAT 300-1000ms	6				
AGAT 250-1000ms	7*				
Non-Weighted MFC 150-1500ms		6	4	5	4
Non-Weighted MFC 200-1100ms		5	5	6	6
Non-Weighted MFC 300-1000ms		5	5	6	5
Weighted MFC 150-1500ms		5	5	6	5
Weighted MFC 200-1100ms		6	6	7*	6
Weighted MFC α = 0 150-1500ms		5	5	6	5
Weighted MFC α = 0.25 150-1500ms		5	5	6	4
Weighted MFC α = 0.5 150-1500ms		5	5	6	4
Weighted MFC α = 0.75 150-1500ms		5	5	5	4
Weighted MFC α = 1 150-1500ms		5	5	5	4
Min	*6*	*5*	*4*	*5*	*4*
Max	*7*	*6*	*6*	*7*	*6*

Both the 250-1000ms AGAT and the dot product measure of the 200-1100ms weighted MFC resulted in 7 significant p values, the highest number of significant results overall (marked by an asterisk in the table).

7.8.1. Statistical Analysis

AGAT – MFC Comparisons

The individual participants' p values were used to compare the AGAT and the MFC. The 200-1100ms weighed MFC had the smallest median p values and the highest number of significant p values of all versions of the MFC, so this version of the MFC was compared to the AGAT. The correlation measure had the smallest median p value and the dot product had the highest number of significant p values, so these two measures of the 200-1100ms weighted MFC were compared against both the 300-1000ms original AGAT and the 250-1000ms extended AGAT.

To compare the individual p values, they were first transformed by a probit function (the inverse cumulative distribution of the standard normal distribution). Some of the original p values were 0 and would become "–infinity" following the probit transformation (which cannot be used in the statistical analyses), therefore, p values of 0 were changed to 0.00001 before undergoing the probit transformation. The resulting values for both versions of the AGAT and the dot product measure of the weighted MFC were found to be normally distributed but the correlation measure was not, so Wilcoxon's signed-rank tests were used to compare the transformed p values. Table 7.15 presents the medians of the original and probit-transformed p values for the 300-1000ms and 250-1000ms AGAT analyses and the dot product and correlation measures of the 200-1100ms weighted MFC.

Table 7.15. Median P Values for the AGAT and MFC Analyses.

	300-1000ms AGAT	*250-1000ms AGAT*	*Weighted MFC Dot Product*	*Weighted MFC Correlation*
Original P Values	0.046	0.026	0.027	0.001
Probit-transformed P Values	-1.688	-1.937	-1.928	-3.291

This table contains the medians of the original and probit-transformed p values from the 300-1500ms and 250-1000ms AGAT analyses and the dot product and correlation measures of the 200-1100ms weighted MFC. All of the original median p values are below 0.05.

The Wilcoxon's signed-rank test did not find a significant difference between the 300-1000ms AGAT and the 200-1100ms weighted MFC's dot product measure, $Z = -.0561$, $p = 0.5751$, or between the 300-1000ms AGAT and the correlation measure, $Z = 0$, $p = 1$.

There was also not a significant difference between the 250-1000ms AGAT and the 200-1100ms weighted MFC's dot product measure, Z = -1.274, p = 0.2026, or the 250-1000ms AGAT and the correlation measure (Z = -0.267, p = 0.7896).

These analyses found no significant differences between the 200-1100ms MFC and both versions of the AGAT. Therefore, no further statistical analyses were conducted to compare the AGAT and MFC methods. While this does not support our hypothesis that the MFC would be able to detect P3s that the AGAT did not (and thus have more significant p values), the results suggest that the AGAT also did not perform significantly better than the MFC and, instead, they performed similarly.

MFC Measure Comparisons

All versions of the MFC used four different measures of distance: covariance, correlation, dot product, and cosine. The dot product measure of the 200-1100ms weighted MFC resulted in the highest number of significant p values (7) out of all measures across all versions of the MFC, but the correlation measure of the 200-1100ms weighted MFC had the smallest median p value (p = 0.001). In order to compare the four measures, all of the individual p values for each measure across every version of the MFC were combined before undergoing the same probit transformation procedure used for the AGAT-MFC comparison. There are 10 versions of the MFC and 11 participants, so there were 110 p values for each measure in the analysis. The transformed p values for the four measures were all found to be normally distributed, so paired t-tests were conducted on the data. Table 7.16 presents the medians of the original and transformed p values and table 7.17 presents the results of the paired t-tests on the transformed p values.

Table 7.16. Median P Values for the Four Measures Across all Versions of the MFC.

	Covariance	*Correlation*	*Dot Product*	*Cosine*
Original P Values	0.066	0.058	0.045	0.065
Probit-transformed P Values	-1.508	-1.569	-1.697	-1.512

This table contains the medians of the original and probit-transformed p values for the four measures across every version of the MFC. Dot product is the only measure with an original median p value below 0.05.

As there are six comparisons, the Holm-Bonferroni correction was applied to adjust the significance level for each comparison accordingly. To perform the Holm-Bonferroni correction, the original p values are first ranked from smallest to largest. The significance level for each comparison is then adjusted as: *significance level = α / (n – rank + 1)*. For example, the significance level for the covariance-correlation comparison is 0.05/(6-1+1) = 0.0083.

Table 7.17. Results of the Paired t-tests on the Four Measures' Transformed P Values.

Comparison	*t*	*p*	*Holm-Bonferroni Corrected Significance Level*
Covariance - Correlation	-5.522	< 0.0001*	0.0083
Covariance - Cosine	-4.455	< 0.0001*	0.01
Covariance - Dot Product	-3.833	0.0002*	0.0125
Dot Product - Cosine	-3.213	0.0017*	0.0167
Correlation - Cosine	-2.449	0.0159*	0.025
Correlation - Dot Product	-1.594	0.1139	0.05

This table contains the results of the paired t-tests on the four measures across every version of the MFC. The degrees of freedom were 109. The Holm-Bonferroni correction was applied and the resulting significance level for each comparison is listed. The only comparison that was not significant was the correlation-dot product comparison.

The results of the paired t-tests in table 7.17 show that there were significant differences for all of the comparisons except the correlation-dot product comparison. These results, combined with the median p values in table 7.16, provide evidence that, overall, the correlation and dot product measures had significantly smaller p values than the covariance and cosine measures, so may be the more appropriate measures for the MFC. There was not a significant difference between the correlation and dot product measures, suggesting that they may be equally appropriate for future MFC research.

7.9. Discussion

This chapter tested several variations of a new MFC analysis and compared it against the AGAT. A significant difference ($p <= 0.05$) between probe and irrelevant is consistent with there being a P3. Therefore, we used the p values to compare the AGAT and the MFC. We expected that the MFC would be able to detect some P3s that the AGAT did not and, as a result, the MFC would find more participants with significant p values. As mentioned earlier in section 7.2.1, the small number of participants (11) limits our ability to report robust statistical differences between the AGAT and the MFC. Therefore, the inferences in this discussion must be considered provisional.

The standard AGAT used a search window of 300-1000ms and found significant results for six participants. This number was used as a baseline against which the MFC was compared. As can be seen in table 7.14, only one version of the MFC, the 200-1100ms weighted MFC, resulted in a higher number of significant results than the original 300-1000ms AGAT. The dot product measure in the 200-1100ms weighted MFC resulted in seven significant results while all other versions of the MFC resulted in six or fewer significant p values. The 200-1100ms weighted MFC also had the smallest median p values for each of the four measures across all versions of the MFC and had the smallest median p value overall (for correlation) as can be seen in table 7.13. As the 200-1100ms MFC used a search window with an earlier start point than the 300-1000ms AGAT and found a P3 for participant 5 that started before 300ms, the AGAT was then run again with a 250-1000ms search window for a fairer comparison. This version of the AGAT also found significant results for the same seven participants. Therefore, the MFC did not appear to demonstrate an increased ability to detect P3s over the AGAT, once the AGAT's search window was extended to account for an earlier onset for P3s. Statistical analyses were conducted on the individual p values comparing the dot product and correlation measures of the 200-1100ms weighted MFC with both the 300-1000ms and 250-1000ms AGAT and found no significant differences. This shows that neither the AGAT nor MFC performed significantly better than the other on this dataset.

Six participants (2, 3, 4, 7, 8, and 9) had clear P3s in their ERPs (figure 7.2) and significant p values in both versions of the AGAT and in at least one measure in every version of the MFC. Two participants, 10 and 11, did not have significant p values in any analyses and had noisy ERPs, so simply may not have had P3s to detect. Participant 1 only had a significant p value ($p = 0.0461$) in the 150-1500ms α weighted MFC where $\alpha = 1$, which was the same as having no weighting. Participant 1's p value had not been significant in the equivalent 150-

1500ms non-weighted MFC version (p = 0.0548), but this was marginal and likely due to the randomisation in the permutation procedure. However, the window of best fit (where the covariance, correlation, dot product or cosine result was highest) selected for this participant was late in the search window (around 740ms to 1440ms), so it is less likely that this was a P3. This participant did not have significant p values in any of the analyses using shorter search windows ending at 1100ms or 1000ms. Participant 6 only had a significant p value for the covariance measure in the 150-1500ms non-weighted MFC (p = 0.0486) but this was not significant in any other measure or version of the MFC.

Participant 5 was the participant whose change of result from non-significant to significant led to the highest number of significant results (seven) and had significant p values for all measures in the 200-1100ms weighted MFC and the 250-1000ms AGAT. In these analyses, the selected window of interest for the AGAT and the window of best fit for the MFC had started between 250ms and 300ms, whereas in versions of the MFC with longer search windows, this participant had a window of best fit that ended later than 1100ms. This suggests that detection of a P3 for this participant using the MFC was aided by the shorter search window and the elimination of later peaks that were unlikely to be P3s. However, this participant's p values in the 200-1100ms non-weighted MFC were just above significance for all measures, with p values between 0.0579 and 0.0771, while their p values in the weighted version with the same search window were between 0.0004 and 0.002. This suggests that the shorter window alone could not account for their significant result in the 200-1100ms weighted MFC, but that the addition of the weighting also aided the detection of the P3 for this participant using the MFC. Both the weighting and the shorter window, (separately) increased the number of significant results found, but it was the combination of the two that led to the greatest improvement and the highest number of significant p values.

While the MFC did not perform significantly better than the AGAT overall on this dataset, the fact that the results of the weighted MFC with a shorter search window matched the results of the AGAT with the extended search window, suggests that the MFC could be at least equal to the AGAT. Further research should replicate these methods using the weighted MFC with a 200-1100ms search window and the AGAT with a 250-1000ms search window on different EEG datasets to see if the overall findings remain the same as this study or if one method will result in the correct classification of more P3s than the other with different data. Additionally, including a larger number of participants in further

research would enable us to report more robust statistical differences between the two methods.

It would also be interesting to compare the AGAT and the weighted MFC to Alsufyani et al's weight template method (2018). In particular, they should be compared on a) datasets similar to the famous faces dataset (Alsufyani et al., 2019), where the probe and the target ERPs do not have similar shapes, and b) on datasets where the probe and target ERPs do have similar shapes. The results of these investigations could help to demonstrate the situations in which the different analyses would be most suitable.

Overall, the dot product measure detected the highest number of significant results in most versions of the MFC. The dot product measure in the 200-1100ms weighted MFC had the highest number of significant p values and the correlation measure had the smallest median p value of the four measures and of all MFC variations. The results of the statistical comparisons of the measures found that both the dot product and correlation measures had significantly smaller p values overall than the covariance and cosine measures, but there was not a significant difference between the correlation and dot product measures. This suggests that the correlation and dot product measures may be the most suitable for use with the MFC. Further research into the MFC could pay extra attention to the 200-1100ms weighted version using the dot product and correlation measures.

While the 200-1100ms weighted MFC variation matched the number of significant p values from the 250-1000ms AGAT, more research needs to be done to demonstrate its reliability and what type of data it is most suitable for before it could be suggested for use in place of the AGAT. Therefore, the AGAT remains our recommended analysis for detecting P3s using the Fringe-P3 method for the moment. Future research using the AGAT method to detect P3s should consider extending the search window to 250-1000ms to account for variability in the onset of a P3 positivity.

8. Conclusions and Future Work

This thesis generalised the Fringe-P3 method by demonstrating its efficacy with a wider variety of stimuli beyond the own-name and famous face stimuli that had been used previously (Alsufyani et al., 2019; Bowman et al., 2013, 2014). This thesis first tested the Fringe-P3 method with famous name (chapter 3), email address (chapter 4), and location image stimuli (chapter 5) using EEG. It then tested the use of the Fringe-P3 method with a behavioural measure, the attentional blink, which could be used as an alternative to or in addition to EEG and the P3 to detect concealed information (chapter 6). This attentional blink experiment also investigated two target search tasks (categorisation and detection) and their impact on the strength of the attentional blink and participants' use of visual search strategies. Finally, chapter 7 tested a potential alternative method of analysing EEG data: the matched filter convolution analysis.

8.1. Generalising the Fringe-P3 Method using EEG

The Fringe-P3 method had previously been demonstrated using own-name (Bowman et al., 2013, 2014) and famous face stimuli (Alsufyani et al., 2019) and detected familiarity with the probe stimuli for 100% of participants. Own-names have particularly strong salience, being perhaps the most frequently rehearsed stimuli we encounter, and there are thought to be areas of the brain specifically dedicated to processing and recognising faces. It could be argued that the own-names and famous faces experiments had such high detection rates simply because of the highly salient stimuli. Therefore, the Fringe-P3 method needed to be tested with other stimuli to demonstrate that it can be used with stimuli that are not as uniquely salient. Therefore, this thesis tested the Fringe-P3 method with three different stimulus types: famous names, email addresses, and location images.

The EEG experiment in chapter 3 tested the Fringe-P3 method with name stimuli, using famous names as the probes. The first aim of this experiment was to show that the strong results from the previous own-name experiments (Bowman et al., 2013, 2014) were not limited to only the highly salient own-name stimuli and the Fringe-P3 method could have similarly strong results using other familiar name stimuli that are still salient, but less so than own-names. The second aim was to show that the Fringe-P3 method could work with multi-item stimuli, in the form of first and last name pairs presented consecutively. The analyses found highly significant results at the group level and significant results for 12 of 14 participants (86%) at the individual participants' level, thereby demonstrating that the Fringe-P3 method can detect familiarity with both famous name stimuli and multi-item

stimuli. This opens up the method for use with other familiar name stimuli and with other multi-item stimuli. Both of these are beneficial for the method's potential real world application in forensic investigations as the method could be used to detect familiarity with, for example, the names of accomplices or victims as well as multi-item information such as addresses and dates.

The EEG experiment in chapter 4 tested the Fringe-P3 method with email address stimuli, using the participant's own email address as the probe. It aimed to demonstrate that the Fringe-P3 method could be used to detect familiarity with online identities, beginning with email addresses. This study found significant results at the group level and for 7 of 11 participants (64%) at the individual participants' level, thereby successfully providing proof of concept that the Fringe-P3 method can be used to detect concealed information of an online identity. This opens up the Fringe-P3 method to be used with other online identifiers, such as usernames. The real world application of this would be in situations where the police already know a person's real world identity and suspect that person may be the user of a specific online identity that is associated with a crime. The Fringe-P3 method could then show that suspect is familiar with that online identity, thereby linking them to it. While this experiment successfully detected familiarity with the participant's own email addresses for 64% of participants, it also highlighted some potential methodological flaws that may have prevented the method working for more participants at the individual participants' level. These included the potential impact of participants using search strategies to help them find the target that also, as a side effect, reduce their chances of seeing the probe and prevents the Fringe-P3 method from working as accurately. The experiment in chapter 6 investigated two target search tasks (categorisation and detection) that may prevent participants from using such search strategies.

The EEG experiment in chapter 5 tested the Fringe-P3 method with pictures of locations as stimuli, using photographs of the participants' own university campus as the probe stimuli. The first aim of this experiment was to demonstrate that the Fringe-P3 method could be used to detect familiarity with pictures of locations. The second aim was to show that the Fringe-P3 method could be used with image stimuli that are more visually complex than the greyscale face stimuli that had been used previously (Alsufyani et al., 2019). This would help forensic investigations link suspects to locations involved in crimes and would open up the method for use with other image stimuli beyond only greyscale faces. The experiment found a large P3 for the target and significant differences between the target and the

irrelevant at the group level and for five of the eight participants (63%) at the individual participant's level. This shows the Fringe-P3 method can detect familiarity with target location stimuli. However, there were no significant differences between the probe and the irrelevant at the group or individual participants' level. There were small positivities for both probe and irrelevant that may have been weak, temporally-jittered P3s, but as these were weak and present for both probe and irrelevant, there was no significant difference between the two. Additionally, there was a negativity following the positivity only for the probe, which could be related to the MERMER (Farwell & Smith, 2001) that has been highlighted as a marker of concealed information, but further research would be needed to investigate if probe location images can indeed generate a detectable MERMER. Since there were significant results for the target stimuli, this shows that location stimuli can generate strong P3s and suggests that the probe stimuli may not have generated significantly stronger P3s than the irrelevant stimuli due to a problem with the method or with the probe or irrelevant stimuli themselves, rather than because they were location images. It was thought that the slow SOA used for this experiment allowed the non-salient irrelevant and distractor stimuli to break through into awareness, which affected the results. It was also thought that participants may have used search strategies to help them find the target which then decreased their chances of seeing the probe, the same as in the email addresses experiment. A future experiment with methodological improvements (e.g., for SOA and stimuli) should be conducted to see if such improvements lead to a more robust test that can detect familiarity with probe location stimuli. An example of a potential future experiment and methodological improvements is described in section 8.4.6.

The finding of significant results at the individual level for the EEG famous names, EEG email addresses, and attentional blink locations experiments is especially important, as the real-world application of the Fringe-P3 method would be for individual suspects in forensic investigations. Additionally, finding significant effects at the group level is not a major challenge, but finding significant results at the individual level is less common and more difficult (due to less data and more susceptibility to noise), so to find significant results for more than half of the participants in these experiments is especially noteworthy.

8.2. The Attentional Blink and Search Strategies

The experiment in chapter 6 used location image stimuli to investigate the use of the attentional blink paradigm with the Fringe-P3 method as an alternative measure for detecting concealed information that could be used instead of or in addition to EEG. The

first aim of this chapter was to show that the attentional blink paradigm could be combined with the Fringe-P3 method to detect familiarity with location image stimuli. This study found significant results at the group level for both the categorisation and detection experiments, which shows that participants did experience attentional blinks when the target was presented shortly after the probe stimulus and, therefore, suggests that they were familiar with the probe. There were also some significant results at the individual participants' level, which suggests that the method could potentially be used to detect familiarity in individuals, however more research with some methodological improvements is needed to demonstrate this more strongly. As with the EEG experiment using location stimuli, it was thought that problems with the SOA had negatively impacted the results.

The second aim of this chapter was to compare the categorisation and detection tasks to see how these affected the accuracy of the Fringe-P3 method and, through this, how these tasks impacted the effectiveness of participants using visual search strategies. It was thought that the categorisation task would prevent participants using search strategies and lead to a stronger attentional blink than the detection task, since participants in the categorisation task were searching for a category of target and did not know in advance what the specific target looked like, so could not search for a specific known feature. The experiment found that there was no significant difference between the two search tasks using the hits data but did find a significant difference using the d' data. The information provided by the d' data, however, was limited as d' scores cannot be associated with specific lags, so can only provide information on differences in overall target accuracy between search tasks and between the probe and irrelevant conditions but cannot be used to compare accuracy inside and outside the attentional blink window. The hits data was also limited by many participants scoring close to ceiling, which reduces our capacity to be able to detect an attentional blink. Therefore, the results of the tasks comparisons were inconclusive and future research would be needed to draw firmer conclusions on any differences between the two tasks.

8.3. The Matched Filter Convolution Analysis

The final study, in chapter 7, tested a potential alternative method for analysing EEG data, the MFC. This was compared against the AGAT analysis, which has been used for previous Fringe-P3 research, including the three EEG experiments in this thesis. It was thought that the MFC would be better able to account for variance in the shape and latency of the P3, as well as any negativities, and, therefore, would be able to detect differential responses that the AGAT analysis had missed. This study found that using the weighted MFC with a search

window of 200-1100ms found a significant p value (which is consistent with there being a differential P3) for one additional participant (seven total) that the standard 300-1000ms AGAT had not found to be significant (six total). However, when the search window for the AGAT was extended to 250-1000ms to allow for more variation in the onset of the P3, the AGAT also found a significant p value for the same additional participant, equalling the weighted MFC in finding seven significant results overall.

The dot product measure of the 200-1100ms weighted MFC had the highest number of significant p values and the correlation measure had the smallest median p value of all measures and versions of the MFC. The 300-1000ms and 250-1000ms AGAT were compared against the correlation and dot product measures of the 200-1100ms weighted MFC and we found no significant differences. Therefore, the MFC did not demonstrate an increased ability to detect P3s over the AGAT on this dataset. More research is needed to investigate if the AGAT and MFC still perform equally or if one performs better with other datasets. The four measures of the MFC were also compared, and the correlation and dot product measures were found to have significantly smaller p values overall than the covariance and cosine measures, but there was not a significant difference between the correlation and dot product measures. This suggests that correlation and dot product may be the more suitable measures for use with the MFC and future research using the MFC should consider focusing on these two measures. It was also suggested that future research using the AGAT analysis use a search window that starts at 250ms instead of 300ms to allow for individual differences in the onset of the P3.

8.4. Limitations and Future Work

The email address and location experiments in this thesis had some limitations which could be investigated and/or overcome in future work. The main limitations were often shared with more than one experiment and included participants using search strategies and problems with the SOA and distractor databases.

8.4.1. Search Strategies

The key limitation that is thought to have affected both the email addresses experiment in chapter 4 and the locations EEG experiment in chapter 5 was participants using search strategies to help them find the target. These search strategies involve looking for a specific feature of the target within the streams (e.g., the first letter of an email address or a turret in the top left corner of a location image) which, as a result, means that all other stimuli that do not share the feature, including the probe, are less likely to breakthrough into

awareness. This, then, decreases our chances of detecting the participant's familiarity with the probe, if they frequently miss the probe due to using this search strategy. This fits with the glance-look model of cognitive control, which posits that stimuli in RSVP streams are first "glanced" at to process the broad information about the stimulus (such as category or if it starts with a specific letter) and stimuli only receive a further "look" to understand the rest of the details if the stimulus matches the target category or feature. We suspect that some participants used this search strategy to help them find the email address and location targets, which caused them to frequently miss the probe stimuli and contributed to why we did not detect differential responses for some participants.

To overcome the problem search strategies pose, a task must be found that forces participants to view the whole of the stimulus space and prevents them focusing on only one feature or area. The attentional blink experiment in chapter 6 attempted to investigate this by comparing two tasks: categorisation and detection. It was thought that the categorisation task would prevent participants using search strategies and lead to a stronger attentional blink than the detection task. However, the results of the comparisons were inconclusive. This experiment had several methodological issues that are likely to have impacted the results, including problems with the SOA and distractor databases (discussed in the next two sections). Future research could replicate this experiment with some methodological improvements, to see if there is indeed any difference in the magnitude of the attentional blink from the two tasks. Additionally, this experiment did not specifically inform participants about search strategies nor did it instruct participants to use or avoid using search strategies. Another future experiment could use the classic attentional blink paradigm (without a CIT) with two targets instead of a probe and a target (to avoid any potential issues with probe salience) to compare three groups: one where participants are given no information or instruction regarding search strategies (the control group), a second group where participants are given information on search strategies but are instructed not to use them, and a third group that are given information on search strategies and are instructed to use them. The group that is instructed to use search strategies could provide information on how much search strategies impact the attentional blink, while the group that is instructed not to use search strategies could provide a guideline for what the attentional blink may look like if a task was found that prevented the use of search strategies. A more detailed description of this potential future experiment can be found in section 8.4.7.

8.4.2. Stimulus Onset Asynchrony

The SOA was 167ms in the EEG locations experiment and 144ms in the attentional blink locations experiment. These SOAs were slower than the famous names experiment in chapter 3 and the previous own-names (Bowman et al., 2013, 2014) and famous faces (Alsufyani et al., 2019) experiments, which used SOAs of 100ms and 133ms. The slower SOAs were chosen for the two locations experiments as it was thought that location images were more visually complex than names and greyscale faces, so would need more time to be perceived enough to be (subliminally) recognised as familiar. However, it is now thought that the slower SOAs were too slow and allowed participants to consciously see the non-salient irrelevant and distractor stimuli as well as the salient stimuli. For the EEG experiment, there was evidence in the recognition tests that the irrelevants had broken through into awareness enough to be given significantly higher recognition scores than unpresented images. There was also evidence of a small positivity in the ERP data for the irrelevant that was similar to the probe ERP, further suggesting that the irrelevants did break through into awareness. For the attentional blink experiment, the slower SOA was thought to have allowed participants to see (almost) all of the target stimuli and resulted in many participants scoring hits near or at ceiling, which reduced our ability to detect an attentional blink. This also meant that 17 participants had to be excluded because the binomial regression analysis cannot complete if participants' hits are at ceiling. Therefore, a faster SOA is needed for future experiments using location image stimuli to prevent these problems from reoccurring.

Future Fringe-P3 experiments should consider using an SOA of at least 133ms (to match previous Fringe-P3 experiments) or using a staircase procedure to select the best SOA for each participant. The staircase procedure works by titrating the SOA during practice trials (featuring a target stimulus and distractor stimuli but no probe or irrelevant stimuli) until that participant's hit rate is, for example, 75% and their correct rejection rate is 80%. By personalising the SOA for each participant, it will ensure that the SOA is not so slow that the participant consciously sees all of the stimuli (including the non-salient stimuli) and their hits reach ceiling, but also not so fast that the task is too hard and the probe is unlikely to break through into conscious awareness. A staircase procedure was successfully used for the email addresses experiment in chapter 4, with SOAs ranging between 100ms and 250ms, and found stronger results at the individual participants' level than the two locations experiments.

8.4.3. Distractor Stimuli

Another methodological issue for both of the locations experiments were the distractor databases. These databases contained around 400 distractors, which is much smaller than the databases used in previous Fringe-P3 research. For example, the email addresses experiment in chapter 4 had a database of 3667 distractors, and the famous and own-names experiments (chapter 3; Bowman et al, 2013, 2014) had databases of 1,000-10,000 distractors. The smaller databases led to distractors repeating more than once in the streams, which, combined with the slow SOA, meant that participants were likely to start seeing and recognising distractors, which could have confounded the results. However, the names experiments used a name generator website to generate thousands of names and the email addresses experiment used code to generate thousands of email addresses. Location image stimuli cannot be randomly generated, so selecting distractors for image experiments is not as easy as for text based experiments. We extracted our target and distractor location images from Konkle et al's (2010) database of over 4000 photographs of scenes, but most of these scenes were not suitable for use as distractors in our experiments. As our critical stimuli were all photographs of the exteriors of buildings, the distractor stimuli also needed to be building exteriors. Therefore, all photographs that were not of the exteriors of buildings had to be removed from the database, which left around 400 suitable distractors. The famous faces experiment by Alsufyani et al (2019) also used photographs as stimuli and had a database of 524 distractors and the same number of trials (225) as the locations EEG experiment. They found strong results at both the group and individual participants' level, which suggests that approximately 500 distractors were enough to prevent interference from repetition. Future research could consider using a distractor database of at least 500 distractors per 200 trials to reduce concerns about the distractors repeating. On the other hand, Avilés et al (2020) found that repeating non-salient stimuli did not increase their chances of being consciously seen for the first time in RSVP streams with SOAs of 84–133ms. Therefore, it may be that the slow SOAs of 144ms and 167ms used in the locations experiments were the main reason the distractors were consciously seen and using a faster SOA would prevent the distractors being consciously seen regardless of the number of repetitions, even with a small distractor database.

8.4.4. Other Limitations and Suggestions for Improvement

The two locations experiments used photographs from two university campuses as the probe and irrelevant stimuli. As such, both experiments would benefit from having the counterbalanced studies run at the other university, with the probe and irrelevant stimuli switched. This would allow us to be more confident that the generation of P3s or

attentional blinks for the probe stimuli was due to them being familiar and not due to another factor, such as higher visual salience. For example, if the Birmingham images generate P3s and attentional blinks when used as probes with Birmingham students but the Kent images do not when used as probes with Kent students, then it would suggest that there is an issue with the Kent stimuli and we cannot confidently conclude that the Birmingham stimuli stood out to Birmingham students solely due to familiarity. On a related note, future Fringe-P3 experiments using image stimuli should attempt to control for visual salience as much as possible, to reduce the possibility of this impacting the results.

The email addresses experiment in chapter 4 was also limited by potential perceptual issues. All email address stimuli were generated according to the rules used to create the University of Birmingham's real email addresses and consisted of three letters followed by three numbers. As a result, all of the email address stimuli were very homogenous. Research has shown that distractors that are too similar to a target can capture attention (Folk et al., 1992; Su et al., 2011), so the lack of variety in the email addresses could have meant that the probe did not stand out more easily than the irrelevant or distractors for some participants. Future research should ensure, where possible, that there is greater heterogeneity in the stimuli used to prevent potential interference from similar non-salient stimuli.

8.4.5. Other Future Work

The MFC needs to be tested on a wider variety of datasets in order to evaluate its accuracy and reliability with other data beyond the email addresses dataset. Several versions of the MFC should be tested, but particular attention should be paid to the weighted MFC using a 200-1100ms search window, as that version detected the highest number of significant results on the email addresses dataset. Future research should continue to compare the MFC to the AGAT to investigate if they continue to perform equally or if one is able to detect more P3s than the other on different datasets. The MFC should also be compared to Alsufyani et al's weight template (2018) on a variety of datasets. It is possible that each of these three analyses may be more appropriate for data with different features. For example, it was thought the MFC would be more suitable than the weight template for detecting P3s in datasets where the target and probe ERPs are very different, as the weight template uses the target as a template to detect P3s for the probe (which relies on them being similar), whereas the MFC uses other participant's probe ERPs as the template. By comparing these three analyses, we may find that each analysis has types of datasets for

which it will be more or less reliable. This, then, could lead to certain analyses being recommended for certain types of data, and, as a result, more reliable detection of concealed information in those experiments.

In the EEG experiment using location stimuli (chapter 5), there was some evidence of a negativity following the small positivity for the probe only. This could be related to the MERMER (Farwell & Smith, 2001), which has been suggested as a marker of concealed information. Future research using location image stimuli could also analyse the MERMER to see if probe location stimuli do indeed generate a MERMER and if this is significantly different to the irrelevant. If future research does find a strong MERMER for the probe but not the irrelevant, then this could potentially be used instead of or in addition to the standard P3 analysis to detect concealed information of location image stimuli using the Fringe-P3 method.

Another important future experiment would be to replicate the countermeasures experiments by Bowman et al (2014) using different stimuli. These experiments instructed participants to use certain countermeasures (such as trying to make a non-salient stimulus salient) while taking part in Fringe-P3 experiments using own-name stimuli and concluded that the countermeasures did not affect the detection of concealed information. Since these experiments, Avilés et al (2020) have investigated potential repetition effects by tasking participants with finding the repeated words in RSVP streams and found that the probability of detecting an unknown repeated stimulus for the first time did not increase with more repetitions. Therefore, we believe that countermeasures should still not be able to confound the detection of concealed information in experiments using other stimuli. However, a future experiment should replicate this study with other stimuli, such as the famous name stimuli, to confirm this.

Finally, the amplitude and latency of P3s has been shown to be affected by individual differences (see Polich, 2007, 2012 for reviews) such as cognitive capability (Houlihan et al., 1998), age (Emmerson et al., 1989; Fjell & Walhovd, 2001; Howard & Polich, 1985; Polich, 1996, 1997; Polich et al., 1990), dementia (Polich et al., 1986), Alzheimer's disease (Jeong, 2004), Parkinson's disease (Hansch et al., 1982; Polich & Criado, 2006; Stanzione et al., 1991), anxiety disorders (Bauer et al., 2001; Enoch et al., 2001), and depression (Gangadhar et al., 1993; Vandoolaeghe et al., 1998). Therefore, the Fringe-P3 method needs to be tested with a much wider variety of people, especially groups with the aforementioned individual differences, in order to confirm whether or not the Fringe-P3 method can be

used effectively with them. Similarly, the AGAT, weight template, and MFC could also be compared using these datasets to see if one of the analyses is more sensitive to detecting P3s that are affected by these individual differences. This information is vital if the Fringe-P3 method were to be used in real forensic situations, as the police would need to know if the test can or cannot be used reliably with their suspect.

8.4.6. Future Experiment: Locations

The EEG experiment in chapter 5 analysed the Fringe-P3 method using location image stimuli. This experiment provided evidence that (target) location image stimuli can breakthrough into awareness, generate a P3, and familiarity with them can be detected. However, these results were only found for the target stimuli and not the probe stimuli. A future experiment should test location image stimuli again to see if the methodological improvements described below result in a more robust test that can detect familiarity with probe location stimuli as well as targets.

Participants

The EEG locations experiment in chapter 5 tested only 8 participants. Preliminary analyses on the data from those participants had already suggested that the experiment had potential methodological problems, so no more participants were tested. However, a future experiment should include a larger number of participants as this will lead to more reliable results. Previous Fringe-P3 research included 14-15 participants and found significant results at the group and individual level (Bowman et al., 2013; Alsufyani et al., 2019). Therefore, it is recommended that at least 14 participants are included for future Fringe-P3 experiments.

Counterbalancing

The experiment should be counterbalanced by running it at both the University of Kent (using Kent images as probes and Birmingham images as irrelevants) and the University of Birmingham (using Birmingham images as probes and Kent images as irrelevants). This will allow us to more robustly test that the probe images are breaking through into awareness and generating P3s because the participant is familiar with them, rather than due to any other aspect of the images (e.g., higher visual salience). If the Kent images only generate P3s for Kent students and not Birmingham students, and the Birmingham images only generate P3s for Birmingham students and not Kent students, then we can more reliably conclude that these P3s were generated due to familiarity. On the other hand, if, for example, only the Birmingham images generate P3s for Birmingham students, but the Kent

images do not generate P3s for Kent students, then it could be that there was some other aspect of the Birmingham images that caused them to generate P3s, rather than familiarity.

Stimuli and Presentation

The following procedure will focus on describing the experiment that could be run at the University of Kent with the Kent images as probes and the Birmingham images as irrelevants. To run the procedure at the University of Birmingham, simply switch the Birmingham images to be probes and the Kent images to be irrelevants.

Some buildings at the University of Kent have been redecorated since the original experiment, including two of the buildings previously used as probe images, so new photographs should be taken for the probe and irrelevant stimuli. A behavioural pilot study should then be run to find the most recognisable stimuli. This pilot study should show participants RSVP streams of buildings, including several photographs of University of Kent buildings, and task participants with detecting images of the University. At the end of each stream, participants should be asked if they saw an image of their university. The three images with the highest detection rates should then be used as probe images for the main experiment. Recognition tests could also be conducted following the streams to further asses which buildings broke through into conscious awareness. These recognition tests could show participants the presented university images and some images of other university buildings that were not presented in the streams. The unpresented images should be included to help us check that participants are not simply giving random ratings or the same rating to all images in the recognition test. Recall tests are not recommended as participants may not know the names of the buildings they saw and may not provide clear enough descriptions that would enable to experimenter to be sure which building they saw and are describing. The same pilot study should be run at the University of Birmingham to find the three most recognisable University of Birmingham buildings to be used as the irrelevant stimuli for University of Kent students and as probe stimuli for the University of Birmingham students.

The three most highly recognised University of Kent buildings (by Kent students) would be used as probe stimuli. The three most recognised University of Birmingham buildings (by Birmingham students) would be used as irrelevant stimuli. The targets would be three unfamiliar buildings chosen from the same database as the distractors.

The experiment in chapter 5 selected targets and created a database of distractors from Konkle et al's (2010) database of scenes. This same database could be used for this future

experiment or additional/replacement images could be used if available. The experiment in chapter 5 had a database of 411 distractors, which led to some distractors repeating, therefore it is recommended that a larger distractor database is used to prevent repetition if possible. However, as discussed earlier in this chapter in sections 8.4.2 and 8.4.3, it was thought that the slow SOA allowed participants to see non-salient items in the streams, including distractors, and that a faster SOA should prevent participants seeing non-salient items even when they are repeated. Therefore, an increase in the SOA is a more important improvement for this future experiment than increasing the number of distractors, so the same database of 411 distractors could be used again if additional/replacements images are not available.

The SOAs used in both chapters 5 (167ms) and 6 (144ms) were slow enough that participants could consciously see non-salient items in the streams. Previous Fringe-P3 research has used SOAs of 133ms and found significant results at the group and individual level. Therefore, an SOA of 133ms could be suggested for this future locations experiment. However, the slower SOAs used for chapters 5 and 6 were chosen as it was thought that complex location stimuli may need more time to be consciously perceived than visually simpler stimuli such as names. Therefore, a more appropriate improvement would be to use a staircase method to select the SOA. Using a staircase method (Cornsweet, 1987) would allow us to find the most suitable SOA for each individual participant, where the SOA is fast enough that non-salient stimuli are not consciously perceived but not so fast that the salient stimuli cannot be consciously perceived. This staircase method could be run during practice trials to find the most suitable SOA within a range of 100-150ms. This range is suggested as 100ms has been successfully used with (visually simpler) name stimuli by Bowman et al (2014) and 150ms is shorter than the 167ms that was found to be too slow in chapter 5. Using this range should allow us to find a faster SOA while still accounting for the complexity of location image stimuli and for participants who may need slower SOAs. The practice trials should contain only distractors and a target (which should be different to the targets used in the main experiment and should not be present in every trial). No probes or irrelevants should be presented in any practice trials. Participants should be tasked with looking for the target and answering “yes” or “no” at the end of each trial when asked if they saw it. An algorithm could be written to automatically adjust the SOA until an SOA is found where the target accuracy rate is closest to 75% and the correct rejection rate is closest to 80%. This SOA would then be used for that participant for the whole experiment.

Tasks

For the main experiment, participants could simply be told to look for the target image. Alternatively, this task could be modified to more directly detect deception by telling participants to pretend that the target is a building from their university and to respond "yes" when asked if they saw their university after seeing the target and to respond "no" if they saw any other buildings, including ones from their real university.

Recognition tests should also be run at the end of each block, similar to the experiment in chapter 5. Participants should be asked to rate the target, probe, irrelevant, and a distractor for that block as well as an unpresented image on a scale from 1 (never saw) to 5 (saw very frequently). This will allow us to further investigate if the probe was recognised more than the other images, and if any irrelevant or distractor images broke through into conscious awareness and were recognised. The unpresented building could be a random building or a building from their university that was not used as a probe. If using distractors in the recognition tests, ensure they are not used as distractors again in any RSVP trials after that test.

Analyses

The main analyses should use the AGAT with a search window of 250-1000ms to analyse the ERP data at the group and individual participants' level. This search window was recommended by the conclusions of chapter 7 and was used in the email addresses experiment in chapter 4.

Paired t-tests or Wilcoxon's signed rank tests could be used to analyse the recognition data (depending on whether the data is normally distributed). The same comparisons used in chapter 5 could be used for this future experiment: probe against an unpresented image, probe against irrelevant, irrelevant against distractor, and irrelevant against an unpresented image.

Further Work

If this improved method successfully detects familiarity with probe location stimuli at the individual participants' level, then further research can be done using this method with these changes. For example, the method could be combined with the attentional blink paradigm, to test the detection of P3s and attentional blinks at the same time. The method for this would be mostly the same, but with two critical stimuli (probe/irrelevant and target) presented in critical trials instead of one, with the target appearing a different lags after the probe/irrelevant stimuli. This method should lead to P3s for the probe (and target) but not the irrelevant and an attentional blink for probe trials but not irrelevant

trials. The target may generate its own distinct P3 if presented at a long lag after the probe (and in irrelevant trials) or may result in a broad P3, merged with the probe P3, when presented shortly after the probe (similar to the broad P3s found for the two-part famous name stimuli in chapter 3).

While the AGAT should be the main analysis, the MFC analysis from chapter 7 could also be run on the data from this future experiment. This would allow us to further investigate the MFC and compare it to the AGAT on another dataset.

Using a staircase procedure to find a suitable SOA for each participant in this experiment will also provide us with the range, mean, and median SOA that can be successfully used with location image stimuli for the Fringe-P3 method. This could then be used for future Fringe-P3 research using location images and other visually complex stimuli.

8.4.7. Future Experiment: Search Strategies

The experiment in chapter 6 investigated the impact of two task types (detection and categorisation) on the attentional blink, with the expectation that the categorisation task would prevent participants from using search strategies to detect the target and, therefore, would lead to a more pronounced attentional blink than the detection task. The results of this comparison were inconclusive. The experiment described below is designed to more directly investigate the impact of search strategies on the attentional blink, which would then allow us to better investigate potential tasks that could prevent participants from using search strategies in future attentional blink experiments.

Participants

The original experiment in chapter 6 tested two groups of participants, with 30 participants in the categorisation group and 22 participants in the detection group (52 overall). However, familiarity with the University of Kent (which was used as irrelevant stimuli) and problems with participants scoring too close to ceiling and preventing the binomial regression analysis from completing, resulted in 14 participants being excluded from the categorisation group and six being excluded from the detection group, leaving 16 participants for each group. An attentional blink experiment is quicker and easier to run than an EEG experiment, so it can test more participants. The proposed future experiment contains three groups rather than two used previously, so it is recommended that 20-30 participants be tested for each group (60-90 overall). The participants should be divided equally into one of three groups, which will be described later in this section.

Stimuli and Presentation

All participants will take part in a typical two-target attentional blink experiment, without probe stimuli or the detection of concealed information.

The stimuli used can be the locations from Konkle et al's (2010) database that were used for chapters 5 and 6. As there are no probe stimuli in this experiment, the stimuli used do not need to be restricted to be of the same type as a probe (unlike chapters 5 and 6 where we could only use the outside of buildings to match the probes, so were limited to 411 distractors). Therefore, the entirety of Konkle et al's (2010) database of locations can be used for this experiment, which contains over 4000 images.

As the targets will be presented at different lags, a staircase method to choose the SOA is not recommended for this experiment, as it would change the lag between the targets for each participant. Therefore, an SOA of 133ms is recommended instead, as this has been used in previous Fringe-P3 experiments, including Alsufyani et al's (2019) experiment using famous face images, which found significant results at the group and individual level. Alternatively, if the future experiment described in section 8.4.6 has been successfully conducted using a staircase procedure to select the SOA, then the mean or median of the SOAs used in that experiment could be used as the SOA for this experiment, as both experiments use location image stimuli.

All participants would be given the same main task – to look for the two targets in the streams and report when they had seen them - but may also be given an additional task depending on their group.

Streams should contain zero, one, or two targets. The two targets would be presented at different lags across the experiment, with target one (T1) presented before target two (T2). The following description of the lags assumes an SOA of 133ms would be used. The lags should be lag 2 (e.g., T2 presented 266ms after T1), 3 (e.g., 399ms after T1), 5 (e.g., 665ms), and 6 (e.g., 798ms). Lags 2 and 3 occur within the typical 250-500ms window where the attentional blink occurs, while lags 5 and 6 are outside of the window, so should not result in attentional blinks. Lag 1 is not recommended as it would present T2 133ms after T1, which is within the 150ms window where lag-1 sparing can occur and could prevent an attentional blink, meaning we could not use that lag to reliably compare the attentional blink between groups. Additionally, lag 4 is not recommended as it would present T2 532ms after T1, which is very close to the end of the typical 250-500ms attentional blink window and, due to individual variance, could be within the attentional blink window for

some participants but outside of it for others, making it unreliable to compare groups. The position of T1 within the streams should be varied so it does not always appear in the same position.

Tasks and Groups

The participants would be divided equally into three groups: a strategy group, a no-strategy group, and a control group.

The strategy group would be given information on a feature-based search strategy where participants look for a specific feature of the target images (e.g., a turret in the top left corner) to help them detect the targets. They would then be tasked with using that search strategy to detect the targets in the experiment. Alternatively, the participants could be given information on a range of search strategies and instructed to use any of them and report at the end of each block which strategy they used.

The no-strategy group would be given information on the same feature-based search strategy but would be instructed *not* to use it or any other strategy, but to simply look for the targets in the streams.

The control group would be given no information on search strategies and no instructions regarding them. They would only be told to look for the targets in the streams. This should serve as a baseline against which the strategy and no-strategy groups can be compared.

The strategy group should allow us to investigate what the results of an attentional blink experiment look like when participants are actively using strategies and how much the attentional blink is impacted by strategies compared to the control and no-strategy groups. The no-strategy group should provide us with an example of what the attentional blink looks like when search strategies have been prevented. This can then be used in comparison with the control and strategy groups but could also be used as a baseline in future experiments to better investigate if tasks that aim to prevent search strategies (such as the categorisation task in chapter 6) do indeed prevent search strategies.

Further Work

Following this experiment, the categorisation and detection tasks from chapter 6 could be compared again in another experiment (with additional methodological improvements as suggested in chapter 6, such as increasing the SOA to 133ms). As we would now have a baseline for what the attentional blink looks like when search strategies are and are not used, we could use this to better investigate if the detection and categorisation tasks influence the use of search strategies. Specifically, we would now have an idea of what the

attentional blink should look like if the categorisation task did prevent the use of search strategies.

8.5. Practical Applications of the Fringe-P3 Method

The overall aim of the Fringe-P3 method is for it to be used in practical situations to help law enforcement detect deception and concealed information. Here, I will outline some example scenarios where the Fringe-P3 method could be used, as well as a response to some ethical concerns about the use of the method in real forensic situations.

Scenario 1: Familiar Names

The famous names experiment in chapter 3 showed that the Fringe-P3 method can detect familiarity with familiar (famous) names. In a real forensic investigation, this method could be used to detect familiarity with the name of an accomplice or victim (instead of a famous name) by using the accomplice/victim's name as the probe stimulus. If the suspect has a significantly stronger P3 for the probe compared to the irrelevant, then we can infer that they are familiar with the accomplice/victim. This, then, would help police link that suspect to the crime and find potential accomplices.

Scenario 2: Online Identities

The email addresses experiment in chapter 4 showed that the Fringe-P3 method can detect familiarity with online identities in the form of email addresses. In a real forensic situation, the Fringe-P3 method would be used when the police already know that a particular email address is associated with a crime and suspect that a specific individual is the user of that email address. They would then use the Fringe-P3 method to detect if that suspect is familiar with that email address, thereby linking them to it. The results of the email addresses experiment combined with the own-names (Bowman et al., 2013, 2014) and famous names (chapter 3) experiments, also suggests that the Fringe-P3 method could detect familiarity with other online (user)names (e.g., twitter handles and forum usernames) as well as email addresses.

Ethics

The main ethical concern raised regarding the Fringe-P3 method is concern that deception detection, especially using brain activity, is an invasion of privacy. However, two methods of deception detection are already widely used by police: fingerprinting and DNA testing. Fingerprints and DNA can prove that someone was at the scene of a crime or touched a crime-related object when they say they haven't, which is analogous to the Fringe-P3 method showing that a suspect recognises a crime-related location or object. P3s are essentially fingerprints within the brain. Furthermore, the Fringe-P3 method is not mind

reading and cannot tell the experimenter what the participant is thinking. Therefore, the ethics of the Fringe-P3 method, fingerprinting, and DNA testing are no different. Additionally, the Fringe-P3 method should be used in practice similar to police questioning, with suspects having the right to refuse to take part in a Fringe-P3 test just as they can refuse to answer a question.

8.6. Overall Conclusions

Overall, this thesis has successfully achieved its main aim of generalising the Fringe-P3 method and has demonstrated that the Fringe-P3 method has the potential to be used with a wider variety of stimuli, including famous names, email addresses, location images, and multi-item stimuli. It has also shown that the attentional blink paradigm can be used with the method to detect concealed information, which, combined with EEG, could result in an even more robust test. Finally, this thesis proposed the MFC, which has the potential to detect P3s that other analyses miss. The results of this thesis are promising for the Fringe-P3 method's real world application for detecting (concealed) recognition of crime-related stimuli by suspects in forensic investigations. The success of the famous names experiment suggests that that method could be used to detect familiarity with the names of accomplices or victims and the email addresses experiment suggests that it could be used to detect familiarity with online identities. The locations experiments, once improved by methodological changes, have the potential to suggest that the method could also be used to detect familiarity with the location of a crime.

Glossary

AGAT – Aggregated Grand Average of Trials. This is the main analysis used for EEG Fringe-P3 research at the group and individual participants' level. It uses a data-driven orthogonal contrast between probe and irrelevant by merging these into one aggregated ERP. The 100ms wide window with the highest mean amplitude is found from this aggregated ERP and selected as the window of interest. This window of interest is then applied separately to the probe and irrelevant ERPs and the difference between their mean amplitudes (within that window) is calculated. A paired-samples t-test is performed on these differences at the group level and a permutation test is performed at the individual participants' level. See section 2.5 for more detail.

Search Window – The window within which the AGAT searches for the window of interest on the aggregated ERP. This is typically 300-1000ms or 250-1000ms, to include most of the P3's positivity (if present). See section 2.5 for more detail.

Window of Interest – A 100ms wide window found by the AGAT analysis. This window is selected from the aggregated ERP and then applied separately to the probe and irrelevant ERPs. The mean amplitudes are calculated within this window for the probe and irrelevant ERPs and used in statistical analyses to see if they are significantly different. See section 2.5 for more detail.

Attentional Blink – The attentional blink is a behavioural phenomenon where the detection rate of a second critical stimulus is significantly lower when it is presented within 200-500ms after a first critical stimulus. This occurs when the two critical stimuli are both targets and when the first critical stimulus is a salient probe and the second is a target. This second case can be used to detect concealed information; if the target is consistently missed when presented shortly after the probe, then it suggests the participant is familiar with the probe and is concealing information about it. See section 2.6 for more detail.

CIT – Concealed Information Test. CITs were formerly known as Guilty Knowledge Tests. CITs are based on the idea that guilty participants have knowledge of a crime that innocent participants do not. It presents a participant with infrequent salient stimuli (probes) amongst frequent non-salient stimuli (distractors). If the participant has a significantly different reaction to the probe compared to the other stimuli, then it suggests they are familiar with the probe stimulus and may be concealing information about it. See section 2.1 for more detail.

EEG – Electroencephalography. EEG is a non-invasive method of detecting electrical activity in the brain via electrodes placed on the scalp. The electrodes detect negative or positive electrical fields generated by the activation of synapses within the brain. These synapses are activated by events, such as the recognition of a stimulus. See section 2.2 for more detail.

Electrodes – The parts of EEG equipment that detect the positive and negative voltages at the scalp. The number of electrodes used for EEG varies depending on the experiment and the system used. See figure 2.1 in section 2.2 for the standard location of electrodes on the scalp for a 32-channel BioSemi system.

ERP – Event-related potentials. ERPs are the waveforms used in most EEG statistical analyses. They are created from EEG data following a processing procedure. First, the EEG data is filtered to reduce noise (activity not related to the signal) and then epoched (trimmed) to the time around the event of interest. Single trials still contain some noise, so the trials are then averaged together to create the ERPs. This results in most of the random noise being averaged out while the consistent signal remains in the ERP. See section 2.2 for more detail.

Noise – This is random activity recorded by EEG that is not related to the event being investigated. The brain is always doing multiple things at once and is not solely focused on the event being investigated, which generates noise. Eye and muscle movements and electrical devices in the room can also be recorded by EEG, causing more noise. However, noise in the data can be reduced through filtering procedures and averaging trials into ERPs before analysis. See section 2.2 for more detail.

Signal – This is the consistent response in the waveforms that occurs following the event being investigated.

Trial – One RSVP stream in the EEG and attentional blink experiments.

Waveform – The timeseries of positive and negative voltages recorded by EEG.

Fringe-P3 Method – The Fringe-P3 method combines EEG (P3) and RSVP with a CIT to detect concealed information. It presents critical stimuli (targets, salient probes, non-salient irrelevants) amongst non-salient distractor stimuli in RSVP streams and records brain responses through EEG. If a participant has a significantly larger P3 response for a probe stimulus compared to an irrelevant stimulus, then it suggests they are familiar with the probe stimulus and are concealing information about it. See section 2.4 for more detail.

MFC – Matched Filter Convolution Analysis. The MFC is a new analysis for detecting P3s at the individual participants' level proposed in this thesis in chapter 7. It uses the average of other participants' probe ERPs (excluding the current participant being analysed) as a filter template for what the probe P3 for that experiment looks like. The analysis can then detect P3s for the current participant by comparing their ERPs to the filter template. This allows us to account for any differences (compared to the classic P3) in the P3's shape, latency, and any negativities that can occur due to the different types of stimuli or methods used in an experiment. Several versions of the MFC are explored in chapter 7.

P3 – The P3 (aka P300 or P3b) is the component of interest in EEG Fringe-P3 research. The P3 is a positive deflection in ERPs around 250-800ms post-stimulus. It is usually recorded from the Pz, Fz, and Cz electrodes and is strongest at Pz. It is typically generated during cognitive tasks where participants are discriminating between stimuli, especially when a stimulus is novel, unexpected, or salient to the participant. As it is generated when a stimulus is salient, it is used to detect familiarity with a stimulus in the Fringe-P3 method. See section 2.2 for more detail on EEG and the P3 and section 2.4 for more detail on the Fringe-P3 method.

RSVP – Rapid Serial Visual Presentation. A method of presenting visual stimuli so rapidly (e.g., 10 stimuli per second) that they appear on the fringe of awareness rather than in conscious awareness. Critical stimuli (e.g., targets) are presented amongst distractor stimuli in streams (e.g., 1 target name presented amongst 19 distractor names). Salient stimuli such as targets and familiar probe stimuli can breakthrough from the fringe into conscious awareness, while non-salient stimuli such as irrelevants and distractors do not breakthrough. When a stimulus breaks through into awareness, it generates a P3 that can be detected by EEG. See section 2.3 for more detail.

SOA – Stimulus Onset Asynchrony. The speed of an RSVP stream, i.e., the time in milliseconds between stimuli in RSVP streams. For example, an SOA of 100ms presents a stimulus every 100ms, resulting in 10 stimuli per second.

Appendix A: Discussion of the Programming Inconsistencies

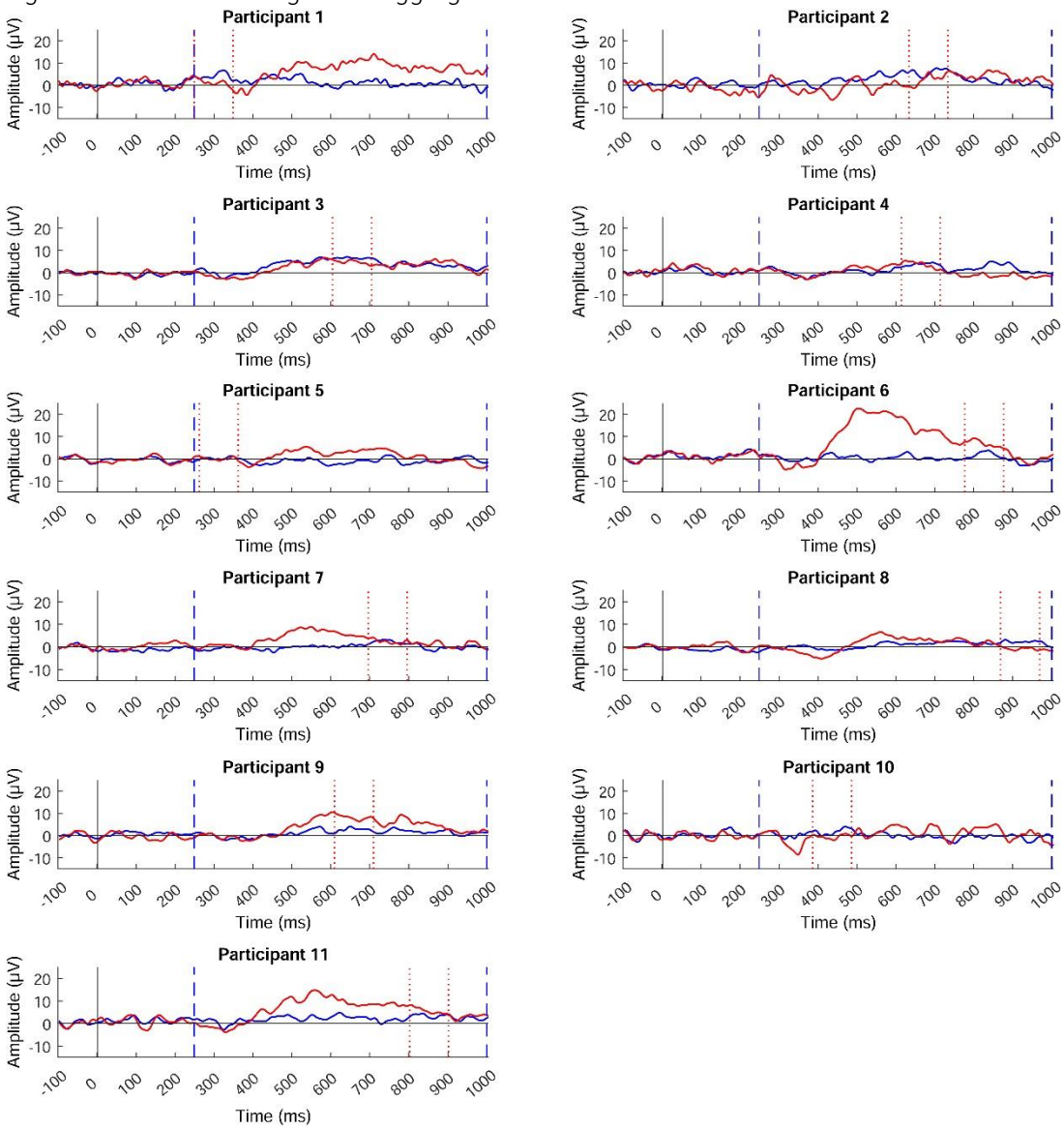
Due to a programming error, Participants 2 and 3 were shown one different probe name pair and Participant 4 was shown two different probe name pairs to the other participants. We do not believe this error affected these participants' results. Both participants 2 and 3 scored 5/5 for probe recall and 0/5 for irrelevant recall, had high recognition scores for probe and low for the irrelevant, and had significant p values, similar to the majority of participants that saw the original probe names. This suggests the error had no effect on their results. Participant 4 did not have a significant p value but still recalled 4 of the 5 probes and none of the irrelevants and scored high in the recognition test for probe (4.4) and low for irrelevant (1.6), showing that they still saw and recognised the probes but not the irrelevants. In comparison, participant 1 was presented with all of the original probe names and had a significant p value but scored lower for recall (2/5) and recognition (3.8) of the probes, so it seems unlikely that the programming error would have had an effect and been the reason behind participant 4's non-significant p value. Instead, noise in the ERP data is the more likely reason.

We also ran the ERP analyses while excluding the participants who saw different probe names in error and the results were the same ($t\ (10)$ = 9.026, p <0.0001, d = 3.528, at the group level) or better (10/11 significant at the individual level). This shows that the programming error and inclusion of these participants did not have a substantive effect on the results or impact our conclusions.

Appendix B: Discussion of the Target Grand Mean Amplitude

Figure B.1 further demonstrates why the target grand mean amplitude was smaller than the probe grand mean amplitude in figure 4.3, despite having a larger peak in figure 4.2. It can be seen the target (red) peaked earlier or later than the aggregated ERP (blue) for most participants and therefore did not peak within the window of interest selected from the aggregated ERP, resulting in a smaller mean amplitude in that window. It can also be seen that participant 6's target ERP peak was much higher than of the other participants, and so may have increased the amplitude in the target grand average in figure 4.2.

Figure B.1. Individual Target and Aggregated ERPs at Pz.



The red ERPs are the target and the blue ERPs are the aggregated ERP used to find the window of interest for the probe-irrelevant AGAT analysis. The vertical blue dashed lines represent the window within which the algorithm searched for the window of interest. The vertical red dotted lines represent the window of interest selected for that participant.

Appendix C: D Prime Scores

The main analyses used hits data only, as this was more appropriate for two reasons. Firstly, false alarms are not associated with specific lags, so d' scores cannot be sorted into the separate or collapsed lags. Therefore, d' scores cannot be used to compare target accuracy inside and outside of the attentional blink window, which is necessary for attentional blink research. Secondly, the binomial regression analysis can only use hits and misses from individual trials, so d' scores cannot be used at the individual participants level, which is vital to the Fringe-P3 method if it is to be used in real forensic investigations, which will focus on individuals. Therefore, d' scores were not used in the main analyses. The d' scores can, however, provide some information on overall target accuracy in the two search task types. Additionally, as a probe or irrelevant stimulus was still presented in target-absent trials, the false alarms and d' scores can be separated into the probe and irrelevant conditions and, therefore, can also provide information on the overall accuracy in the two stimulus type conditions.

The overall d' scores for categorisation can be found in table C.1 and the d' scores for detection can be found in table C.2. It can be seen that there were fewer false alarms and higher d' scores in the detection group compared to the categorisation group. This is expected, as participants in the detection group were shown a specific target to search for, which is an easier task than looking for a category of targets, as the categorisation group did.

A 2x2 mixed ANOVA was performed on the search task type and stimulus type d' data. There was a significant main effect of search task type ($F(1,30) = 32.421$, $p < 0.0001$), showing that participants' target accuracy was significantly higher in the detection group compared to the categorisation group. This fits with our expectation that the categorisation task would be more difficult than the detection task. There was also a significant main effect of stimulus type ($F(1,30) = 8.776$, $p = 0.0059$), showing that participants' target accuracy was significantly higher in the irrelevant condition compared to the probe condition. This fits with our expectation that there would be an attentional blink that would cause the participant to miss the target in the probe condition. There was not a significant interaction between search task type and stimulus type ($F(1,30) = 0.152$, $p = 0.6996$).

These results are different to the 2x2x2 and 2x2x4 mixed ANOVAs performed on the hits data. Both the hits and d' analyses found a significant main effect of stimulus type, however, the hits analysis did not find a significant main effect of search task type while the

d' analysis did find this to be significant. As the hits comparison was not significant, the significant difference between d' scores in the detection and categorisation groups must be caused by there being significantly more false alarms in the categorisation group. The hits analysis also found a significant interaction between search task type and stimulus type, which was not significant in the d' analysis, showing that the different search task types did not affect the difference in target accuracy between the probe and irrelevant conditions even though it affected the number of hits.

Table C.1. Overall D Primes for Categorisation.

	Probe			*Irrelevant*		
Participant	*Hits*	*False Alarms*	*d'*	*Hits*	*False Alarms*	*d'*
1	52	6	2.392	43	2	2.407
2	48	4	2.343	43	2	2.407
3	45	6	1.956	48	2	2.676
4	45	11	1.577	44	9	1.659
5	42	9	1.561	41	6	1.759
6	50	0	3.361	55	0	3.777
7	34	24	0.421	32	13	0.867
8	43	7	1.765	47	5	2.166
9	49	11	1.805	51	15	1.711
10	48	22	1.182	53	25	1.402
11	40	4	1.932	40	7	1.623
12	52	13	1.894	50	16	1.590
13	34	7	1.360	36	3	1.898
14	43	18	1.097	37	19	0.774
15	39	15	1.060	50	14	1.695
16	44	15	1.297	50	10	1.935
Mean	*44.250*	*10.750*	*1.688*	*45.000*	*9.250*	*1.897*
Median	*44.500*	*10.000*	*1.671*	*45.500*	*8.000*	*1.735*
Std Dev	*5.592*	*6.728*	*0.674*	*6.593*	*7.188*	*0.713*

This table shows the overall hits, false alarms, and d' scores for probe and irrelevant in the categorisation group. There were some cases with a false alarm rate of 0 so the Macmillan and Kaplan (1985) correction was applied.

Table C.2. Overall D Primes for Detection.

	Probe			*Irrelevant*		
Participant	*Hits*	*False Alarms*	*d'*	*Hits*	*False Alarms*	*d'*
1	44	0	3.017	47	3	2.428
2	38	1	2.469	44	2	2.457
3	42	1	2.652	48	0	3.236
4	51	0	3.430	52	1	3.239
5	39	1	2.513	42	0	2.918
6	44	1	2.751	53	2	3.026
7	55	1	3.511	51	0	3.430
8	34	3	1.813	39	2	2.219
9	48	1	2.970	54	0	3.676
10	41	0	2.871	44	1	2.751
11	38	0	2.735	46	1	2.856
12	44	1	2.751	52	1	3.239
13	31	0	2.436	38	1	2.469
14	38	0	2.735	41	0	2.871
15	50	1	3.095	54	0	3.676
16	55	1	3.511	56	2	3.335
Mean	*43.250*	*0.750*	*2.829*	*47.563*	*1.000*	*2.989*
Median	*43.000*	*1.000*	*2.751*	*47.500*	*1.000*	*2.972*
Std Dev	*7.066*	*0.775*	*0.440*	*5.785*	*0.966*	*0.447*

This table shows the overall hits, false alarms, and d' scores for probe and irrelevant in the detection group. There were some cases with a false alarm rate of 0 so the Macmillan and Kaplan (1985) correction was applied.

Appendix D: Lag 1 v Lag 7 Analyses

The separate lag blink curves in figure 6.4 show that the mean hits for the probe condition at lag 1 were noticeably smaller than the later lags, but lag 3 was not much smaller than lags 5 and 7. Since lag 3 is 432ms after the target, it is close to end of the 200-500ms attentional blink window, so it is possible that participants may have recovered from an attentional blink by then and were able to detect the target more easily again. Therefore, additional analyses were conducted comparing only lags 1 and 7 to see if there were more significant differences between these two lags and, as a result, see if the inclusion of lag 3 had affected the main collapsed lag and four separate lag analyses.

Group Level

Categorisation

A 2x2 repeated measures ANOVA was performed on the stimulus type and lag data. There were significant main effects of lag ($F(1,15) = 17.245$, $p = 0.0009$) and stimulus type ($F(1,15) = 6.486$, $p = 0.0223$), and a significant interaction between stimulus type and lag ($F(1,15) = 31.134$, $p = 0.0001$). These results mostly match the collapsed lag and separate lag ANOVAs, except for the stimulus type main effect, which was significant for the lag 1 v 7 analysis but was not significant for the separate lags or collapsed lag analyses. This means that there was a significant difference between probe and irrelevant in the lag 1 v 7 analysis that was not significant when including lags 3 and 5. This matches our original hypotheses.

A two-tailed binomial regression analysis was also conducted on the stimulus type and lag data. There was a significant main effect of lag ($F(1,60) = 6.343$, $p = 0.0145$) and a significant interaction between stimulus type and lag ($F(1,60) = 8.905$, $p = 0.0041$), which matches the previous analyses. There was not a significant main effect of stimulus type ($F(1,60) = 2.553$, $p = 0.1153$), which does not match the 2x2 ANOVA in the lag 1 v 7 analysis above but does match the four separate lag and collapsed lag analyses.

Detection

A 2x2 repeated measures ANOVA was performed on the stimulus type and lag data. There were significant main effects of lag ($F(1,15) = 6.846$, $p = 0.0195$) and stimulus type ($F(1,15) = 9.304$, $p = 0.0081$), and a significant interaction between stimulus type and lag ($F(1,15) = 9.164$, $p = 0.0085$).

A two-tailed binomial regression analysis was conducted on the stimulus type and lag data and also found significant main effects of stimulus type ($F(1,60) = 5.019$, $p = 0.0288$) and lag ($F(1,60) = 4.614$, $p = 0.0358$), and a significant interaction between stimulus type and lag ($F(1,60) = 6.817$, $p = 0.0114$).

The lag 1 v 7 ANOVA and binomial regression results match each other and match the results from the separate lag and collapsed lag analyses.

Three-way interaction

A 2x2x2 mixed ANOVA was conducted on the search task type, stimulus type, and lag data. There were significant main effects of stimulus type ($F(1,30) = 15.760$, $p = 0.0004$) and lag ($F(1,30) = 22.027$, $p = 0.0001$), and a significant interaction between stimulus type and lag ($F(1,30) = 31.655$, $p < 0.0001$). There was not a significant main effect of search task type ($F(1,30) = 0.050$, $p = 0.8250$), or significant interactions between search task type and stimulus type ($F(1,30) = 0.288$, $p = 0.5956$), search task type and lag ($F(1,30) = 0.533$, $p = 0.4710$), or search task type, lag, and stimulus type ($F(1,30) = 0.363$, $p = 0.5512$).

Most of these results were similar to the four separate lag and collapsed lag three-way analyses, with the only differences being that the interaction between search task type and stimulus type was non-significant in the lag 1 v 7 analysis when it had been significant in the four separate lag and collapsed lag analyses.

Individual Participants' Level

Since the lag 1 v 7 analyses only used hits from two lags instead of four, the maximum number of hits in any of the four bins used in the binomial regression analyses was half the size (15) of the collapsed lag analyses (30), as, for example, the ProbeIn bin now only had hits from lag 1, whereas previously it had hits from lags 1 and 3. Therefore, there was less chance of variation in scores and more chance of participants scoring at ceiling in the lag 1 v 7 analyses. As discussed in section 6.2.4, the capacity to observe an attentional blink is impacted by proximity to ceiling, and the binomial regression analysis cannot complete if any of the four bins are at ceiling. By using half the data, more participants were at ceiling and three more participants had to be excluded from the lag 1 v 7 analysis as the binomial regression analysis could not complete: participant 15 for categorisation and participants 6 and 7 for detection. This left 15 participants in the categorisation group and 14 in the detection group. If the lag 1 v 7 analyses were to be used in future research, more trials would need to be conducted in order to have a greater capacity to observe the attentional blink and include more participants in the analyses.

Categorisation

There were three significant results in the two-tailed binomial regression analysis. One participant had a significant main effect of lag (M = 0.384, Med = 0.439, SD = 0.269): participant 13 (F(1,56) = 5.522, p = 0.0223). Two participants had a significant interaction between stimulus type and lag (M = 0.398, Med = 0.287, SD = 0.362): participant 2 (F(1,56) = 5.377, p = 0.0241) and 8 (F(1,56) = 5.070, p = 0.0283). No participant had a significant main effect of stimulus type (M = 0.506, Med = 0.439, SD = 0.280). This is a decrease of one significant result (for stimulus type) compared to the two-tailed collapsed lag analysis.

There were seven significant results in the one-tailed binomial regression analysis. One participant had a significant main effect of stimulus type (M = 0.361, Med = 0.269, SD = 0.254): participant 16 (F(1,56) = 3.663, p = 0.0304). Two participants had a significant main effect of lag (M = 0.253, Med = 0.220, SD = 0.233): participant 13 (F(1,56) = 5.522, p = 0.0112) and 16 (F(1,56) = 3.663, p = 0.0304). Four participants had a significant interaction between stimulus type and lag (M = 0.200, Med = 0.144, SD = 0.183): participant 2 (F(1,56) = 5.377, p = 0.0120), 4 (F(1,56) = 3.003, p = 0.0443), 8 (F(1,56) = 5.070, p = 0.0141), and 12 (F(1,56) = 3.752, p = 0.0289). This is an increase of four significant results compared to the one-tailed lag 1 v 7 analysis (one for stimulus type, one for lag, and two for the interaction) and an increase of two significant results (for the interaction) compared to the one-tailed collapsed lag analysis.

Table D.1 presents the number of p values below certain values where the outcome of the equation was in the expected direction for the categorisation group in the lag 1 v 7 analysis. Overall, there were more p values below 0.05, 0.1, and 0.2 in the one-tailed test compared to the two-tailed test. The number of equation outcomes in the expected direction overall for the interaction shows that almost all of the equation outcomes (14/15) were in the expected direction.

Table D.1. One and Two-Tailed Results for Categorisation in the Lag 1 v 7 Analysis.

	One-Tailed			*Two-Tailed*		
Number of results in the expected direction	*Stim. Type*	*Coll. Lag*	*Interaction*	*Stim. Type*	*Coll. Lag*	*Interaction*
With significant p values (p <= 0.05)	1	2	4	0	1	2
With p values <= 0.1	1	5	6	1	2	4
With p values <= 0.2	5	7	10	1	5	6
Num. equation outcomes in the expected direction overall	10	13	14	10	13	14

This table presents the number of p values below certain values where the outcome of the equation was in the expected direction for the categorisation group in the lag 1 v 7 analysis. Each cell is out of 15 participants.

Detection

There were four significant results in the two-tailed binomial regression analysis. One participant had a significant main effect of stimulus type (*M* = 0.409, *Med* = 0.335, *SD* = 0.321): participant 14 (*F*(1,56) = 7.195, p = 0.0096). One participant had a significant main effect of lag (*M* = 0.480, *Med* = 0.576, *SD* = 0.324): participant 2 (*F*(1,56) = 5.687, p = 0.0205). Two participants had a significant interaction between stimulus type and lag (*M* = 0.350, *Med* = 0.335, *SD* = 0.264): participant 1 (*F*(1,56) = 4.045, p = 0.0491) and 2 (*F*(1,56) = 5.687, p = 0.0205). There was a significant main effect of stimulus type for one participant in the lag 1 v 7 analyses whereas there had been none in the collapsed lag analysis, however there were two less significant results for the main effect of lag in the lag 1 v 7 analysis compared to the collapsed lag analysis, meaning that there was one less significant result overall for the lag 1 v 7 analysis compared to the two-tailed collapsed lag analysis.

There were eight significant results in the one-tailed binomial regression analysis. Two participants had a significant main effect of stimulus type (*M* = 0.276, *Med* = 0.188, *SD* = 0.281): participant 9 (*F*(1,56) = 2.845, p = 0.0486) and 14 (*F*(1,56) = 7.195, p = 0.0048).

Three participants had a significant main effect of lag (*M* = 0.375, *Med* = 0.288, *SD* = 0.369): participant 2 (*F*(1,56) = 5.687, p = 0.0102), 3 (*F*(1,56) = 3.486, p = 0.0336) and 14 (*F*(1,56) = 2.958, p = 0.0455). Three participants had a significant interaction between stimulus type and lag (*M* = 0.319, *Med* = 0.167, *SD* = 0.374): participant 1 (*F*(1,56) = 4.045, p = 0.0246), 2 (*F*(1,56) = 5.687, p = 0.0102), and 13 (*F*(1,56) = 3.345, p = 0.0364). This is an increase of four significant results compared to the two-tailed lag 1 v 7 analysis (one for stimulus type, two for lag, and one for the interaction) and an increase of two significant results compared to the one-tailed collapsed lag analysis (one for stimulus type and one for the interaction).

Table D.2 presents the number of p values below certain values where the outcome of the equation was in the expected direction for the detection group in the lag 1 v 7 analysis. Overall, there were more p values below 0.05, 0.1, and 0.2 in the one-tailed test compared to the two-tailed test. The number of equation outcomes in the expected direction overall for the interaction shows that most of the equation outcomes (11/14) were in the expected direction.

Table D.2. One and Two-Tailed Results for Detection in the Lag 1 v 7 Analysis.

	One-Tailed			*Two-Tailed*		
Number of results in the expected direction	*Stim. Type*	*Coll. Lag*	*Interaction*	*Stim. Type*	*Coll. Lag*	*Interaction*
With significant p values (p <= 0.05)	2	3	3	1	1	2
With p values <= 0.1	5	5	6	2	3	3
With p values <= 0.2	7	5	8	5	5	6
Num. equation outcomes in the expected direction overall	12	11	11	12	11	11

This table presents the number of p values below certain values where the outcome of the equation was in the expected direction for the detection group in the lag 1 v 7 analysis. Each cell is out of 14 participants.

Table D.3 presents a summary of the number of significant p values found in the two-tailed and one-tailed binomial regression analyses for the collapsed lag and lag 1 v 7 analyses at the individual participants' level. There were more significant results in the one-tailed analyses compared to the two-tailed analyses for both the collapsed lag and lag 1 v 7 analyses. There were more significant results in the collapsed lag analyses than the lag 1 v 7 analyses when using the two-tailed binomial regression, but more significant results in the lag 1 v 7 analyses when using the one-tailed binomial regression. However, in both analyses, the majority of participants did not have significant results, so statistical analyses

comparing the results of the collapsed lag and lag 1 v 7 analyses were not conducted. Overall, the exclusion of lags 3 and 5 in the lag 1 v 7 analyses did not lead to a greatly improved number of significant results compared to the collapsed lag analyses.

Table D.3. Number of Significant Results in the Collapsed Lag and Lag 1 v 7 Analyses.

		Two-tailed Collapsed Lag	*One-tailed Collapsed Lag*	*Two-tailed Lag 1 v 7*	*One-tailed Lag 1 v 7*
Categorisation	*Stim. Type*	1	1	0	1
	(Coll.) Lag	1	2	1	2
	Interaction	2	2	2	4
	Total	*4*	*5*	*3*	*7*
Detection	*Stim. Type*	0	1	1	2
	(Coll.) Lag	3	3	1	3
	Interaction	2	2	2	3
	Total	*5*	*6*	*4*	*8*

This table presents the number of significant results found in the two-tailed and one-tailed binomial regression analyses for the collapsed lag and lag 1 v 7 analyses for comparison. Excluding the total rows, the cells in the collapsed lag analyses are out of 16 and the cells in the Lag 1 v 7 analyses are out of 15 for categorisation and 14 for detection.

Discussion

For the detection task, the group level lag 1 v 7 analyses were the same as the four separate lag and collapsed lag analyses. For the categorisation task, there were some differences between the group level lag 1 v 7 analyses and the separate lag and collapsed lag analyses, with a significant main effect of stimulus type in the lag 1 v 7 2x2 ANOVA that was not present in the collapsed lag or separate lags ANOVAs. This same main effect, however, was not significant in the lag 1 v 7 binomial regression. The three-way lag 1 v 7 analysis also found similar results to the collapsed lag analyses, with the only difference being that there was not a significant two-way interaction between search type and stimulus type, when this had been significant in the collapsed lag analyses.

At the individual participants' level, there were two additional significant results compared to the collapsed lag analyses for both the categorisation and detection tasks when using the one-tailed binomial regression, but there was one less significant result compared to the collapsed lag analyses for both categorisation and detection when using the two-tailed binomial regression. The exclusion of lags 3 and 5 did not lead to a substantial change in the number of significant results for individuals and the majority were still non-significant.

While there were some differences between the lag 1 v 7 analyses and the four separate lag and collapsed lag analyses, there were less differences than anticipated, especially at the individual participants' level. This suggests that lag 3 being close to the end of the 200-500ms attentional blink window did not have a substantial impact on the group or individual participants' level analyses. Therefore, lag 3 at 432ms does not need to be excluded from the current study or future studies using this SOA.

References

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *The Forensic Examiner*, *15*(February), 6–11.

Acqualagna, L., & Blankertz, B. (2013). Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP). *Clinical Neurophysiology*, *124*(5), 901–908. https://doi.org/10.1016/j.clinph.2012.12.050

Alsufyani, A. (2018). A comparison of classification algorithms for Event Related Potentials. *Proceedings of the 2017 4th International Conference on Computer Applications and Information Processing Technology, CAIPT 2017*, *2018-Janua*, 1–5. https://doi.org/10.1109/CAIPT.2017.8320734

Alsufyani, A., Hajilou, O., Zoumpoulaki, A., Filetti, M., Alsufyani, H., Solomon, C. J., Gibson, S. J., Alroobaea, R., & Bowman, H. (2019). Breakthrough percepts of famous faces. *Psychophysiology*, *56*(1), e13279. https://doi.org/10.1111/psyp.13279

Alsufyani, A., Harris, K., Zoumpoulaki, A., Filetti, M., & Bowman, H. (2021). Breakthrough Percepts of Famous Names. *Cortex*. https://doi.org/10.1016/j.cortex.2021.02.030

Alsufyani, A., Zoumpoulaki, A., Filetti, M., Janssen, D., & Bowman, H. (2018). Countering Cross-Individual Variance in Event Related Potentials with Functional Profiling. *BioRxiv*, 455030. https://doi.org/10.1101/455030

Ambach, W., Bursch, S., Stark, R., & Vaitl, D. (2010). A Concealed Information Test with multimodal measurement. *International Journal of Psychophysiology*, *75*(3), 258–267. https://doi.org/10.1016/j.ijpsycho.2009.12.007

Avilés, A., Bowman, H., & Wyble, B. (2020). On the limits of evidence accumulation of the preconscious percept. *Cognition*, *195*, 104080. https://doi.org/10.1016/j.cognition.2019.104080

Balas, B., & Momsen, J. L. (2014). Attention “blinks” differently for plants and animals. *CBE Life Sciences Education*, *13*(3), 437–443. https://doi.org/10.1187/cbe.14-05-0080

Barnard, P. J., Ramponi, C., Battye, G., & Mackintosh, B. (2005). Anxiety and the deployment of visual attention over time. *Visual Cognition*, *12*(1), 181–211. https://doi.org/10.1080/13506280444000139

Barnard, P. J., Scott, S., Taylor, J., May, J., & Knightley, W. (2004). Paying Attention to Meaning. *Psychological Science*, *15*(3), 179–186. https://doi.org/10.1111/j.0956-7976.2004.01503006.x

Barry, R. J., Steiner, G. Z., De Blasio, F. M., Fogarty, J. S., Karamacoska, D., & MacDonald, B. (2020). Components in the P300: Don't forget the Novelty P3! *Psychophysiology*, *57*(7). https://doi.org/10.1111/psyp.13371

Bauer, L. O., Costa, L., & Hesselbrock, V. M. (2001). Effects of alcoholism, anxiety and depression on P300 in women: A pilot study. *Journal of Studies on Alcohol*, *62*(5), 571–579. https://doi.org/10.15288/jsa.2001.62.571

Ben-Shakhar, G., & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: Effects of mental countermeasures. *Journal of Applied Psychology*, *81*(3), 273–281. https://doi.org/10.1037/0021-9010.81.3.273

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: A meta-analytic review. *Journal of Applied Psychology*, *88*(1), 131–151. https://doi.org/10.1037/0021-9010.88.1.131

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2

Bowman, H., Brooks, J. L., Hajilou, O., Zoumpoulaki, A., & Litvak, V. (2020). Breaking the Circularity in Circular Analyses: Simulations and Formal Treatment of the Flattened Average Approach. *PLoS Computational Biology*, *16*(11). https://doi.org/10.1371/journal.pcbi.1008286

Bowman, H., Filetti, M., Alsufyani, A., Janssen, D., & Su, L. (2014). Countering countermeasures: Detecting identity lies by detecting conscious breakthrough. *PLoS ONE*, *9*(3), e90595. https://doi.org/10.1371/journal.pone.0090595

Bowman, H., Filetti, M., Janssen, D., Su, L., Alsufyani, A., & Wyble, B. (2013). Subliminal Salience Search Illustrated: EEG Identity and Deception Detection on the Fringe of Awareness. *PLoS ONE*, *8*(1), e54258. https://doi.org/10.1371/journal.pone.0054258

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review*, *114*(1), 38–70. https://doi.org/10.1037/0033-295X.114.1.38

Bowman, H., Wyble, B., Chennu, S., & Craston, P. (2008). A reciprocal relationship between bottom-up trace strength and the attentional blink bottleneck: Relating the LC-NE and ST2 models. *Brain Research*, *1202*, 25–42. https://doi.org/10.1016/j.brainres.2007.06.035

Bradley, M. T., Barefoot, C. A., & Arsenault, A. M. (2011). Leakage of information to innocent suspects. In *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 187–199). Cambridge University Press. https://doi.org/10.1017/CBO9780511975196.011

Braun, J., Joseph, J. S., Chun, M. M., & Nakayama, K. (1998). Vision and attention: The role of training [5] (multiple letters). *Nature*, *393*(6684), 424–425. https://doi.org/10.1038/30875

Brett, A. S., Phillips, M., & Beary, J. F. (1986). Predictive Power of the Polygraph: Can the "Lie Detector" Really Detect Liars? *The Lancet*, *327*(8480), 544–547. https://doi.org/10.1016/S0140-6736(86)90895-0

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*(2), 105–113. https://doi.org/10.3758/BF03210498

Brooks, J. L., Zoumpoulaki, A., & Bowman, H. (2017). Data-driven region-of-interest selection without inflating Type I error rate. *Psychophysiology*, *54*(1), 100–113. https://doi.org/10.1111/psyp.12682

Chennu, S., Alsufyani, A., Filetti, M., Owen, A. M., & Bowman, H. (2013). The cost of space independence in P300-BCI spellers. *Journal of NeuroEngineering and Rehabilitation*, *10*(1), 82. https://doi.org/10.1186/1743-0003-10-82

Cherry, E. C., & Taylor, W. K. (1954). Some Further Experiments upon the Recognition of Speech, with One and with Two Ears. *Journal of the Acoustical Society of America*, *26*(4), 554–559. https://doi.org/10.1121/1.1907373

Chun, M. M., & Potter, M. C. (1995). A Two-Stage Model for Multiple Target Detection in Rapid Serial Visual Presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 109–127. https://doi.org/10.1037/0096-1523.21.1.109

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, *110*(1), 24–30. https://doi.org/10.1016/S0168-5597(98)00033-1

Cornsweet, T. M. (1987). The Staircase-Method in Psychophysics. *The American Journal of Psychology*, *75*(3), 485–491. https://doi.org/https://doi.org/10.2307/1419876

Craston, P., Wyble, B., Chennu, S., & Bowman, H. (2009). The attentional blink reveals serial working memory encoding: Evidence from virtual and human event-related potentials. *Journal of Cognitive Neuroscience*, *21*(3), 550–566. https://doi.org/10.1162/jocn.2009.21036

Curran, T., & Hancock, J. (2007). The FN400 indexes familiarity-based recognition of faces. *NeuroImage*, *36*(2), 464–471. https://doi.org/10.1016/j.neuroimage.2006.12.016

Cutmore, T. R. H., Djakovic, T., Kebbell, M. R., & Shum, D. H. K. (2009). An object cue is more effective than a word in ERP-based detection of deception. *International Journal of Psychophysiology*, *71*(3), 185–192. https://doi.org/10.1016/j.ijpsycho.2008.08.003

Di Lollo, V., Kawahara, J. I., Shahab Ghorashi, S. M., & Enns, J. T. (2005). The attentional blink: Resource depletion or temporary loss of control? *Psychological Research*, *69*(3), 191–200. https://doi.org/10.1007/s00426-004-0173-x

Eimer, M. (2000). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neurophysiology*, *111*(4), 694–705. https://doi.org/10.1016/S1388-2457(99)00285-0

Emmerson, R. Y., Dustman, R. E., Shearer, D. E., & Turner, C. W. (1989). P3 latency and symbol digit performance correlations in aging. *Experimental Aging Research*, *15*(3), 151–159. https://doi.org/10.1080/03610738908259769

Enoch, M. A., White, K. V., Harris, C. R., Rohrbaugh, J. W., & Goldman, D. (2001). Alcohol use disorders and anxiety disorders: Relation to the P300 event-related potential. *Alcoholism: Clinical and Experimental Research*, *25*(9), 1293–1300. https://doi.org/10.1111/j.1530-0277.2001.tb02350.x

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1476–1492. https://doi.org/10.1037/0096-1523.31.6.1476

Farwell, L. A., & Smith, S. S. (2001). Using Brain MERMER Testing to Detect Knowledge Despite Efforts to Conceal. *Journal of Forensic Sciences*, *46*(1), 14925J. https://doi.org/10.1520/JFS14925J

Fiedler, K., Schmid, J., & Stahl, T. (2002). What is the current truth about polygraph lie detection? In *Basic and Applied Social Psychology* (Vol. 24, Issue 4, pp. 313–324). Lawrence Erlbaum Associates Inc. https://doi.org/10.1207/S15324834BASP2404_6

Fischler, I., Jin, Y. S., Boaz, T. L., Perry, N. W., & Childers, D. G. (1987). Brain potentials related to seeing one's own name. *Brain and Language*, *30*(2), 245–262. https://doi.org/10.1016/0093-934X(87)90101-5

Fjell, A. M., & Walhovd, K. B. (2001). P300 and neuropsychological tests as measures of aging: Scalp topography and cognitive changes. *Brain Topography*, *14*(1), 25–40. https://doi.org/10.1023/A:1012563605837

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary Covert Orienting Is Contingent on Attentional Control Settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 1030–1044. https://doi.org/10.1037/0096-1523.18.4.1030

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2006). A critique of functional localisers. *NeuroImage*, *30*(4), 1077–1087. https://doi.org/10.1016/j.neuroimage.2005.08.012

Furedy, J. J., Davis, C., & Gurevich, M. (1988). Differentiation of Deception as a Psychological Process: A Psychophysiological Approach. *Psychophysiology*, *25*(6), 683–688. https://doi.org/10.1111/j.1469-8986.1988.tb01908.x

Gallai, D. (1999). Polygraph evidence in federal courts: Should it be admissible? *American Criminal Law Review*, *36*(1), 87.

Gangadhar, B. N., Ancy, J., Janakiranaiah, N., & Umapathy, C. (1993). P300 amplitude in non-bipolar, melancholic depression. *Journal of Affective Disorders*, *28*(1), 57–60. https://doi.org/10.1016/0165-0327(93)90077-W

Ganis, G., & Patnaik, P. (2009). Detecting concealed knowledge using a novel attentional blink paradigm. *Applied Psychophysiology Biofeedback*, *34*(3), 189–196. https://doi.org/10.1007/s10484-009-9094-1

Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A., & Schendan, H. E. (2011). Lying in the scanner: Covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *NeuroImage*, *55*(1), 312–319. https://doi.org/10.1016/j.neuroimage.2010.11.025

Gannon, T. A., Wood, J., Pina, A., Tyler, N., Barnoux, M. F. L., & Vasquez, E. A. (2014). An Evaluation of Mandatory Polygraph Testing for Sexual Offenders in the United Kingdom. *Sexual Abuse: Journal of Research and Treatment*, *26*(2), 178–203. https://doi.org/10.1177/1079063213486836

Gödert, H. W., Rill, H. G., & Vossel, G. (2001). Psychophysiological differentiation of deception: The effects of electrodermal lability and mode of responding on skin conductance and heart rate. *International Journal of Psychophysiology*, *40*(1), 61–75. https://doi.org/10.1016/S0167-8760(00)00149-5

Grandison, T. D., Ghirardelli, T. G., & Egeth, H. E. (1997). Beyond similarity: Masking of the target is sufficient to cause the attentional blink. *Perception and Psychophysics*, *59*(2), 266–274. https://doi.org/10.3758/BF03211894

Hansch, E. C., Syndulko, K., Cohen, S. N., Goldberg, Z. I., Potvin, A. R., & Tourtellotte, W. W. (1982). Cognition in Parkinson disease: An event-related potential perspective. *Annals of Neurology*, *11*(6), 599–607. https://doi.org/10.1002/ana.410110608

Harris, K., Miller, C., Jose, B., Beech, A., Woodhams, J., & Bowman, H. (2020). Breakthrough percepts of online identity: Detecting recognition of email addresses on the fringe of awareness. *European Journal of Neuroscience*, *53*(3), 895–901. https://doi.org/10.1111/ejn.15098

Holeckova, I., Fischer, C., Giard, M. H., Delpuech, C., & Morlet, D. (2006). Brain responses to a subject's own name uttered by a familiar voice. *Brain Research*, *1082*(1), 142–152. https://doi.org/10.1016/j.brainres.2006.01.089

Hommel, B., & Akyürek, E. G. (2005). Lag-1 sparing in the attentional blink: Benefits and costs of integrating two events into a single episode. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(8), 1415–1433. https://doi.org/10.1080/02724980443000647

Honts, C. R., Hodes, R. L., & Raskin, D. C. (1985). Effects of Physical Countermeasures on the Physiological Detection of Deception. *Journal of Applied Psychology*, *70*(1), 177–187. https://doi.org/10.1037/0021-9010.70.1.177

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Journal of Psychophysiology*, *1*(3), 241–247.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1994). Mental and Physical Countermeasures Reduce the Accuracy of Polygraph Tests. *Journal of Applied Psychology*, *79*(2), 252–259. https://doi.org/10.1037/0021-9010.79.2.252

Houlihan, M., Stelmack, R., & Campbell, K. (1998). Intelligence and the effects of perceptual processing demands, task difficulty and processing speed on P300, reaction time and movement time. *Intelligence*, *26*(1), 9–25. https://doi.org/10.1016/S0160-2896(99)80049-X

Howard, L., & Polich, J. (1985). P300 Latency and Memory Span Development. *Developmental Psychology*, *21*(2), 283–289. https://doi.org/10.1037/0012-1649.21.2.283

Hu, X., Bergström, Z. M., Bodenhausen, G. V., & Rosenfeld, J. P. (2015). Suppressing Unwanted Autobiographical Memories Reduces Their Automatic Influences: Evidence From Electrophysiology and an Implicit Autobiographical Memory Test. *Psychological Science*, *26*(7), 1098–1106. https://doi.org/10.1177/0956797615575734

Iacono, W. G. (2008). Effective Policing: Understanding How Polygraph Tests Work and Are Used. *Criminal Justice and Behavior*, *35*(10), 1295–1308. https://doi.org/10.1177/0093854808321529

Iidaka, T. (2014). Role of the fusiform gyrus and superior temporal sulcus in face perception and recognition: An empirical review. *Japanese Psychological Research*, *56*(1), 33–45. https://doi.org/10.1111/jpr.12018

Jeong, J. (2004). EEG dynamics in patients with Alzheimer's disease. In *Clinical Neurophysiology* (Vol. 115, Issue 7, pp. 1490–1505). Elsevier. https://doi.org/10.1016/j.clinph.2004.01.001

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, *76*(5), 349–364. https://doi.org/10.1037/h0031617

Jones, W., Pincham, H. L., Gootjes-Dreesbach, E. L., & Bowman, H. (2020). Fleeting Perceptual Experience and the Possibility of Recalling Without Seeing. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-64843-2

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a "preattentive" feature search task. *Nature*, *387*(6635), 806–807. https://doi.org/10.1038/42940

Junghöfer, M., Bradley, M. M., Elbert, T. R., & Lang, P. J. (2001). Fleeting images: A new look at early emotion discrimination. *Psychophysiology*, *38*(2), 175–178. https://doi.org/10.1017/S0048577201000762

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/jneurosci.17-11-04302.1997

Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology*, *16*(2), 177–216. https://doi.org/10.1016/0010-0285(84)90007-0

klein Selle, N., Verschuere, B., Kindt, M., Meijer, E., Nahari, T., & Ben-Shakhar, G. (2017). Memory detection: The effects of emotional stimuli. *Biological Psychology*, *129*, 25–35. https://doi.org/10.1016/j.biopsycho.2017.07.021

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551–1556. https://doi.org/10.1177/0956797610385359

Kozel, F. A., Padgett, T. M., & George, M. S. (2004). A replication study of the neural correlates of deception. *Behavioral Neuroscience*, *118*(4), 852–856. https://doi.org/10.1037/0735-7044.118.4.852

Kugelmass, S., & Lieblich, I. (1966). Effects of realistic stress and procedural interference in experimental lie detection. *The Journal of Applied Psychology*, *50*(3), 211–216. https://doi.org/10.1037/h0023324

Lancry-Dayan, O. C., Nahari, T., Ben-Shakhar, G., & Pertzov, Y. (2018). Do You Know Him? Gaze Dynamics Toward Familiar Faces on a Concealed Information Test. *Journal of Applied Research in Memory and Cognition*, *7*(2), 291–302. https://doi.org/10.1016/j.jarmac.2018.01.011

Levenson, J. S. (2009). Sex offender polygraph examination: An evidence-based case management tool for social workers. *Journal of Evidence-Based Social Work*, *6*(4), 361–375. https://doi.org/10.1080/15433710902911147

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique* (Second Edi, Vol. 33). MIT Press. https://doi.org/10.1073/pnas.0703993104

Lukács, G., Weiss, B., Dalos, V. D., Kilencz, T., Tudja, S., & Csifcsák, G. (2016). The first independent study on the complex trial protocol version of the P300-based concealed information test: Corroboration of previous findings and highlights on vulnerabilities. *International Journal of Psychophysiology*, *110*, 56–65. https://doi.org/10.1016/j.ijpsycho.2016.10.010

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*(6), 385–388. https://doi.org/10.1037/h0046060

Lykken, D. T. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology*, *44*(4), 258–262. https://doi.org/10.1037/h0044413

Lykken, D. T. (1998). *A Tremor in the Blood: Uses and Abuses of the Lie Detector* (2nd ed.). Plenum Press.

Mack, A., & Rock, I. (1998). Inattention: Faces and Other “Meaningful” Stimuli. In *Inattentional Blindness*. The MIT Press. https://doi.org/10.7551/mitpress/3707.003.0009

Macmillan, N. A., & Kaplan, H. L. (1985). Detection Theory Analysis of Group Data. Estimating Sensitivity From Average Hit and False-Alarm Rates. *Psychological Bulletin*, *98*(1), 185–199. https://doi.org/10.1037/0033-2909.98.1.185

Martens, S., Munneke, J., Smid, H., & Johnson, A. (2006). Quick minds don’t blink: Electrophysiological correlates of individual differences in attentional selection. *Journal of Cognitive Neuroscience*, *18*(9), 1423–1438. https://doi.org/10.1162/jocn.2006.18.9.1423

Matsuda, I., Nittono, H., & Allen, J. J. B. (2013). Detection of concealed information by P3 and frontal EEG asymmetry. *Neuroscience Letters*, *537*, 55–59. https://doi.org/10.1016/j.neulet.2013.01.029

McAuliffe, S. P., & Knowlton, B. J. (2000). Dissociating the effects of featural and conceptual interference on multiple target processing in rapid serial visual presentation. *Perception and Psychophysics*, *62*(1), 187–195. https://doi.org/10.3758/BF03212071

Meijer, E. H., Selle, N. K., Elber, L., & Ben-Shakhar, G. (2014). Memory detection with the Concealed Information Test: A meta analysis of skin conductance, respiration, heart rate, and P300 data. *Psychophysiology*, *51*(9), 879–904. https://doi.org/10.1111/psyp.12239

Meijer, E. H., Verschuere, B., Gamer, M., Merckelbach, H., & Ben-Shakhar, G. (2016). Deception detection with behavioral, autonomic, and neural measures: Conceptual and methodological considerations that warrant modesty. *Psychophysiology*, *53*(5), 593–604. https://doi.org/10.1111/psyp.12609

Meixner, J. B., & Rosenfeld, J. P. (2010). Countermeasure mechanisms in a P300-based concealed information test. *Psychophysiology*, *47*(1), 57–65. https://doi.org/10.1111/j.1469-8986.2009.00883.x

Meixner, J. B., & Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology*, *48*(2), 149–154. https://doi.org/10.1111/j.1469-8986.2010.01050.x

Moray, N. (1959). Attention in Dichotic Listening: Affective Cues and the Influence of Instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. https://doi.org/10.1080/17470215908416289

Munsterberg, H. (1908). *On the Witness Stand*. Clark, Boardman.

Nieuwenstein, M. R., Chun, M. M., Van Der Lubbe, R. H. J., & Hooge, I. T. C. (2005). Delayed attentional engagement in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1463–1475. https://doi.org/10.1037/0096-1523.31.6.1463

Nieuwenstein, M. R., & Potter, M. C. (2006). Temporal Limits of Selection and Memory Encoding. *Psychological Science*, *17*(6), 471–475. https://doi.org/10.1111/j.1467-9280.2006.01730.x

O'Sullivan, M., Frank, M. G., Hurley, C. M., & Tiwana, J. (2009). Police lie detection accuracy: The effect of lie scenario. *Law and Human Behavior*, *33*(6), 542–543. https://doi.org/10.1007/s10979-009-9191-y

Olivers, C. N. L., Van Der Stigchel, S., & Hulleman, J. (2007). Spreading the sparing: Against a limited-capacity account of the attentional blink. *Psychological Research*, *71*(2), 126–139. https://doi.org/10.1007/s00426-005-0029-z

Ouimet, C., & Jolicœur, P. (2007). Beyond Task 1 difficulty: The duration of T1 encoding modulates the attentional blink. *Visual Cognition*, *15*(3), 290–304. https://doi.org/10.1080/13506280600693741

Peth, J., Vossel, G., & Gamer, M. (2012). Emotional arousal modulates the encoding of crime-related details and corresponding physiological responses in the Concealed Information Test. *Psychophysiology*, *49*(3), 381–390. https://doi.org/10.1111/j.1469-8986.2011.01313.x

Pickering, E. C., & Schweinberger, S. R. (2003). N200, N250r, and N400 Event-Related Brain Potentials Reveal Three Loci of Repetition Priming for Familiar Names. *Journal of Experimental Psychology: Learning Memory and Cognition*, *29*(6), 1298–1311. https://doi.org/10.1037/0278-7393.29.6.1298

Pincham, H. L., Bowman, H., & Szucs, D. (2016). The experiential blink: Mapping the cost of working memory encoding onto conscious perception in the attentional blink. *Cortex*, *81*, 35–49. https://doi.org/10.1016/j.cortex.2016.04.007

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. In *Experimental Brain Research* (Vol. 209, Issue 4, pp. 481–493). Springer. https://doi.org/10.1007/s00221-011-2579-1

Polich, J. (1996). Meta-analysis of P300 normative aging studies. *Psychophysiology*, *33*(4), 334–353. https://doi.org/10.1111/j.1469-8986.1996.tb01058.x

Polich, J. (1997). EEG and ERP assessment of normal aging. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials*, *104*(3), 244–256. https://doi.org/10.1016/S0168-5597(97)96139-6

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J. (2012). Neuropsychology of P300. In *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0089

Polich, J., & Criado, J. R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, *60*(2), 172–185. https://doi.org/10.1016/j.ijpsycho.2005.12.012

Polich, J., Ehlers, C. L., Otis, S., Mandell, A. J., & Bloom, F. E. (1986). P300 latency reflects the degree of cognitive decline in dementing illness. *Electroencephalography and Clinical Neurophysiology*, *63*(2), 138–144. https://doi.org/10.1016/0013-4694(86)90007-6

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological Psychology*, *41*(2), 103–146. https://doi.org/10.1016/0301-0511(95)05130-9

Polich, J., Ladish, C., & Burns, T. (1990). Normal variation of P300 in children: Age, memory span, and head size. *International Journal of Psychophysiology*, *9*(3), 237–248. https://doi.org/10.1016/0167-8760(90)90056-J

Potter, M. C. (1975). Meaning in visual search. *Science*, *187*(4180), 965–966. https://doi.org/10.1126/science.1145183

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology. Human Learning and Memory*, *2*(5), 509–522. https://doi.org/10.1037/0278-7393.2.5.509

Potter, M. C., Banks, B. S., Muckenhoupt, M., & Chun, M. M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(4), 979–992. https://doi.org/10.1037/0278-7393.24.4.979

Potter, M. C., Nieuwenstein, M., & Strohminger, N. (2008). Whole report versus partial report in RSVP sentences. *Journal of Memory and Language*, *58*(4), 907–915. https://doi.org/10.1016/j.jml.2007.12.002

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, and Psychophysics*, *76*(2), 270–279. https://doi.org/10.3758/s13414-013-0605-z

Potter, M. C., Wyble, B., Pandav, R., & Olejarczyk, J. (2010). Picture Detection in Rapid Serial Visual Presentation: Features or Identity? *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1486–1494. https://doi.org/10.1037/a0018730

Proverbio, A. M., Mariani, S., Zani, A., & Adorni, R. (2009). How are "Barack Obama" and "President Elect" differentially stored in the brain? An ERP investigation on the processing of proper and common noun pairs. *PLoS ONE*, *4*(9), 7126. https://doi.org/10.1371/journal.pone.0007126

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary Suppression of Visual Processing in an RSVP Task: An Attentional Blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860. https://doi.org/10.1037/0096-1523.18.3.849

Reid, R. E. (1947). A revised questioning technique in lie-detection tests. *The Journal of Criminal Law and Criminology, Including the American Journal of Police Science*, *37*(6), 542–547. https://doi.org/10.2307/1138979

Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A Modified, Event-Related Potential-Based Guilty Knowledge Test. *International Journal of Neuroscience*, *42*(1–2), 157–161. https://doi.org/10.3109/00207458808985770

Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., & Chedid, E. (2008). The Complex Trial Protocol (CTP): A new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. *Psychophysiology*, *45*(6), 906–919. https://doi.org/10.1111/j.1469-8986.2008.00708.x

Rosenfeld, J. P., Nasman, V. T., Whalen, R., Cantwell, B., & Mazzeri, L. (1987). Late Vertex Positivity in Event-Related Potentials as a Guilty Knowledge Indicator: A New Method of Lie Detection. *International Journal of Neuroscience*, *34*(1–2), 125–129. https://doi.org/10.3109/00207458708985947

Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*, 205–219. https://doi.org/10.1111/j.1469-8986.2004.00158.x

Rosenzweig, G., & Bonneh, Y. S. (2019). Familiarity revealed by involuntary eye movements on the fringe of awareness. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-39889-6

Rosenzweig, G., & Bonneh, Y. S. (2020). Concealed information revealed by involuntary eye movements on the fringe of awareness in a mock terror experiment. *Scientific Reports*, *10*(1), 1–15. https://doi.org/10.1038/s41598-020-71487-9

Ross, N. E., & Jolicœur, P. (1999). Attentional blink for color. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1483–1494. https://doi.org/10.1037/0096-1523.25.6.1483

Saxe, L., & Ben-Shakhar, G. (1999). Admissibility of polygraph tests: The Application of Scientific Standards Post-Daubert. *Psychology, Public Policy, and Law*, *5*(1), 203–223. https://doi.org/10.1037/1076-8971.5.1.203

Saxe, L., Dougherty, D., & Cross, T. (1985). The Validity of Polygraph Testing. Scientific Analysis and Public Controversy. *American Psychologist*, *40*(3), 355–366. https://doi.org/10.1037/0003-066X.40.3.355

Schweinberger, S. R., Pickering, E. C., Burton, A. M., & Kaufmann, J. M. (2002). Human brain potential correlates of repetition priming in face and name recognition. *Neuropsychologia*, *40*(12), 2057–2073. https://doi.org/10.1016/S0028-3932(02)00050-7

Shapiro, K. L., Arnell, K. M., Raymond, J. E., & Arnell, K. M. (1997). The Attentional Blink. *Trends in Cognitive Sciences*, *1*(8), 291–296. https://doi.org/10.1016/S1364-6613(97)01094-2

Shapiro, K. L., Caldwell, J., & Sorensen, R. E. (1997). Personal Names and the Attentional Blink: A Visual "Cocktail Party" Effect. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 504–514. https://doi.org/10.1037/0096-1523.23.2.504

Simione, L., Akyürek, E. G., Vastola, V., Raffone, A., & Bowman, H. (2017). Illusions of integration are subjectively impenetrable: Phenomenological experience of Lag 1 percepts during dual-target RSVP. *Consciousness and Cognition*, *51*, 181–192. https://doi.org/10.1016/j.concog.2017.03.004

Simons, R. F., Graham, F. K., Miles, M. A., & Chen, X. (2001). On the relationship of P3a and the Novelty-P3. *Biological Psychology*, *56*(3), 207–218. https://doi.org/10.1016/S0301-0511(01)00078-3

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *NeuroReport*, *12*(13), 2849–2853. https://doi.org/10.1097/00001756-200109170-00019

Stanzione, P., Fattapposta, F., Giunti, P., D'Alessio, C., Tagliati, M., Affricano, C., & Amabile, G. (1991). P300 variations in parkinsonian patients before and during dopaminergic monotherapy: a suggested dopamine component in P300. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials*, *80*(5), 446–453. https://doi.org/10.1016/0168-5597(91)90093-D

Stenberg, G., Hellman, J., Johansson, M., & Rosén, I. (2009). Familiarity or conceptual priming: Event-related potentials in name recognition. *Journal of Cognitive Neuroscience*, *21*(3), 447–460. https://doi.org/10.1162/jocn.2009.21045

Su, L., Bowman, H., & Barnard, P. J. (2011). Glancing and then looking: On the role of body, affect, and meaning in cognitive control. *Frontiers in Psychology*, *2*(DEC), 1–48. https://doi.org/10.3389/fpsyg.2011.00348

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-Potential Correlates of Stimulus Uncertainty. *Science*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Tacikowski, P., Jednoróg, K., Marchewka, A., & Nowicka, A. (2011). How multiple repetitions influence the processing of self-, famous and unknown names and faces: An ERP study. *International Journal of Psychophysiology*, *79*(2), 219–230. https://doi.org/10.1016/j.ijpsycho.2010.10.010

Tacikowski, P., & Nowicka, A. (2010). Allocation of attention to self-name and self-face: An ERP study. *Biological Psychology*, *84*(2), 318–324. https://doi.org/10.1016/j.biopsycho.2010.03.009

Tateuchi, T., Itoh, K., & Nakada, T. (2012). Neural mechanisms underlying the orienting response to subject's own name: An event-related potential study. *Psychophysiology*, *49*(6), 786–791. https://doi.org/10.1111/j.1469-8986.2012.01363.x

Tourangeau, R., Smith, T. W., & Rasinski, K. A. (1997). Motivation to Report Sensitive Behaviors on Surveys: Evidence From a Bogus Pipeline Experiment. *Journal of Applied Social Psychology*, *27*(3), 209–222. https://doi.org/10.1111/j.1559-1816.1997.tb00629.x

Touryan, J., Gibson, L., Horne, J. H., & Weber, P. (2011). Real-time measurement of face recognition in rapid serial visual presentation. *Frontiers in Psychology*, *2*(MAR), 42. https://doi.org/10.3389/fpsyg.2011.00042

Trippe, R. H., Hewig, J., Heydel, C., Hecht, H., & Miltner, W. H. R. (2007). Attentional Blink to emotional and threatening pictures in spider phobics: Electrophysiology and behavior. *Brain Research*, *1148*(1), 149–160. https://doi.org/10.1016/j.brainres.2007.02.035

Valsecchi, M., & Turatto, M. (2009). Microsaccadic responses in a bimodal oddball task. *Psychological Research*, *73*(1), 23–33. https://doi.org/10.1007/s00426-008-0142-x

Vandoolaeghe, E., Van Hunsel, F., Nuyten, D., & Maes, M. (1998). Auditory event related potentials in major depression: Prolonged P300 latency and increased P200 amplitude. *Journal of Affective Disorders*, *48*(2–3), 105–113. https://doi.org/10.1016/S0165-0327(97)00165-1

Versace, F., Bradley, M. M., & Lang, P. J. (2010). Memory and event-related potentials for rapidly presented emotional pictures. *Experimental Brain Research*, *205*(2), 223–233. https://doi.org/10.1007/s00221-010-2356-6

Vincent, A., & Furedy, J. J. (1992). Electrodermal differentiation of deception: potentially confounding and influencing factors. *International Journal of Psychophysiology*, *13*(2), 129–136. https://doi.org/10.1016/0167-8760(92)90052-D

Visser, T. A. W. (2007). Masking T1 Difficulty: Processing Time and the Attentional Blink. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 285–297. https://doi.org/10.1037/0096-1523.33.2.285

Visser, T. A. W., Bischof, W. F., & Di Lollo, V. (1999). Attentional switching in spatial and nonspatial domains: Evidence from the attentional blink. *Psychological Bulletin*, *125*(4), 458–469. https://doi.org/10.1037/0033-2909.125.4.458

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological Evidence for a Postperceptual Locus of Suppression during the Attentional Blink. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1656–1674. https://doi.org/10.1037/0096-1523.24.6.1656

Vrij, A., Fisher, R. P., & Blank, H. (2017). A cognitive approach to lie detection: A meta-analysis. *Legal and Criminological Psychology*, *22*(1), 1–21. https://doi.org/10.1111/lcrp.12088

Wolford, G., & Morrison, F. (1980). Processing of unattended visual information. *Memory & Cognition*, *8*(6), 521–527. https://doi.org/10.3758/BF03213771

Wood, J., Alleyne, E., Ó Ciardha, C., & Gannon, T. A. (2020). *An Evaluation of Polygraph Testing by Police to Manage Individuals Convicted or Suspected of Sexual Offending*.

Wood, N., & Cowan, N. (1995). The Cocktail Party Phenomenon Revisited: How Frequent Are Attention Shifts to One's Name in an Irrelevant Auditory Channel? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 255–260. https://doi.org/10.1037/0278-7393.21.1.255

Wyble, B., Bowman, H., & Nieuwenstein, M. R. (2009). The Attentional Blink Provides Episodic Distinctiveness: Sparing at a Cost. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 787–807. https://doi.org/10.1037/a0013902

Yang, H., Wang, F., Gu, N., Gao, X., & Zhao, G. (2013). The cognitive advantage for one's own name is not simply familiarity: An eye-tracking study. *Psychonomic Bulletin and Review*, *20*(6), 1176–1180. https://doi.org/10.3758/s13423-013-0426-z

Zoumpoulaki, A., Alsufyani, A., & Bowman, H. (2015). Resampling the peak, some dos and don'ts. *Psychophysiology*, *52*(3), 444–448. https://doi.org/10.1111/psyp.12363