

Electrophysiological correlates of correct and incorrect eyewitness identification: The  
role of the N250 and P300 in real-world face recognition

by

Krista B. Friesen  
Bachelor of Arts (hons), University of Winnipeg, 2007

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

MASTER OF SCIENCE

in the Department of Psychology

© Krista B. Friesen, 2010  
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy  
or other means, without the permission of the author.

## **Supervisory Committee**

Electrophysiological correlates of correct and incorrect eyewitness identification: The role of the N250 and P300 in real-world face recognition

by

Krista B. Friesen  
Bachelor of Arts (hons), University of Winnipeg, 2007

### **Supervisory Committee**

James W. Tanaka, Psychology  
**Co-Supervisor**

D. Stephen Lindsay, Psychology  
**Co-Supervisor**

J. Don Read, Psychology  
**Departmental Member**

## Abstract

### **Supervisory Committee**

James W. Tanaka, Psychology  
**Co-Supervisor**

D. Stephen Lindsay, Psychology  
**Co-Supervisor**

J. Don Read, Psychology  
**Departmental Member**

This set of studies used event-related potentials (ERPs) to investigate the electrophysiology of face recognition as it may occur in real-life circumstances including eyewitness identification. Research using photographs has identified two ERP components as being associated with memory function, the N250 (an early indicator of perceptual recognition) and the P300 (associated with stimulus evaluation, task-relevance, and probability), however, there is no published research examining the relationship between these components and memory for live events, and recognizing a face seen briefly in real-life is somewhat different than recognizing a face only seen in a photograph. The present set of three studies employed a “target / non-target” paradigm that used a live interaction at each encoding stage. In Experiment 1, participants were introduced to and asked to identify their study experimenter as the target. In the second experiment, participants were introduced to the same experimenter and asked to imagine that they witnessed her committing a convenience-store robbery. Participants were asked to “cover” for the experimenter by choosing a different photo from the line-up to accuse as the culprit – they are asked to lie. In Experiment 3, participants witnessed a live simulated theft and were asked to identify the culprit from a line-up. In all three

experiments the line-up paradigm was identical – participants were shown a repeated series of sequentially-presented photographs and were asked to correctly identify one target among nine foils (non-targets), while brainwaves were recorded. Results showed that across all three studies, both the N250 and P300 were attenuated for the person selected as the target, in comparison to correct rejections of foils. Additional results from Experiment 2 showed that, compared to rejections of foils, both the N250 and P300 were enhanced for rejections of the experimenter-photo when participants were “covering” for her. In Experiment 3, participants who were unable to correctly identify the culprit showed that the N250 to incorrect rejections of the culprit was larger than correct rejections of foils. Finally, a comparison of participants who correctly identified the culprit and those who incorrectly identified a foil showed that the amplitude of the N250 and P300 to the selected culprit were equivalent regardless of eyewitness accuracy. Collectively, results from Experiments 1, 2 and 3 provide support that the memory effects indicated by the N250 and P300 components reported in laboratory studies generalize to person memories acquired during live interactions.

## Table of Contents

Supervisory Committee .....	ii
Abstract .....	iii
Table of Contents .....	v
List of Figures .....	vi
Acknowledgments .....	viii
Dedication .....	ix
Introduction .....	1
Experiment 1 .....	14
Experiment 2 .....	23
Experiment 3 .....	33
General Discussion .....	54
Bibliography .....	60

## List of Figures

Figure 1. Figure 1A (left) shows all the stimuli used for all three experiments. The target photo (top) is the experimenter, the nine foils are below. Figure 1B represents the line-up task. Participants viewed a blank screen (jittered between 2500 and 2500 ms), followed by a 500 ms fixation cross, followed by a second blank screen (500 ms), followed by a photograph which remained on the screen for 2500 ms or until the participant responded (whichever was first).....	16
Figure 2 Experiment 1 N250 waves. ERP data from PO7 (left posterior channel) and PO8 (right posterior channel), separated by experiment half (first and second). The average waveform from target trials is presented in black, the average waveform for foil trials is presented in pink. Baselines shown are -100 ms pre-stimulus onset.....	19
Figure 3 P300 wave data and topographical maps for Experiment 1. Wave data (left) are separated by half (first, second). Average wave data for target trials is shown in black, average wave data for foils is shown in pink. Topographical maps show positive (red) and negative (blue) impulses as measured across the scalp. The large area of centro-parietal positivity shown on the map for the target is representative of a typical P300 response.....	21
Figure 4. Experiment 2 N250 waves. ERP data from PO7 (left posterior channel) and PO8 (right posterior channel), separated by experiment half (first and second). The average waveform from target trials is presented in black, the average waveform for experimenter-probe (Probe) trials is presented in blue, the average waveform for foil trials is presented in pink. Baselines show -100 ms pre-stimulus onset.....	28
Figure 5. P300 wave data and topographical maps for Experiment 2. Wave data (left) are separated by half (first, second). Average wave data for scapegoat trials is shown in black, average wave data for the probe is shown in blue, and average wave data for foils is shown in pink. Topographical maps show positive (red) and negative (blue) impulses as measured across the scalp within the second half. A large parieto-central positivity can be seen for the target and probe conditions only (top and centre, respectively), typical of the P300 component. This activity is not seen for foils (bottom).....	30
Figure 6. Mean behavioural data for the Experiment 3 post-study questionnaire. Results are reported as a mean on a scale of 1-7. Dark bars represent the Correct ID group, light bars represent the Incorrect ID group. Error bars are 95% within-subjects confidence intervals.....	40
Figure 7. N250 wave data for the Incorrect ID group from PO7 (left hemisphere) and PO8 (right hemisphere) shown separated by experiment half. The average wave data for Incorrect ID trials is presented in black, the average of trials for incorrect rejections of the experimenter-culprit is presented in blue, the average waveforms for correctly rejected foils is presented in pink.....	42
Figure 8. N250 mean amplitude data at PO7 for both Correct and Incorrect ID groups. Data are averaged across experiment half. Error bars are 95% within-subjects confidence intervals.....	43
Figure 9. P300 waveform data and topographical maps at channel Pz for the Correct ID group. Data are separated by half. Topographical maps show scalp activity in the second	

half for the Correct ID (top) and Foil (bottom) during the most positive point in the P300 window (300 ms to 600 ms).....	44
Figure 10. P300 waveform data and topographical maps at channel Pz for the Incorrect ID group. Data are separated by half. Topographical maps show scalp activity in the second half for the Incorrect ID (top) and Incorrect Rejections of the Culprit (bottom) during the most positive point in the P300 window (300 ms to 600 ms). Topographical maps for foils were similar in correct and incorrect ID groups (see Figure 9 for map of Foil trials).....	46
Figure 11. P300 amplitude data plotted for both Correct and Incorrect ID groups, averaged across experiment half. Error bars are 95% within subjects confidence intervals. .....	47

## Acknowledgments

I would like to thank my supervisors Drs. Jim Tanaka and Steve Lindsay for their ideas, and time, as well as for their incredible support and encouragement throughout my Masters work. I would also like to thank lab co-ordinators Sean Butler, Laura Dixon, and Chelsea Durber for their hard work and assistance throughout the lengthy data collection process. I would also like to thank Tanya Hutchinson, Tamara Meixner, Sam Ennis, as well as all the other research assistants who contributed their time and effort to collecting the best data possible while giving participants a pleasant experience. Finally I would like to thank my office-mate Justin Kantner, for always being there with an encouraging word and a pot of chocolate fondue, when times got particularly tough.

## **Dedication**

I would like to dedicate this work to my grandfather, Clarence Toews, who passed away on August 25<sup>th</sup>, 2010. With his unconditional love, encouragement, and unwavering support, he taught me all the things about life that a university Masters education could not. I love and miss you.

## Introduction

When a witness in court points at a defendant and says “That’s him!” jurors often feel confident in convicting that defendant. In contrast, empirical research has consistently shown that people are not as accurate at recognizing faces as they think they are (Bruce, Burton & Hancock, 2007; Megreya & Burton, 2008; Sporer, Penrod, Read, & Cutler, 1995). This misplaced trust in identification evidence leads to some serious applied problems. Mistaken eyewitness identification is a leading cause of wrongful convictions (Huff, 2004; Wells & Olson, 2003). Because *exclusion* of eyewitness evidence is both impractical and unlikely, it is important for forensic experts, fact-finders, and researchers to understand the mechanisms involved in eyewitness recognition, and to develop more reliable and sensitive measures to collect and evaluate eyewitness evidence. The goal of the current set of studies was to investigate the electrophysiological components related to person-recognition within different types of live interactions, including a simulated crime. This research was designed to contribute both to the applied field of eyewitness memory research, as well as to basic memory research into underlying processes involved in face and identity processing.

A common criticism of the eyewitness literature is a lack of “real-world” generalizability (e.g. McCloskey, Egeth, & McKenna, 1986; Turtle, Read, Lindsay, & Brimacombe, 2008). Eyewitness identification memory is generally studied in the laboratory using a variety of methods, each including some form of study or witness stage, followed after some time by a memory test. Some studies of eyewitness identification have included orchestrated live mock crimes (e.g. Brigham, Maass, Snyder,

& Spaulding, 1982; Lindsay, Wells, & Rumpel, 1981; Luus & Wells, 1994), but the vast majority have used either slide shows or videos ranging from low resolution closed-circuit TV videos (Liu, Seetzen, Burton, & Chaudhuri, 2003) to realistic high-resolution staged-crime videos (Lieppe, Eisenstadt, Rauch, & Stambush, 2006). Recorded methods have the benefit of consistency – ensuring that all participants see exactly the same thing and are easier to employ than repeatedly staging a live mock crime. A drawback to these approaches is that they are not representative of a witness's actual experiences; researchers cannot be sure that watching a video elicits the same behavioural results as witnessing a live crime. Further, these methods may make witnesses appear more accurate than they actually are. One study showed that participants who viewed crime videos consistently recalled *more* details about the crime than participants who saw the same crime presented live (Ihleback, Love, Eilersten, & Magnussen, 2003). This finding provides compelling evidence that memory data found using videotaped crimes, though useful for testing hypotheses, may provide an overly optimistic estimate of eyewitness recall performance. Although there are difficulties with staging a mock event, there are benefits that outweigh the disadvantages including increased external validity. Specifically, witnesses of a mock-crime have to rely on incidental memory which takes into account emotion, context, surprise, and attention, in addition to basic levels of encoding such as depth, contour and colour, which are often absent in memory tasks that use photographs. In the third of three experiments presented in this paper, participants witnessed a surprise staged theft, in order to best approximate the experience of being a “real” eyewitness.

Few researchers test eyewitness memory conditions using real-world live scenarios, and even fewer have studied the electrophysiological underpinnings of eyewitness identification. Event-related potentials (ERPs) are post-synaptic impulses that originate in the cortex, time-locked to a stimulus event and recorded by electrodes placed on the scalp. These impulses yield distinct patterns of neural activity associated with specific cognitive functions (Luck, 2005). Many researchers have used ERPs to investigate the mechanisms underlying short term memory for static photographs of faces shown in a single pose (eg. Tanaka, Curran, Porterfield, and Collins, 2006; Pfutze, Sommer, & Schweinberger, 2002) or across multiple poses (Kaufmann, Schweinberger, & Burton, 2008; Schweinberger, Pickering, Jentsch, Burton & Kaufmann, 2002), and have found that the earliest ERP component associated with recognition is seen over inferior temporal electrode sites and is characterized by a negative deflection within a positive-going waveform approximately 200-300 ms after a familiar face appears. This component is referred to as the N250 (sometimes called the N250r, or the early repetition effect (ERE) in early literature). This N250 repetition effect is the strongest when identical photographs are shown in close temporal proximity and the magnitude of the N250 repetition effect decays as the number of faces between prime and target increases (Schweinberger, Pickering, Burton, & Kaufmann, 2002; Schweinberger, Huddy, & Burton, 2004). The N250 is sensitive to multiple viewpoints (Miyakoshi, Kanayama, Nomura, Iidaka, & Ohira, 2008) and weakened by the use of different images of the same previously unfamiliar faces (Kaufmann et al., 2009). Other research shows that a strong N250 component is elicited by highly familiar stimuli such as a participant's own face (Tanaka, et al. 2006) other personally familiar faces (Herzmann, Schweinberger,

Sommer, & Jentsch, 2004) and highly familiar celebrity faces (Boehm & Sommer, 2005; Herzmann et al. 2004). Recent research also suggests that faces familiarized in a laboratory setting will elicit an N250 effect equal to that of celebrity faces, providing a more controlled alternative when researching the acquisition of face representations, and providing evidence of N250 sensitivity to familiarity in the absence of contextual or biographical information (Herzmann & Sommer, 2010).

One important issue is whether the findings from these ERP studies are transferable to recognition of real people in the real-world. Faces in the world are encoded with depth, colour, motion, and fluid emotional expression. Importantly, live interactions with a person contain context including where and how many times you have met, the nature of your relationship, and the sound of his or her voice. Encoding a live face is much more complex than encoding a photograph in the laboratory, and no studies have been published that examine the N250 in response to a real-world stimulus. This is an important question because there are a number of studies suggesting that differences between 2-dimensional and 3-dimensional processing (Caharel, Jiang, Blanz, & Rossion, 2009) and movement processing (Lander & Chuang, 2005; Lander & Davies, 2007; but also see Lee, Habak, & Wilson, 2010) can influence a person's ability to recognize a face. In addition, our relationship with a person can influence how easily we recognize them as well as the magnitude of our N250 response to viewing their face. Results from one study showed that famous target faces and famous non-target faces (an incidental versus intentional memory paradigm) elicited an N250 equivalent in magnitude (Herzmann & Sommer, 2010). However, studies using previously *unfamiliar* target and non-target faces show that even after 400 trials using sequential presentation of the same

ten photographs, unfamiliar non-target faces do not elicit an N250 (Gordon & Tanaka, submitted; Tanaka et al. 2006). In these cases, only the previously unfamiliar target elicited an N250 response over time, suggesting a relationship between the N250 and experience as well as a relationship between the N250 and salience or task relevance. These differences in data based on degree of familiarity highlight important differences between familiar and unfamiliar face-processing. It is sometimes suggested that humans are “face-experts” because of a universal experience with faces and a typically robust ability to discriminate between faces with similar features. However, research has frequently shown that humans are not face-experts in general, but rather that they are *familiar* face experts (see Johnston & Edmonds, 2009, for a review), and are surprisingly poor at both recognizing (Bruce, Burton & Hancock, 2007; Sporer, et al., 1995) and matching previously unfamiliar faces (Bruce, et al., 1999; Bruce, Henderson, Newman, & Burton, 2001; Megreya & Burton, 2008).

One of the most obvious applications of unfamiliar face recognition research is eyewitness memory, in which jurors and fact finders rely on a witness’s ability to recognize a face that they may have never seen before. In this domain, mistaken identifications can have serious consequences including both convicting an innocent person, and allowing the real culprit to remain free. Unfortunately, research has shown time and again that eyewitness memory, while mediated by many different factors, is very often not reliable (see Memon, Mastroberardino, & Fraser, 2008, for a review).

Because the N250 can be used to distinguish between familiar and unfamiliar non-target faces, it is the ideal component to use to study eyewitness recognition, as differences in its amplitude should differentiate between correct identifications and

rejections of line-up photos, as well as possibly being able to distinguish between incorrect identifications, an incorrectly rejected culprit (who should be previously familiar) *and* correctly rejected foils (previously unfamiliar), if the participant is incorrect in their identification decision.

A second important ERP component in memory and target identification literature is the P300. The P300 is a late, slow, positive differential brain impulse beginning approximately 300 ms after stimulus onset. This component is seen at central parietal midline electrode sites, and is an index of target detection for tasks in which the probability of a target occurring is much lower than that of a non-target occurring (Donchin, 1981). Tasks such as these are referred to as “oddball” tasks, in which low frequency stimuli (called oddballs) are embedded within a series of high frequency non-target items. The P300 is elicited by both active responding to the target stimuli and to passive viewing (see Donchin & Coles, 1988). The effect of stimulus probability on the amplitude of the P300 is mediated by the degree of rarity of the oddball – that is, as frequency of target items decreases, the P300 amplitude to these targets increases (Duncan-Johnson & Donchin, 1977).

Additionally, it is suggested that P300 latency is indicative of stimulus evaluation time (Verleger, 1997), and that P300 latency increases with increased reaction time when making a stimulus category decision (Kutas, McCarthy, & Donchin, 1977). However, P300 peak latency occurs somewhat later (peaking between 300 and 400 ms, typically) than initial decisions are thought to be made. For example, categorization of a face occurs at approximately 170 ms, indicated by a very reliable posterior component called the N170 (Bentin Allison, Puce, Perez, & McCarthy, 1996; Jeffreys, 1996; Rossion et al,

2000). There is even some evidence showing that individual face discrimination (differentiating between same or different face stimuli) can happen in this same time window, as early as 130 ms after the onset of a stimulus (Jacques & Rossion, 2006). In addition, as discussed above, the earliest categorization of familiarity occurs prior to the P300, in the time window of the N250. Because of its delayed latency, the P300 is thought to be related to a cognitive category decision, but is not a direct index of a perceptual category decision (Pritchard, 1981). The P300 is also shown to be independent of basic decision-making, partly because in almost any task, the P300 is elicited by relevant or target stimuli, even though a decision must still be made about non-target events. It is also independent of response selection. In a study by McCarthy and Donchin (1981) using compatible (right hand response to the word “right”) and incompatible responses (right hand response to the word “left”) with stimuli that were either embedded in visual noise or not, they found that the P300 was modulated by noise (which increased reaction time by making the stimulus more difficult to categorize), but was not affected by response type. That is, when controlling for noise, increases in the reaction time required by making an incompatible response did not affect P300 amplitude. This suggests that P300 is unrelated to response selection, and provides support for the theory that the P300 latency is modulated by evaluation time.

The P300 is also sensitive to the motivational significance of a stimulus. This idea is supported by research showing that task-relevant target stimuli which are found to elicit an enhanced P300 compared to non-targets, even when targets and non-targets are equated for frequency (Duncan-Johnson & Donchin, 1977; Fabiani, Karis, & Donchin, 1983). The motivational significance approach to the P300 is also supported by research

showing that non-target items that are salient such as positively or negatively valenced stimuli will also elicit an attenuated P300 response compared to neutral non-target items (Johnston, Miller & Burleson, 1986; Keil, et al., 2002).

The idea that personally relevant non-targets can elicit a P300 response has motivated the use of ERPs in a forensic setting, particularly within deception detection. The most frequently employed task for deception detection in ERP research is the Guilty Knowledge Test (GKT) (Lykken, 1960). The earliest published work with the GKT used it in conjunction with galvanic skin response (GSR) measures to detect “guilt” among participants, some of whom had knowledge of a mock-crime (Lykken, 1959). The idea of this task is to embed crime-relevant probes among target and non-target items. Some participants have knowledge of the crime (the guilty group) and others do not (the innocent group). The idea is that participants with guilty knowledge will respond physiologically (eg: heightened GSR) to crime-relevant items, even when they do not require an overt response. This is precisely what was found in these early studies, and led to the development of a traditional polygraph test. More recently, the GKT was used with ERP measurements to develop a more sophisticated lie detection technique (Farwell & Donchin, 1991; Rosenfeld, et al., 1988). Probe items in the GKT have been shown repeatedly to produce an enhanced P300 response to crime relevant non-target probes, only in the guilty condition (Johnson & Rosenfeld, 1992, Mertens & Allen, 2008). Accuracy rates for categorizing participants correctly as “guilty” or “not guilty” based on the amplitude of the P300 component in response to crime-relevant probes is consistently quite high, ranging from about 70% to as high as 95% (Allen & Iacono, 1997; Farwell & Donchin, 1991; Rosenfeld et al. 1988). In fact, one researcher created and marketed the

procedure as a 100% effective method of detecting guilt (Farwell's Brain Fingerprinting: see [www.brainwavescience.com](http://www.brainwavescience.com)). Since the development of this technology, a number of researchers have criticized Farwell's methods (eg. Rosenfeld, 2005) and tested several possible counter-measures effective in "tricking" the test by diminishing the amplitude of the P300 response (Ben-Shakhar & Dolev, 1996; Rosenfeld, Soskins, Bosh, & Ryan, 2004). For this reason, the present paper makes no claims toward using ERPs to detect guilt or to identify reliable witnesses in applied settings.

However, differences in the P300 component may be used experimentally to identify accurate and inaccurate witnesses within an eyewitness line-up identification task in that the real culprit may be more motivationally salient or personally relevant in correct recognitions than an incorrectly identified foil in cases in which the participant-witness is wrong. Additionally, amplitude differences in the P300 may be able to detect a familiarity response to the real culprit when he or she is incorrectly identified as an innocent foil. The information provided by ERP recordings may be used to improve methods of questioning eyewitnesses in a way that would increase correct explicit recall of familiar faces and thereby reduce instances of misidentification.

In the literature, there is currently one published paper examining eyewitness memory using ERPs. Lefebvre, Marchand, Smith, and Connolly (2007) presented videotapes of mock crimes and asked participants to identify the culprit from the video immediately after viewing it, one hour later, or the following week. They found that the P300 was largest in participants who correctly identified the culprit and that the amplitude of the P300 differentiated the culprit from foils, even at a one-week delay. These data are consistent with previously discussed research suggesting a relationship

between the P300 and recognition (Fabiani, Karis, & Donchin, 1983); however, the researchers did not examine ERPs for incorrect witnesses. This has left a substantial gap in the present literature, regarding the manifestation of ERPs in witnesses who cannot correctly identify the culprit from a line-up. A second limitation of Lefebvre and colleagues' study is that all participants watched four videos and made four respective line-up decisions. Therefore, to ensure that all conditions were equal, participants were told prior to watching the first video that they would be asked to identify the culprit from a line-up. This reduces the real-world generalizability of the findings, as does using videos rather than employing a live mock-crime. One of the goals of the present set of studies was to investigate the generalizability of previous research by using a live event at the encoding stage of each experiment. Each of the three studies builds on the previous, from a simple design using an intentional memory paradigm to a more difficult incidental design in which the participants unexpectedly witness a live mock-crime.

The goal of the first experiment was to conduct a basic memory task using a simple real-world interaction, and well-established ERP memory task. Because of the intention to use a mock-crime as the "study" phase of later experiments, Experiment 1 used a personal interaction in the study phase. Participants were introduced to their experimenter, who explained the ERP set-up procedure. Participants were then informed that the target for the task *was* the experimenter, and their role is to correctly identify her photograph among a series of foils with a button-press keyboard response. Once the participant felt confident that s/he would recognize the experimenter in a photograph, the experimenter left the room and the task began. This design was to provide ERP evidence of basic recognition, in order to produce an estimate of what was to be expected from a

correct identification of a culprit, in later versions of the task. This was designed to be analogous to a real-world eyewitness situation in which the culprit would be seen in real-life and the police line-up would require the witness to identify the culprit among photographs. Photograph line-ups are the common practice in Canada.

Experiment 2 was designed and implemented in exactly the same manner as Experiment 1, with a modification in the instruction. For the second experiment, the condition of interest was the ERP response when the participant is lying. Participants were told to imagine they were witness to a crime committed by the experimenter. They were then told to imagine themselves as complicit and that during the line-up task they should try to “cover” for the experimenter by accusing an innocent foil. The reason for this design was to increase the saliency of the photo of the experimenter *without* making the photo task-relevant. Therefore, any differences found in ERP waveforms between rejection of the experimenter-culprit and rejection of innocent foil may be due to a cognitive difference (recognizing the experimenter), rather than a response difference.

Finally, in Experiment 3 the mock-crime component was implemented. The experimental procedure was again the same, however, instead of being told that the experimenter was the target, participants witnessed an unexpected crime (the theft of an iPod from a chair near them) and were asked to identify the culprit from a photo line-up. In this case, participants were not expected to have 100% accuracy (as was expected in the previous two designs), because the interaction with the experimenter culprit was brief and incidental. This allowed participants to be divided into two separate groups based on eyewitness accuracy which could then be analyzed with separate within-subjects

comparisons as well as across accuracy conditions in a between-subjects design (correct witnesses and incorrect witnesses).

Based on previous research, we expected correct identifications of the experimenter in Experiment 1 to result in increased negativity during the time window of the N250, and that this negativity would increase from the first half to the second half of trials (Tanaka et al, 2006). Additionally, because the experimenter was the target, was task-relevant, and had a low probability of occurring compared to foils, we expected that the P300 component would be much larger in amplitude to the target photograph than to previously unfamiliar foil photographs.

In Experiment 2 we were interested in the manifestation of the N250 and P300 to the experimenter photo when the experimenter was not the target. If the saliency manipulation was effective, the P300 should be amplified for the experimenter photo than to other foils. We also thought that incidental recognition of the experimenter might result in an attenuated N250 response, as recognized non-target faces have been shown to elicit an N250 (Herzmann & Sommer, 2010).

In Experiment 3 we predicted that participants who correctly identified the culprit would show a distinct electrophysiological response to the culprit's face compared to the foils' faces. We expected to find heightened activity in posterior regions (N250) as a part of a perceptual recognition response to the culprit. In addition, we predicted an enhanced P300 in response to *whomever* the participant identified as the culprit, regardless of whether he/she had chosen correctly. That is, when participants chose correctly we expected an enhanced P300 to the culprit, and when they were incorrect we expected an enhanced P300 to whomever they chose instead. This hypothesis is based on research

showing that the P300 can be elicited by a low probability, novel stimulus and is likely to appear to any face that requires a unique response (Donchin, 1981, Donchin & Coles, 1988).

A final question in which we were interested was what would happen when participants saw a photo of the culprit, but did not identify her. If correct rejections of foils and incorrect rejections of the culprit could be differentiated by either the P300 or the N250, then we might suspect that the culprit's face was encoded, and that the rejection of her photograph was a retrieval error. That is, a mnemonic representation might have existed but was not sufficiently accessible to support recognition. In this case, it may be that the culprit's face was familiar, but did not reach a threshold at which the participant felt comfortable accusing the person as being the culprit. An alternative possibility is that the correct and incorrect rejections are *not* differentiated by the brainwaves. If this is the case, then we might blame an encoding failure as the reason for the misidentification. We hoped to be able to use this paradigm to probe this question.

## Experiment 1

### *Method*

#### *Participants*

Twenty-two undergraduate students (10 men) from the University of Victoria were recruited from a psychology research subject pool and participated in the study for bonus course credit. Two additional participants were tested but one participant missed the target in 38% of trials and another displayed physiological artefacts in nearly 50% of trials. These two participants were excluded from analyses. The mean age of participants was 20.6 (SD = 3.0) and twenty participants self-reported that they were right-handed, two self-reported that they were left-handed. All reported normal or corrected-to-normal vision with no reported brain injury or trauma. Informed consent was obtained from all participants prior to testing. This paradigm was approved by and conducted in accordance with the University of Victoria's Human Research Ethics Committee.

#### *Materials*

Test stimuli consisted of ten colour photographs (nine foils and one target), taken with a digital camera under fluorescent lighting. The target (experimenter) and all foils were Caucasian females in their mid-20s. Photographs were formatted using Adobe PhotoShop software. All images were frontal views set on a light grey background and cropped to 288 by 288 pixels (see Figure 1A).

#### *Procedure*

Upon arrival to the session, each participant was greeted by the experimenter and led to the ERP lab. This interaction served as the initial familiarization phase and lasted

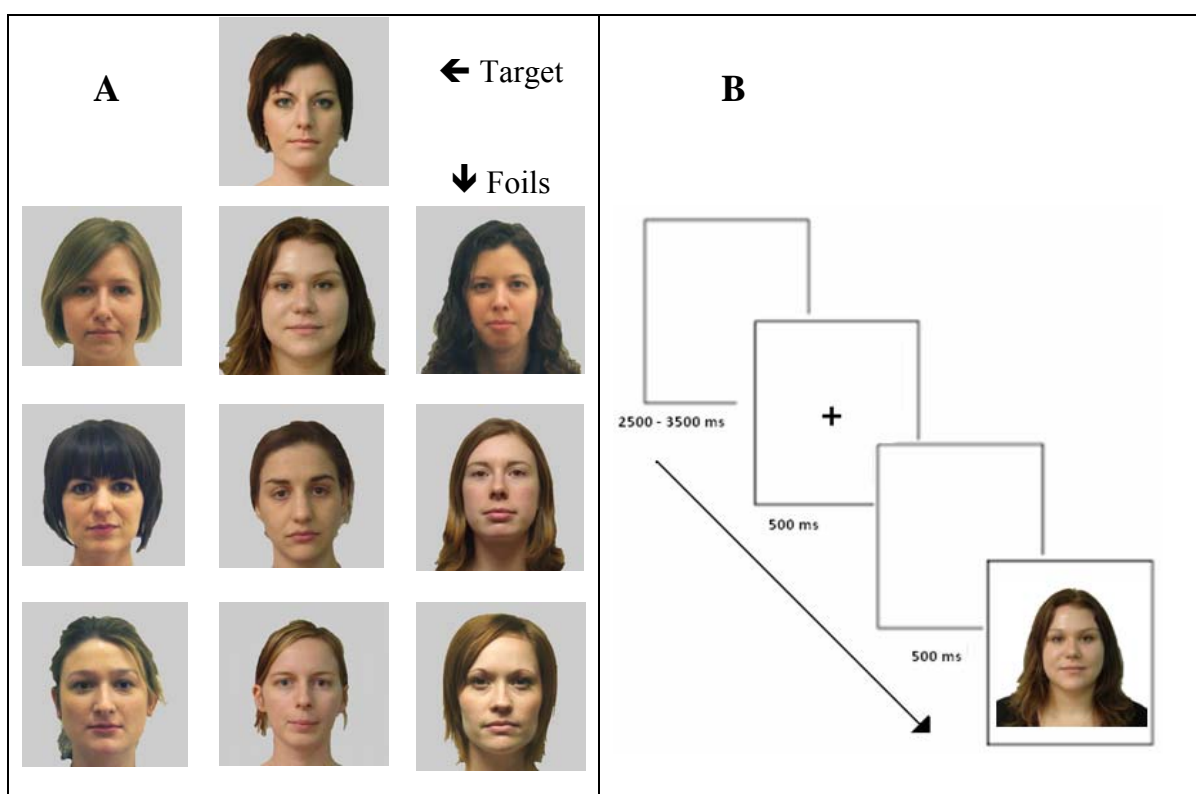
approximately 2 minutes. During this time the experimenter introduced herself and shook hands with participants. As participants were not made aware that this was part of the experiment, the nature of the conversation differed for each interaction, but generally included introductory questions such as “have you ever participated in an ERP study before?”. The participant was then seated in an ERP testing chamber and the apparatus (including cap, electrodes and amplifier) was presented and explained. Detailed information about the apparatus and recording procedure is presented below. After the EEG set-up was completed, the experimenter returned and sat facing the participant and explained the EEG procedure and the task (approximately 5 minutes). Each participant was told the following instructions (which were also displayed on the monitor): “Please look closely at the experimenter. Next, imagine she has stolen your wallet and you have been asked to identify her from a police line-up. When you see the culprit-experimenter’s photo, please press ‘1’ (one) on the keyboard. When you see any other face, please press ‘0’ (zero).” Participants were given as much time as needed in order to be sure of recognizing the experimenter before she left the room and the task began.

Participants viewed stimuli from a distance of approximately 70 cm, at a viewing angle of approximately 13.9° wide and 9.8° high. Stimuli were presented on a 17-inch computer screen using E-Prime software (Psychological Software Tools, Pittsburg, PA). The task consisted of four blocks of 100 trials. In each block, the set of 10 photos was randomly presented 10 times. In total, each individual stimulus was shown 40 times. Rest breaks were included after each block. During each break a research assistant entered the room and asked the participant how (s)he was doing and whether they had any questions or concerns. During the second break electrode impedances were checked and adjusted if

necessary. Once the research assistants left the room, participants were told that they could begin the next block whenever they felt ready.

Prior to each trial, participants viewed a blank screen with a variable length between 2500ms and 3500ms. Each trial began with a fixation cross centered on the screen (500 ms) followed by a second blank screen (500 ms) for baseline correction, then the stimulus was presented. Participants had up to 2500ms to respond, while the photograph was on the screen. See Figure 1B for a summary of the procedure.

Finally, participants were debriefed and thanked.



**Figure 1.** Figure 1A (left) shows all the stimuli used for all three experiments. The target photo (top) is the experimenter, the nine foils are below. Figure 1B represents the line-up task. Participants viewed a blank screen (jittered between 2500 and 2500 ms), followed by a 500 ms fixation cross, followed by a second blank screen (500 ms), followed by a photograph which remained on the screen for 2500 ms or until the participant responded (whichever was first).

### *Electrophysiological Recordings*

EEG was recorded from 41 electrode sites in accordance with the 10-20 electrode system (Jasper, 1958). A cap mounted with 36 Ag/AgCl electrodes was placed on the participant's head, referenced to a common ground: Fp1, Fpz, Fp2, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FCz, FC2, FC6, FT10, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CPz, CP2, CP6, TP10, P7, P3, Pz, P4, P8, PO7, POz, PO8, Oz. Horizontal electro-oculograms were recorded with electrodes placed on the outer canthi of each eye (LHEOG and RHEOG). Vertical electro-oculograms were recorded with an electrode placed above and below the right eye (Fp2, VEOG). All electrode impedances were kept below 10k $\Omega$ . EEG data were sampled at 250 Hz, amplified using QuickAmp equipment (Brain Products, GmbH, Munich, Germany) and filtered through a passband of 0.017 Hz – 67.5 Hz. Following data collection EEG data were filtered through a 0.1-20 Hz passband phase shift free Butterworth filter, and re-referenced to linked earlobes (M1, M2). Ocular artifacts were removed using the algorithm described by Gratton, Coles, and Donchin (1983). Trials with voltages exceeding 35 $\mu$ V per sampling point were discarded. Continuous EEG waveforms were segmented into 800-ms time epochs (200 ms before stimulus onset to 600 ms after onset) for each experimental condition (target and foil). Segments were baseline corrected from -200 ms to 0 ms. Fewer than 5% of total data were discarded.

## Results

### *Behavioural Results*

For each participant in all analyses a single random foil was selected to compare with the culprit (or the Incorrect ID). This is the standard procedure for ERP analyses, as it is important that conditions are matched for number of trials.

Overall accuracy was near ceiling at 96.4% correct. Reaction time was significantly faster in the foil condition ( $M = 585.64$  ms,  $SE = 30.51$ ) than in the target condition ( $M = 686.84$  ms,  $SE = 40.63$ ),  $F(1,21) = 30.20$ ,  $p < .001$ ,  $h_p^2 = .590$ .

### *Electrophysiological Results*

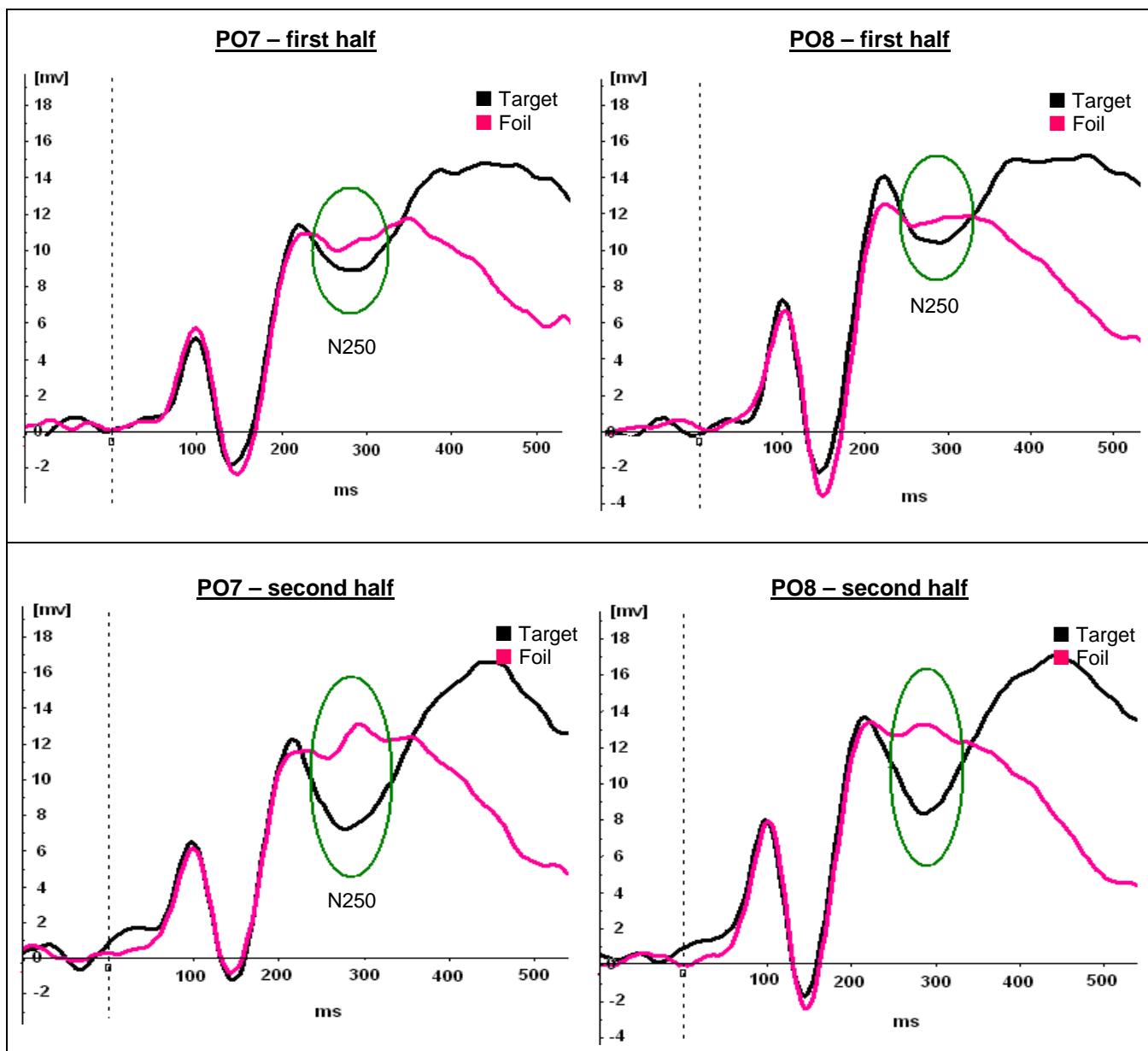
All statistical analyses for ERP data were conducted using average waveforms. Waveforms were segmented by condition (target, foil) and only correct trials were included. For each participant, one random foil was chosen as a comparison group for the target condition. This was done to yield an approximately equal number of trials per condition. Note that in all figures of ERP wave data, positive is plotted up.

### *N250 Results*

Analyses of the N250 focused on posterior channels, specifically PO7 and P08, where these components are most well-defined, consistent with previous research (Miyakoshi et al, 2008; Schweinberger, Pfütze, & Sommer, 1995; Tanaka et. al. 2006). The N250 does not always exhibit a well defined peak (refer to Figure 2), therefore, latency analyses and peak amplitude analyses were not possible with this component. Instead, each amplitude point within a window of 230 ms to 320 ms was averaged together to create a mean amplitude across the entire epoch. This window was chosen based on visual inspection of the waves, and is approximately the same time window used in previous research with this component (Pfützte, Sommer, & Schweinberger, 2002). N250 waveplots can be seen in Figure 2.

These data were subjected to a 2 (condition: target, foil) x 2 (hemisphere: left, right) x 2 (half: first, second) repeated measures Analysis of Variance (ANOVA). The mean amplitude across participants for the target condition was  $9.63 \mu\text{V}$  ( $SE = 1.11$ ) and

was more negative than the mean amplitude for the foil condition ( $M = 11.76 \mu\text{V}$ ,  $SE = 1.19$ ). This main effect was significant,  $F(1,21) = 7.252$ ,  $p = .014$ ,  $h_p^2 = .257$ , but was qualified by a Condition x Half interaction,  $F(1,21) = 17.264$ ,  $p < .001$ ,  $h_p^2 = .451$ .



**Figure 2** Experiment 1 N250 waves. ERP data from PO7 (left posterior channel) and PO8 (right posterior channel), separated by experiment half (first and second). The average waveform from target trials is presented in black, the average waveform for foil trials is presented in pink. Baselines shown are -100 ms pre-stimulus onset.

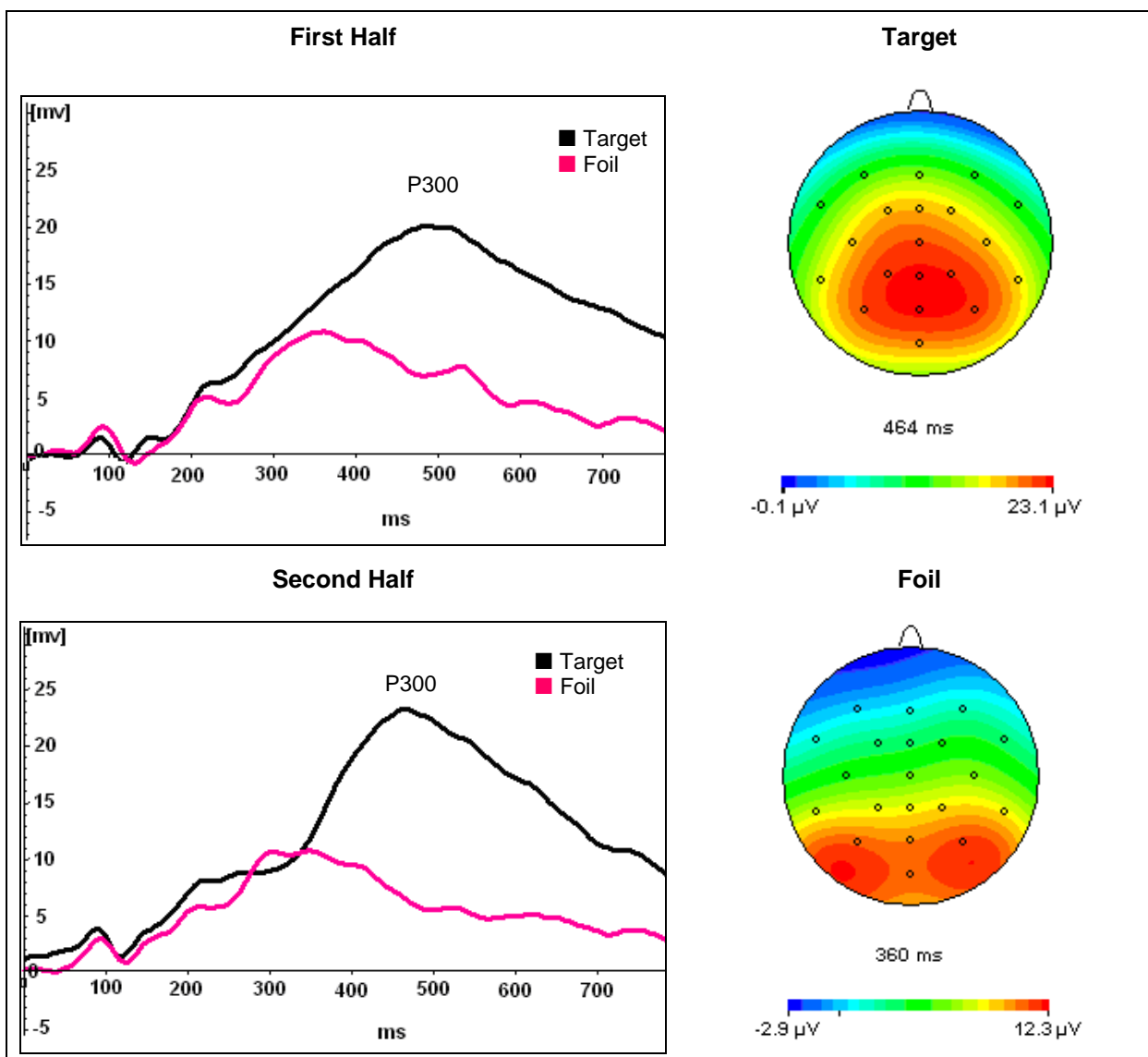
The mean amplitude for the target decreased from the first half ( $M = 10.3 \mu\text{V}$ ,  $SE = 1.17$ ) to the second ( $M = 8.95 \mu\text{V}$ ,  $SE = 1.26$ ), while the mean amplitude for the foil stayed relatively constant from the first half ( $M = 11.1 \mu\text{V}$ ,  $SE = 1.15$ ) to the second ( $M = 12.43 \mu\text{V}$ ,  $SE = 1.28$ ). That is, the target showed a significant N250 in the second half, while the foil generated no N250 response at all, in either half of trials. This result was consistent with research using a similar paradigm in which the N250 for the target condition became more negative as the participants gained experience with the face, while the N250 for foil faces did not show this effect (Tanaka et al. 2006). There was no main effect of hemisphere ( $p = .155$ ) and no other interactions approached significance (smallest  $p = .183$ ).

### *P300 Results*

The P300 is a central posterior component which can be seen throughout channels Cz, Cpz, and Pz and may be analyzed at any of these channels, depending on where the signal is maximal (Ishizu, Ayabe, & Kojima, 2008; Lefebvre et al, 2006). Analysis of the present data showed a maximal P300 amplitude at channel Pz ( $M = 17.24 \mu\text{V}$ ,  $SE = 1.10$ ) compared to channels Cpz ( $M = 16.14 \mu\text{V}$ ,  $SE = 1.17$ ) or Cz ( $13.81 \mu\text{V}$ ,  $SE = 1.39$ ), therefore, P300 analyses were conducted at channel Pz. Much like the N250, the P300 peak was not discernable in the foil condition, so latency analyses were carried out within the target condition based on peak amplitude. A oneway ANOVA determined that experiment half did not affect the latency of the P300 component ( $F(1,21) = .38$ ,  $p = .543$ ,  $h_p^2 = .018$ ). P300 data and topographical maps can be seen in Figure 3.

To analyze amplitude differences between target and foil, mean amplitudes were calculated from 300 ms to 600 ms post stimulus onset. Mean amplitudes were subjected

to a 2 x 2 repeated measures ANOVA with condition (target, foil) and half (first, second) as within-subjects variables. The target condition had a significantly larger mean amplitude ( $M = 17.24 \mu\text{V}$ ,  $SE = 1.10$ ) than the foil condition ( $M = 7.82 \mu\text{V}$ ,  $SE = 1.15$ ),  $F(1,21) = 91.97$ ,  $p < .001$ . There was no main effect of experiment half ( $F = .274$ ,  $p =$



**Figure 3** P300 wave data and topographical maps for Experiment 1. Wave data (left) are separated by half (first, second). Average wave data for target trials is shown in black, average wave data for foils is shown in pink. Topographical maps show positive (red) and negative (blue) impulses as measured across the scalp. The large area of centro-parietal positivity shown on the map for the target is representative of a typical P300 response.

.606). However, the main effect of condition was qualified by an interaction between condition and half,  $F(1,21) = 7.60$ ,  $p = .012$ ,  $h_p^2 = .266$ . Direct comparisons showed that the mean amplitude for the target condition increased from first half ( $M = 16.57 \mu\text{V}$ ,  $SE = 1.21$ ) to second half ( $M = 17.91 \mu\text{V}$ ,  $SE = 1.11$ ) though this was only a trend,  $p = .069$ , while in the foil condition, mean amplitude within the specified epoch did not change from first half ( $M = 8.16 \mu\text{V}$ ,  $SE = 1.31$ ) to second half ( $M = 7.45$ ,  $SE = 1.10$ ),  $p = .387$ .

### *Discussion*

Results from Experiment 1 replicated those of other N250 research showing the acquisition of representations of newly familiarized faces over time (Tanaka et al, 2006, Gordon & Tanaka, submitted). Data demonstrated main effects of condition that were driven by a condition by half interaction. This interaction showed that the N250 was more negative in the second half than in the first for the target condition only, replicating previous research connecting the N250 with familiarity (Itier & Taylor, 2004; Schweinberger et al., 2002; Tanaka et al., 2006), and provided evidence that face representation acquisition for a person encountered in a live interaction elicits an ERP pattern similar to that shown in research using photographs.

Additionally, Experiment 1 replicated P300 findings, in which the target face elicited an enhanced P300 response in comparison to foils (Donchin & Coles, 1988; Fabiani, Karis, & Donchin, 1983). However, because the enhanced P300 component is elicited by both low-probability (Duncan-Johnson & Donchin, 1977) and personally-salient stimuli (Farwell & Donchin, 1991; Keil, et al., 2002), the effect cannot be attributed to recognition.

## Experiment 2

Experiment 2 was designed for several reasons. First, in Experiment 1 the probability of a target occurring was .1, while the probability of a foil occurring was .9. This probability difference is enough to evoke the P300 response we found (Donchin & Coles, 1988). Donchin (1977) showed a negative relationship between target probability and P3 amplitude. A second motivation for Experiment 2 is that in Experiment 1 the only face in the line-up that was salient also required a unique response – any differences in the N250 and P300 results may have been due to their task relevance. In order to be able to associate the N250 and P300 to memory processes, a condition was needed that would be cognitively salient, but not require a different response than any other foil. To do this, participants were told that the experimenter had committed a fictitious crime, but they were asked to “cover” for the experimenter, rather than accuse her. This manipulation used to make the experimenter’s face salient, but not a target.

This manipulation also created an opportunity to examine a second forensic event – deception. Participants were asked to lie. Of course being told to lie is not the same as being intrinsically motivated to lie, and because participants had not actually witnessed a crime, a concern was that they would not be invested enough in the story for the photo of the “culprit” to become salient. In order to increase the salience of the photograph of the culprit-experimenter, it was made very clear that a photo of the experimenter would be in the line-up and if they were being honest they would accuse her. However, when her face appeared they should be sure to say that she was innocent. In this way, the culprit-experimenter would be relevant to the story, but not overtly task-relevant.

Previous research on lie-detection with ERPs has implemented a task called the Guilty Knowledge Test. In it, a group of participants are given inside knowledge of a mock-crime. This is the guilty group. A second control group does not get the additional information – the innocent group. A yes/no task was given in which crime-relevant probes were shown among photos of non-relevant target and non-target items. Results showed an enhanced P300 to the crime-relevant non-target probes in the guilty group only (Farwell & Donchin, 1991). Recent research showed that asking a participant to respond deceptively about their name improved memory detection in a similar task. That is, the P300 was found to be enhanced to memory-related questions if the participant had to be deceptive (Verschuere, Rosenfeld, Winograd, Labkovsky, & Wiersema, 2009). In this case, the respondent was instructed to lie, providing evidence that even sanctioned deception with no fear of consequence can enhance a P300 response.

For this we used a modified Guilty Knowledge Test (GKT), in which participants were told that the experimenter was the culprit of a fictitious crime but that they were to lie and select a different foil to accuse (this non-guilty target is referred to as the “scapegoat” condition). The forced choice was “guilty” or “not guilty”, but in this experiment, unlike Experiment 1, the target was whichever face the participant chose to accuse, and the experimenter’s face was to be treated like the other foils. This is similar to the “probe” condition in the GKT, and therefore, the experimenter’s face is referred to as the experimenter-probe condition. We hypothesized that the N250 would serve as indices of familiarity to both the experimenter-probe face and the scapegoat face even though the experimenter’s face is ostensibly not the target in this paradigm. We also predicted a P300 to both the scapegoat and the experimenter probe. A P300 to both faces

would serve as evidence that both were recognized as relevant, and that the manipulation used to differentiate the experimenter-probe face from other foil faces was successful in spite of both conditions requiring identical overt (“not guilty”) responses.

### *Method*

#### *Participants*

This paradigm was approved by and conducted in accordance with the University of Victoria’s Human Research Ethics Committee. Twenty-one undergraduate students (10 men) from the University of Victoria were recruited from a psychology research subject pool and participated in the study for bonus course credit. Five additional participants were tested but were excluded from analyses, one due to failure to complete the task correctly and four for having fewer than sixteen artefact-free trials in a given condition. The mean age of participants was 22.76 years ( $SD = 5.97$ ). Two participants self-reported that they were left-handed and one reported being ambidextrous. All reported normal or corrected-to-normal vision with no reported brain injury or trauma. Informed consent was obtained from all participants prior to testing.

#### *Materials*

The materials used in Experiment 2 were the same as those used in Experiment 1.

#### *Procedure*

The procedure for Experiment 2 was nearly the same as that for Experiment 1 (refer back to Figure 1), with the exception of a difference in task instruction. Rather than being told to identify the experimenter as the culprit, the participant was told the following, by the experimenter: “Please look closely at the experimenter (me). Next, imagine you have witnessed me robbing a convenience store – you were there and you

know it was me. Now you have been asked to identify the culprit in a police line-up. The problem is that you were there *with* me and you don't want me to be caught because I could implicate you. So what you are going to do instead is accuse someone else.”

Following this, participants were shown all the foil faces simultaneously on a computer screen and asked to select one person that they would accuse. Participants were given as much time as needed to ensure that they would recognize his/her “scapegoat” in the task. They were then given the instructions to press “1” (one) to accuse the scapegoat, and “0” (zero) to say that the person is innocent. Emphasis was placed on ensuring that the participant understood that when the experimenter’s face appeared on the screen, they were to press zero, just like all the other non-target faces. The three conditions that are analyzed are referred Scapegoat (who is the target) Experimenter-probe (salient but non-target), and Foil.

Electrophysiological recordings were taken in the same manner as described above, and when the experiment was complete, participants were debriefed and thanked.

### *Results*

Data were analyzed in a similar way to Experiment 1, in that one random foil was chosen as a comparison (control) condition in order to maintain a consistent number of trials across conditions. The major difference in analysis is that condition now includes three levels: scapegoat (the target), experimenter-probe, and foil. The conservative Bonferroni correction was used to control for Type 1 errors, and the Greenhouse-Geisser method was used to control for sphericity. These methods were not used in the first experiment because no post-hoc comparisons were needed.

### *Behavioural Results*

Mean accuracy of all participants (given that they were considered “correct when they correctly selected the scapegoat and correctly rejected the experimenter-probe) was 97% (with a minimum of 91%, so no participants were excluded based on poor accuracy).

There was a significant main effect of reaction time for condition,  $F(2,21) = 30.196, p < .001, h_p^2 = .602$ . Post-hoc comparisons showed that participants correctly categorized the foil ( $M = 564.1$  ms,  $SE = 36.24$ ) more quickly than they did the experimenter-probe ( $M = 654.92$  ms,  $SE = 33.97$ ),  $p = .001$ . Also, they categorized the experimenter-probe even more quickly than scapegoat, which participants took the longest to categorize ( $M = 710.87$ ,  $SE = 36.24$ ),  $p = .007$ .

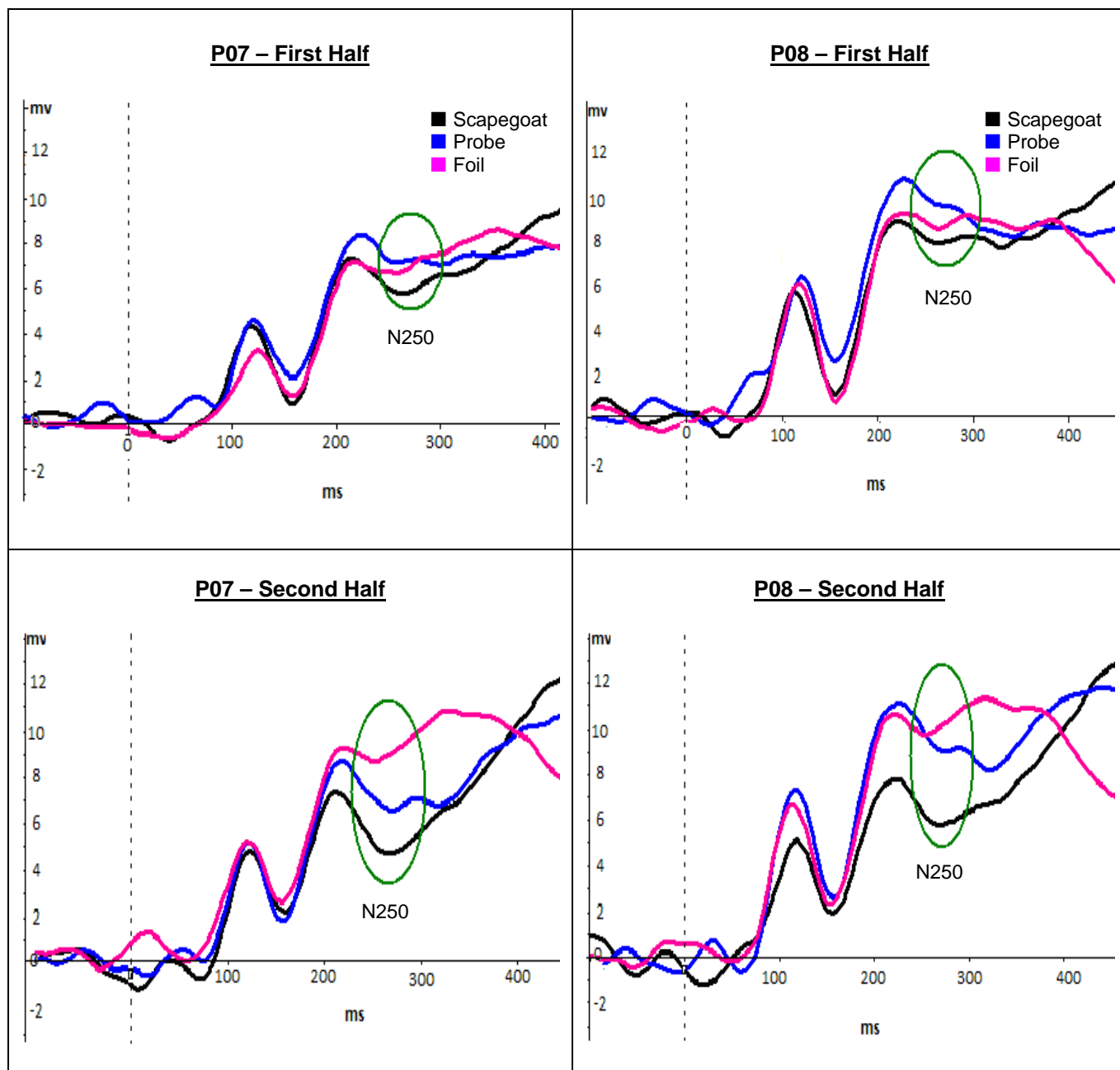
### *N250 Results*

Analyses for the N250 component were performed similarly to the N250 analyses for Experiment 1. The window used to create the mean amplitude was widened to 230-380 ms to accommodate a slightly delayed N250. Wave data for PO7 and PO8 can be seen in Figure 4.

A 2 (hemisphere: left, right) by 3 (condition: scapegoat, experimenter-probe, foil) by 2 (half: first, second) repeated measures ANOVA was conducted. N250 mean amplitude was more negative in the left hemisphere ( $M = 7.48$   $\mu\text{V}$ ,  $SE = .99$ ) than the right ( $M = 8.83$   $\mu\text{V}$ ,  $SE = 1.26$ ),  $F(1,20) = 4.6, p = .044, h_p^2 = .187$ .

There was also a significant main effect of condition,  $F(2,20) = 7.57, p = .002, h_p^2 = .275$ . The scapegoat condition elicited the most negative N250 ( $M = 6.99$   $\mu\text{V}$ ,  $SE = 1.13$ ), followed by the experimenter-probe ( $M = 8.27$   $\mu\text{V}$ ,  $SE = 1.05$ ), and the foil was

least negative ( $M = 9.214 \mu\text{V}$ ,  $SE = 1.23$ ). Post-hoc comparisons showed that while the scapegoat and foil differed significantly ( $p = .007$ ), the experimenter-probe condition did not differ in amplitude from either the scapegoat ( $p = .13$ ) or the foil ( $p = .186$ ).



**Figure 4.** Experiment 2 N250 waves. ERP data from PO7 (left posterior channel) and PO8 (right posterior channel), separated by experiment half (first and second). The average waveform from target trials is presented in black, the average waveform for experimenter-probe (Probe) trials is presented in blue, the average waveform for foil trials is presented in pink. Baselines show -100 ms pre-stimulus onset.

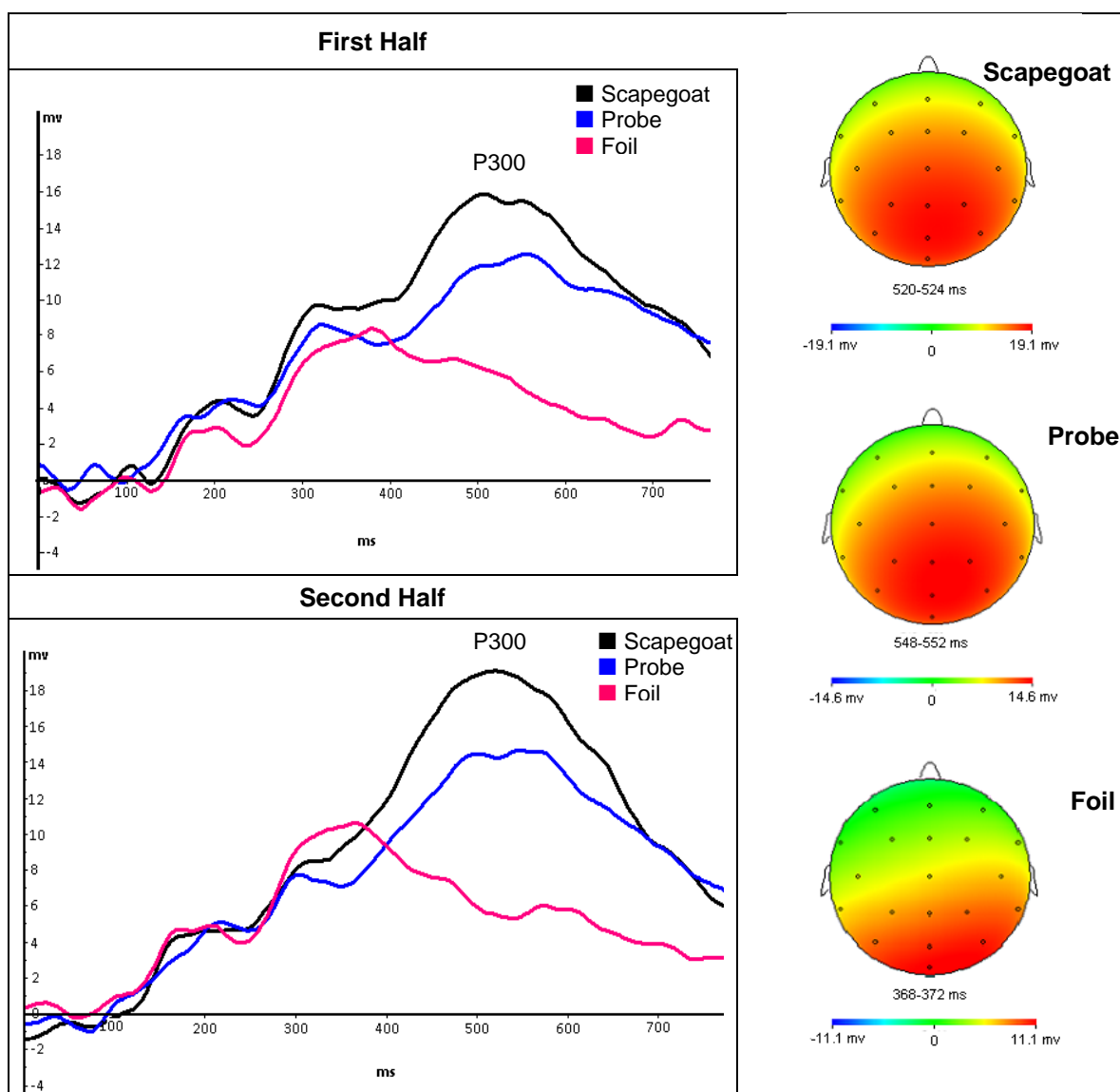
The main effect of condition was qualified by a Condition x Half interaction ( $F(2,20) = 8.144, p = .001, h_p^2 = .289$ ). Based on this, a oneway ANOVA was conducted to test the simple effect of condition within the second half only. Results indicated that the scapegoat ( $M = 7.48 \mu\text{V}, SE = 1.16$ ) and experimenter-probe ( $M = 8.26 \mu\text{V}, SE = 1.08$ ) were not significantly different ( $p = .119$ ) but the N250 for the experimenter-probe condition was significantly more negative than that for the foil condition ( $M = 10.14 \mu\text{V}, SE = 1.28$ ),  $p = .04$ . That is, the scapegoat and experimenter-probe both elicited an equivalent N250, significantly more negative than that elicited by foils.

### *P300 Results*

For the same reasons stated for Experiment 1, peak latency analyses for the P300 were restricted to the scapegoat and experimenter-probe conditions. A 2 x 2 repeated measures ANOVA with condition (scapegoat, experimenter-probe) and half (first, second) as within-subjects variables was conducted and no significant effects of P300 latency were found (largest  $F = 1.588$ , smallest  $p = .222$ ). Wave and topographical data for the P300 at channel Pz can be found in Figure 5.

Mean amplitude analyses within a time epoch of 300-600 ms were conducted using a 3 (condition: scapegoat, experimenter-probe, foil) by 2 (experiment half: first, second) repeated measures ANOVA. The main effect of condition was significant,  $F(2,20) = 30.045, p < .001, h_p^2 = .6$ . Mean amplitudes were  $13.58 \mu\text{V}$  ( $SE = 1.29$ ) for the scapegoat condition,  $10.66 \mu\text{V}$  ( $SE = .85$ ) for the experimenter-probe condition, and  $7.22 \mu\text{V}$  ( $SE = .98$ ) for the foil condition. Post-hoc analyses were conducted and showed that the amplitude for the scapegoat was significantly greater than that for the experimenter-probe ( $p = .004$ ) which, in turn, was significantly greater than that for the foil ( $p < .001$ ).

There was also a significant main effect of experiment half. The second half ( $M = 11.27 \mu\text{V}$ ,  $SE = .93$ ) was significantly more positive than the first ( $M = 9.72 \mu\text{V}$ ,  $SE = .99$ ),  $F(1,20) = 19.1$ ,  $p < .001$ ,  $h_p^2 = .489$ . There was no significant interaction between condition and half ( $F = .241$ ,  $p = .765$ ).



**Figure 5.** P300 wave data and topographical maps for Experiment 2. Wave data (left) are separated by half (first, second). Average wave data for scapegoat trials is shown in black, average wave data for the probe is shown in blue, and average wave data for foils is shown in pink. Topographical maps show positive (red) and negative (blue) impulses as measured across the scalp within the second half. A large parieto-central positivity can be seen for the target and probe conditions only (top and centre, respectively), typical of the P300 component. This activity is not seen for foils (bottom).

### *Discussion*

In Experiment 2, the addition of the experimenter-probe face (which is different from the experimenter target from Experiment 1 because in Experiment 2 the experimenter is a non-target probe item) provided a unique finding. The experimenter was made salient by interacting with the participant in the real world, and by the story that she was the actual culprit of the fictitious crime described in the instructions. The experimenter face was to be treated the same as any other foil face, but the results of Experiment 2 suggest that she was not perceived the same way. The N250 in response to the experimenter-probe condition fell in between the scapegoat and foil conditions, not significantly different from either. The P300 data are similar in that the amplitude of the P300 in response to the experimenter-probe was significantly smaller than that elicited by the scapegoat, but also significantly *larger* than that elicited by foils. Because every photograph in the experiment was shown an equal number of times, and foils required the same button-press response as the experimenter-probe, the ERPs results cannot be explained by either an oddball effect (because as a foil, the experimenter-probe is shown an equal number of times compared to other individual foils) or task relevance (because the experimenter-probe is not a target).

The experimenter-probe face, in spite of being seen in real-life and encoded with contextual information as well as motion information, was treated the same as both the scapegoat and the foils at the start of the experiment. This might indicate that the real-world interaction did not affect the N250 any differently than a photo. However, after a series of trials, our results suggest that the experimenter was not physiologically identical to the other foils. When restricting N250 analyses to the second half of the experiment,

the experimenter-probe condition does differ from the scapegoat, and is significantly enhanced compared to foils. This result is not necessarily due only to a saliency effect. Recent research using a similar “target / non-target” task has shown that a female non-target face embedded within a series of male faces (one of the males is the target face) does not elicit an enhanced N250. A stronger manipulation using a highly visually distinctive face (for example, a black non-target face embedded within a series of white faces (one of which is the target) or vice versa) also found no effect of distinctiveness on the N250 (Kantner, et al., 2009). That an N250 effect was seen for the experimenter-probe suggests more than just photo saliency. Given that the N250 is sensitive to familiarity, it seems probable that the N250 effect seen in Experiment 2 is indexing the participants’ recognition of the experimenter, in the same way that the probe items in the GKT will elicit a P300 (Allen & Iacono, 1997; Farwell & Donchin, 1991; Johnson & Rosenfeld, 1992).

### Experiment 3

Experiment 3 was a first effort to use ERPs to examine real-world memory in the applied setting of eyewitness memory. The results of Experiments 1 and 2 were promising and motivated this next step, in which participants had to rely on their incidental memory rather than being told that they would see a photo of the experimenter. In this experiment, instead of telling the participant to imagine witnessing a crime, they actually *did* witness a crime (albeit a simulated one). As each participant waited outside the lab, a confederate-thief walked by and stole an mp3 player from a seat near him/her. They were told moments later that the theft was part of the study and that their job would be to identify the culprit from a photo line-up while their brainwaves were recorded.

Based on the previous two experiments, we expected that when participants correctly identified the thief (Correct ID group), their brainwaves would show an enhanced N250 and P300 to their correct identification of the culprit (the confederate-thief), compared to a correct rejection of a foil.

In the present study, because participants were only exposed to the thief for a shorter duration and required incidental memory as they did not know in advance that they would need to remember her, accuracy was not expected to be at ceiling. We expected that some people would make an incorrect identification (Incorrect ID group). In this case, like in Experiment 2, we expected to see enhanced N250 and P300 components to their selected foil (incorrect ID trial). For these incorrect ID cases, a third condition is available for analyses: we planned to examine brainwaves in response to the incorrect rejection of the culprit. This condition might be similar to the experimenter-

probe condition in Experiment 2 because the representation of the actual culprit may be stored and the participant's ERP data (N250 and/or P300) may still present with a memory trace. Based on data from the GKT and from Experiment 2, we hypothesized that incorrect identifiers would still show an enhanced N250 and P300 to the confederate-thief, even in the absence of an overt recognition decision, and that the amplitude of these two components in response to the confederate-thief would likely fall somewhere between the amplitudes for incorrect ID trials and the correct rejection of foils. On the other hand, if we fail to find a difference between the incorrect rejection of the confederate-thief and the correct rejection of a foil, it seems likely that the participant experienced an encoding failure; that he/she did not encode and cannot possibly remember the face of the culprit. Note that incorrect ID trials are analogous to a false identification in a real-world line-up, which is a condition of particular interest because of the significant applied consequences of making such an error – incarcerating an innocent person and allowing a guilty person to remain free.

### *Method*

#### *Participants*

Sixty-three undergraduate students from the University of Victoria participated in the study, in exchange for optional course credit. This paradigm was approved by and conducted in accordance with the University of Victoria's Human Research Ethics Committee. Nine participants selected multiple culprits, one failed to complete the task correctly, two were unable to complete the task because of equipment failure, and three had movement artefacts in more than half of trials. These fifteen participants were excluded from analyses. The remaining 48 participants (31 female) ranged between 17

and 27 years, with a mean age of 19.98 ( $SD = 2.08$ ). Forty-one participants reported right-hand dominance; six reported left-hand dominance, and one reported being ambidextrous. None reported any serious history of brain injury or trauma (one reported a mild concussion in the past year but we did not find this serious enough to warrant exclusion), and all claimed to have normal or corrected-to-normal vision.

### *Materials*

The materials used in Experiment 3 were the same as those used in Experiments 1 and 2.

### *Procedure*

Once again, the experimental procedure for Experiment 3 was nearly the same as that for Experiments 1 and 2 (refer back to Figure 1), with the exception of a difference in the exposure phase. In experiments 1 and 2, the experimenter introduced herself to the participants and explained how her photo would be among the photo array (thus engaging intentional memory – the participant had the opportunity to study the experimenter’s face after being told the task instructions). The present experiment was somewhat more complex. When the participant arrived for his/her appointment s/he was asked to sit in a chair outside the lab and wait for their experimenter. Shortly after, a confederate-participant sat down next to them and pretended to also be waiting for an experiment. After a few minutes the confederate-participant stood up, placed her belongings (a backpack and an mp3 player) on the chair and said the following line: “I have to run upstairs for a minute but will be right back, in case someone comes out to get me” (referring to an experimenter). Care was taken not to ask the real participant to watch the belongings, as we did not want the participant to feel protective over the items (which

might result in the participant intervening in the theft). Additionally, we wanted to minimize any feelings of guilt or fault in the participant, once the theft occurred. A few minutes after the confederate-participant exited, the confederate-thief walked past the participant, stopped, and asked the participant if the things on the chair belong to her/him. When the participant replied that they do not (as they always did), the confederate-thief grabbed the mp3 player and ran away. The conversation is included to ensure the participant actually looked directly at the confederate-thief's face. The length of the conversation was approximately 10-15 seconds, based on the length of participant responses. The confederate-participant then immediately returned to inform the participant that the theft was part of the study and that the participant's job is to identify the culprit. The time that participants are led to believe that the theft is real is approximately 5 seconds – the time it takes to run down a length of hallway, approximately 10 meters. The brevity of the deception is intentional, because we wanted to elicit a feeling of surprise in order to approximate a real-world witness scenario, while alleviating any feelings of distress or fear as quickly as possible. This event is similar to mock crime scenarios used in previous eyewitness identification research (Lindsay, Wells, & Ferguson, 1979).

Following the mock-crime, participants were led to the ERP lab and were fitted with the electrode cap. Electrophysiological recordings were taken in the same manner as described above for Experiments 1 and 2. Participants were shown 10 photos sequentially (randomized within sets of 10), in four blocks of 100 trials, for a total of 400 trials – 40 per photo. This is the same method as was done in both previous experiments. Participants were asked to press “1” to identify the thief, and “0” to indicate that the

photo is *not* the thief. After learning that some participants accused more than one of the people in the line-up, instructions were modified to include an explanation that each of the photos was of a different person and that there were no further tricks in the study. They were asked to make a selection and then, while it would be okay for the participant to change his/her mind early in the experiment, s/he should try to select one person as early as possible and to commit to choosing that person for the remainder of the session.

Upon completion of the ERP task, participants were asked several follow-up questions. They were asked to rate their exposure to the confederate thief at the time of the crime (“on a scale of 1-7, how good of a look did you get at the thief, 1 being didn’t look at all, 7 being a good, clear look), their confidence in their choice (1 being a complete guess, 7 being 100% certain), as well as how distressed they felt when they witnessed the event (1 being not at all distressed, 7 being extremely distressed). Finally, participants were debriefed and thanked. The debriefing included a description of the motivation and hypotheses of the study, and well as a complete explanation as to why it was important that they did not know about the theft in advance. They were then asked not to discuss the event with other students.

### *Results*

For each participant in all analyses a single random foil was selected to compare with the culprit (or the Incorrect ID). This is the standard procedure for ERP analyses, as it is important that conditions are matched for number of trials. In order to keep all analyses consistent, behavioural data were analyzed in the same way, with the same foils for each participant. Like analyses for Experiment 2, where post-hoc tests were used, the

Bonferroni correction was used to control for Type 1 errors in multiple comparisons, and the Greenhouse-Geisser statistic was used, to control for violations of sphericity.

### *Behavioural Results*

Most participants ( $N = 39$  or 81%) reportedly believed that the theft was real. Only two participants thought it was part of the study, and the remaining seven believed that the thief was a friend of the confederate and was not actually stealing the mp3 player. Several participants actively pursued the thief, but were quickly stopped by the confederate-participant and assured that no theft had occurred and that the event was part of the study.

Generally, participants found the task to be quite difficult. Of the 48 participants, 21 (44%) correctly identified the culprit (correct ID group) and the remaining 27 mistakenly identified an innocent foil (incorrect ID group). Oneway ANOVAs indicated that on a scale of 1-7, participants in the incorrect ID group reported less exposure or a worse view of the culprit ( $M = 3.65$ ,  $SD = 1.06$ ) than the correct ID group ( $M = 4.5$ ,  $SD = 1.05$ ),  $F(1,46) = 7.67$ ,  $p = .008$ . The correct ID group also reported greater confidence in their choice ( $M = 4.95$ ,  $SD = 1.56$ ) than the incorrect ID group ( $M = 3.37$ ,  $SD = 1.29$ ),  $F(1,45) = 14.38$ ,  $p < .001$ . Reports of exposure and confidence were marginally correlated, Pearson  $R = .275$ ,  $p = .061$ .

Reaction time also differed as a function of accuracy. Correct identifiers responded more quickly than incorrect identifiers when making identifications (correct ID:  $M = 688.21$  ms,  $SD = 156.74$ ; incorrect ID:  $M = 789.47$  ms,  $SD = 144.56$ ),  $F(1,46) = 5.39$ ,  $p = .025$ . and when correctly rejecting foils (correct ID group:  $M = 539.52$  ms,  $SD = 91.17$ ; incorrect ID group:  $M = 612.03$  ms,  $SD = 132.87$ ),  $F(1,46) = 4.57$ ,  $p = .038$ .

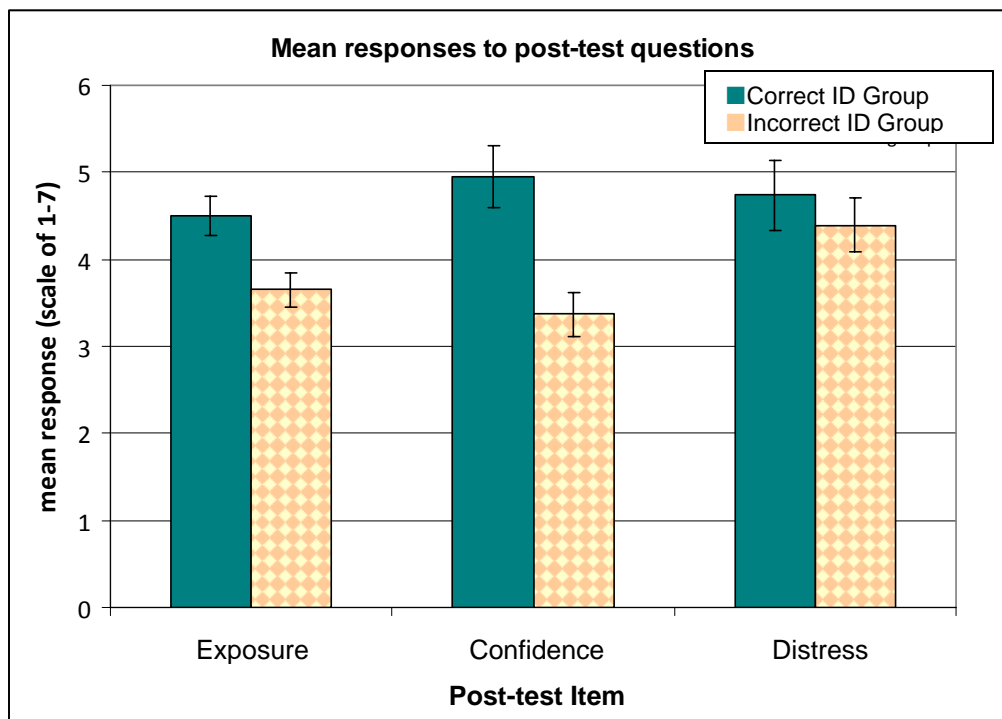
Within the incorrect ID group there were three possible responses in a given trial: an incorrect identification of an innocent foil (Incorrect ID), a correct rejection of a foil (foil), and an incorrect rejection of the actual culprit, these three types of trials are referred to hereafter as “stimulus type”. A 3 level (stimulus type: incorrect ID, foil, incorrect rejection of the culprit) repeated measures ANOVA was conducted on RT and showed a significant main effect,  $F(2,52) = 21.68, p < .001, h_p^2 = .455$ . Post-hoc comparisons indicated that participants took significantly more time to incorrectly reject the actual culprit ( $M = 728.67$  ms,  $SE = 35.1$ ) than to correctly reject a foil ( $M = 612.03$  ms,  $SD = 132.87$ ),  $p = .001$ . RT for incorrect ID trials and for incorrect rejection trials did not differ, taking a similar amount of time to mistakenly accuse an innocent foil as to incorrectly reject the culprit,  $p = .110$ .

A oneway ANOVA comparing reported level of distress (reported on a scale of 1-7) showed that distress level was approximately equal in both the correct ID ( $M = 4.74$ ,  $SD = 1.83$ ) and incorrect ID ( $M = 4.39$ ,  $SD = 1.6$ ) conditions,  $F(1,46) = .496, p = .485$ . Correlational analyses suggested a positive relationship between distress level and reported exposure. Those that reported higher levels of distress were also more likely to report greater exposure to the culprit, Pearson  $R = .323, p = .027$ . Distress was not correlated with foil RT, choice RT, or confidence (lowest  $p = .541$ ). Behavioural data for exposure, confidence, and distress level can be found in Figure 6.

### *N250 Results*

The N250 component was calculated as a mean-amplitude within a time window of 230-330 ms, as was done in Experiment 1 (Herzmann, Schweinberger, Sommer, &

Jentsch, 2004; Tanaka & Pierce, 2009). Separate analyses were conducted for the Correct ID group and the Incorrect ID group.



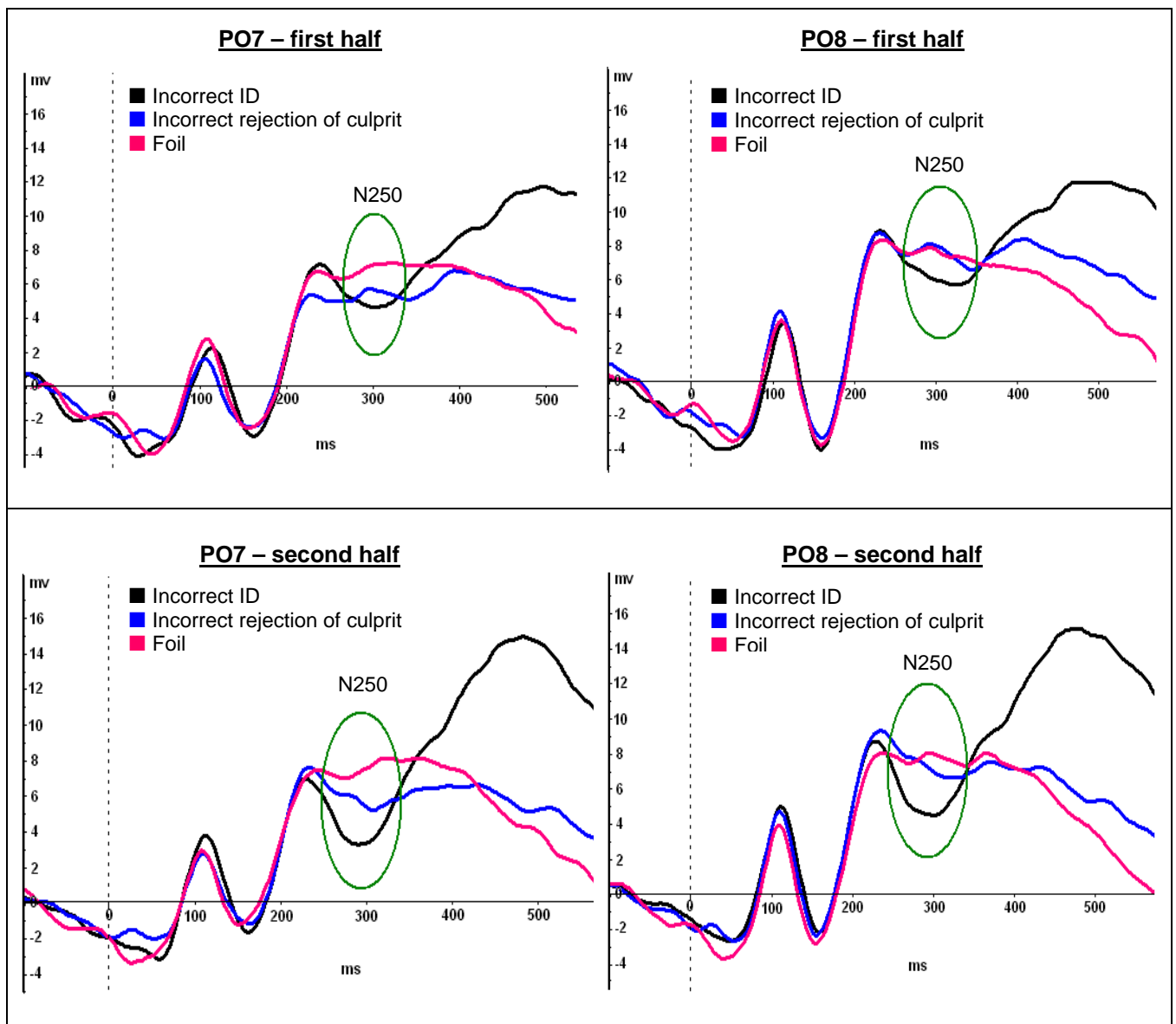
**Figure 6.** Mean behavioural data for the Experiment 3 post-study questionnaire. Results are reported as a mean on a scale of 1-7. Dark bars represent the Correct ID group, light bars represent the Incorrect ID group. Error bars are 95% within-subjects confidence intervals.

*Correct ID Analyses.* Within the Correct ID group, a 2 (stimulus type: correct ID, foil) by 2 (hemisphere: right, left) by 2 (experiment half: first, second) repeated measures ANOVA was conducted. A significant main effect of stimulus type was found in that the time epoch for the N250 was more negative for the Correct ID ( $M = 6.84 \mu\text{V}$ ,  $SE = .923$ ) than for foils ( $M = 8.6 \mu\text{V}$ ,  $SE = 1.07$ ),  $F(1,20) = 8.752$ ,  $p = .008$ ,  $h_p^2 = .304$ . The analyses also showed the left hemisphere was more negative ( $M = 6.61 \mu\text{V}$ ,  $SE = 1.02$ ) than the right ( $M = 8.83 \mu\text{V}$ ,  $SE = 1.09$ ),  $F(1,20) = 5.751$ ,  $p = .026$ ,  $h_p^2 = .223$ . There was no main effect of experiment half ( $F(1,20) = .021$ ,  $p = .887$ ) and there were no significant

interactions between any of the variables (largest  $F = 1.817$ , smallest  $p = .193$ ). The N250 mean amplitude data at channel PO7 (where the effect was strongest) for both correct ID and incorrect ID groups are presented in graph format in Figure 8.

*Incorrect ID Analyses.* Within the Incorrect ID group, a 3 x 2 x 2 repeated measures ANOVA was conducted with stimulus type (incorrect ID trials, incorrect rejection of culprit, foil), hemisphere (left, right), and experiment half (first, second) as within-subjects variables. Wave data for the N250 can be seen in Figure 7. There was a main effect for stimulus type,  $F(2,52) = 6.174$ ,  $p = .005$ ,  $h_p^2 = .192$  and a main effect of hemisphere, with more negativity found in the left hemisphere ( $M = 5.9 \mu\text{V}$ ,  $SE = .89$ ) than the right ( $M = 7.3 \mu\text{V}$ ,  $SE = .94$ ),  $F(1,26) = 5.018$ ,  $p = .034$ ,  $h_p^2 = .162$ . There was no effect of experiment half ( $F(1,26) = .161$ ,  $p = .691$ ). Post-hoc comparisons of stimulus type (using the Bonferroni correction for multiple comparisons) indicated that incorrect ID trials ( $M = 5.64 \mu\text{V}$ ,  $SE = .94$ ) were significantly more negative than correct rejections of foils ( $M = 7.42 \mu\text{V}$ ,  $SE = .96$ ),  $p = .004$ . Incorrect rejections of the culprit ( $M = 6.73 \mu\text{V}$ ,  $SE = .815$ ) were not found to differ significantly from the incorrect ID trials ( $p = .163$ ) or from the foil rejections ( $p = .424$ ). The significant main effects of hemisphere and stimulus-type were qualified by a significant interaction,  $F(2,52) = 7.289$ ,  $p = .002$ ,  $h_p^2 = .219$ . Motivated by the significant interaction and based on the hypothesis for this experiment in which we were interested in the N250 effect for incorrect rejections of the culprit, further analyses were conducted within the left hemisphere alone, where the N250 effect was determined to be more pronounced. A single variable, three-level (stimulus type: incorrect ID trial, incorrect rejection of culprit, foil) repeated measures

ANOVA within the left hemisphere again showed a significant main effect of stimulus type,  $F(2,52) = 8.22, p = .001, h_p^2 = .24$ . Post-hoc comparisons showed again that the

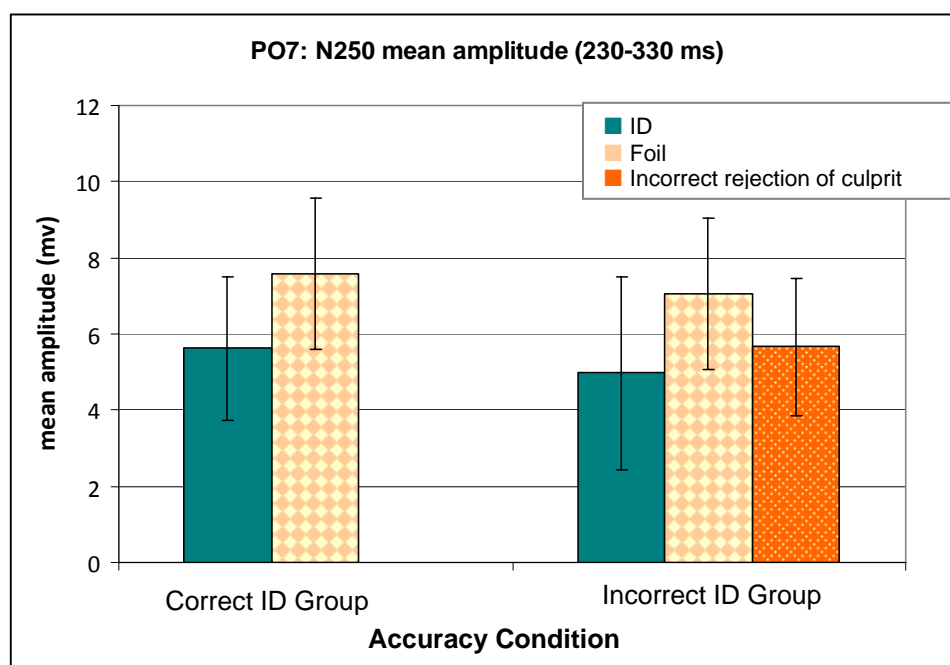


**Figure 7.** N250 wave data for the Incorrect ID group from PO7 (left hemisphere) and PO8 (right hemisphere) shown separated by experiment half. The average wave data for Incorrect ID trials is presented in black, the average of trials for incorrect rejections of the experimenter-culprit is presented in blue, the average waveforms for correctly rejected foils is presented in pink.

incorrect ID trial evoked the most negative N250 ( $M = 4.97 \mu\text{V}, SE = .97$ ), significantly more negative than that evoked for foils ( $M = 7.06 \mu\text{V}, SE = .98$ ),  $p = .001$ . Incorrect

rejections of the culprit were also found to evoke a more negative N250 than the foils ( $M = 5.66\mu\text{V}$ ,  $SE = .88$ ),  $p = .027$ . Incorrect ID trials and the incorrect rejection of the culprit did not differ significantly,  $p = .664$ .

No other interactions reached significance, though hemisphere and half trended toward significance,  $F(1,26) = 5.037$ ,  $p = .093$ . In the left hemisphere there is a slight increase from the first half ( $M = 5.84\mu\text{V}$ ,  $SE = .955$ ) to the second half ( $M = 5.96\mu\text{V}$ ,  $SE = .864$ ), while in the right hemisphere there is a slight *increase* in amplitude from the first half ( $M = 7.5\mu\text{V}$ ,  $SE = .991$ ) to the second half ( $M = 7.11\mu\text{V}$ ,  $SE = .911$ ) of experimental trials.



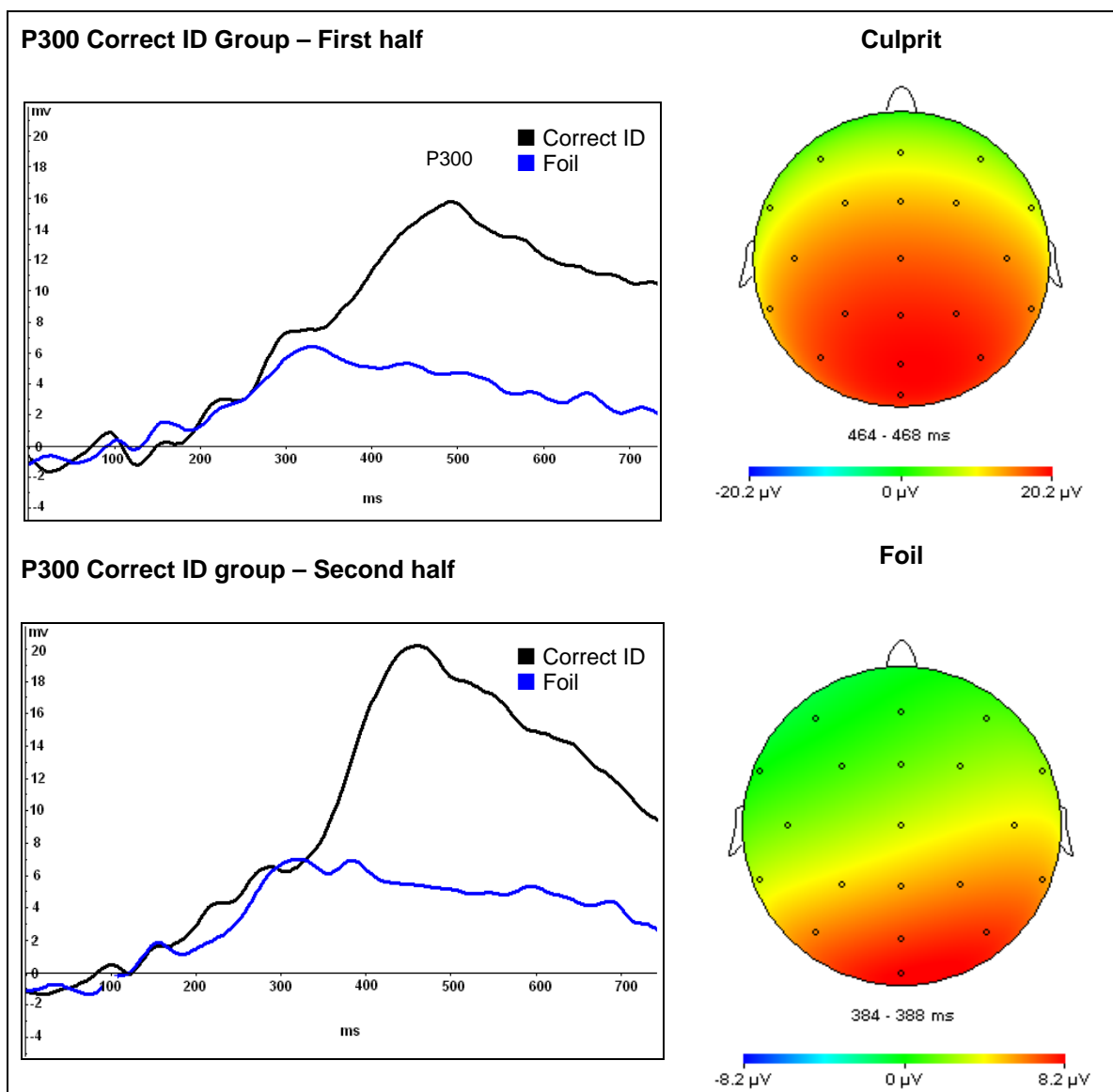
**Figure 8.** N250 mean amplitude data at PO7 for both Correct and Incorrect ID groups. Data are averaged across experiment half. Error bars are 95% within-subjects confidence intervals.

### *P300 Results*

*Within-subjects analyses.* The P300 component was calculated as a mean-amplitude within a time window of 300 to 600 ms, at channel Pz (Lefebvre et al. 2006). Like

analyses for the N250, data for the Correct ID group and the Incorrect ID group were analyzed separately. P300 wave and topographical data are displayed in Figure 9.

*Correct ID Analyses.* Within the Correct ID group, results of a 2 x 2 repeated measures ANOVA, with stimulus type (correct ID, foil) and experiment half (first, second) as within-subjects variables. Results indicated a significant effect of stimulus type, in that



**Figure 9.** P300 waveform data and topographical maps at channel Pz for the Correct ID group. Data are separated by half. Topographical maps show scalp activity in the second half for the Correct ID (top) and Foil (bottom) during the most positive point in the P300 window (300 ms to 600 ms).

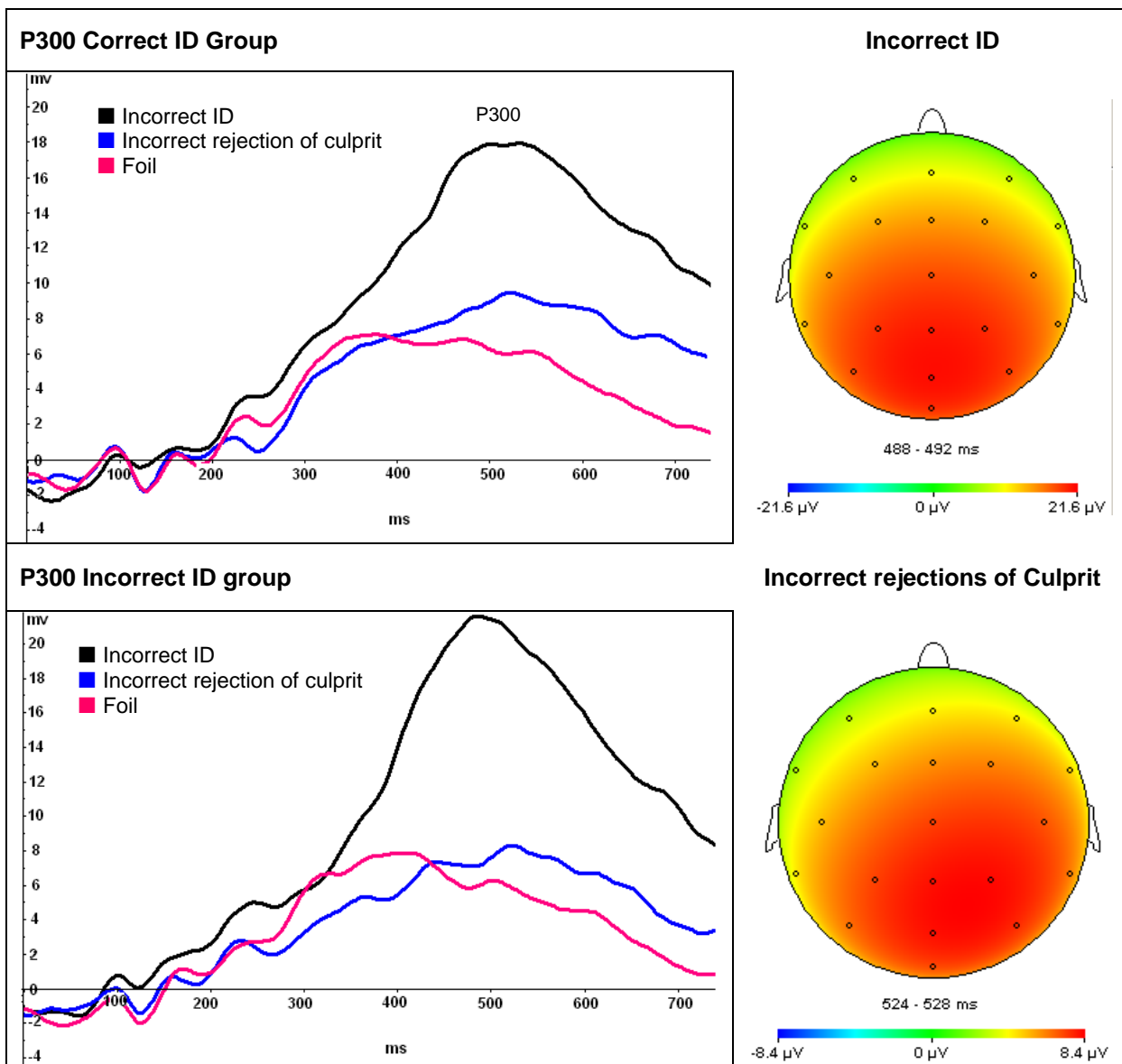
P300 amplitude for correct IDs ( $M = 13.59 \mu\text{V}$ ,  $SE = 4.39$ ) was much greater than that elicited by foils ( $M = 5.37 \mu\text{V}$ ,  $SE = 3.68$ ),  $F(1,20) = 110.48$ ,  $p < .001$ ,  $h_p^2 = .847$ .

Analyses also revealed a main effect of experiment half. The P300 was significantly greater in the second half ( $M = 10.38 \mu\text{V}$ ,  $SE = .79$ ) than in the first ( $M = 8.54 \mu\text{V}$ ,  $SE = .865$ ),  $F(1,26) = 16.928$ ,  $p = .001$ ,  $h_p^2 = .458$ . Wave and topographical data are presented in Figure 9.

The interaction between stimulus type and half did not reach significance ( $F = 3.077$ ,  $p = .095$ ), but was directional, suggesting that the increase in P300 amplitude from first to second half was greater for the correct ID condition than for foils.

*Incorrect ID Analyses.* Within the Incorrect ID group, a 3 x 2 repeated measures ANOVA was conducted, with stimulus type (incorrect ID, incorrect rejection of the culprit, foil) and experiment half (first, second) as within-subjects variables. Analyses indicated a main effect of stimulus type,  $F(2,52) = 59.626$ ,  $p < .001$ ,  $h_p^2 = .696$ . Post-hoc comparisons showed that this effect was motivated by the incorrect ID, which was significantly larger in amplitude ( $M = 14.81 \mu\text{V}$ ,  $SE = 1.11$ ) than either the incorrect rejection of the culprit ( $M = 7.06 \mu\text{V}$ ,  $SE = .86$ ),  $p < .001$ , or the correct rejection of foils ( $M = 6.45 \mu\text{V}$ ,  $SE = .75$ ),  $p < .001$ . The incorrect rejection of the culprit and correct rejection of foils did not differ,  $p = 1.00$ .

There was no significant main effect of experiment half ( $F(1,26) = .040$ ,  $p = .844$ ). However, there was a significant interaction between stimulus type and half, showing that the Incorrect ID trials developed a stronger P300 response from the first ( $M = 13.998 \mu\text{V}$ ,  $SE = 1.08$ ) to the second half ( $M = 15.624 \mu\text{V}$ ,  $SE = 1.223$ ) of trials while foils (first:  $M = 6.47 \mu\text{V}$ ,  $SE = .948$ ; second:  $M = 6.415 \mu\text{V}$ ,  $SE = .71$ ) and incorrect

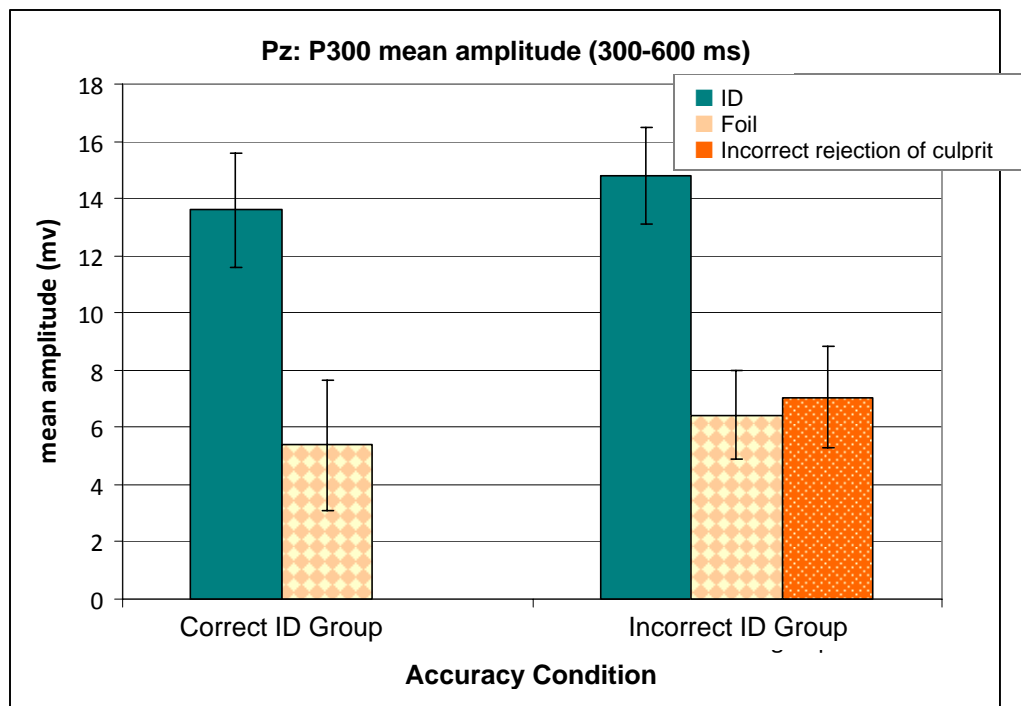


**Figure 10.** P300 waveform data and topographical maps at channel Pz for the Incorrect ID group. Data are separated by half. Topographical maps show scalp activity in the second half for the Incorrect ID (top) and Incorrect Rejections of the Culprit (bottom) during the most positive point in the P300 window (300 ms to 600 ms). Topographical maps for foils were similar in correct and incorrect ID groups (see Figure 9 for map of Foil trials).

rejections of the culprit (first:  $M = 7.763 \mu\text{V}$ ,  $SE = .86$ ; second:  $M = 6.43 \mu\text{V}$   $SE = 1.03$ )

did not increase,  $F(2,52) = 4.042$ ,  $p = .024$ ,  $h_p^2 = .135$ .

P300 data for Correct and Incorrect ID groups averaged across half are presented in graph format in Figure 11.



**Figure 11.** P300 amplitude data plotted for both Correct and Incorrect ID groups, averaged across experiment half. Error bars are 95% within subjects confidence intervals.

*Between-subjects analyses.* Oneway ANOVAS of both N250 and P300 amplitude data for line-up targets showed no difference based on accuracy. That is, P300 and N250 amplitudes for correct identifications (correct ID group) and mistaken identifications (incorrect ID group), or were roughly equivalent, largest  $F = .654$ , smallest  $p = .426$ .

An analysis of P300 latency differences between the Correct and Incorrect ID groups was conducted based on visual inspection of the waves. A direct comparison (oneway ANOVA) of P300 peak latency showed that the P300 for the Incorrect ID group was delayed ( $M = 512.89$  ms,  $SE = 11.1$ ) compared to the Correct ID group ( $M = 477.71$  ms,  $SE = 11.83$ ),  $F(1,46) = 4.627$ ,  $p = .037$ . Based on previous research, correlational analyses were conducted on P300 latency of each participant's choice (either a correct ID or an incorrect ID) and RT (for the participant's choice), as well as reported exposure,

and confidence. The only correlation that reached significance was RT and P300 latency for participant choice, Pearson  $R = .379$ ,  $p = .008$ .

### *Discussion*

Experiment 3 was somewhat different in that the exposure paradigm (the theft) was brief and unexpected, and therefore accuracy was not assured, and fewer than half of participants accurately identified the culprit. The ERP data from the Correct ID group replicate those seen for targets and foils in Experiments 1 and 2, with the exception of the half effect, typically seen in the N250 for target items. Enhanced N250 and P300 components were seen for the correct ID in comparison to foils. These data provide evidence that the N250 and P300 effects seen in previous research using photographs (Itier & Taylor 2002, 2004; Miyakoshi et al., 2008; Schweinberger et al., 2002; Tanaka et al., 2006) is generalizable to real-life interactions in which recognition is correct. The mean amplitudes of the N250 and P300 both differentiate newly familiarized faces from unfamiliar foils.

In the Incorrect ID group, both the N250 and P300 amplitudes again distinguished between identifications (which then become the participant's target item) and foils (non-target items), replicating Experiments 1 and 2 and the data for correct identifiers in Experiment 3. However, only the N250 amplitude could be used to differentiate incorrect rejections of the culprit from correct rejections of foils, and this only occurred in the left hemisphere. This is interesting because research has typically found that the N250 effect is dominant in the right hemisphere (Pfütze, Sommer, & Schweinberger, 2002; Schweinberger et al., 2002; Schweinberger, Huddy, & Burton, 2004). Some research has suggested that the left-hemisphere dominant N250 occurs as a familiarity response, while

the right hemisphere N250 occurs in response to self-relevant items (Miyakoshi et al. 2008). This was suggested because the researchers observed a left dominant N250 to famous faces, and a right dominant N250 to one's own face. However, follow-up research on self-relevant object recognition failed to find right hemisphere lateralization for self-relevant items (Miyakoshi, Nomura, & Ohira, 2007). The reason for right or left lateralization within the N250 time window is an open question. The right laterality could be explained by the hypothesis that the N250 face familiarity effect can be localized to the right fusiform gyrus, a region of the brain implicated by fMRI research in the processing of face information (Kanwisher, McDermott, & Chun, 1997). However, because laterality effects differ across studies, it is difficult to make hypotheses regarding hemispheric dominance. No published research has attempted to systematically explain this variability and this area remains a gap in the literature.

A particularly interesting and unexpected result is that the N250 did not change significantly from the first to the second half of trials for target items (identification trials) in either the correct or incorrect conditions. A stimulus type by half effect was anticipated, that would have shown that the N250 amplitude was becoming more negative for target items across time, while not affecting non-target items. This result would have shown that the acquisition of a familiarity representation was occurring in a similar way to representations acquired in research using photographs. That experiment half does not affect the N250 regardless of accuracy is a surprising result as the research demonstrating a stronger N250 effect with increased experience with an item is reasonably well-established. This finding might suggest that the mock-crime component of Experiment 3 (perhaps the brevity of the encounter, the increased stress or surprise

associated with the event, or the fact that participants did not know they would have to remember the thief's face until after she was out of sight) might play a role in the speed of representation acquisition, in that the N250 was already present in the first half of trials. However, visual inspections of the N250 waveforms (see Figure 7) still suggests that the N250 is more pronounced in the second half of trials, and the non-significant effect of half might be a product of the time window required by N250 analyses – because it is not a peak amplitude, but rather a mean amplitude within 100 ms, the half effect may be hidden by other activity within the time epoch. Additional studies replicating this result would be needed before making any concrete hypotheses as to the cause.

Arguably the most interesting result in Experiment 3 is the finding that the N250 in left posterior regions significantly differentiates the two different types of rejection conditions: the correct rejection of foils, and the incorrect rejection of the actual culprit. This finding supports the first hypothesis for the experiment – that a significant N250 effect for rejections of the actual culprit might suggest an error in retrieval. If the alternative hypothesis was true and the culprit's face was never encoded, there should have been no N250 effect for incorrect rejections of the culprit. This result suggests that the participants who *say* the culprit is innocent do in fact retain some internal representation of the face. This is not necessarily to say that an internal representation exists without the participants knowledge, but perhaps that the culprit's face feels familiar, but does not reach the critical threshold of familiarity that is required before a participant would say that an item is “old” (seen before). Instead, participants may decide to use a more featural style of discrimination, choosing their “culprit” based on

something they do remember, like hairstyle. Qualitatively, some participants offered explanations for their choice and some of the explanations were feature-driven such as “I remember her hair being longer”, suggesting that they were not selecting the culprit based on a holistic sense of familiarity. These data also suggest that differences in the amplitude of the N250 are not necessarily driven by a conscious and confident recognition of a face.

Additionally, correct identifiers made their responses more quickly than incorrect identifiers, a result that is consistent with eyewitness research indicating that correct lineup IDs tend to be faster than incorrect IDs (Smith, Lindsay, & Pryke, 2000; Sauerland & Sporer, 2009; Weber, Brewer, Wells, Semmler, & Keast, 2004).

That the P300 does not differentiate incorrect rejections of the culprit and correct rejections of foils indicates that in this case, the P300 cannot be used to detect subtle instances of recognition. That there was no distinguishable P300 to the incorrect identifications of the culprit also suggest that the photo of the culprit was not salient enough to evoke an oddball or personal saliency response. It is very interesting that the component most associated with familiarity (the N250) *does* dissociate incorrect rejections of the culprit from correct rejections of foils, while the component associated with saliency, task relevance, and probability does not. This dissociation of the N250 and P300 provides evidence that the N250 is the better indicator of recognition of faces. Additionally, on a basic science level, these data reveal a dissociation between the N250 and P300, which have been suggested to be a related complex, and that when one increases, so does the other (Warren, Tanaka & Holroyd, 2010). The present data suggest that at least in some cases, the N250 and P300 are independent.

Post-test data showed that confidence and exposure were reported to be higher in the correct ID group than in the incorrect ID group. These differences could be explained by the fact that the confederate-thief was always present for the follow-up and debrief, so in many cases the participant was aware of whether or not s/he had made a correct choice before the follow-up questions were asked. This knowledge may have been responsible for the lower scores in the incorrect ID group – knowing that they were incorrect might have caused them to display a hindsight bias; retrospectively deciding that they must not have been confident or must have had a poor view of the culprit (Bradfield & Wells, 2005; Wells & Bradfield, 1998, 1999). However, the incorrect identifiers also showed increased reaction times in both target and foil conditions, suggesting that correct identifiers were in fact, more easily able to evaluate and categorize each stimulus as target or non-target, perhaps indicating that they did get a better look at the culprit, and therefore were likely to be more confident in their identification. The relationship between confidence and accuracy is shown to be moderate in choosers and weak in non-choosers (people who reject a line-up, saying the culprit is not present) (Brewer & Wells, 2006; Sporer, et al., 1995). In the present study, rejecting the line-up was not given as an option. It is likely that some of the incorrect identifiers would have rejected the line-up if they had been told it was an option. This might have resulted in greater confidence in the incorrect ID group, as some people may have made a somewhat arbitrary decision, feeling like none of the line-up members looked like the culprit they had seen.

The increased confidence and better exposure reported by correct identifiers compared to incorrect identifiers is likely related to their shortened reaction times, which in turn are significantly correlated with P300 latency differences. Previous research has

established this relationship between reaction time and a P300 response, and suggested that it is related to stimulus evaluation time (Kutas et al., 1977). Pritchard (1981) explains that it cannot be an indicator of perceptual decision making, because the P300 occurs too late for that. This would suggest that the P300 in the present data is related to the level of confidence – the participants who incorrectly identify a foil (incorrect ID) are less confident, take more time to make an overt (behavioural) decision, and this is because they are taking longer on each trial to evaluate the stimulus. Those participants who have a greater degree of confidence appear to be able to make their decisions more quickly both behaviourally, and physiologically. Based on previous research, these data would suggest that incorrect identifiers require more time to evaluate each stimulus, resulting in a delayed P300 (Duncan-Johnson & Kopell, 1981; Verleger, 1997). It has also been suggested that the P300 is an integrative component, providing a connection between stimulus processing and response (Verleger, Jaśkowski, & Wascher, 2005), which is supported by the present data that demonstrate a relationship between confidence, P300 peak latency and physical button-press reaction time.

A final finding that merits discussion is that the N250 and P300 mean amplitudes did not differ as a function of accuracy. That is, electrophysiologically, correct IDs and incorrect ID trials manifest in the same way, especially over time, as these components stabilize. Even when participants in the present study were less confident in their choice, they still developed ERP components equal in amplitude to a more confident, correct responder. This might help to explain why witnesses, especially by the time of a trial, can still be very believable, even if they are wrong.

## General Discussion

In the present set of studies, ERP signals were recorded while participants viewed faces and identified the study researcher after a variety of different personal interactions, including a simple intentional identification task (Experiment 1), a deception task in which the experimenter was among the images but was not the target (Experiment 2), and in an eyewitness memory line-up task after witnessing a mock-crime (Experiment 3). Results from the three studies all demonstrate a strong N250 and P300 effect for target items, requiring a unique response. In Experiment 1 this is the experimenter, for Experiment 2 it is the scapegoat, and in Experiment 3 this is the person they identify as the culprit, regardless of accuracy. Experiment 2 also showed these effects for the experimenter-probe face, a familiar face included in the stimulus set which did not require a unique response. This finding suggests that the N250 and P300 are not target-dependent. However, because participants were asked to “lie” and *pretend* that they did not know the experimenter-probe this additional use of cognitive resources as well as the response inhibition that participants were asked to exercise (to say that the experimenter was innocent when they were previously informed that she actually was the “culprit”), might also be related to the attenuation of the N250 and P300 compared to other previously unfamiliar foil (non-target) faces. Finally, Experiment 3 showed an N250 effect for the face of the perpetrator of a live mock-crime, even when the participant did not correctly identify her as the culprit. That is, the N250 may be sensitive to differences in degree of familiarity, as the participant might have felt that the real culprit was still somewhat familiar, even though she was not correctly identified as the culprit. However,

this effect was not seen in the P300 amplitude, which was equivalent for rejections of true foils and rejections of the actual culprit. This evidence suggests that the N250 and P300 are not measures of the same type of familiarity. The N250 is sensitive to minor degrees of familiarity, while eliciting a P300 response seems to require a definitive decision that something has been recognized (even if the recognition decision is not linked to a response or an overt declaration of recognition).

Effect of hemisphere, in the case of the N250 was also examined in each of the three experiments. Experiment 1 results showed a bilateral N250 effect, with no main effect of hemisphere, while both Experiments 2 and 3 demonstrated left-hemisphere dominance. As discussed previously, the lateral dominance of the N250 effect has been shown primarily in the right hemisphere, though it is occasionally also seen in the left or in both hemispheres. This result needs additional probing, as there is no consistent research showing what kinds of experiments or stimuli will elicit a left or right dominant waveform. Experiments 2 and 3 were more cognitively involved and more difficult than Experiment 1 and perhaps different attribution of cognitive resources is partly responsible for the left-hemisphere effect, but this is just speculation.

The third variable examined in all three studies was experiment half. The results of the first two studies provided the expected result – the experiment half would interact with stimulus type, in that target items evoke a greater N250 and P300 response in the second half than in the first. Experiment 3 showed this effect for the P300, but failed to elicit an effect of half (or any interactions involving half) in either the correct or incorrect ID groups for the N250. As mentioned in the Experiment 3 discussion, this is a particularly unexpected result because of the ubiquitous data showing that the N250 is

associated with increased familiarity and experience and that the N250 response gets increasingly stronger with increasing presentations of a stimulus. Like the hemisphere effects, this result should be examined further, as it may be suggestive of an acquisition time difference between intentional and incidental memory (as experiments 1 and 2 rely on intentional memory, participants are told in advance who the target will be and are given extra time to look at the target, while Experiment 3 is a very brief incidental memory paradigm), or may be related to the surprise aspect of the mock-crime. Because both of the first two experiments still employ a live encounter during the exposure phase, the non-significant effect of half should not be related to the real-world aspect of the encounter.

Overall, results show that N250 and P300 components develop similarly in real-world situations as they do in lab studies using photographs. Of course, a concern is that the culprit (or experimenter) is the same in every session across all three experiments, and the photographs are identical images from trial to trial throughout the study. Future versions of the study should try to use multiple culprits, so that in some versions a confederate might be the culprit and in others s/he would be an innocent foil. This would allow some degree of counterbalancing between subjects, and would help to minimize any potential stimulus effects. Given that in reality police line-ups are not administered 40 consecutive times, using multiple images of the confederate-culprit may provide useful information about whether the N250 and P300 results are strictly photograph dependent, or if they would be sustained across multiple varying images of the same person. From an applied standpoint, the results do suggest that exposure to the wrong suspect even in a line-up can cause the development of an electrophysiological familiarity

response, to the correct culprit *or* to an innocent suspect. That is, the test has effects, and it is important to be aware of the fact that this physiological sense of recognition can develop quite quickly, as was shown in Experiment 3 in which the N250 for the target was already quite strong in the first half of trials, regardless of accuracy.

Additionally, in order to better generalize these findings to real-world eyewitness memory situations, it may be useful to include target absent line-ups and inform participants that rejecting the line-up is an acceptable response. This would inevitably lead to a higher degree of accuracy across participants. It is likely that some participants were making complete guesses during this task because no one in the line-up seemed familiar to them. In these cases it would be easier to interpret the N250 responses if they were allowed to reject the line-up. It would still be possible to look at incorrect rejections of the culprit in this case, but while maintaining a higher degree of generalizability to a real police line-up experience.

An important consideration in the present set of studies is foil selection and the similarity of foils to the confederate thief or experimenter. Foils in this project were selected based on availability and a very basic physical description (Caucasian females in their mid-twenties). Inevitably, some foils were more similar in appearance and basic physical characteristics (such as hairstyle) to the experimenter than others, and these foil options were chosen more frequently than other less similar foils. If participants were choosing incorrectly based on simple external characteristics like hairstyle, then it could be useful to have a measure of degree of similarity between the foils and the culprit. This would allow for some level of interpretation as to why certain foils are chosen more often. It would be interesting to examine the N250 and P300 responses to the most

similar foils. If the N250 and/or P300 responses develop in cases of incorrect rejections of the culprit as well as for the most similar foils, it could be interpreted that it is a sense of similarity that is causing the electrophysiological response, rather than an internal familiar representation. There might be a non-biased way to get a measure of foil similarity such as showing a target face and then showing two foils and asking which is more similar to the target and developing a familiarity face-space web, but it is quite likely that examining which foils are incorrectly chosen most frequently would also provide a reasonably accurate measure of similarity. Another option is to simply ask people outright to rank each foil on their level of similarity to the target and see what the consensus is. This would be less scientific than building a face-space, but would be much easier to implement. Finally, asking participants to rate their level of confidence in each individual identification or rejection decision would provide additional information about the sensitivity of the N250 and P300 components. It is possible that the data would show that the N250 or P300 would be related to lower confidence, because trials with lower confidence may be indicative of a more difficult recognition decision in which the photo “feels” familiar but the participant is not sure. The ERP components might be sensitive to this uncertain level of familiarity.

Finally, another important consideration is how the N250 and P300 would manifest in a control situation for Experiment 3. The distinct absence of a half effect for Experiment 3 but not for the first two would suggest that the consistency in the N250 from first to second half is related to either the surprise of the crime itself, or to the incidental memory paradigm. One way to investigate this result further would be to include a situation in which everything about the “crime” remains the same, with the

exception that the “thief” in the control situation merely asks about the ipod, but then does *not* steal it and just runs down the hall as if they suddenly realized they were late. This would maintain the incidental paradigm and the whole conversation would be the same in both theft and non-theft trials, but the increased distress or surprise associated with the crime would be controlled. Another idea for a control would be to carry out the crime exactly the same as was done here, but to include the confederate’s photograph in the line-up as a non-target and, test the N250 and P300 in response to the confederate photo, which might evoke electrophysiological indicators of familiarity in cases in which the confederate is recognized among the photos.

While further research is needed both to replicate the present results, and investigate the boundary conditions of the N250 and P300 amplitude effects, overall, the data support the idea of the N250 and P300 being representative of perceptual familiarity and cognitive decision-making respectively, in real-world situations. In addition, these data provide evidence that a mental representation of a person encountered in real-life may be maintained, even when that person is not correctly identified in successive viewings. This suggests that mistaken identifications are not a product of an encoding error, but rather they result from an error at the retrieval stage. Of course, while more study is needed, these results are an initial step in the direction of understanding correct and incorrect eyewitness identifications from an electrophysiological viewpoint.

## Bibliography

- Allen, J., & Iacono, W. (1997). A comparison of methods for the analysis of event-related potentials in deception detection. *Psychophysiology*, *34*, 234-240.
- Ben-Shakhar, G., & Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: Effect of mental countermeasures. *Journal of Applied Psychology*, *81*(3), 273-281.
- Bentin, S., Allison, T., Puce, A., & Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*(6), 551-565.
- Boehm, S., & Sommer, W. (2005). Neural correlates of intentional and incidental recognition of famous faces. *Cognitive Brain Research*, *23*(2-3), 153-163.
- Bradfield, A., & Wells, G. L. (2005). Not the same old hindsight bias: Outcome information distorts a broad range of retrospective judgments. *Memory & Cognition*, *33*, 120-130.
- Brewer, N., & Wells, G. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*(1), 11-30.
- Brigham, J., Maass, A., Snyder, L., & Spaulding, K. (1982). Accuracy of eyewitness identification in a field setting. *Journal of Personality and Social Psychology*, *42*(4), 673-681.
- Bruce, V., Burton, M., & Hancock, P. (2007). Remembering faces. *The handbook of eyewitness psychology, Vol II: Memory for people* (pp. 87-100). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P., Burton, A., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, *5*(4), 339-360.
- Bruce, V., Henderson, Z., Newman, C., & Burton, A. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, *7*(3), 207-218.
- Caharel, S., Jiang, F., Blanz, V., & Rossion, B. (2009). Recognizing an individual face: 3D shape contributes earlier than 2D surface reflectance information. *NeuroImage*, *47*(4), 1809-1818.

- Donchin, E. (1981). Surprise! Surprise? *Psychophysiology*, *18*, 493-513.
- Donchin, E., & Coles, M. (1988). Is the P300 component a manifestation of context updating?. *Behavioral and Brain Sciences*, *11*(3), 357-427.
- Duncan-Johnson, C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*(5), 456-467.
- Duncan-Johnson, C., & Kopell, B. (1981). The Stroop effect: Brain potentials localize the source of interference. *Science*, *214*, 938-940.
- Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. *Psychophysiology*, *3*, 298-308.
- Farwell, L., & Donchin, E. (1991). The truth will out: Interrogative polygraphy ('lie detection') with event-related brain potentials. *Psychophysiology*, *28*(5), 531-547.
- Gordon, I. & Tanaka, J.W., (submitted). Putting a name to a face: The role of name labels in the formation of face memories. *Journal of Cognitive Neuroscience*.
- Gratton, G., Coles, M.G., Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468-484.
- Herzmann, G., Schweinberger, S., Sommer, W., & Jentsch, I. (2004). What's special about personally familiar faces? A multimodal approach. *Psychophysiology*, *41*(5), 688-701.
- Herzmann, G., & Sommer, W. (2010). Effects of previous experience and associated knowledge on retrieval processes of faces: An ERP investigation of newly learned faces. *Brain Research*, *1356*, 54-72.
- Huff, G. (2004). Wrongful Convictions: The American Experience. *Canadian Journal of Criminology and Criminal Justice*, *46*(2), 107-120.
- Ihlebaek, C., Love, T., Eilertsen, D. E., & Magnussen, S. (2003). Memory for a staged criminal event witnessed live and on video. *Memory*, *11*(3), 319-327.
- Ishizu, T., Ayabe, T., & Kojima, S. (2008). Configurational factors in the perception of faces and non-facial objects: An ERP study. *International Journal of Neuroscience*, *118*, 955-966.
- Itier, R.J., Taylor, M.J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. *NeuroImage*, *15*, 353-372.

- Itier, R. J., & Taylor, M. J. (2004). Effects of repetition and configural changes on the development of face recognition processes. *Developmental Science*, 7, 469-487.
- Jacques, C., & Rossion, B. (2006). The Speed of Individual Face Categorization. *Psychological Science*, 17(6), 485-492.
- Jeffreys, D. A. (1996). Evoked potential studies of face and object processing. *Visual Cognition*, 3, 1-38.
- Johnson, M., & Rosenfeld, J. (1992). Oddball-evoked P300-based method of deception detection in the laboratory: II. Utilization of non-selective activation of relevant knowledge. *International Journal of Psychophysiology*, 12(3), 289-306.
- Johnston, R., & Edmonds, A. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, 17(5), 577-596.
- Johnston, V., Miller, D., & Burseson, M. (1986). Multiple P3s to emotional stimuli and their theoretical significance. *Psychophysiology*, 23(6), 684-694.
- Kantner, J., Gordon, I., Friesen, K.B., Breuer, A.T., Warren, C.M., & Tanaka, J.W. (2009). The electrophysiological correlates of distinctiveness processing in a Joe-No Joe task. Paper presented at the annual meeting of Northwest Cognition and Memory, Bellingham, WA.
- 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, 4302-4311.
- Kaufmann, J.M., Schweinberger, S.R., & Burton, A.M. (2009). N250 ERP correlates of the acquisition of face representations across different images. *Journal of Cognitive Neuroscience*, 21, 625-641.
- Keil, A., Bradley, M., Hauk, O., Rockstroh, B., Elbert, T., & Lang, P. (2002). Large-scale neural correlates of affective picture-processing. *Psychophysiology*, 39(5), 641-649.
- Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, 197, 792-795.
- Lander, K., & Chuang, L. (2005). Why are moving faces easier to recognize?. *Visual Cognition*, 12(3), 429-442.
- Lander, K., & Davies, R. (2007). Exploring the role of characteristic motion when learning new faces. *The Quarterly Journal of Experimental Psychology*, 60(4), 519-526.

- Langeslag, S. J. E., Morgan, H. M., Jackson, M. C., Linden, D. E. J., & Van Strien, J. W. (2009). Electrophysiological correlates of improved short-term memory for emotional faces. *Neuropsychologia*, *47*, 887-896.
- Lee, Y., Habak, C., & Wilson, H. (2010). Seeing an unfamiliar face in rotational motion does not aid identity discrimination across viewpoints. *Vision Research*, *50*(9), 854-859.
- Lefebvre, C. D., Marchand, Y., Smith, S. M., & Connolly, J. F. (2007). Determining eyewitness identification accuracy using event-related brain potentials (ERPs). *Psychophysiology*, *44*, 894-904.
- Leippe, M. R., Eisenstadt, D., Rauch, S. M., & Stambush, M. A. (2006). Effects of social-comparative memory feedback on eyewitnesses' identification confidence, suggestibility, and retrospective memory reports. *Basic and Applied Social Psychology*, *28*(3), 201-220.
- Lindsay, R., Wells, G., & Rumpel, C. (1981). Can people detect eyewitness-identification accuracy within and across situations?. *Journal of Applied Psychology*, *66*(1), 79-89.
- Liu, C. H., Seetzen, H., Buton, A. M., & Chaudhuri, A. (2003). Face recognition is robust with incongruent image resolution: Relationship to security video images. *Journal of Experimental Psychology: Applied*, *9*(1), 33-41.
- Luck, Steven J., (2005). *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.
- Luus, C., & Wells, G. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology*, *79*(5), 714-723.
- Lykken, D. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, *43*(6), 385-388.
- Lykken, D. (1960). The validity of the guilty knowledge technique: The effects of faking. *Journal of Applied Psychology*, *44*(4), 258-262.
- Johnson, M.M., Rosenfeld, J.P. (1992). Oddball-evoked P300-based method of deception detection in the laboratory: utilization of non-selective activation of relevant knowledge. *International Journal of Psychophysiology*, *12*, 289-306.
- McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, *211*, 77-80.
- McCloskey, M., Egeth, H., & McKenna, J. (1986). The experimental psychologist in court: The ethics of expert testimony. *Law and Human Behavior*, *10*, 1-13.

- Megreya, A., & Burton, A. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, *14*(4), 364-372.
- Memon, A., Mastroberardino, S., & Fraser, J. (2008). Münsterberg's legacy: What does eyewitness research tell us about the reliability of eyewitness testimony?. *Applied Cognitive Psychology*, *22*(6), 841-851.
- Mertens, R., Allen, J.J.B. (2008). The role of psychophysiology in forensic assessments: deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology*, *45*, 286-298.
- Miyakoshi, M., Kanayama, N., Nomura, M., Iidaka, T., & Ohira, H. (2008). ERP study of viewpoint-independence in familiar-face recognition. *International Journal of Psychophysiology*, *69*, 119-126.
- Miyakoshi, M., Nomura, M., & Ohira, H. (2007). An ERP study on self-relevant object recognition. *Brain and Cognition*, *63*, 182-189.
- Pfütze, E., Sommer, W., & Schweinberger, S. R. (2002). Age-related slowing in face and name recognition: Evidence from event-related brain potentials. *Psychology and Aging*, *17*, 140-160.
- Pritchard, W. (1981). Psychophysiology of P300. *Psychological Bulletin*, *89*(3), 506-540.
- Rosenfeld, J. (2005). 'Brain Fingerprinting': A Critical Analysis. *The Scientific Review of Mental Health Practice: Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work*, *4*(1), 20-37.
- 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*, 157-161.
- Rosenfeld, J., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology*, *41*(2), 205-219.
- Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*, 69-74.

- Sauerland, M., & Sporer, S. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied*, *15*(1), 46-62.
- Schweinberger, S. R., Huddy, V., & Burton, A. M. (2004). N250r: A face-selective brain response to stimulus repetitions. *NeuroReport*, *15*, 1501–1505.
- Schweinberger, S. R., Pfütz, E., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 722-736.
- 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*, 2057-2073.
- Schweinberger, S., Pickering, E., Jentsch, I., Burton, A., & Kaufmann, J. (2002). Event-related potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research*, *14*(3), 398-409.
- Smith, S., Lindsay, R., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed?. *Journal of Applied Psychology*, *85*(4), 542-550.
- Sporer, S., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*(3), 315-327.
- Tanaka, J. W., Curran, T., Porterfield, A. L., & Collins, D. (2006). Activation of preexisting and acquired face representations: The N250 event-related potential as an index of face familiarity. *Journal of Cognitive Neuroscience*, *18*, 1488-1497.
- Tanaka, J., & Pierce, L. (2009). The neural plasticity of other-race face recognition. *Cognitive, Affective & Behavioral Neuroscience*, *9*(1), 122-131.
- Turtle, J., Read, J., Lindsay, D., & Brimacombe, C. (2008). Toward a more informative psychological science of eyewitness evidence. *Applied Cognitive Psychology*, *22*(6), 769-778.
- Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology*, *34*(2), 131-156.
- Verleger, R., Jaśkowski, P., & Wascher, E. (2005). Evidence for an Integrative Role of P3b in Linking Reaction to Perception. *Journal of Psychophysiology*, *19*(3), 165-181.

- Verschuere, B., Rosenfeld, J., Winograd, M., Labkovsky, E., & Wiersema, R. (2009). The role of deception in P300 memory detection. *Legal and Criminological Psychology, 14*(2), 253-262.
- Weber, N., Brewer, N., Wells, G., Semmler, C., & Keast, A. (2004). Eyewitness Identification Accuracy and Response Latency: The Unruly 10-12-Second Rule. *Journal of Experimental Psychology: Applied, 10*(3), 139-147.
- Warren, C.M., Tanaka, J.W. & Holroyd, C.B. (2010). *Topological changes in the electrophysiological response to infrequent targets implicate the locus coeruleus' norepinephrine system.* Paper presented at the meeting of the Canadian Society for Brain Behaviour and Cognitive Science, Halifax, NS, Canada.
- Wells, G. L., & Bradfield, A. L. (1998). 'Good, you identified the suspect': Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83*, 360-376.
- Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the postidentification-feedback effect be moderated?. *Psychological Science, 10*, 138-144.
- Wells, G., & Olson, E. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295.