Document Title:	Information Exposure, Presentation Modality, and Cognitive Mechanisms of Countermeasures in P300 Concealed Information Tests
Author(s):	Michael Ross Winograd
Document No.:	244532
Date Received:	January 2014
Award Number:	2012-IJ-CX-0022

NORTHWESTERN UNIVERSITY

Information Exposure, Presentation Modality, and Cognitive Mechanisms of Countermeasures in P300 Concealed Information Tests

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Psychology / Brain, Behavior, and Cognition

By

Michael Ross Winograd

EVANSTON, ILLINOIS

August 2013

ABSTRACT

Exposure to Information, Detection of Central and Peripheral Items in Multiple Modalities, and Cognitive Mechanisms of Countermeasures in P300 Concealed Information Tests

Michael Ross Winograd

The Concealed Information Test (CIT) is a credibility assessment tool designed for use in criminal investigations. In contrast to other methods that attempt to detect signs of deception, the CIT detects when a suspect recognizes crime-relevant information. The P300 event-related potential CIT is one version of this test. While it has been the subject of study for over 25 years, much is yet to be discovered about how certain variables affect its accuracy and effectiveness. Here, three studies tested various aspects of the P300-CIT. The first study examined the impact of exposure to crime-relevant information on innocent and guilty participants. In the second study, a P300-CIT was used to detect details both central and peripheral in a mock crime using both pictorial and visual presentation modalities. In the final study, two experiments were conducted to attempt to elucidate the cognitive mechanism responsible for evoking P300s during the use of countermeasures, which are covert responses used to attempt to defeat the test. Results from the first study showed that exposing innocent participants to crime-relevant details made them indistinguishable from guilty participants, leading to a high false-positive rate. Additionally, a trend was observed that prior exposure also biased the P300-CIT toward enhanced sensitivity in guilty participants. In the second study, details central to a crime were better recalled than peripheral details. Evidence was found suggesting that probe – irrelevant P300 differences

were larger for a central item, even when rates of explicit memory were equal. Other data suggested that there might be a benefit to using pictorial stimuli over verbal stimuli for the central item. In the final study, different countermeasure methods were used in a P300-CIT in order to manipulate the stimulus evaluation process, complexity, and meaningfulness of countermeasures, to determine their effects on evoked P300 amplitudes and latencies. No differences were found between any countermeasure methods, suggesting that a simple recognition process is responsible for evoking P300s to the countered stimuli. The results build on our current knowledge of the P300-CIT and provide new data that can help toward the development of a field-ready test and methods for its use.

ACKNOWLEDGEMENTS

I am grateful for my family for their unconditional support over the years and for their encouragement when things didn't always go as planned. I would like to thank the other members of my lab, John Meixner, Xiaoqing Hu, and Elena Labkovsky for the help and feed feedback they have given when discussing my research. Thank you to the members of my committee, Dr. Steven Franconeri, Dr. Ken Paller, and Dr. Karl Rosengren for their guidance and feedback on my proposal. To the University: Northwestern has been a wonderful home since I first set foot on campus ten years ago. Finally, I owe an enormous debt of gratitude to my advisor, Dr. J. Peter Rosenfeld for everything since I joined his lab at the end of my junior year at Northwestern in 2006. There is not enough room here to concisely state all that he has done to help me. As he knows, written communication is not my strong suit, so it will have to suffice it to say that I am sincerely grateful for his support, guidance, and everything he has taught me as both an undergraduate and graduate student over the past seven years.

TABLE OF CONTENTS

<u>LIST OF TABLES AND FIGURES</u>

CHAPTER 1 – INTRODUCTION

The Innocence Project, a public policy and litigation organization that works to exonerate wrongfully convicted people, reports that there have been 310 post-conviction exonerations in United States history based solely on DNA evidence. Further, they state that the single greatest cause of wrongful convictions is mistaken eyewitness identification (by the victim, a secondary witness, or both), which was a factor in over 72% of these cases ("Facts on post-conviction DNA exonerations.," 2012). Often, these mistaken identifications were compounded by faulty forensics, biased police methods (e.g. improperly designed photo lineups), or false confessions given after excessive interrogation. Both the phenomena of mistaken eyewitness identifications and false confessions have been replicated in controlled experimental settings (Kassin & Kiechel, 1996; Sporer, Penrod, Read, & Cutler, 1995). In situations where a suspect is identified through eyewitness identification, there is a clear tendency to accept the word of a victim or eyewitness over that of a potential suspect. While it is logical for police and prosecutors to pursue bringing a person to trial based purely upon this kind of evidence, it is a mistake that can and has led to wrongful convictions. One potential tool that can be used to help prevent these errors and aid in criminal investigations is the concealed information test (CIT; Lykken, 1959).

The CIT is a credibility assessment tool that allows an examiner to determine if a person has knowledge of specific details of a situation or crime. Using the CIT methodology, a suspect is asked a question about an aspect of the crime (e.g. *'What was the murder weapon?')* and is then typically presented with a series of, at least five potential answers (e.g. *pistol, knife, shotgun*). When a guilty suspect recognizes the correct (*probe*) answer, it has been suggested (Gamer, 2011) that it elicits a differential orienting response (OR; Sokolov, 1990) compared to the incorrect (*irrelevant*) items, which, in its autonomic nervous system (ANS) version, is detected through changes in physiological signals such as skin conductance, pulse volume, and respiration rate or amplitude suppression (Ben-Shakhar & Elaad, 2002).

The observation of this enhanced OR to the probe items indicates that a suspect recognized the correct answer to the question, and thus has specific knowledge of the crime that he should only know if he was involved. While such a finding is not sufficient to infer guilt, the CIT could be useful for law enforcement agents to help determine whether or not a suspect otherwise implicated solely through eyewitness identification or circumstantial evidence may have actually been involved in the crime. In contrast, an innocent person with no specific knowledge of the crime would show no such effect. Unlike in the United States, the CIT is routinely used in Japan to help guide criminal investigations (Nakayama, 2002; Osugi, 2011). Additionally, in contrast to the well-known and commonly used comparison question test (CQT) polygraph (the preferred method of testing for law enforcement in the United States), which suffers from poor specificity and has been widely criticized in academic circles (Ben-Shakhar, 2002; Vrij, 2008), the CIT has significantly lower theoretical and observed false-positive rates (Vrij, 2008). The probability of an innocent person reacting the strongest to the probe item in a single question is $1/n$, with n being the total number of possible answers. This potential false-positive rate decreases with the addition of each possible item within a block and additionally with each added block. The exact probability is dependent on a decision criterion (e.g. a suspect is only diagnosed as “guilty” if they react the strongest to the probe item on three of five blocks).

Using binomial probabilities, if an investigator tested for knowledge of five details of a crime with five items in each block, the probability that an innocent person would be diagnosed as guilty based on this requirement is just .051. The addition of a sixth item in each block reduces this probability to .032. This is one reason why the CIT has strong support within the academic community (National Research Council, 2003; Iacono, 2011; Iacono & Lykken, 1997). While the original CIT was designed for use with an ANS polygraph (see Ben-Shakhar & Elaad, 2003, for a review), it can also be applied for use with the P300 event-related potential (ERP; Farwell & Donchin, 1991; Rosenfeld et al., 1988).

P300

The P300 complex consists of two main subcomponents, the P3a and the P3b. The P3a component peaks earlier (240 – 300 ms) and is maximal more frontally than the P3b (Polich, 2007; Rushby, Barry, & Doherty, 2005). It is known to be sensitive to stimulus novelty (Combs & Polich, 2006; Courchesne, Hillyard, & Galambos, 1975), is related to the orienting response (Polich, 2007), and is elicited even when tasks do not require explicit stimulus processing (Squires, Squires, & Hillyard, 1975). In P300-CITs, and for the purposes of these studies, the component of interest is the P3b (referred to throughout here as “P300”). It is an endogenous ERP that is elicited through an oddball recognition response and is thought to be indicative of context updating in working memory when a person perceives a stimulus (Donchin & Coles, 1988; Polich, 2007). It generally peaks between 350-600ms, and is maximal over the parietal midline. A number of variables are known influence P300 amplitude, such as attention (Wickens, Kramer, Vanasse, & Donchin, 1983), objective global and local stimulus probabilities (Duncan-Johnson & Donchin, 1982; Squires, Wickens, Squires, & Donchin, 1976), subjective probability (Johnson Jr., 1986;

Rosenfeld, Biroschak, Kleschen, & Smith, 2005), and meaningfulness (Johnson Jr., 1986). Lower true or perceived stimulus probabilities, increased meaningfulness, and increased attention all increase the size of evoked P300 responses. The influences of stimulus probability and meaningfulness are known to be additive in nature, suggesting the P300 is not indicative of a singular process, but rather is a combination of multiple cognitive sub-processes, and that this combination is influenced by attention and information processing demands both during recall and encoding (Curran, 2004; Johnson Jr., 1993; Polich, 2007). In addition to variables that influence P300 amplitude, the component's latency increases with greater stimulus complexity and difficulty of subsequent stimulus evaluation and categorization processes, but it is not affected by increased difficulty of response selection (Magliero, Bashore, Coles, & Donchin, 1984; McCarthy & Donchin, 1981) such as in a Stroop task (Rosenfeld & Skogsberg, 2006)

P300 CIT

Given the P300's sensitivity to both stimulus probability and meaningfulness, it lends itself well for use in a CIT. When a suspect recognizes a relevant answer among the incorrect irrelevants, this probe stimulus evokes a larger P300 response compared to the irrelevant items, which indicates that the suspect has knowledge of the crime. In Winograd and Rosenfeld (2011), we applied our novel P300-based CIT, the Complex Trial Protocol (CTP), to a mock crime scenario. Overall, 82% of guilty and 92% of innocent participants were correctly classified based on their P300 responses, suggesting that the CTP has both good sensitivity and specificity at detecting incidentally acquired information, and compares favorably to other similar tests and methods (see Table 1). This finding is significant given that the majority of P300-CIT studies to report high detection rates were conducted using self-referring autobiographical information as stimuli (e.g. Rosenfeld et al., 2008). This information is much more deeply encoded and recognizable than episodic information (Rosenfeld, Biroschak, & Furedy, 2006; Rosenfeld, Shue, & Singer, 2007). Further, in contrast to the standard three-stimulus method of P300-CITs (where each trial is either a probe, target, or one of multiple irrelevants), the CTP has been shown to be highly resistant to countermeasures (CMs), which are covert responses a person can execute in an attempt to defeat the test (Rosenfeld, Soskins, Bosh, & Ryan, 2004).

Table 1: P300 Mock Crime Studies

		Correct Detection Rates			
Authors	**Block**	**Guilty**	**Innocent**	**A'**	**AUC**
Abootalebi, Moradi, & Khalilzadeh (2006)		0.79	0.79	0.87 [b]	
Abootalebi, Moradi, & Khalilzadeh (2009)		-	-		0.88 [b]
Farwell & Donchin (1991)	Study 1	0.9	0.85		0.99 [c]
Hu, Pornpattannanangkul, & Rosenfeld (2013)	High Aware	-	-		0.79 [a]
	Low Aware	-	-		0.55 [a]
Hu & Rosenfeld (2012)	Immediate	0.67	1		0.89 [a]
	1-month Delay	0.75	1		0.95 [a]
Lui & Rosenfeld (2008)	2 probe	0.87	0.71	0.87 [d]	
	3 probe	0.71	0.64	0.76 [e]	
Mertens & Allen (2008)		0.47	1	0.87 [a]	
Rosenfeld et al. (2004)		0.73	0.91	0.89 [a]	
Rosenfeld, Shue, & Singer (2007)	1 probe	0.55 [f]			
Winograd & Rosenfeld (2011)		0.82	0.92	0.93 [a]	

Note: Only studies which reported either area under the curve (AUC) or detection rates are presented. Authors used varying statistical methods of classification (a = bootstrap amplitude difference, b = wavelet classifier, c = bootstrapped cross-correlation, d = bootstrapped spatial-temporal PCA on fronto-central site, e = bootstrapped spatial-temporal PCA on parietal-occipital site, f = bootstrap amplitude difference with multiple blocks (2 of 3 needed for guilty diagnosis). AUC given for papers not reporting separate group hit rates, and for studies including a receiver operating characteristic (ROC) analysis or classification data for each participant. Reported correct detection rates and calculated A’ (Grier, 1971) values are given for studies that reported only correct classification rates. See original studies for “block” and other details. Hu et al. studies used other conjoint measures besides P300, but only P300 data reported here.

The end goal in this line of research is to develop a test that can be used as a tool to aid in criminal investigations, specifically one that can discriminate between guilty and innocent people at an early stage in an investigation. Given a guilty suspect's motivation for self-preservation that is manifested by engaging in activities to avoid being convicted of a crime, an ecologically valid and accurate CIT that is resistant to CMs would be a valuable tool for law enforcement to help focus investigations on the correct suspects. For example, when an eyewitness identifies a person as a suspect, they can be given a CIT to test for knowledge of details relevant to the crime. In the case of a burglary, possible probes could include the location of entry, a tool used to break a window, and the identity of various stolen items. If the suspect is found to have larger P300 responses to the probe items on the CIT, it can be reliably inferred that he has knowledge of the crime, and thus the investigation of him as a suspect should continue. The alternative finding that the suspect did not react to the probe items would then be an indication that he was likely not involved, and the investigation should then focus on other suspects. Such a test could operate as a measure to help prevent the false confessions and convictions that have ended with hundreds of (known) innocent people in prison. However, before a P300-CIT can be used in the field, it must first be thoroughly vetted in the laboratory, and threats to its validity need to be examined. This is especially necessary if the CIT is to ever meet the criteria of the Daubert Standard ("*Daubert v. Merrell Dow Pharmaceuticals, Inc*," 1993), the United States Supreme Court ruling that established criteria for the inclusion of scientific evidence and expert witness testimony in courts, or the Frye standard ("*Frye v. United States*," 1923), a previous ruling still followed in some states in trials below the federal level. The studies presented here build upon the previous research conducted on P300-CITs over the past 25 years and examine three issues, prior knowledge and exposure to information, the detection of central and peripheral crime details, and the cognitive mechanisms of countermeasures, that have potentially direct impacts on the effectiveness of the P300-CIT in the field

The first of the three studies (Chapter 3 - The Impact of Prior Knowledge from Participant Instructions in a Mock Crime P300 Concealed Information Test) examines the issue of information leakage (when details of a crime are revealed to the public) on the accuracy of a P300-CIT. It is primarily focused on the impact of prior knowledge on innocent participants, and whether simple knowledge of a crime-relevant detail (in absence of actual guilt) is sufficient enough to evoke large enough P300s to the probe item, leading to false-positives. Secondarily, issues of external validity of laboratory studies are discussed in regards to potential artificial enhancement of detection rates of guilty participants when they are exposed to crime-relevant details prior to the commission of mock crimes. The second study (Chapter 4 - Effects of Presentation Modality on the Detection of Central and Peripheral Details in a P300 CIT) was completed as an attempted replication, extension, and combination of two prior studies that manipulated methodological aspects of the P300-CIT and their impacts on P300 amplitudes, and thus, its ability to detect concealed knowledge. The first studied question is whether or not the P300-CIT is effective at detecting details that are either centrally (directly involved) or peripherally (also present at the scene) relevant to the commission of a mock crime. This was investigated while simultaneously manipulating the presentation modality of the stimuli in the P300-CIT to determine if there was an advantage of using actual pictures of items involved in a mock crime over their verbal representations. The crossing of these two variables also made it possible to test for differential effects of presentation modality on central and peripheral details. Finally, in the third study (Chapter 5 – Cognitive Mechanism of Countermeasures in P300 Concealed Information Tests), the theoretical question what cognitive mechanism is responsible for evoking of P300s during CM use in P300-CITs was examined for the first time. Taken together, the research and data presented here provide insights into important methodological and theoretical issues in P300-CITs. The findings from these studies can help aid in the development of an ecologically valid and field ready test for use in criminal investigations.

CHAPTER 2 – GENERAL EEG METHODS

Data Acquisition

Electroencephalogram (EEG) recording in each study was taken using Ag/AgCl electrodes at the Fz, Cz, and Pz sites connected to a mesh cap (Electro-Cap International, Inc.) and referenced to linked mastoids. Electrooculogram (EOG) was recorded referentially using a reference electrode placed on the forehead above the inner corner of the right eye. A ground electrode connected the forehead to the chassis of the isolated side of the amplifier system. Signals were passed through 30-Hz low-pass and 0.3 Hz (3 db) high-pass filters (Contact Precision Instruments EEG 8 system) and digitized with a 16-bit A/D converter sampling at 500 Hz.

Data Processing

Stimuli were presented and reaction times measured by a PC running Psytask (Mitsar Co., LTD) in Windows XP. EEG data were recorded and processed using WinEEG (Mitsar Co., LTD) on a PC running Windows XP. Any trials with an observed value greater than 75uV on any EOG or EEG channel were automatically rejected. Due to individual differences, further artifacting was conducted on a participant-by-participant basis, setting a rejection criterion based on visual inspection of EOG artifacts. Any participants with fewer than 20 accepted single sweeps for any individual stimulus were eliminated from all analyses (Polich, 1991). After artifacting, the EEG data were further processed off-line in WinEEG with an additional 6.0 Hz high-cut filter. The data were than exported to allow for the measurement and bootstrapping of P300s using a custom Matlab-based program. All statistical analyses were completed using IBM SPSS Statistics, v. 20 (IBM).

The Complex Trial Protocol P300-CIT

The CIT protocol used in each study was the Complex Trial Protocol (CTP; Rosenfeld et al., 2008), with a few modifications in each (e.g. the number and type of stimuli or the responses participants made). To date, there have been over a dozen peer-reviewed studies of the CTP (for a review, see: Rosenfeld, 2011; Rosenfeld, Hu, Labkovsky, Meixner, & Winograd, *in review*). For the current studies, each trial consists of two parts. The first, lasting 1800ms, begins with a 100ms baseline period followed by either a probe or irrelevant stimulus which is presented for 300ms (after which it is replaced with a blank black screen). After a randomly varying delay of 0 – 450 ms, the second half of the trial then consists of either a target (*11111*) or non-target (*22222*, *33333*, *44444*, or *55555*) stimulus that is also presented for 300ms. The second half of each trial lasted 2700ms, making the total trial length vary from 4500 – 4950 ms. Targets and non-targets followed every first stimulus with an equal 0.5 probability. In an effort to focus attention on the probe and irrelevant stimuli, because the CTP requires no differential responses, as in the original three-stimulus P300-based CIT (Farwell & Donchin, 1991; Rosenfeld et al., 1988), we randomly paused the protocol every 30-40 trials (after the first stimulus of the trial) and asked the participant to identify the stimulus that had just been presented. No participants in any of the studies were rejected for failing to properly identify the stimuli on more than one trial.

Individual Bootstrap Analysis

To classify each participant as either guilty or innocent in each study, we compared probe and irrelevant P300 amplitudes using a bootstrap procedure (Wasserman & Bockenholt, 1989) on the Pz site, where P300 is known to be maximal (Fabiani, Gratton, Karis, & Donchin, 1987). The bootstrap procedure consists of randomly selecting (with replacement) a sample of individual probe sweeps equal to the total number of non-rejected probe trials (*N*), and then making an average probe ERP from the selected sweeps. The same number (*N*) of individual sweeps are then sampled, with replacement, from all irrelevant (Iall) sweeps and averaged into a bootstrapped Iall ERP. The amplitudes of these two ERPs are then compared to see if the probe is arithmetically larger than the irrelevant. This process is then repeated 1000 times. A participant is diagnosed as knowledgeable if the proportion of significant bootstrap iterations (where probe > irrelevant) is above a specific cutoff, which traditionally has been set at 0.9 (Rosenfeld et al., 2008; Winograd & Rosenfeld, 2011), however this cutoff can vary depending on changes in experimental methods.

CHAPTER 3 - THE IMPACT OF PRIOR KNOWLEDGE FROM PARTICIPANT INSTRUCTIONS IN A MOCK CRIME P300 CONCEALED INFORMATION TEST

Introduction

Many concealed information test (CIT; Lykken, 1959) studies that employ mock crime scenarios inform participants in the experimental instructions of exact details for which they will later be tested for knowledge. Since some crimes (though not all) in the field expose perpetrators to some crime details only during commission of the crime, studies such as the majority of those P300-based CIT studies in Table 1 may poorly represent some actual field conditions (i.e. they have low ecological validity). To date, only Winograd and Rosenfeld (2011) and Hu, Pornpattananangkul, and Rosenfeld (2013) focused on the detection of purely incidentally acquired knowledge. The simple act of revealing crime details to participants may be sufficient to induce them to recognize probe details in a CIT, evoking large responses to probe items, thus leading to participants being classified as “guilty.”

Indeed, many studies have shown this to be true for autonomic CITs. Gamer, Gödert, Keth, Rill, and Vossel (2008), Gamer (2010), and Gamer, Kosiol, and Vossel (2010) demonstrated that both the standard ANS-CIT and the ANS-guilty actions test (GAT) were unable to discriminate between truly guilty and informed innocent participants. A similar result was reported by Nahari and Ben-Shakhar (2011), though, it should be noted that Gamer et al. (2010) found more forgetting and greater physiological response decreases in innocent informed participants tested after a time delay. There is, however, some contradictory evidence. In a study by Bradley and Warfield (1984), three different groups of informed-innocent participants were created by having the participants either witness a crime, be told the crime details, or carry out a different activity using the same crime details. They found that guilty detection scores were higher for true guilty participants, and only the innocent participants who executed another activity with the crime-relevant details differed from the sample of truly naïve innocent participants. Most of these results, however, demonstrate the potential danger of leakage of crime information to the public in situations where a CIT could be employed, as innocent people could appear guilty on a CIT, or truly guilty suspects could claim that information learned through legitimate means could be responsible for their knowledge of probe items. Based on these results, we sought to determine if informed innocents would also be indistinguishable from true guilty participants using a P300-based CIT.

One might simply assume that if informing innocent participants of mock crime details makes them appear guilty on an ANS-based CIT, then the same effect would occur in a P300-based CIT. However, several effects seen in ANS-CITs do not occur in P300 versions. For example, time delays between a mock crime and CIT have been found to decrease response magnitudes for items in ANS-based CITs (Gamer et al., 2010; Nahari & Ben-Shakhar, 2011), an effect that was not found in a recent P300-CIT (Hu & Rosenfeld, 2012). Gamer and Berti (2012) found differential effects for detection of central and peripheral items using the SCR, but no differences between the two based on P300 amplitudes. Similarly, Gamer and Berti (2010) found differential effects of task relevance and recognition on SCR and P300. Further, in contrast to P300 CIT studies that utilized multiple blocks of testing (Meixner & Rosenfeld, 2011; Rosenfeld et al., 2007; Rosenfeld et al., 2004), previous research has demonstrated that the response magnitudes of ANS orienting responses are significantly affected by habituation (Ben-Shakhar, Frost, Gati, & Kresh, 1996), an effect not seen with P300-CITs (Rosenfeld et al., 2008).

While the CIT has significant support in the scientific community (Iacono & Lykken, 1997), the ecological validity of laboratory tests (i.e. how well results would translate to field use) has been largely unexamined. The majority of P300 CIT studies to date have used autobiographical details or previously studied and rehearsed information as probes. One limitation in using this information to assess the accuracy of the CIT is that these familiar items are well rehearsed and thus much more deeply encoded into memory than details one might notice and encode during the commission of a crime (some of which may be purely unplanned and incidentally acquired). This is problematic for ecological validity, as Rosenfeld et al. (2007) found larger probe-irrelevant P300 amplitude differences for self-referring information (e.g. names, birthdates, area codes, social security numbers) than mock crime information. A related effect was reported by Rosenfeld et al. (2006), who found that a participant's own name elicited a larger P300 than that of the experimenter's name, even one to which participants were exposed numerous times and rehearsed to a 100% recall criterion.

Within mock crime studies, one element that can have a significant impact on validity is the instruction set given to the participant. In many of the studies in Table 1, participants were tested for knowledge of details that were explicitly revealed through instructions and/or reinforced through an interrogation (Hu & Rosenfeld, 2012; Mertens & Allen, 2008; Rosenfeld et al., 2007), reinforced by having participants write a detail of the item (Abootalebi, Moradi, & Khalilzadeh, 2006, 2009), or even learned through rote memorization (Farwell & Donchin, 1991; Rosenfeld et al., 2004). The latter two of these procedures have limited ecological validity. In a field CIT examination, an examiner would not be likely to disclose the identity of a probe item (at most he would present it, along with all the other stimuli, to a suspect in order to review the items they will see on a test) just prior to an examination, nor ask a suspect to write down or reveal an aspect of a stolen item (which they are denying having taken). Based on the previous findings that well-rehearsed information evokes larger P300s than less salient information (Rosenfeld et al., 2006; Rosenfeld et al., 2007), we predicted that any procedure that reveals or reinforces the identity of the probe item prior to the P300 CIT would bias the results towards higher sensitivity by making the probe more salient than it would be without the disclosure. However, it should be noted that some mock crime situations (e.g. Farwell & Donchin, 1991) may require participants to learn specific details through instructions in order to properly execute the mock crime.

To test whether prior knowledge of probe items would affect the P300-CIT, we employed a fully counterbalanced 2 X 2 factorial design. We manipulated participant's guilt (guilty vs. innocent) and knowledge (informed vs. naïve) of the probe detail given in the mock crime instructions. While the primary focus of the current study was to determine if revealing crime-relevant knowledge to innocent participants would cause them to appear guilty in the subsequent CIT, the full 2 X 2 design was chosen in order to also determine if information given through experimental instructions might have an additional effect on guilty participants. We predicted that correct detection rates would be higher for the guilty-informed group than for the guilty-naïve group, and lowest for the innocent-naïve (true innocent). The critical group, however, would be the participants who were knowledgeable as to the identity of the stolen item, but who did not actually take it (innocent-informed). We predicted that more participants in this group would be falsely diagnosed as guilty (false positive) than in the innocent-naive group, and that they would be indistinguishable from the guilty-naïve group based on both P300 amplitudes and detection rates. Additionally, we expected to find larger probe P300 amplitudes and larger probe – irrelevant P300 differences in both the guilty-informed versus the guilty-naïve group and in the innocent-informed versus innocent-naïve groups due to the informed groups' additional exposure to the probe item prior to executing the mock crime. Importantly, we sought to demonstrate this effect using a much less salient form of exposure to the crime details than the rehearsal or memorization procedures previously used in related experiments (such as those in Table 1).

Method

Participants

Participants were undergraduate or graduate students at Northwestern University who took part in the experiment for either course credit or monetary compensation ($10). There were 63 total participants. Six were rejected due to excessive EEG artifacts, two due to improper execution of the mock crime, and one for not following directions for the button press responses during the CIT examination. The remaining 54 participants (32 female) ranged in age from 18 – 24 (M = 19.5, SD = 1.7). Participants were randomly assigned to one of four conditions, innocent-naïve (n = 14), innocent-informed (n = 13), guilty-naïve (*n* = 14) or guilty-informed (*n* = 13). The local institutional review board (IRB) approved all methods.

Procedure

Mock Crime. After giving informed consent, all participants were given instructions for how to commit the mock crime. The mock crime and instructions were identical to those from Winograd and Rosenfeld (2011), with the notable exception of the manipulation of participant knowledge of the to-be-stolen item between the naïve and informed conditions. Participants were told to go to the main office of the psychology department and tell the secretaries they were putting a test (a completed multiple choice answer sheet) in Dr. Rosenfeld's mailbox, and that there was an envelope in the mailbox labeled "For Dr. Rosenfeld" with either an "item" (naïve condition) or a "ring" (informed condition) inside, which they were to steal by removing it from the envelope, and bring back to the laboratory. The instructions for the naïve and informed conditions were identical except for these two words. In the entire set of instructions, the words "item" or "ring" (depending on condition) appeared just twice and were presented in identical typeface as the rest of the instructions. The experimenters did nothing else to inform participants about the identity of the item. Participants who were assigned to a guilty condition then executed the mock crime. Participants who were assigned to an innocent condition were told, upon finishing reading the instructions that they would not actually be committing the crime, but were to simply walk down to the same office and return. This was in an effort to equalize the time between exposure to the instructions and the CIT examination with that of the guilty groups. Upon returning to the laboratory, all participants were informed that they were then going to be given a brain-based lie detector test. They were told that they would be tested for knowledge of an item that was stolen out of a mailbox in the department office, and that they should do their best to try to appear innocent on the test.

P300-CIT. The CIT protocol used was the CTP (Rosenfeld et al., 2008), with a few modifications. The word *ring* was used as the probe for participant. Additionally, we used six similar irrelevant items (*watch, necklace, wallet, locket, bracelet,* and *cufflink*), each presented with equal probability as the probe. Participants immediately pressed one of five buttons on a keypad (one button for each finger) with their left hand in response to the first stimulus of every trial, as an indicator that they had seen and read the stimulus. They were instructed to randomize which button they pressed on each trial and to not follow any sort of pattern. On the second part of the trial, participants pressed a “yes” button with their right hand to all targets and a “no” button to all non-targets. A total of 360 trials were collected for each participant. Upon completion of the CIT, all guilty and innocent-informed participants who were classified as “guilty” were asked if they knew what the stolen item was. All of these participants correctly recalled or recognized the ring as the critical item.

P300 Analysis

P300 amplitude was calculated using the peak-to-peak (p-p) method, which has been shown to be more sensitive for the detection of concealed information than standard baseline-to-peak measurement (Meijer, Smulders, Merckelbach, & Wolf, 2007; Soskins, Rosenfeld, & Niendam, 2001). The P300 peak was defined as the mean amplitude of the maximum positive 100ms segment between 350 and 800ms post-stimulus, with P300 latency defined as the midpoint of this 100ms segment. The p-p measure of P300 is calculated by taking the difference in amplitude between this positive P300 peak and the mean of the most negative 100ms segment between P300 latency and 1700ms post-stimulus (this time being the earliest start of the target/non-target portion of the trial based on the randomly varied 0 – 450 ms delay between the first and second halves of the trial). No ERP analyses were conducted on the target/non-target portion of the trial.

Bootstrapped Detection Rates

Since previous studies have shown that highly rehearsed self-referring information is more readily detected than incidentally acquired or mock crime information (Rosenfeld et al., 2006; Rosenfeld et al., 2007), the standard .9 cutoff that has been used previously to diagnose participants as guilty or innocent (Rosenfeld et al., 2008) may not be ideal for the current study which is focused on detecting information acquired through the commission of a mock crime. To determine the best cutoff for the current study, the number of significant iterations between the guilty-naïve and innocent-naïve groups was compared. At each possible value of the number of significant iterations (ranging from 0-1000), we calculated the combined proportion of correct diagnoses. The smallest number of significant iterations which was found to result in the best overall detection rate was then used as the cutoff to determine hit rates in each condition of the study. While determining an ideal cutoff based on this data set may result in an overestimate of the CTP's true accuracy in the current study, the same cutoff can be used on similar future studies to allow for a better comparison of hit rates.

Receiver Operating Characteristic

In order to give an unbiased measure of a test's ability to discriminate between guilty and innocent participants, many studies have alternatively or additionally used receiver operating characteristic (ROC) analyses (e.g. Ben-Shakhar & Elaad, 2003; Hu et al., 2013; Meixner & Rosenfeld, 2011) to determine the classification efficiency of CITs. The ROC method is taken from signal detection theory (Green & Swets, 1966), and is used to estimate the degree of separation on a dependent measure between two groups. In the case of P300-CITs, it serves as an indicator of a study's sensitivity and specificity at all cutoffs. The analysis gives the area under the ROC curve (AUC), which ranges from 0.5, representing chance, and 1.0 representing perfect discrimination (no overlap) between two groups. Using the number of significant bootstrapped iterations as the dependent measure, AUC values were calculated for the guilty-informed, guilty-naïve, and innocent-informed groups, each compared to the innocent-naïve group, to give an estimation of the CTP's ability to properly discriminate between guilty and innocent participants. Finally, we compared AUC values between the guilty-informed, guilty-naïve, and innocent-informed groups (for methods, see: Hanley & McNeil, 1983; McNeil & Hanley, 1984)

N-1 Chi-Squared Analyses

In the current study, we employed a comparative trial design (commonly used in medical research) by controlling the proportions of participants in guilty-naïve, guilty-informed, innocent-naïve, and innocent-informed groups, and then experimentally classified them as "guilty" or "innocent" based upon their bootstrap test. This results in a series of 2x2 contingency tables that can be used to compare detection rates between conditions. It is traditionally advised to employ chi-squared tests only when the minimum expected count in any given cell in a 2x2 contingency table is at least five. However, Campbell (2007) showed that alternative tests, such as the Fischer-Irwin (Fisher, 1922) are too conservative in some designs where the experimenter controls one of the two variables and is interested only in differences of participant outcomes on a second variable (such as the comparison trial method. Through a series of simulations, Campbell (2007) demonstrated that within these designs, when the minimum expected count in any given cell is greater than one, the optimal test to use is the "*N* – 1 chi-squared test" (where the term *N* – 1 is used in place of *N* when calculating expected cell counts). Based on this, we employed a series of N – 1 chi-squared tests to determine if detection rates differed between the groups in the current study.

Results

Grand average ERPs are presented in *Figure 1*. Differences in p-p P300 amplitude between probe and irrelevant stimuli are evident in each condition except for the naïve-innocents, the only group to have no knowledge of the identity of the probe item

Figure 1: Chapter 2 Grand Average Probe and Iall ERPs

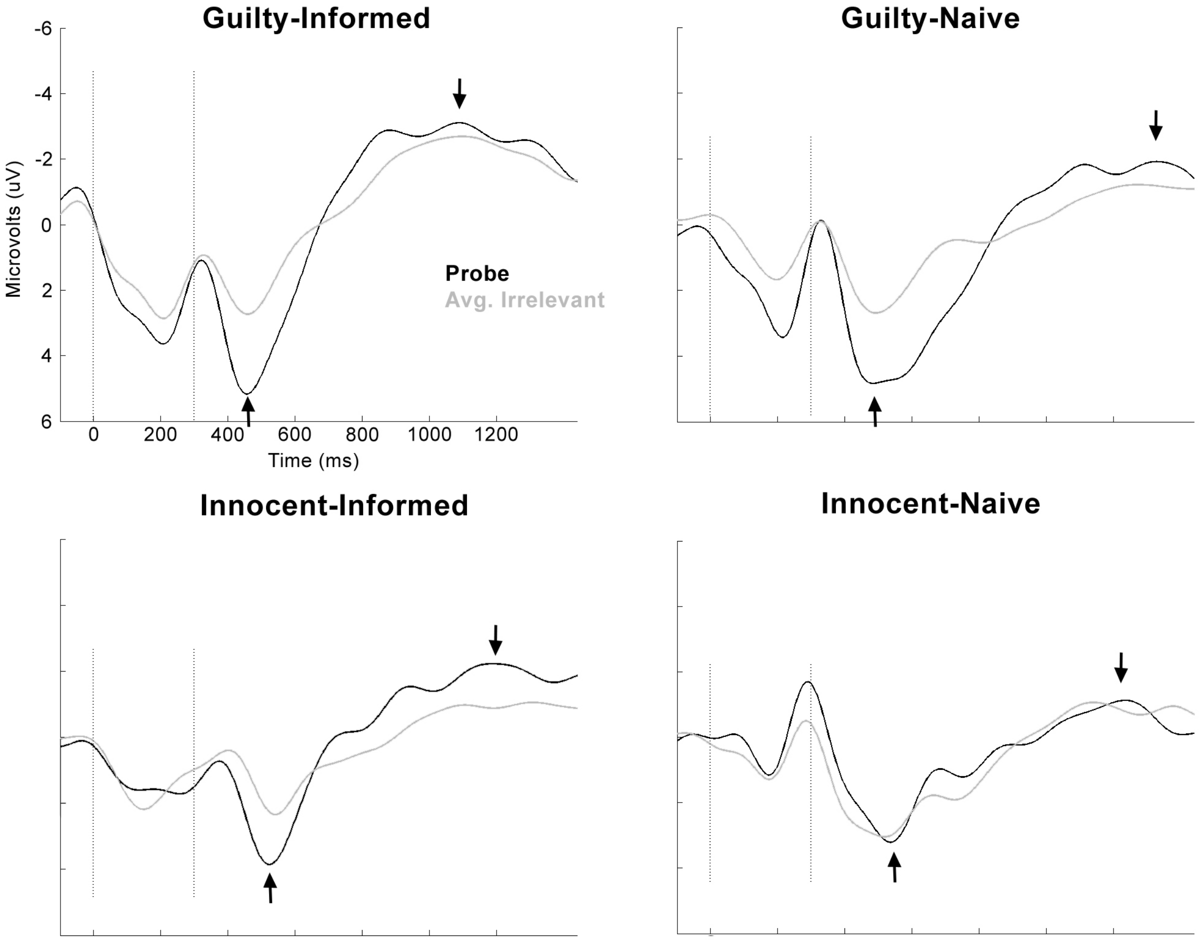

Figure 1: Grand average *probe* and *Iall* (average irrelevant) ERPs at Pz site. Simple knowledge of the identity of the *probe* item was sufficient for evoking large P300s. *Note:* positive is plotted down on the y-axis. Arrows represent the peaks of P300 and the subsequent maximum negativity used in the peak-to-peak P300 measurement.

P300 Amplitudes

Figure 2 shows computed p-p P300 amplitudes for probe and irrelevant stimuli for all four conditions. Here too, only the innocent-naïve group did not appear to have larger P300s to probe than irrelevant stimuli. We ran a 2 (Stimulus: probe, Iall) X 2 (Knowledge: informed, naive) X 2 (Guilt: guilty, innocent) mixed-model ANOVA on peak-peak P300 amplitude. Partial η^2 values are reported for significant results. There was a main effect of stimulus, $F(1, 50) = 62.71$, $p < .001$, $\eta^2 = .56$, with the mean probe amplitudes larger than Iall. The main effects of both knowledge ($p = .63$) and guilt ($p = .83$) were not significant. There were two significant interactions for guilt X stimulus, $F(1,50) = 6.64$, $p = .013$, $\eta^2 = .12$, and knowledge X stimulus, $F(1,60) = 8.98$, $p = .004$, $\eta^2 = 1.5$. Participant's guilt led to greater probe – irrelevant differences than in innocent participants. Knowledgeable participants also had larger probe – irrelevant differences than naïve ones.

Regarding the interaction of knowledge X stimulus, as can be seen in both *Figure 1* and Figure 2, in the three groups where participants had knowledge of the identity of the stolen item (either through instructions, committing the crime, or both), probe amplitudes were clearly larger than Iall. However, in the innocent-naïve group, there was no such difference between the two stimulus types. This result suggests that the three other conditions may be indistinguishable from one another in terms of mean probe and Iall amplitudes. Since this study was primarily focused on the effect of crime-relevant knowledge in innocent participants, we conducted a follow-up 2 (stimulus: probe vs. Iall) X 3 (Group: innocent-informed, guilty-informed, guilty-naïve) mixed-model ANOVA to examine whether the innocent-informed group would be distinguishable from the two guilty groups. As before, there was a an expected main effect of stimulus, $F(1,37) = 73.88$, $p < .001$, $\eta^2 = .67$. In contrast, the results for the main effect of group ($p = .93$) and a stimulus X group interaction ($p = .59$) were both not significant, suggesting that the reason why there was no significant main effect of guilt in the previous analysis was because the innocent-informed group was indistinguishable from the two guilty groups. It also suggests that the lack of a probe – irrelevant difference in the innocent-naïve group was probably responsible for the significant interaction in the prior analysis.

Figure 2: Chapter 2 P300 Amplitudes

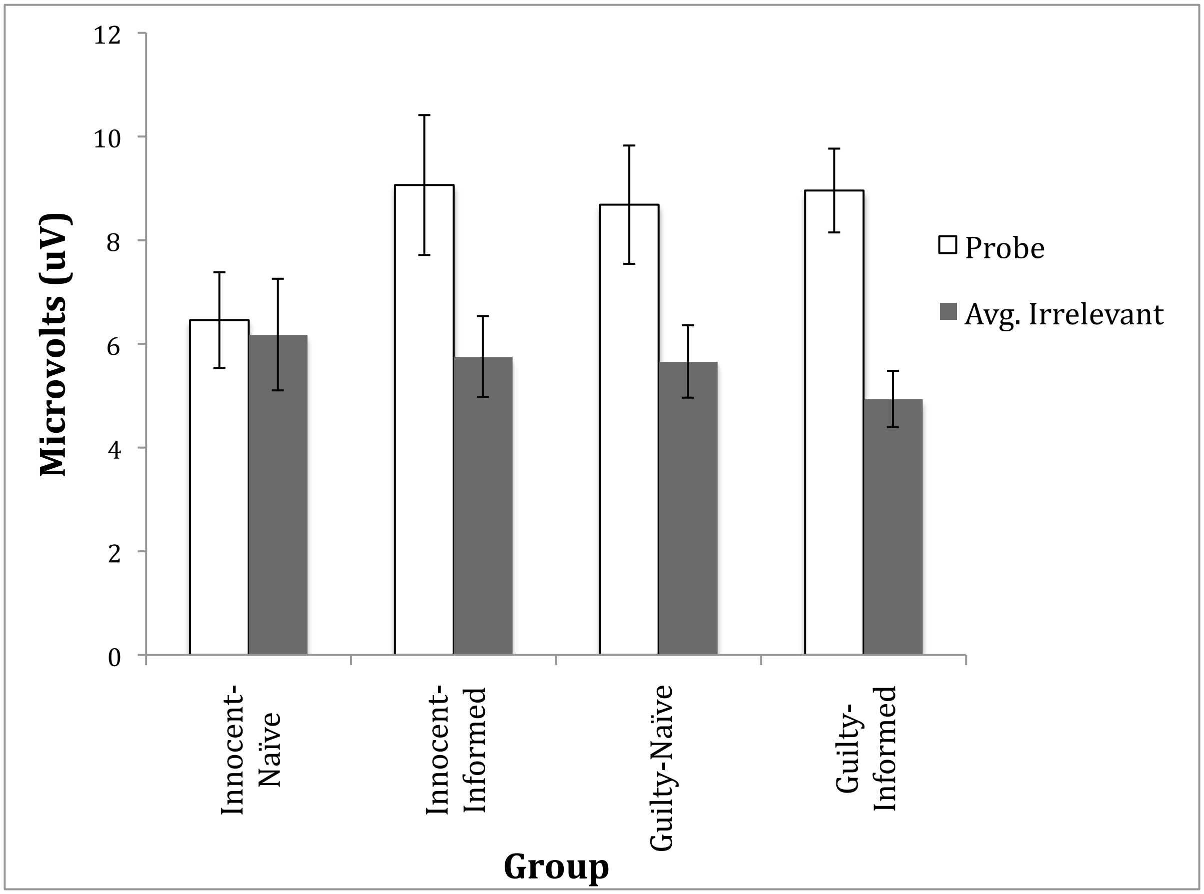

Figure 2: Probe and Iall P300 amplitudes for each group. Significant differences were found between probe and irrelevant amplitudes for all conditions but the innocent-naïve participants.

Individual Bootstrapped Detection Rates

The best discrimination between the two guilty groups and innocent-naïve participants was found using a cutoff of 800 of 1000 bootstrapped iterations where probe > Iall as a criterion for discriminating guilty and innocent participants. Thus, we report our detection rates here using an 80% bootstrap confidence interval (see Table 2). Overall, the observed detection rates are consistent with our hypothesis that exposure to guilty knowledge items would result in a bias toward classifying participants as guilty. The detection rate in the guilty-informed group was 13/13 (100%) compared to 11/14 (79%) in the guilty-naïve group. In the innocent conditions, 12/14 (86%) of the naïve participants (innocent-naïve group) were correctly classified as innocent, yielding a false-positive rate of 14%. In the critical group of informed-innocents (innocent-informed), only 4/13 (31%) of participants were correctly classified as innocent.

Table 2: Chapter 2 Correct Bootstrapped Detection Rates and AUCs

Group	*N*	Correct	Prop	AUC
Innocent-naïve	14	12	0.86	X
Innocent-informed	13	4	0.31	.797
Guilty-naïve	14	11	0.79	.852
Guilty-informed	13	13	1.00	.956

Note: Detection rates based on a bootstrap criterion of .8. AUC was calculated for each condition versus the innocent-naïve group.

We ran a series of *N* – 1 chi-squared tests to identify significant differences in detection rates between these groups (results are reported with the adjusted *N)*. As predicted, as with the P300 results, the detection rates between the guilty-naïve and innocent-informed groups did were not different, $\chi^2(1, 26) = 0.33$, $p = .56$. Further, innocent-informed participants were more likely to be diagnosed as "guilty" than their naïve counterparts, $\chi^2(1,26) = 8.15$, $p = .004$. In contrast to the guilty-naïve participants, those in the guilty-informed group were more often classified as guilty than participants in the innocent-informed group, $\chi^2(1,25) = 4.59$, $p = .03$. Compared to the innocent groups, the effect of prior exposure to probe details was less pronounced in guilty participants. We found a marginally significant trend that guilty-informed participants were more often correctly classified as guilty than those in the guilty-naïve group, $\chi^2(1,26) = 3.06$, $p = .08$. Collapsing across guilty and innocent conditions, there was a strong overall effect of knowledge. Informed participants were more likely to be classified as guilty than those who were naïve, $\chi^2(1,51) = 6.96$, $p = .008$.

Receiver Operating Characteristic

To examine the overall ability of the CTP to discriminate between guilty and innocent individuals, we first performed a series of ROC analyses on the number of bootstrap iterations where probe > Iall using both the innocent-naive and innocent-informed groups for measures of specificity (see Table 2). This was done in an effort to show how exposure to crime-relevant details negatively impacts the P300-CIT's ability to discriminate between truly guilty and simply knowledgeable participants. The best discrimination accuracy was found for guilty-informed versus innocent-naïve groups (AUC

= .956). This value was larger than for both the guilty-naïve group (.852), and the innocent-informed groups (.797) versus the innocent-naïve group, however, these differences in AUC between the guilty-informed and the two other groups were not found to be significantly different (both *p*'s > .10). There was significant overlap in the number of significant bootstrap iterations between the guilty-naïve and innocent-informed groups. The AUC comparing these two groups was just .519, only slightly higher than chance, so these individuals are essentially indistinguishable from one another based on the analysis of differences between probe and irrelevant P300 amplitudes.

Reaction Times

Mean probe and irrelevant RTs are presented in Figure 3. Previous studies have used reaction time measures as an indicator or guilty knowledge (Seymour & Kerlin, 2008; Verschuere, Crombez, Degrootte, & Rosseel, 2010), and in a an earlier mock crime P300-CIT study (Winograd & Rosenfeld, 2011), we found significant group differences between guilty and innocent participants using RTs to probe and Iall stimuli (but did not use this data for the purpose of individual diagnostic tests). To examine potential RT effects in the current study, we ran a 2 (Stimulus: probe, Iall) X 2 (Knowledge: informed, naive) X 2 (Guilt: guilty, innocent) repeated-measures ANOVA. None of the main effects or interactions was significant (all *p*'s > .19).

Figure 3: Chapter 2 Reaction Times

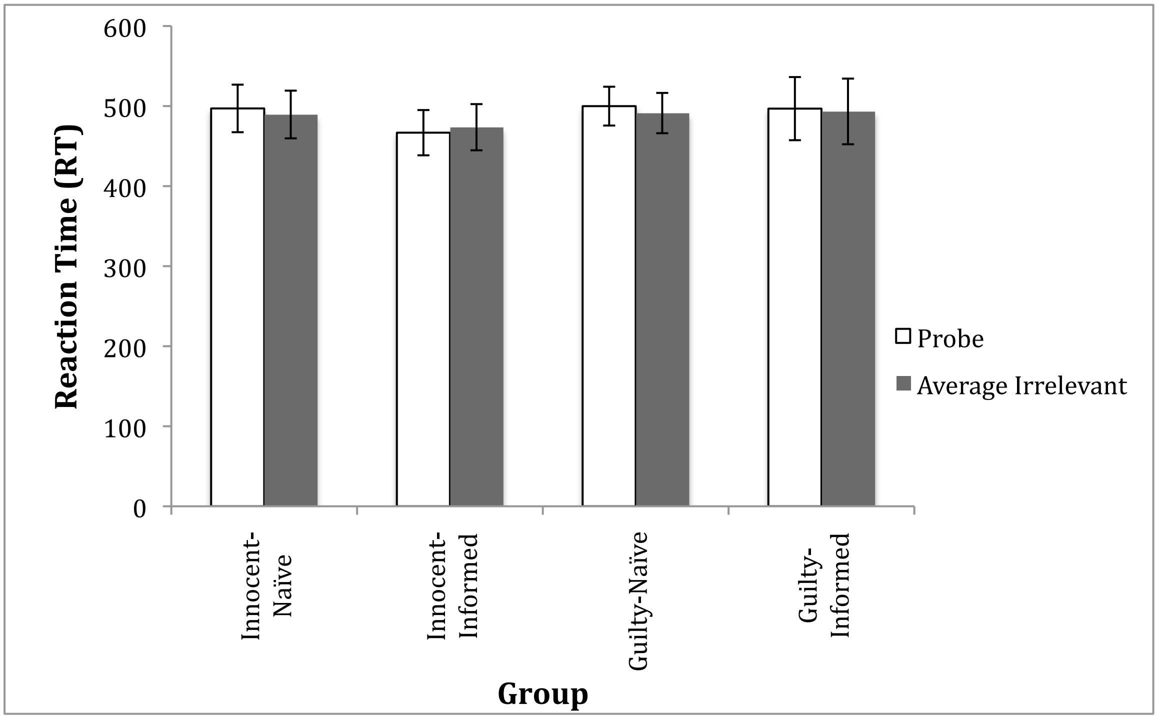

Figure 3. Mean probe and Iall reaction times (RTs) for each group. No differences were found between groups or stimulus types.

Discussion

The present results demonstrate a potential flaw in the experimental designs of many previous P300 CIT studies that utilized mock crimes (see Table 1). Simple knowledge of a correct detail (gained through instructions, rather than through commission of a mock crime) was found to be sufficient to elicit large P300s to probe stimuli in innocent-informed participants, making them indistinguishable (based on P300 amplitudes) from those who actually committed the crime. It appears to also reinforce the experiential knowledge of a crime in guilty participants, leading to a trend that guilty-informed participants were better detected than guilty-naïve ones.

As expected, when the identity of the probe was revealed to participants through mock crime instructions, this disclosure had an impact on their P300 amplitudes and bootstrapped detection rates. Despite being innocent of committing the mock crime, participants in the innocent-informed condition recognized the probe, which evoked larger P300s than those to irrelevant items. Further supporting our hypothesis, the innocent-informed participants were indistinguishable from those in the two true guilty conditions based on P300 amplitudes. Additionally, the AUC values showed the discrimination accuracy between informed-innocent and guilty-naïve participants were not different, again supporting our prediction. These results are consistent with those found previously with ANS-CITs and GATs in immediate testing conditions (Gamer, 2010; Gamer et al., 2008; Gamer et al., 2010; Nahari & Ben-Shakhar, 2011).

By using a full 2 x 2 factorial design in the current study, we were able to also extend the prior knowledge effect to guilty participants (Nahari & Ben-Shakhar, 2011, had a guilty-informed group but did not test a guilty-naïve condition.). Along with a strong effect in innocent participants, the results of our chi-square tests revealed a trend toward enhanced sensitivity even in guilty participants when crime-relevant information was revealed to them prior to the commission of the mock crime. While some aspects of a crime may be rehearsed (e.g. targeting a specific house for a robbery), other details may be only incidentally acquired through the commission of the crime (e.g. a ring found and stolen while burglarizing a house). As demonstrated here, when a detail like this is not rehearsed or learned prior to the commission of the crime, the ability to detect a participant or suspect's knowledge of it is reduced compared to when the details are rehearsed. Importantly, in this case, the probe we chose to test for was one that was central to the crime. In a real-life investigation, some probe items may be peripheral details that a suspect may not directly interact with during the commission of the crime (e.g. a potentially noticeable painting or poster on a wall). Gamer et al. (2010) defined peripheral details as those that a participant might or might not have perceived and encoded during a mock crime, in contrast to central details, which must have been perceived in order to complete the task. They found that for guilty participants, the rate of forgetting peripheral details was higher than that for central details (while the rates decreased similarly for informed innocents). Since peripheral details are not encoded as deeply or remembered as well as central details, it is possible that the enhancement created by revealing details may be more pronounced for peripheral than for central details, an issue examined in the next chapter.

The results from this study have two main implications. First, leakage of crime information may be a serious threat to the use of CITs in the field. If certain aspects of a crime become public knowledge, they may no longer be valid details to use to test for a suspect's involvement, as truly innocent suspects may have this specific knowledge of the crime through legitimate means. This finding is critical to control for, because if the CIT is to ever become a commonplace part of criminal investigations in the United States, it is necessary that law enforcement agencies do not leak critical information about a crime to the general public, as a suspect could argue that their knowledge of crime details was due to news reports or some other information leak (controlling for this can be done, as in Japan; Osugi, 2011). The second implication from the results in this study is that mock crime studies that reveal details to participants through instructions or memorization procedures may be artificially inflating detection rates for these details in a CIT, since guilty informed participants were more likely to be diagnosed as guilty than guilty naïve participants. This point is more cautionary than critically important. As discussed above, there are many aspects of a crime that may be planned in advance or rehearsed that could be used as a probe on a CIT. However, if one is conducting a mock crime CIT study, special attention should be paid to avoid artificially reinforcing details (such as peripheral ones), that would more likely be simply incidentally acquired through the actual act of committing a crime.

It should be noted that while we obtained detection rates in this current study that are similar to those of Winograd and Rosenfeld (2011), the bootstrap criterion that we needed to use (so as to maximize the function of overall accuracy = A (correct detections + correct rejections) / N) was different (80% in the current study and 90% in the previous). We attribute this need to a few changes in our protocol since the previous study was published (such changes may produce ROCs of differing shapes in which the A-function is maximum at different values) First, in Winograd and Rosenfeld (2011), both the "I Saw It" and Target/Non-Target responses were made with the same hand on a single two-button box. In the current study, these were changed to a five button random "I Saw It" response on the left hand and a target/non-target response on the right hand. This may have distracted participants from the implicit probe/irrelevant recognition task. Additionally, we have changed from using a CTP protocol with an asymmetric target probability (targets followed probes more often than they followed irrelevants) to one where targets and non-targets follow probes and irrelevants with equal probability. In the asymmetric CTP protocol, guilty participants may notice that the probe stimulus is more likely to be followed by a target than the irrelevant stimuli, enhancing the salience of the probe. When compared directly (Rosenfeld, Tang, Meixner, Winograd, & Labkovsky, 2009), the symmetric and asymmetric protocols did not result in differential detection rates or P300 amplitudes. However, Rosenfeld et al. (2009) utilized self-referring information rather than mock crime details in the CIT. Since self-referring information is detected with higher accuracy than episodic information (Rosenfeld et al., 2006), it is possible that the effect of the asymmetric protocol was overshadowed in Rosenfeld et al. (2009) by the large P300s evoked to self-referring probes. In the case of episodic information from mock crimes, the asymmetric protocol as used in Winograd and Rosenfeld (2011) may have enhanced the salience of the probe, increasing the discriminability of probe and irrelevant stimuli, allowing for the use of a more stringent bootstrap criterion. We have recently moved to consistently using a symmetric protocol, because in field use, a suspect or defense attorney could argue that with an asymmetric protocol, the only reason a person reacted to a probe was because he noticed it was followed more often by targets. This use of a lower bootstrap criterion is, however, not necessarily a serious problem, as the previous use of a 90% bootstrap criterion is an arbitrary choice from early P300 CIT studies (Farwell & Donchin, 1991). A criterion for future real-life testing would need to be determined from an analysis of valid field tests, and could simply be whatever cutoff yields an institutionally desired level of sensitivity and specificity (see: Rosenfeld, 2011).

Overall, the results of the current study were consistent with recent findings with the ANS-CIT and GAT (Gamer, 2010; Gamer et al., 2008; Gamer et al., 2010; Nahari & Ben-Shakhar, 2011). Prior knowledge of crime details was sufficient for inducing high rates of false positives in a P300-CIT. Additionally, we found that revealing the identity of the probes to guilty participants prior to a CIT also led to a trend toward enhanced sensitivity. Indeed, there was a highly significant (p = .004) effect of prior knowledge on combined guilty and innocent participants. While most prior P300-CITs for mock crimes used methods that explicitly revealed probe items to participants, the recent results do not necessarily de-value any previous findings, as many crimes may indeed require planning that would enhance a suspect's memory for certain details. Instead, the results suggest that researchers should carefully consider the amount of detail given in instructions to participants in accordance with what kind of situation they are choosing to model in a mock crime in order to keep ecological validity high.

CHAPTER 4 - EFFECTS OF PRESENTATION MODALITY ON THE DETECTION OF CENTRAL AND PERIPHERAL DETAILS IN A P300 CIT

Introduction

In criminal interrogations, guilty suspects are often uncooperative or lie to investigators in an effort to hide their involvement in a crime. The ability to accurately detect a signal that indicates when a suspect is being deceptive would be a major boon for law enforcement. While the autonomic nervous system (ANS)-based comparison question test (CQT) and its derivatives (Raskin & Honts, 2002) are preferred by the law enforcement community, the CQT has been much criticized by many in the academic community due in part to its questionable theoretical foundation (see Ben-Shakhar, 2002, for a review). As an alternative to the CQT, Lykken (1959) created the ANS concealed information test (CIT), which, rather than detecting signs of deception, instead detects when a suspect recognizes crime-relevant details during questioning. The CIT methodology can also be applied for use with the P300 event-related potential (ERP), and has been since its earliest applications more than two decades ago (Farwell & Donchin, 1991; Rosenfeld et al., 1988; Rosenfeld, Nasman, Whalen, Cantwell, & Mazzeri, 1987).

Since the ANS-CIT has been the subject of study for around 30 more years than the P300 version, much more is known about how issues such as the modality of stimulus presentation, subject responding, or number of and repetition of details affect its accuracy (see the meta-analysis by Ben-Shakhar & Elaad, 2003, for a review). A number of factors are known to influence the P300-CIT's efficacy. Highly rehearsed self-referring information, such as a participant’s name is more readily detected than incidentally acquired (e.g. an experimenter’s name) information (Rosenfeld et al., 2006; Rosenfeld et al., 2007). Further, detection rates and efficiency (as measured by ROC analyses) for participants in studies that test for highly rehearsed self-referring information (Hu, Hegeman, Landry, & Rosenfeld, 2012; Labkovsky & Rosenfeld, 2012a; Rosenfeld et al., 2008) are generally better than in studies testing for mock crime information (Abootalebi et al., 2006, 2009; Hu et al., 2013; Hu & Rosenfeld, 2012; Mertens & Allen, 2008; Rosenfeld et al., 2004; Winograd & Rosenfeld, 2011). Less clear is the influence of whether the tested details are central (required to be perceived in order to successfully commit a crime) or peripheral items (those present at a scene but not required to be perceived to successfully commit a crime). While the results from ANS-CITs are fairly consistent in finding that central details are better detected than peripheral ones, even when no differences were observed in an explicit memory test (Carmel, Dayan, Naveh, Raveh, & Ben-Shakhar, 2003; Gamer et al., 2010; Nahari & Ben-Shakhar, 2011; Peth, Vossel, & Gamer, 2012), the one ERP-CIT to examine this issue found, surprisingly, that P300 amplitudes did not differ between item types (Gamer & Berti, 2012), while ANS measures did. This result was unexpected given the strong effects found in the ANS-CIT. Gamer and Berti (2012) attributed this finding to the P300 being sensitive to “mere recognition” (Meijer, Smulders, & Wolf, 2009) but not depth of processing.

Their counterintuitive finding could suggest one of three possible hypotheses: 1) There truly is no effect of central versus peripheral details and depth of processing on P300 amplitudes in a CIT, 2) There is an effect of depth of processing on P300, but it is less pronounced than with ANS measures, or 3) The items Gamer and Berti (2012) designated as peripheral in their mock crime may have been more significant or meaningful than they had planned, and thus were more akin to central ones. The two stimuli they designated as peripheral were the *office* where the mock crime took place and a *keyring pendant* affixed to a key used to unlock a desk drawer. Given their own definition of a central detail as being ones that "were directly related to the execution of the crime or actively handled during the course of the mock crime," one could reasonably argue that both of the stimuli they designated as peripheral to be actually more central in nature. The participants *must* have encoded details about the office in which they committed the crime in order to navigate around it (unless they had their eyes closed) and *did* actively handle the *keyring pendant* since it was attached to a key that they used to open a drawer. In fact, good evidence for this third possibility comes from the fact that after a one-week delay, 18 of 20 participants correctly recalled both of these peripheral items. The remaining two participants each recalled one item, leading to an overall 95% correct recall rate of these peripheral items. In contrast, Nahari and Ben-Shakhar (2011) found better memory for central over peripheral details (such as the identity a soda can or the name of a newspaper located on a desk where the mock crime took place).

Along with item type, another aspect of a CIT that could affect its accuracy is the modality in which stimuli are presented. In word-list learning paradigms, congruence between the stimulus modalities in study and test phases facilitates more accurate recognition; items studied as pictures are better and more efficiently (as indexed by shorter RTs) recognized when presented pictorially versus verbally (Stenberg, Radeborg, & Hedman, 1995). Applied to a CIT context, this could suggest that pictorial stimuli may enhance recognition for items that were encoded as episodic memories visually during the commission of a mock crime. Results in ANS-CITs have been mixed, with some showing better classification accuracy with verbal than pictorial stimuli (Ben-Shakhar & Gati, 1987; Gati & Ben-Shakhar, 1990) while Ben-Shakhar et al. (1996) found no differences. To our knowledge, Ambach, Bursch, Stark, and Vaitl (2010) conducted the only study directly comparing the presentation of pictorial and verbal stimuli in a P300 and found no differences between the two presentation modalities (i.e. no differential effect on probe and irrelevant stimuli). However, they defined the positive P300 peak as the average amplitude of the largest 100ms segment between 400-1000ms. A quick look at their grand average ERPs shows a clear P300 peak occurring prior to 400ms in both the pictorial and verbal stimuli. Thus, their defined look window did not properly capture the P300 peak, and visual inspection of the ERPs suggests that the probe – irrelevant P300 difference may have indeed been larger in the pictorial condition. Since the CIT is an applied test, when comparing either the item type or presentation modality, this difference in amplitudes between probe and irrelevant stimuli is the critical measure when comparing either the item type or presentation modality. While a finding of a main effect of modality, which would indicate larger P300s overall to either verbal or pictorial stimuli, may be theoretically interesting in regards to the cognitive processes involved, an examination of this is outside scope of the current research.

In the current study, we sought to replicate and expand the previous P300-CIT studies examining the effects of item type (central vs. peripheral) and stimulus modality (verbal vs. pictorial). In two blocks of testing, participants were tested for knowledge of a central detail (a stolen item) and a peripheral detail (another non-stolen item located at the scene). In a partial-Latin square design, participants also were presented with verbal stimuli in one block and pictorial stimuli in the other. Our primary prediction was that, in line with the ANS-CIT evidence, there would be an interaction in which the probe – irrelevant P300 amplitude difference would be larger for central than peripheral items. As a result, the probe-irrelevant difference comparison would also favor central items, leading to more probe > Iall bootstrap iterations. Given that the P300 in a CIT has been shown to be sensitive to mere recognition (Meijer et al., 2009), and that memory research has found a facilitative effect of modality congruence between encoding and recognition (Stenberg et al., 1995), we predicted that pictorial stimuli would also enhance P300s to probes (leading to a larger probe – irrelevant difference) and the number of probe > Iall bootstrap iterations. Additionally, we reasoned that this effect would likely be more pronounced for the peripheral detail, since we expected to be near a ceiling effect for P300 amplitudes and probe > Iall bootstrap iterations using the central detail.

Method

Participants

A total of 27 participants were tested, three of whom were rejected from analysis (two due to excessive eye artifacts leaving fewer than 20 non-artifacted trials of each type and one for not executing any of the assigned button press responses). The final sample consisted of 24 participants (10 male) who ranged in age from 18-22 years old (*M* = 19.21, *SD* = 1.22). All participants were undergraduate students at Northwestern University who completed the experiment for course credit. All participants gave informed consent and filled out an autobiographical information sheet prior to beginning the experiment. The local IRB approved all methods.

Materials

Prior to beginning the study, we gathered 14 items to be used in the mock crime. We then randomly assigned seven to be the central detail and the other seven to be the peripheral detail (Figure 4). For each participant, one item was chosen from each set to be used in the mock crime by randomly selecting from those which had not been used yet by other participants (so, after the first participant, we chose from six items, after two participants five items, etc.). Once each item had been used, this process was repeated starting with the seven original items. In the end, each item was used at least three times, with three items being repeated a fourth time. In each block of testing, the non-chosen items were used as irrelevants during the CIT. A 15" LCD monitor (refresh rate = 60Hz) operating at a resolution of 1280 X 1024 pixels was located approximately 2 ft in front of the participants. In the verbal blocks, stimuli consisted of words presented in all letters in Arial font in white on a black background. The words were 0.5 inches tall. Pictorial stimuli were 800 x 800 pixels with each item scaled to approximately the same size on a white background.

Figure 4: Chapter 3 Pictures of Stimuli

Figure 4: Items used for the central and peripheral details and their names. When an item was randomly chosen to be a probe, the other six items served as irrelevants in the same block.

Procedure

Testing conditions. Since we chose to examine two dichotomous variables (item type and presentation modality), there were four possible combinations of blocks: central-pictorial (CP), central-verbal (CV), peripheral-pictorial (PP), and peripheral-verbal (PV). Employing a partial Latin square design, participants completed two blocks of testing each. In a process similar to randomly choosing the items, participants were randomly assigned to which testing condition they completed on the first block based on which conditions prior participants had been in. So, if the first participant had been assigned to the CP condition for their first block, the second participant was randomly assigned to one of the three other blocks. For the second block, participants completed a block consisting of the two conditions that were not used in the first. For example, if a participant was assigned to CP in the first block, the second block was then PV. Six participants were assigned to each testing condition in the initial block, resulting in 12 total blocks for each condition.

Mock crime. Participants were seated in the testing room in a large chair in front of a table. On the table were two large identical envelopes. Each participant read instructions that explained that they were to choose one of the two envelopes, memorize the instructions in it, and when ready, complete their assigned task. They were told that each envelope contained a different set of instructions and that they were to withhold revealing to the experimenter which set they chose until the end of the experiment. In reality, both sets of instructions were the same.

The mock crime instructions told participants to steal an item out of a small white container in a specific mailbox in the graduate student mailbox area. The identities of the stolen item and the peripheral detail were not revealed to the participants through the instructions, as in Winograd and Rosenfeld (2011), to control for exposure effects (as seen in Chapter 3). The mailbox was one cubby in a large array, measuring 12” wide X 8” high, and was located approximately four feet above the ground. The white container (5” X 5” X 3”) was always placed on the left side of the mailbox, approximately 1” from the opening, while the peripheral item was placed on the right side of the mailbox. Upon stealing the item from the container, participants hid it somewhere on their person and returned to the lab for testing.

P300-CIT. We used the CTP (Rosenfeld et al., 2008) version of the P300-CIT. Participants made two button-press responses to the stimuli that were presented. For each probe or irrelevant in the first half of the trial, they pressed a button (the “I saw it” response) with the index finger of the left hand. For the second part, with their right hand, participants clicked the right mouse button for each target and the left mouse button for each non-target stimulus. For this study, we chose to discontinue the use of the five-button response (as in Chapter 3) after multiple participants stated that maintaining random responses throughout the test was difficult and distracting. The probe and each of the irrelevants were presented 50 times, for a total of 350 trials with the whole block taking approximately 28 minutes. Upon completion of their first block of testing, participants were given a short two-minute break, and then the procedure was repeated for the next block.

Memory test. Upon completion of the two blocks of testing, participants completed a brief memory test to determine if they knew the two details from committing the crime. First, they were simply asked if they could recall what the stolen item was and if they also knew the identity of the peripheral item located in the mailbox. If they could not recall either item, the participant was shown the seven items (the pictures with the verbal name superimposed) again and asked if they recognized one. Gamer and Berti (2012) examined depth of processing between central and peripheral details when rates of explicit memory were statistically equal. To do the same here, some analyses were conducted including only participants who correctly recalled both the central and peripheral items (i.e. we excluded those participants who did not recall the peripheral item, interpreting this to mean that they did not encode any information about it during the mock crime). These data are presented in the Results section as the "Correct Recall Group."

P300 analysis. We measured P300 with using both the standard baseline—peak method and the peak—peak method. We chose to do both since the baseline—peak is a pure measure of P300 amplitude (Polich, 1991) whereas the peak—peak measure has been shown to be more sensitive for applied purposes in concealed memory detection (Soskins et al., 2001). For both the baseline—peak and peak—peak measures, the positive-going P300 peak was defined as the mean amplitude of the maximum 100ms segment between 300-800ms post-stimulus. The latency of the peak was defined as the midpoint of this 100ms segment. To determine the peak—peak measure of P300, the difference between the amplitude of the positive P300 peak and the mean of the most negative 100ms segment between P300 latency and 1700ms (the earliest possible start time of the target/non-target portion of the trial) was calculated.

Individual bootstrap analysis. As in Chapter 3, a participant was diagnosed as guilty if the proportion of probe > Iall bootstrapped iterations (when probe > Iall) was above 0.8, as this cutoff was found to yield the best overall accuracy with guilty and innocent participants. Hit rates between different conditions were again compared using an *N*-1 chi-squared procedure (where the term *N*-1 is used in place of *N* in the denominator when calculating expected counts). This procedure has been shown to be more sensitive than Fischer-Irwin or chi-square with Yates's correction in comparison trial experimental designs where the experimenter controls the counts in either the rows or the columns and is interested in detecting differential effects of outcomes on the second variable (Campbell, 2007), and is the same method used in Chapter 3 (p. 31). All bootstrap analyses were conducted solely using the peak—peak measure of P300, which has been shown to be more sensitive to the detection of concealed information (Meijer et al., 2007; Soskins et al., 2001).

Results

Memory Test

All participants were able to correctly recall the central detail upon completion of both blocks of testing. Eighteen of the 24 participants also correctly recalled the peripheral item. The six who did not also did not recognize it when shown pictures of the items again, reporting they had not noticed it in the mailbox. This difference in recall rates between the central and peripheral blocks was significant, χ^2 (1,23) = 6.74, $p = .009$.

Reaction Times

Reaction times (RTs) to the "I Saw It" response are presented in Table 3. These RTs were analyzed in a 2 (item type: central vs. peripheral) X 2 (stimulus: probe vs. Iall) repeated-measures ANOVA. Results showed a main effect of stimulus, $F(1, 23) = 5.21$, $p = .032$, $\eta^2 = .19$, with longer RTs following probes than irrelevants. The effect of item type (*p*

= .270) and the item type X stimulus interaction (p = .405) were both not significant. The RTs were then analyzed in a 2 (presentation modality: pictorial vs. verbal) X 2 (stimulus) repeated-measures ANOVA. While there was no main effect of presentation modality (p = .134), we did find an interaction between presentation modality and stimulus, $F(1, 23)$ = 4.41, p = .047, η^2 = .16. Post-hoc t-tests revealed that in the pictorial block, probe RTs (M = 345.63, SD = 104.6) were longer than for Iall (M = 335.92, SD = 96.70), p = .007, while no differences existed in the verbal blocks (p = .846).

Table 3: Chapter 3 Probe and Iall RTs for Each Item Type and Modality

	Probe		**Iall**	
Block	RT	*SD*	RT	*SD*
Central	344.71	104.38	341.54	103.87
Peripheral	360.46	112.58	353.33	101.30
Pictorial	345.63	104.61	335.92	96.70
Verbal	359.54	112.49	358.96	107.22

Note: All values calculated collapsing across blocks 1 and 2

Full Sample P300 Amplitudes

Grand average ERPs are presented in Figure 5. All analyses were completed with an alpha level of 0.05. Partial eta squared effect sizes are reported for significant effects. Since our two dependent variables, the baseline—peak and peak--peak measures of P300, are highly correlated (Soskins et al., 2001), the preferred multivariate analysis of variance (MANOVA) procedure is passed over in favor of separate univariate analyses of variance (ANOVAs) to avoid problems of multicollinearity. Additionally, because the order of blocks and habituation were potential nuisance variables given the experimental design, we had to first determine if there was an effect of block before conducting any analyses on the effects of our variables of interest. To do this, two separate 2 (block: 1 vs. 2) X 2 (stimulus: probe vs. Iall) repeated-measures ANOVAs were conducted on baseline—peak and peak—peak P300 measures. No effects were found for block or a block X stimulus interaction for either (baseline—peak or peak—peak) dependent measure (all *p*'s > .26), so all further analyses are collapsed across block. As was expected, a significant main effect was found for stimulus type with both the baseline—peak $F(1, 23) = 24.30$, $p < .001$, $\eta^2 = .514$, and peak—peak P300 amplitudes, $F(1, 23) = 34.94$, $p < .001$, $\eta^2 = .60$. In both cases, collapsing across central/peripheral and pictorial/verbal blocks, mean probe P300 amplitudes were larger than Iall. For any following analyses at this level, this redundant main effect of stimulus type won't be reported again.

Figure 5: Chapter 3 Grand Average ERPs

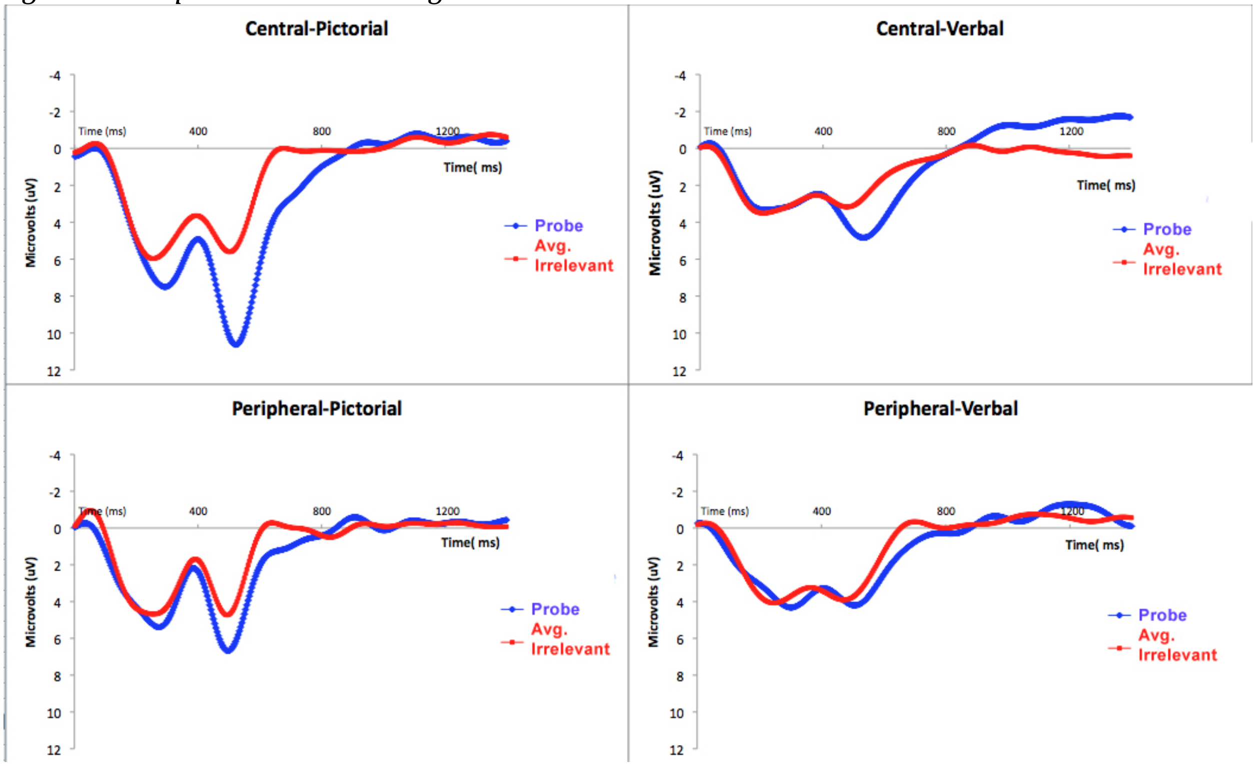

Figure 5: Grand average ERPs for probe and irrelevant stimuli in for each detail X modality combination.

Baseline—Peak P300 amplitudes. Base-peak P300 amplitudes are presented in Figure 6. The results from all following analyses can be found in Table 4. To examine the effect of item type, we conducted 2 (item type: central vs. peripheral) X 2 (stimulus) repeated-measures ANOVA. The results yielded a trend for the main effect of item type, $F(1,23) = 2.86$, $p = .104$, $\eta^2 = .11$, with central details evoking overall larger P300s than peripheral details. As predicted, we found a significant interaction between item type and stimulus, $F(1,23) = 6.09$, $p = .021$, $\eta^2 = .21$. Follow-up t-tests revealed that while there was no difference in Iall amplitudes between the central ($M = 4.67$, $SD = 2.30$) and peripheral ($M = 4.55$, $SD = 2.16$) details, $p = .81$, the mean probe amplitude for the central detail ($M = 8.11$, $SD = 4.40$) was larger than in the peripheral detail ($M = 5.82$, $SD = 3.55$), $p = .044$.

Next, a similar 2 (modality: pictorial vs. verbal) X 2 (stimulus) repeated-measures ANOVA was conducted. Collapsing across the central and peripheral details, there was a main effect of modality, $F(1, 23) = 16.47$, $p < .001$, $\eta^2 = .42$. Pictorial stimuli evoked larger P300s than verbal stimuli. The results also yielded a significant modality X stimulus interaction, $F(1, 23) = 5.61$, $p = .027$, $\eta^2 = .20$. Supporting our predictions, while probe amplitudes were larger than Iall for both modalities, this difference was larger for pictorial stimuli (3.04uV) than for verbal stimuli (1.30uV).

Since the experimental design, which resulted in a partial Latin square, made a full omnibus test impossible, to test for interactions between item type and presentation modality, 2 (modality) X 2 (stimulus) mixed-model ANOVAs were needed with modality as a between-subjects factor both the central and peripheral items. For the central detail, the main effect of stimulus showed that probes evoked larger P300s than Iall, $F(1, 23) = 27.11$,

$p < .001$, $\eta^2 = .55$. This analysis also yielded a significant effect of modality, $F(1, 23) = 15.50$, $p = .001$, $\eta^2 = .41$. As in the prior analysis collapsing across central and peripheral blocks, pictorial stimuli evoked larger P300s than verbal stimuli. Additionally, we found a significant interaction between modality and stimulus, $F(1, 23) = 6.20$, $p = .021$, $\eta^2 = .22$. Again, supporting our prediction that pictorial stimuli would result in larger probe—Iall P300 differences, the mean difference in the pictorial group ($M = 5.10$, $SD = 4.09$) was larger than that in the verbal group ($M = 1.80$, $SD = 2.05$).

The data were then analyzed similarly, but with the intention of comparing the effect of modality solely within the peripheral block in another 2 (modality) X 2 (stimulus) mixed-model ANOVA. While, once again, the main effect of stimulus showed that probes elicited larger P300s than irrelevants, $F(1, 23) = 5.09$, $p = .034$, $\eta^2 = .19$. There was no main effect of stimulus modality, $p = .404$. Additionally, contrary to our hypothesis that pictorial stimuli would enhance the probe—Iall difference, especially in the peripheral block, the test for the interaction of modality X stimulus was not significant, $p = .419$.

Peak—peak P300 amplitudes. The same analyses conducted on baseline—peak amplitudes were repeated using the peak—peak P300 measure (see Figure 7). Refer to Table 4 for a summary of the following ANOVA results. The test for a main effect of item type in the 2 (item type) X 2 (stimulus) repeated-measures ANOVA was not significant ($p = .289$). As with the baseline—peak measure, there was a significant item type X stimulus interaction, $F(1, 23) = 6.39$, $p = .019$, $\eta^2 = .217$. The probe—Iall difference in the central blocks (4.71uV) was larger than in the peripheral blocks (2.37uV), shown by follow-up *t*-tests to be due to a trend for enhanced P300s to the probes in the central ($M = 10.81$, $SD =$

5.26) versus the peripheral block (*M* = 8.65, *SD* = 5.06), *p* = .105, while Iall amplitudes did not differ between the central (*M* = 6.09, *SD* = 3.07) and peripheral (*M* = 6.29, *SD* = 2.82) blocks, *p* = .775.

The 2 (modality) X 2 (stimulus) repeated-measures ANOVA revealed a significant main effect for modality, *F*(1, 23) = 10.95, *p* < .001, η^2 = .32. As with the baseline—peak measure, collapsing across the central and peripheral details, peak—peak P300 amplitudes were larger in the pictorial block than the verbal block. Interestingly, in contrast to with the baseline—peak measure, the test for a modality X stimulus interaction was not significant, *p* = .251.

Next, we repeated the two 2 (modality) X 2 (stimulus) mixed-model ANOVAs individually within the central and peripheral blocks. Once again, probe P300s were also larger than Iall P300s, *F*(1, 23) = 35.68, *p* < .001, η^2 = .62. In the central block, there was also a significant main effect of modality, *F*(1, 23) = 7.77, *p* = .011, η^2 = .21. Again, pictorial stimuli elicited larger P300s than verbal stimuli, though the effect size was smaller than with the base—peak measure (.41). As in the previous analysis collapsing across item types, and in contrast to the base—peak measure, there was no interaction between modality and stimulus with the peak—peak measure of P300, *p* = .328.

Within the peripheral block, there was an effect of stimulus type, with probes evoking larger P300s than irrelevants, *F*(1, 23) = 10.28, *p* = .004, η^2 = .32, which was, once again, a larger effect than seen with the base—peak measure (.19). As with the base—peak measure, there was no effect of modality, *p* = .432, or a modality X stimulus interaction, *p* = .580.

Figure 6: Chapter 3 Full Sample Baseline—Peak P300 Amplitudes

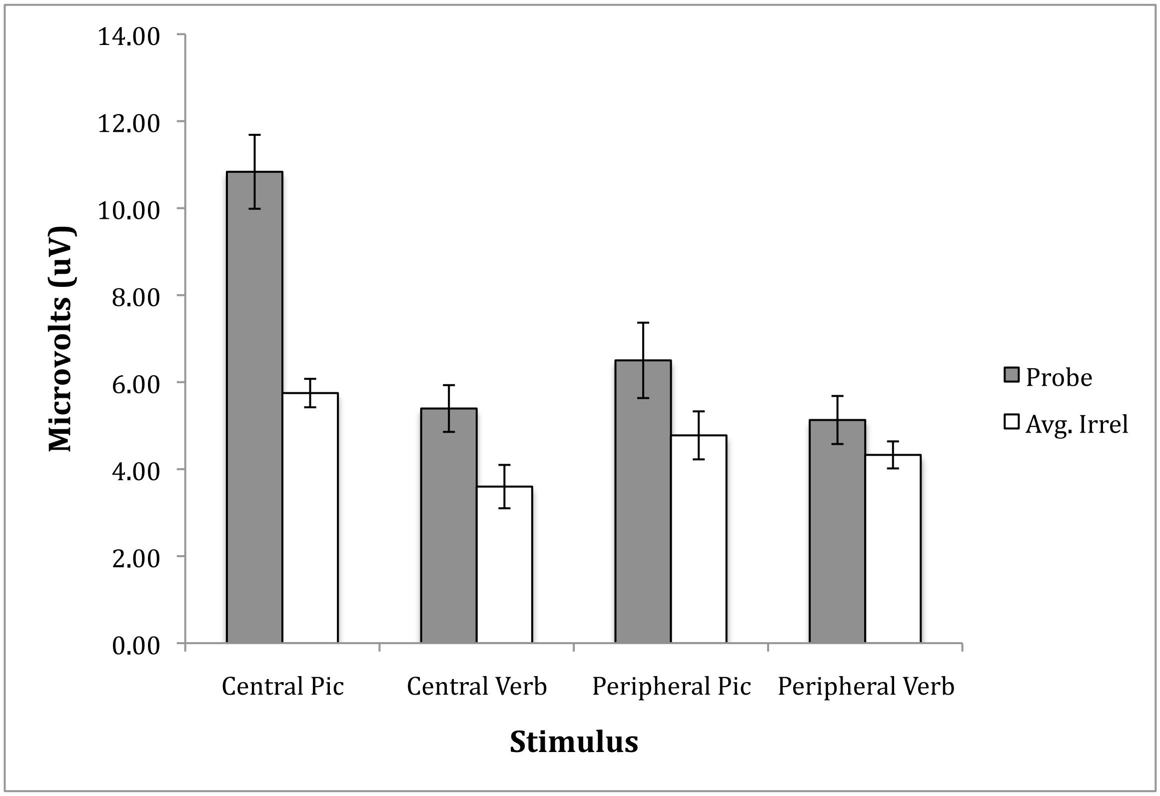

Figure 6: Baseline—peak probe and irrelevant P300 amplitudes for the full sample separated by condition.

Figure 7: Chapter 3 Full Sample Peak—Peak P300 Amplitudes

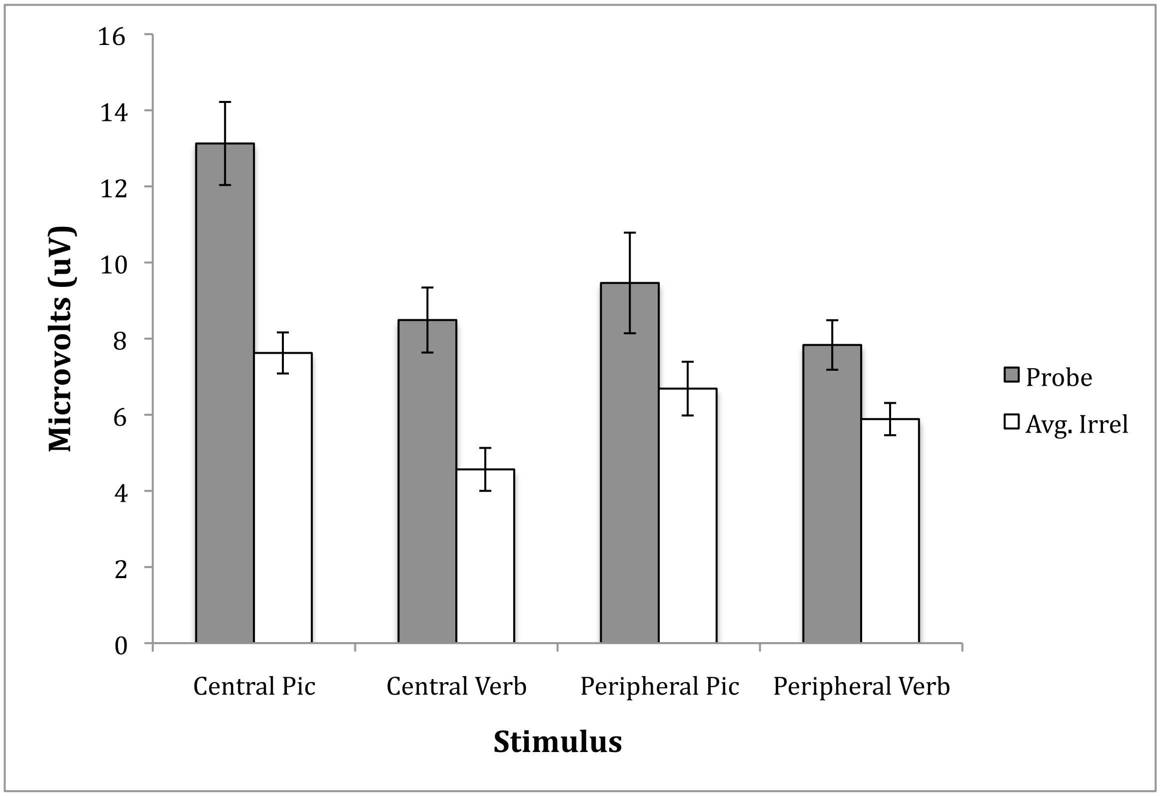

Figure 7: Peak—peak probe and irrelevant P300 amplitudes for the full sample separated by condition.

Table 4: Chapter 3 Full Sample ANOVA Results

Block	**Variables**	**Effect Test**	**P300 Measure**	***F***	***p***	**η^2**
Overall	2 (Item) X 2 (Stim)	Item	B-P	2.86	0.104	0.11
			P-P		NS	
		Stim	B-P	24.30	< .001*	0.51
			P-P	34.94	< .001*	0.6
		Item X Stim	B-P	6.09	0.021*	0.21
			P-P	6.39	0.019*	0.22
	2 (Mod) X 2 (Stim)	Mod	B-P	16.47	< .001*	0.42
			P-P	10.95	< .001*	0.32
		Stim	BP	24.30	< .001*	0.51
			P-P	34.94	< .001*	0.6
		Mod X Stim	B-P	5.61	0.027*	0.2
			P-P		NS	
Central	2 (Mod) X 2 (Stim)	Mod	B-P	15.50	0.001*	0.89
			P-P	7.77	0.011*	0.22
		Stim	B-P	27.11	< .001*	0.55
			P-P	35.68	< .001*	0.62
		Mod X Stim	B-P	6.20	0.021*	0.22
			P-P		NS	
Peripheral	2 (Mod) X 2 (Stim)	Mod	B-P		NS	
			P-P		NS	
		Stim	B-P	5.09	0.034*	0.19
			P-P	10.28	0.004*	0.32
		Mod X Stim	B-P		NS	
			P-P		NS	

Note: Results with p-values above .15 are presented as NS. Degrees of freedom for all analyses were (1, 23). Partial eta-squared effect sizes are reported for significant effects and trends. * Indicates significant at α = .05. Mod = Modality. Stim = Stimulus.

Correct recall sample P300 amplitudes

Baseline—peak P300 amplitudes. See Table 5 for a summary of the following analyses. P300 amplitudes are presented in Figure 8. For the participants who correctly recalled both central and peripheral items, the 2 (item type) X 2 (stimulus) repeated-measures ANOVA yielded a main effect of stimulus, $F(1, 17) = 20.14$, $p < .001$, $\eta^2 = .54$, as probe P300s were larger than irrelevants. There was no main effect of item type ($p = .280$). In contrast to Gamer and Berti (2012), even with the subjects who did not encode the peripheral stimulus excluded, we found a trend for an interaction between item type and stimulus, $F(1, 17) = 3.13$, $p = .095$, $\eta^2 = .16$. The probe – Iall difference in the central block ($M = 3.36$, $SD = 3.35$) was larger than in the peripheral block ($M = 1.85$, $SD = 2.72$).

The results for the 2 (modality) X 2 (stimulus) ANOVA were similar to the results from the full sample. There was a main effect of modality, with pictorial stimuli evoking larger P300s than verbal stimuli, $F(1, 17) = 23.39$, $p < .001$, $\eta^2 = .58$, and a modality X stimulus interaction, $F(1, 17) = 9.17$, $p = .008$, $\eta^2 = .35$. Again, the probe – Iall difference was larger for the pictorial stimuli ($M = 3.73$ uV) than the verbal stimuli ($M = 1.47$ uV).

Within the central block, a 2 (modality) X 2 (stimulus) mixed-model ANOVA (with modality as the between-subjects factor) revealed main effects of stimulus, $F(1, 16) = 20.06$, $p < .001$, $\eta^2 = .56$, with probe P300s being larger than Iall, and of modality, $F(1, 16) = 8.14$, $p = .012$, $\eta^2 = .34$. As in the test with the whole sample, the pictorial stimuli evoked larger P300s than the verbal stimuli (though the effect size of .89 in the previous analysis was much larger). Further, we again observed a modality X stimulus interaction, $F(1, 16) = 5.39$, $p = .034$, $\eta^2 = .25$. Post-hoc *t*-tests found that while Iall amplitudes did not differ between the pictorial (*M* = 5.65, *SD* = 1.74) and verbal (*M* = 4.22, *p* = 2.69) modalities (*p* = .190), probe amplitudes to pictorial stimuli (*M* = 10.47, *SD* = 3.21) were significantly larger than those for verbal ones (*M* = 5.57, *SD* = 3.07), *t*(16) = 3.16, *p* = .006).

The results of the same 2 X 2 ANOVA in the peripheral block were slightly different. Probes elicited larger P300s than irrelevants, *F*(1, 16) = 8.48, *p* = .010, η^2 = .35. There was a trend for a significant effect of modality, *F*(1, 16) = 2.78, *p* = .115, η^2 = .15 with pictorial stimuli evoking somewhat larger P300s than verbal stimuli. There was no significant modality X stimulus interaction (*p* = .483)

Peak—peak P300 amplitudes. Peak—peak P300 amplitudes from the correct recall group are presented in Figure 9. The following analyses are summarized in Table 5. As with the full sample, the 2 (item type) X 2 (stimulus) repeated-measures ANOVA yielded an effect where probe stimuli were larger than irrelevants, *F*(1, 17) = 27.60, *p* < .001, η^2 = .62. As with the baseline—peak measure, and again in contrast to the full sample analysis, there was no effect of item type (*p* = .602) and the item type X stimulus interaction only trended toward significance, *F*(1, 17) = 2.56, *p* = .128, η^2 = .131.

The 2 (modality) x 2 (stimulus) repeated-measures ANOVA on the correct recall sample again found that probe P300s were larger than Iall P300s, *F*(1,17) = 27.60, *p* < .001, η^2 = .62. As with the baseline—peak measure, pictorial stimuli evoked larger P300s than verbal stimuli, *F*(1, 17) = 11.16, *p* = .004, η^2 = .40, but this effect was not as large as with the base—peak measure (η^2 = .58). The test for a modality X stimulus interaction trended toward significance, *F*(1, 17) = 2.88, *p* = .108, η^2 = .145, with probe – irrelevant P300 differences being larger in the pictorial block (M = 4.52, SD = 3.50) than the verbal block (M = 3.23, SD = 3.62).

Within the central block, a 2 (modality) x 2 (stimulus) mixed-model ANOVA yielded a significant main effect of stimulus $F(1, 17) = 27.85$, $p < .001$, $\eta^2 = .64$ (probe P300s were larger than irrelevants), and a marginally significant effect of modality, with pictorial stimuli evoking larger P300s than verbal stimuli, $F(1, 16) – 3.43$, $p = .083$, $\eta^2 = .18$. Again, this effect was smaller than with the base—peak measure ($\eta^2 = .34$). Unlike the base—peak measure, the modality X stimulus interaction was not significant ($p = .477$). The same analysis conducted on the peripheral block found only a main effect of stimulus. As in each previous analysis, probes evoked larger P300s than irrelevants, $F(1, 17) = 14.32$, $p = .002$, $\eta^2 = .47$. Neither the main effect of modality ($p = .163$) nor the modality X stimulus interaction was significant ($p = .494$).

Figure 8: Chapter 3 Correct Recall Sample Baseline—Peak P300 Amplitudes

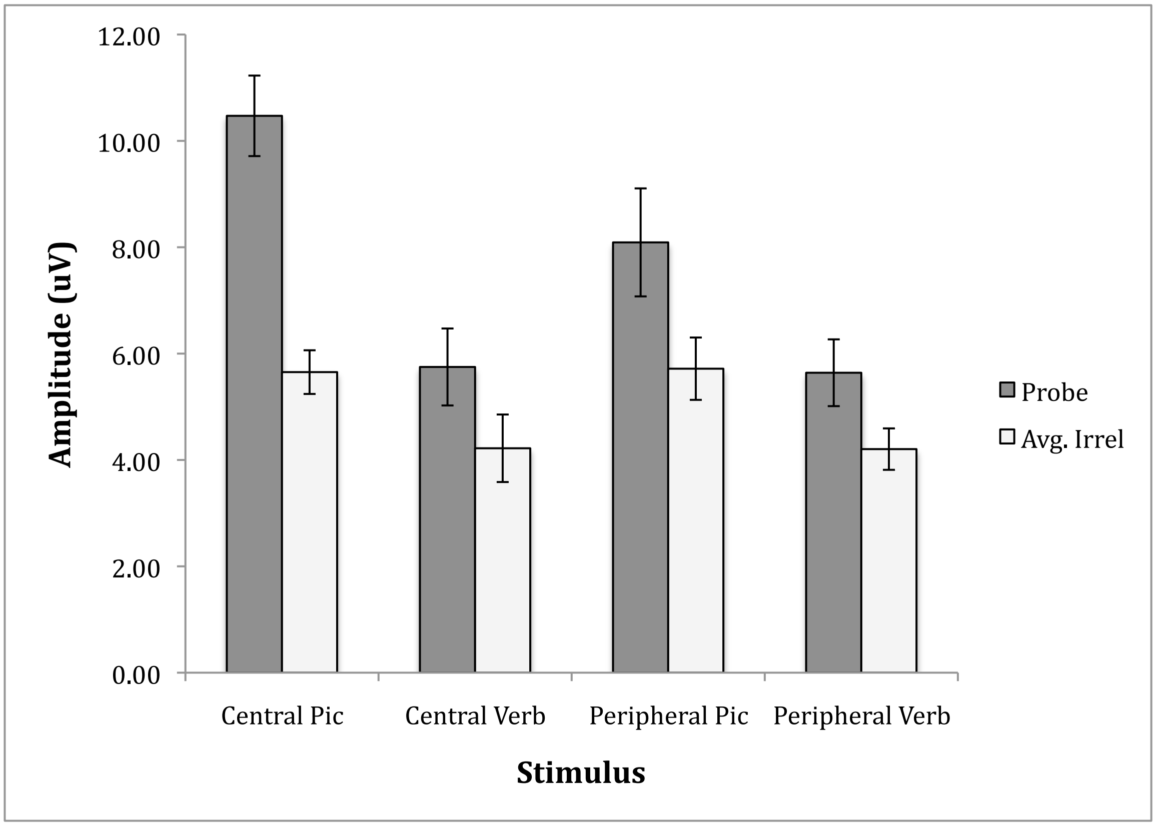

Figure 8: Baseline—peak probe and irrelevant P300 amplitudes separated by condition.

Figure 9: Chapter 3 Correct Recall Sample Peak—Peak P300 Amplitudes

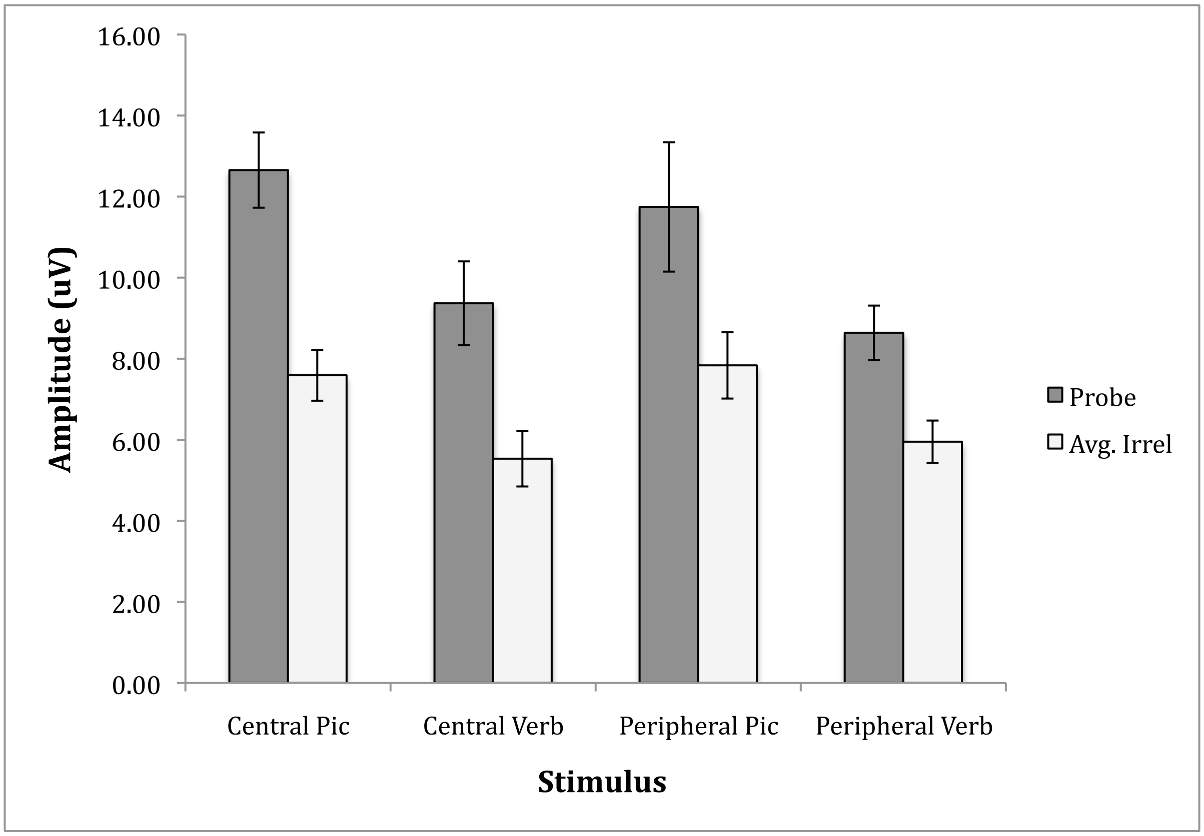

Figure 9: Peak—peak probe and irrelevant P300 amplitudes separated by condition.

Table 5: Chapter 3 Correct Recall P300 Amplitude ANOVA Results

Block	Variables	Effect Test	P300 Measure	*F*	*p*	h^2
Overall	2 (Item) X 2 (Stim)	Item	B-P		NS	
			P-P		NS	
		Stim	B-P	20.14	< .001*	0.54
			P-P	27.6	< .001*	0.62
		Item X Stim	B-P	3.13	0.095	0.16
			P-P	2.56	0.128	0.13
	2 (Mod) X 2 (Stim)	Mod	B-P	23.39	< .001*	0.58
			P-P	11.16	0.004*	0.4
		Stim	B-P	20.14	< .001*	0.54
			P-P	27.6	< .001*	0.62
		Mod X Stim	B-P	9.17	0.008*	0.35
			P-P	2.88	0.108	0.15
Central	2 (Mod) X 2 (Stim)	Mod	B-P	8.14	0.012*	0.34
			P-P	3.43	0.083	0.18
		Stim	B-P	20.06	< .001*	0.56
			P-P	27.85	< .001*	0.64
		Mod X Stim	B-P	5.39	0.034*	0.25
			P-P		NS	
Peripheral	2 (Mod) X 2 (Stim)	Mod	B-P	2.78	0.115	0.15
			P-P		NS	
		Stim	B-P	8.48	0.01*	0.35
			P-P	14.32	0.002*	0.47
		Mod X Stim	B-P		NS	
			P-P		NS	

Note: Results with p-values above .15 are presented as NS. Degrees of freedom for all analyses were (1, 17). Partial eta-squared effect sizes are reported for significant effects and trends. * Indicates significant at $\alpha = .05$. Mod = Modality. Stim = Stimulus.

Full Sample Bootstrap Results

Number of significant bootstrapped iterations. To test for potential effects of block, two independent-samples *t*-tests were run comparing the number of significant iterations in the central and peripheral blocks between participants who had each in the first and second block, respectively. Neither of these comparisons was significant (both *p*'s > .650). So, for further analyses, the number of significant iterations was collapsed across block. Supporting our prediction and collapsing across presentation modality, a paired-samples *t*-test found that the number of significant bootstrapped iterations was higher in the central block (M = 890.21, SD = 162.94) than in the peripheral block (M = 737.38, SD = 242.30), $t(23) = 2.47$, $p = .022$. As with the P300 analyses, we were interested in potential differential effects of pictorial and verbal stimuli on the central and peripheral items. We found no difference between the number of significant bootstrapped iterations between the pictorial and verbal modalities for the central item, $t(22) = .075$, $p = .941$, or the peripheral item, $t(22) = 4.54$, p = .654.

Bootstrapped detection rates. In Chapter 3, the point that yielded the best-combined sensitivity and specificity (measured by total proportion of correct diagnoses) was 800 probe > Iall iterations. The hit rates (at 800 here) for the full sample in the current study are reported in Table 6 (we report hit rates only using the peak—peak comparison method). Overall, the differences in hit rates for central and peripheral blocks were marginally significant, χ^2 (1,23) = 3.32, $p = .068$, with a greater proportion of guilty diagnoses in the central block. No differences were found when comparing pictorial versus verbal presentation modalities within the central and peripheral blocks (both *p*'s > .59).

Correct Recall Sample Bootstrap Results

Number of significant bootstrapped iterations. When the participants who did not recall the peripheral item were eliminated from the analysis, the number of significant bootstrapped iterations between the central (*M* = 877.56, *SD* = 180.10) and peripheral items (*M* = 827.67, *SD* = 176.23) did not differ, *t*(17) = 0.85 p = .407. In agreement with the results including all participants, with this sample, there was also no difference in the number of significant iterations between pictorial and verbal presentation modalities in the central, *t*(16) = 0.49, *p* = .634, or peripheral, *t*(16) = 0.85, *p* = .408, blocks.

Bootstrapped hit rates. Hit rates for the sample including only participants who recognized both details are presented in Table 7. While with the full sample we found a marginally significant difference in detection rates between central and peripheral blocks, when participants who did not recognize the peripheral item in the memory test were eliminated, this difference disappeared (p = .678). As before, detection rates between the peripheral verbal (8/10, 80%) and peripheral pictorial (5/8, 63%) blocks still did not differ (p = .403).

Table 6: Chapter 3 Diagnostic Hit Rates for the Full Sample

	Hit Rates					
	Block Type				Overall	
	Central Pic	Central Verb	Peripheral Pic	Peripheral Verb	Central	Peripheral
Correct	9/12	10/12	5/12	8/12	19/24	13/24
Prop	0.75	0.83	0.42	0.67	0.79	0.54

Note: Detection rates are based on the use of a .8 bootstrap criterion.

Table 7: Chapter 3 Diagnostic Hit Rates for the Correct Recall Sample

	Hit Rates					
	Block Type				Overall	
	Central Pic	Central Verb	Peripheral Pic	Peripheral Verb	Central	Peripheral
Correct	8/10	6/8	5/8	8/10	14/18	13/18
Prop	0.80	0.75	0.63	0.80	0.78	0.72

Note: Detection rates are based on the use of a .8 bootstrap criterion.

Discussion

In the current study, we examined the effectiveness of a P300-CIT at detecting central and peripheral items from a mock crime while manipulating the stimulus presentation modality. Using our full sample, with both the baseline—peak and peak—peak measures of P300, we found the predicted interaction between item type and stimulus. P300 amplitudes to probe stimuli were augmented in the central block over the peripheral block, leading to greater probe – irrelevant differences. This effect still trended toward significance when including only those participants who recalled both items. We also observed the predicted effect of enhanced probe amplitudes using pictorial versus verbal stimuli, however this effect was only found with the baseline—peak measure. While this effect was found when collapsing across the central and peripheral blocks, when analyzed separately, it was observed only in the central block, the opposite of our predicted effect. These findings stand in contrast to those reported by Gamer and Berti (2012), who found no differences in P300 amplitudes between central and peripheral details, and Ambach et al. (2010) who reported that presentation modality had no effects on P300 in a CIT.

As mentioned in the introduction, the way P300 was measured in Ambach et al. (2010) appears to be problematic as the true maximum positivity of P300 may have been missed. In the current study, to avoid this potential problem, and altering what we have done in many previous studies, we began our search for the positive-going P300 peak at 300ms. Despite this, when using the peak—peak measure of P300, as Ambach et al. (2010) did, we did not observe a differential effect of stimulus modality on probes versus irrelevants. However, when limiting our analysis to just those participants who correctly recalled both the central and peripheral details, we found a trend (*p* = .108) toward this effect. Taken together, these results suggest that when using a baseline—peak measure of P300, pictorial stimuli offer an advantage over verbal stimuli in the context of a CIT, potentially due to the congruence between the modalities of encoding and recall (Stenberg et al., 1995).

Our other results regarding central versus peripheral item types that conflict with those reported by Gamer and Berti (2012) are somewhat harder to reconcile. Using the full sample, we found a clear effect of larger probe – irrelevant P300 amplitude differences for the central versus peripheral items. This comparison, however, is not the critical one. The important point made by Gamer and Berti (2012) was their finding that P300 amplitudes between central and peripheral details did not differ when rates of explicit memory recall were equal, an effect previously demonstrated numerous times with ANS measures (Carmel et al., 2003; Gamer et al., 2010; Nahari & Ben-Shakhar, 2011; Peth et al., 2012). Consistent with Nahari and Ben-Shakhar (2011), participants in our study were more likely to not recall the peripheral item than the central item. When these participants were excluded from the analysis, our item type X stimulus interaction still trended toward significance with both the baseline—peak and peak—peak measures. If one assumes that the central item (which was stolen) was encoded and processed more deeply than the peripheral item (which was simply observed incidentally), then these results may suggest that P300 is indeed sensitive to depth of processing in a CIT context. Still, this may not be as prominent as with ANS measures or as powerful an effect as simple recognition of the probe stimuli (Meijer et al., 2009). The true effect of central versus peripheral details and depth of processing on P300 amplitudes thus remains somewhat unclear. It is possible that our observed trends may have become significant with increased statistical power. This is, of course, an empirical question, and further examination is necessary to more clearly elucidate these effects. One possible explanation for our differential findings is the fact that we utilized the CTP (Rosenfeld et al., 2008) whereas Gamer and Berti (2012) used the more common three-stimulus protocol method, which may better focus participants' attention by requiring differential responses on each trial depending on the presented stimulus. Additionally, participants in our study completed the CIT examination immediately after the mock crime, rather than after a week delay. However, one would assume that any effect on memory from a time delay would be a deleterious one, as observed in ANS-CITs (Carmel et al., 2003; Gamer et al., 2010; Nahari & Ben-Shakhar, 2011; Peth et al., 2012), so their lack of an effect, even after a one-week delay remains significant. Regardless of these differences, the main finding by Gamer and Berti (2012) that P300 is *less* sensitive to depth of processing than ANS measures remains important and significant.

Another point to consider based on these results is that the terms "central" and "peripheral," in regards to an item or detail's role in a crime, are somewhat arbitrary. While it is well known that items that are processed more deeply are later better remembered (in fact, ERP recordings sensitive to this effect during encoding can be predictive of later memory performance; Paller, Kutas, & Mayes, 1987), items in a crime should not be treated dichotomously as if they are processed either deeply or shallowly. In reality, the depth of processing of an individual item from a crime would vary along a continuum ranging from shallow (or not at all) to deep depending on characteristics such as the item's visibility, novelty, and whether or not the suspect interacted with it. Taking this into consideration, it is possible that the two items used as peripheral details by Gamer and Berti (2012) may have been more deeply encoded and processed than the peripheral detail in the current study. This difference could account for why we observed a trend for an interaction between item type and stimulus type (when explicit memory rates were equal) and Gamer and Berti (2012) did not. Given this, it is critical for an examiner in the field to identify details of a crime that are most likely to have been not only seen, but processed deeply by a suspect for use in a CIT. The potential difficulty in identifying sufficient numbers of such details in a crime is one reason that Podlesny (1993) was skeptical of the applicability of the CIT in real-life cases. This concern, however, is somewhat lessened by the ANS-CIT's consistent and successful use in Japan (Nakayama, 2002; Osugi, 2011).

It may be useful in the future to attempt to quantify depth of processing of mock crime stimuli through subjective participant ratings (or some other measure) in order to better elucidate its impact on P300 amplitudes in a CIT. Additionally due to time constraints with each participant, we were only able to complete two blocks of testing per person, resulting in a partial Latin square design. A fully counterbalanced study in which participants complete four blocks of testing on two central and two peripheral details, with one block each using pictorial and verbal presentation modalities may better address the questions of the influences of depth of processing and presentation modality.

In summary, the results of the current study showed clearly that details central to a mock crime are more likely to be remembered and detected in a CIT (similar to effects found between self-referring and incidentally acquired information; Rosenfeld et al., 2006; Rosenfeld et al., 2007). Further, the data suggest that even when explicit memory rates are held constant, P300 may still be sensitive to depth of processing, though this effect appears to be less potent than with ANS measures. Additionally, mere recognition of a probe (Meijer et al., 2009) is generally sufficient for evoking larger P300s than to irrelevant stimuli. Further, we found initial evidence that the use of pictorial stimuli enhances this difference perhaps because the presentation of a picture of an actual item related to a crime may enhance a suspect's recognition of it over a verbal representation. All together, these data emphasize the importance of picking proper CIT details (Meijer, Verschuere, & Ben-Shakhar, 2011) that are central to a crime and likely to have been deeply encoded and processed. The additional use of pictorial stimuli may prove to enhance the sensitivity of the test even further. Further examination of these issues will lead to the development of a more accurate and useful CIT.

CHAPTER 5 – COGNITIVE MECHANISM OF COUNTERMEASURES IN P300 CONCEALED INFORMATION TESTS

Introduction

Since the ANS-CIT is routinely used in police investigations in Japan (Osugi, 2011), the P300-CIT could also be a potentially valuable tool for law enforcement around the world. One threat to the field application of the P300-CIT, however, is the use of countermeasures (CMs), covert responses a suspect can use in an effort to defeat the test. If it were to be found that relatively simple CMs could defeat the test, its usefulness in the field could be compromised. Elucidating the underlying cognitive mechanisms of effective CMs could be one step toward identifying and eventually preventing, these CMs. With this focus, we completed two studies examining whether or not a simple recognition process is responsible for evoking P300s during CM use.

The most common version of the P300-CIT (referred to from this point on as the "three-stimulus" version) consists of three stimuli: a single guilty knowledge item (probe) and a number of generally related non-meaningful items (at least four irrelevants), and a single target stimulus (of the same type as the irrelevants). The participant makes one button press to the probe and all of the irrelevants (e.g. left mouse click), and a unique response (e.g. right mouse click) to the target in order to force attention. While the explicit task involved in the three-stimulus CIT is identification of the target stimulus, the meaningful probe stimulus is also recognized by the participant, and evokes a P300 response. One common method for diagnosing a participant or suspect as guilty or innocent is the bootstrapped amplitude difference (BAD) method, in which probe and irrelevant P300s are subjected to a bootstrapping procedure (Wasserman & Bockenholt, 1989) to determine if the amplitude of the P300 to the probe is reliably larger than the P300 to the irrelevants. This is determined by whether the number of bootstrapped iterations where this is the case is above a specified cutoff (Rosenfeld, 2011). Another method is a bootstrapped cross-correlation (BCC) analysis that compares whether the probe ERP is more highly cross-correlated with the target ERP versus the irrelevant ERPs (Farwell & Donchin, 1991), with higher probe-target correlation taken as an indication of guilt. (For a summary of the two methods, see Rosenfeld, 2011; Rosenfeld et al., 2004.) More recently, Rosenfeld et al. (2008) introduced the CTP version of the P300-CIT and have exclusively used the BAD method for individual diagnostics.

Regardless of the method of comparison used, the critical requirement needed to diagnose a participant as guilty in the P300-CIT is that the probe P300 is distinguishable from the irrelevant P300s (using either an amplitude or correlation comparison method). When they are not, the participant is diagnosed as innocent or non-knowledgeable. So, to defeat a P300-CIT, a participant can execute CMs that are designed to make the irrelevant P300s more closely resemble those evoked by probes. This is generally done by assigning covert responses to some of the irrelevant stimuli, essentially turning them into covert concealed target stimuli, which makes them meaningful and capable of evoking P300s. Rosenfeld et al. (2004) accomplished this by having participants execute a mix of physical and mental CMs (e.g. a covert finger press, a covert toe wiggle, an imagined face slap) to all but one of the irrelevant stimuli presented. Both Rosenfeld et al. (2004) and Mertens and Allen (2008) found these CMs to be highly effective in four experiments, lowering detection rates (based on either the BAD and BCC methods) to between 7 and 54%, in comparison to simple guilty sensitivity rates as high as 92% in Rosenfeld et al. (2004). The CTP has also been tested using a number of different CMs, and so far, has been found to be more resistant to CM use than is the three-stimulus protocol, as seen Table 8, in which simple guilty and CM hit rates for recent CTP studies are presented.

Table 8: Detection rates in CTP CM studies

Authors	Block	Detection Rate Guilty	Detection Rate CM	CM Type
Hu et al. (2012)	2/8	12/12 (100%)	12/13 (92.3%)	Mental
	4/8		10/12 (83.3%)	
	6/8		10/14 (71.4%)	
Labkovsky and Rosenfeld (2012b)	1/4	13/13 (100%)	11/12 (91.7%)	Mental
	2/4		12/12 (100%)	
	3/4		12/12 (100%)	
	4/4		11/12 (91.7%)	
	5/4*		11/12 (91.7%)	
Meixner, Haynes, Winograd, Brown, and Rosenfeld (2009)	Exp 1 – 2/4	12/12 (100%)	5/11 (45.5%)	Mental
	Exp 2 – 2/4	14/15 (93.3%)	10/16 (62.5%)	
Rosenfeld and Labkovsky (2010)	2/4	13/13 (100%)	12/12 (100%)	Mental
Rosenfeld et al. (2008)	Study 1 – 2/4	12/12 (100%)	11/12 (91.7%)	Physical/Mental
	Study 2 – 2/4	12/12 (100%)	12/12 (100%)	
Sokolovsky, Rothenberg, Labkovsky, Meixner, and Rosenfeld (2011)	Lumping – 2/4	11/12 (91.7%)	11/13 (84.6%)	Mental
	Splitting – 2/4		10/12 (83.3%)	

Note: See original studies for specific methods. * Participants also executed a CM to the probe stimulus.

Along with eliciting P300s to irrelevant stimuli, CMs also have effects on other physiological responses to probes. Ganis, Rosenfeld, Meixner, Kievit, and Schendan (2011) applied the same method of CMs as in Rosenfeld et al. (2004) to a CIT using functional magnetic resonance (fMRI) imaging and found that the same CMs that were effective against the three-stimulus P300 CIT were also effective against an fMRI-CIT. They speculated that the CMs worked because assigning a covert CM to irrelevant items reduced the relative salience and significance of the probe, thus making the neural and haemodynamic response to the stimulus less strong than in a simple guilty participant. Supporting this, Hu et al. (2012) found that probe P300 amplitudes during CM use were smaller than in a simple guilty group, and also were significantly smaller in a group that executed CMs to six of eight irrelevants compared to participants who executed them to just two or four irrelevants. Along with changes in salience, another potential explanation for decreases in responses (specifically with P300) to probe stimuli during CM use is increased task demand (Meixner et al., 2009). However, in contrast to the reduction in probe P300 amplitudes seen in Meixner et al. (2009), Sokolovsky et al. (2011) simply found increases in irrelevant P300 amplitudes when they had participants execute CMs to two of four irrelevant stimuli. So, the reduction of probe P300 amplitudes caused by increased task demand is not the sole mechanism by which CMs reduce the probe – Iall difference.

In an attempt to create a CM that would be easier for participants to execute (and thus would be less detectable based on an analysis of reaction times), we conducted a pilot study in which we had two groups of participants employ different methods of CMs in a P300-CIT (Winograd, Labkovsky, Noriega, Lamano, & Rosenfeld, 2011). The first group executed unique CMs to each of the countered irrelevants (*one-for-each*), the same method used in previous CM studies, while the second group executed only a single CM, but did it to all of the countered irrelevants (*one-for-all*). We found that the one-for-all method was as effective at eliciting P300s to countered stimuli as the one-for-each. Based on this, we speculated that the critical mechanism responsible for evoking P300s to countered irrelevant stimuli was simple recognition of the stimuli as meaningful (since they have a CM response assigned to them), rather than the actual execution of the CM itself. If this is the case, various aspects about a CM itself in a P300 CIT, such as the particular CM's complexity or relative salience, which are known to generally affect P300 latency or amplitude, should be able to be manipulated without changing the evoked P300 response, since the execution of the CM occurs after the initial recognition of the stimulus.

In the current study, we sought to further examine this hypothesis. In two experiments, we altered different aspects of CMs that participants executed during a P300-CIT. In *Experiment 1*, we replicated our pilot project by manipulating the method of CMs; participants completed blocks of both the one-for-all and one-for-each methods of CMs. In *Experiment 2*, participants completed two blocks of one-for-all CMs in which we manipulated the salience of the mental CM.

Experiment 1 Introduction

In *Experiment 1*, along with P300 amplitudes, we were particularly interested in the latency of P300s to countered irrelevant stimuli. Previously, P300 latency has been demonstrated to be sensitive to stimulus evaluation processes, but not to response selection or compatibility (Kutas, McCarthy, & Donchin, 1977; Magliero et al., 1984; McCarthy & Donchin, 1981). Given this, if the P300s evoked during CM use were dependent on the actual execution of the CMs, we expected to observe longer latencies in the more difficult one-for-each block, which involves a more complex stimulus evaluation process (the identification of whether or not any given stimulus needs to be countered along with determining which CM goes with it) than the easier one-for-all block which involves a simple stimulus evaluation process (in which the only decision is identifying whether or not a stimulus needs to be countered since the same CM response is used for all countered stimuli). A lack of difference in latency would suggest instead that an earlier process, in which the participant recognizes the to-be-countered stimulus as meaningful, evokes the P300s to the countered irrelevants. Our hypothesis is that it is a simple recognition process of the to-be-countered stimuli that evokes P300s, and thus, there would be no differences in P300 latency between the one-for-all and one-for-each methods. Similarly, we expected to find no difference in P300 amplitudes between the two CM methods (and as a direct result, no difference in correct detection rates). Within participants, however, we expected to see the largest P300s to probe stimuli as the CTP has previously been shown to be resistant to CMs (Hu et al., 2012; Labkovsky & Rosenfeld, 2012a; Meixner et al., 2009; Rosenfeld & Labkovsky, 2010; Rosenfeld et al., 2008). For the irrelevant stimuli, we expected to see larger P300 amplitudes to countered irrelevants than non-countered irrelevants since the assigning of CMs to the countered stimuli makes them meaningful and evokes P300s.

Method

Participants

Participants were 22 undergraduate students enrolled in an introductory psychology course who completed the experiment for course credit. Two were removed from all analyses due to having excessive EEG artifacts (fewer than 20 accepted probe trials) during a block of testing. The remaining 20 participants (14 female) ranged in age from 18-21 (*M* = 18.75, *SD* = 0.85). Each participant was randomly assigned to one of four conditions based upon which method of CM and detail they would be tested for knowledge of in the first block. Each condition had a total of five participants. The local IRB approved all methods.

Procedure

Study design. We utilized a partial Latin-square design in order to counter-balance the orders of both the two methods of CMs (one-for-all and one-for-each) and the two guilty knowledge details used (birthdate and hometown). Each participant completed two blocks of testing. So, if in the first block a participant was tested for their birthdate and used the one-for-all CM, in the second block they would be tested for knowledge of their hometown and would use the one-for-each method of CMs. Thus, all participants here were knowledgeable guilty CM users.

After giving informed consent, participants completed an information sheet to collect demographic information including their birthdate and hometown. After completing this form, participants read the experiment instructions. The instructions informed participants that they would be completing two blocks of testing in which the experimenter would attempt to detect autobiographical details (their birthdate and hometown). To be sure that none of the irrelevant dates or hometowns were meaningful, each participant was shown lists of 20 dates and 20 town names. They were asked to inform the experimenter if any of the details were "personally significant or meaningful." The first eight dates/towns from each list were the default stimuli used as irrelevant details in each block. If the participant identified any of these eight stimuli as meaningful, they were replaced with one of the remaining non-meaningful stimuli. For the block testing for the participant's birthdate, irrelevant dates were replaced with other non-meaningful dates if they were in the same month or shared the same number as the participant's birthdate. Participants then completed a two-minute practice block where they practiced the specific responses they had to make to the presented stimuli. After the practice block, participants were given the instructions for their first CM block of testing. The instructions for both the first and second blocks explained that the test worked by detecting a brain wave elicited when a person recognizes meaningful information in order to later emphasize the importance of assigning and executing CMs correctly. At the end of the first block, participants took a short break before receiving the instructions for and completing the second block of testing.

P300-CIT. We again used the CTP P300-CIT (Rosenfeld et al., 2008). Probe and irrelevant stimuli were presented in capital letters in all capital letter white text on a black background (e.g. "BOULDER" or "MAR 17") about 0.5" high. The monitor was placed approximately two feet in front of the participant. Participants made button-press responses to each stimulus as it was presented. To every probe and irrelevant stimulus, the participants pressed a single button with the index finger of their left hands (the "I saw it" response – meant to indicate that the participant has seen and read the stimulus). As in Chapter 4, we had participants execute the single button press because we felt that the five-button random response was too distracting for participants. For targets, the participants clicked the right button on a mouse with their right hands, and the left button to all non-targets. There were a total of 360 trials (40 each for the probe and the eight irrelevants).

Countermeasures. Participants executed CMs to four of the eight irrelevant stimuli. The so-called "lumping" CM responses were done simultaneously with the "I Saw It" response, as in Sokolovsky et al. (2011). After the initial practice, participants received the CM instructions for the first block of full testing. They were informed that since the test worked by detecting their recognition of a meaningful date/town name, the best way to defeat the test was to make some of the irrelevant items meaningful by assigning them a CM. The instructions for each block were identical, except for the specific CM used and which self-referring detail (birthdate or hometown) would be presented as a probe. Participants were presented with the four irrelevant dates or town names to which they would execute CMs and were instructed to memorize which CM went with each irrelevant item.

One-for-all. Participants were instructed to say their first name silently to *all* of the four countered-irrelevant items each time they were presented simultaneously with the "I Saw It" button-press response. The instructions given to the participants prior to the start of the block listed each of the irrelevant dates or town names they would see and indicated which of them were to be countered.

One-for-each. Participants were instructed to say a unique CM (meaningful first names) silently in their head to *each* of the four countered-irrelevant items each time they were presented simultaneously with the "I Saw It" button-press response. As in the *one-for-all* block, the instructions prior to beginning the block listed each of the irrelevant dates and which ones were to be countered with which names. The CMs were executed to the first four of eight irrelevant dates or town names. The specific CM assignments were the participant's first name, last name, a parent or sibling's name, and a different parent or sibling's name, respectively.

ERP analysis. We utilized the peak-to-peak (p-p) method for calculating P300 amplitudes since prior research has shown it to be more effective for detecting concealed knowledge than a baseline-to-peak measure (Meijer et al., 2007; Soskins et al., 2001). The P300 peak was defined as the mean amplitude of the maximum positive 100ms segment at the Pz site between 350 and 800ms post-stimulus. The latency of P300 was defined as the midpoint of this 100ms segment. Peak-to-peak amplitude is determined by calculating the difference in amplitude between the positive P300 peak and the subsequent maximum negative point of the ERP between P300 latency and the earliest possible presentation of a target/non-target stimulus (1700ms).

Bootstrap analysis. Again, we used a bootstrap procedure (Wasserman & Bockenholt, 1989) to determine, on an individual basis, if participants are guilty or innocent based on their p-p P300 amplitudes at the Pz site (where they are known to be largest; Fabiani et al., 1987). In another recent study examining the effectiveness of various CMs in the CTP using self-referring information, Hu et al. (2012) used a criterion of 90% probe > Iall bootstrap iterations as the cutoff point to declare a participant as "guilty." Given their high sensitivity and specificity, we chose to use the same 90% cutoff in the current study. This was also the same cutoff used in the first study on the CTP (Rosenfeld et al., 2008), which also utilized self-referring information. (We have found 80% to be better with incidentally acquired information.)

Results

P300 Amplitudes and Latencies

Grand average ERPs from the Pz site are presented in Figure 10. Probe stimuli in both CM conditions evoked large P300s. In both blocks, clear P300s are visible to the countered irrelevants, which are larger than to the non-countered irrelevants, suggesting that participants successfully executed the CMs. All analyses were completed using an alpha level of 0.05. Partial eta squared effect sizes are reported for significant effects. Greenhouse-Geisser correction was applied in ANOVAs when sphericity was violated.

Figure 10: Chapter 4 – Experiment 1 Grand Average ERPs

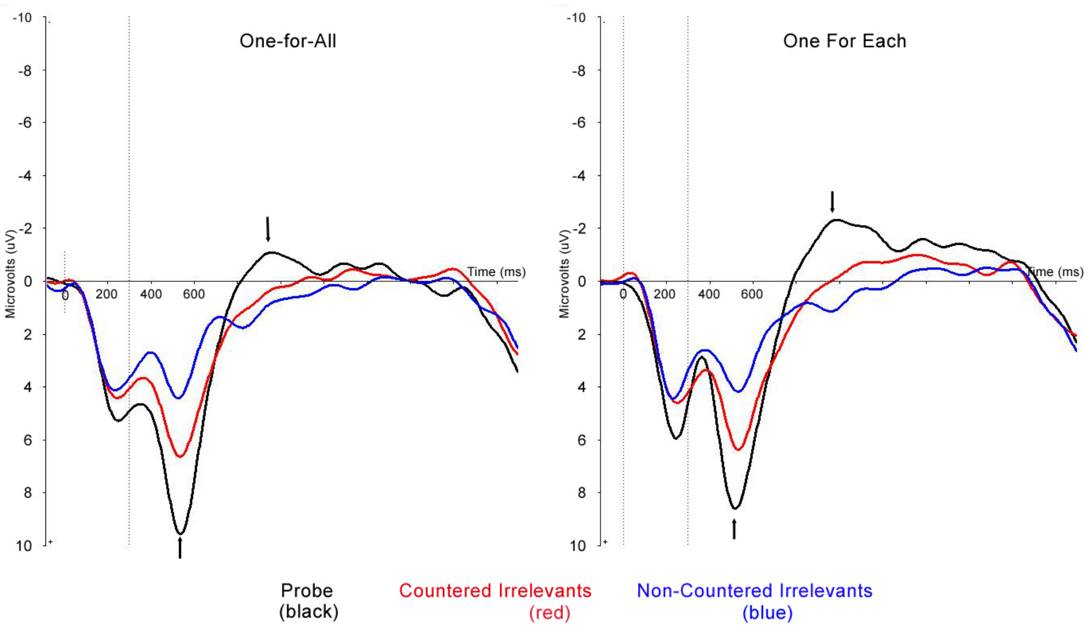

Figure 10: Grand average ERPs at the Pz site for probe, countered-irrelevant and non-countered irrelevant stimuli in the *one-for-all* and *one-for-each* blocks. Probes elicited the largest P300s. The P300 to countered irrelevants was larger than to the non-countered irrelevants using both CM methods, suggesting that participants were able to complete the assigned task. As predicted, no differences in P300 amplitude or latency were found when comparing the one-for-all and one-for-each methods of CMs.

To first determine if the order of blocks or the guilty knowledge details used had any effects on any of the stimuli, we ran two 2 (block: first, second; or guilty knowledge detail: hometown, birthdate) X 3 (stimulus: probe, countered irrelevant, non-countered irrelevant) repeated-measures analysis of variance (ANOVA). There was no significant main effect of block (p = .831), however there was a trend for an interaction between block and stimulus, $F(1.25,23.82) = 2.967$, p(g-g) = .09, $\eta^2 = .135$. Probe amplitudes were larger in the first block than in the second while countered irrelevant and non-countered irrelevant amplitudes were smaller in the first block. The 2 (guilty knowledge detail: hometown/birthdate) X 3 (stimulus) ANOVA yielded neither a significant main effect of detail (p = .838) nor an interaction between detail and stimulus (p(g-g) = .386). Given that there were no significant main effects or interactions involving block or guilty knowledge detail, all following analyses collapsed across detail and block.

Next, we conducted two 2 (CM type: one-for-all, one-for-each) X 3 (stimulus: probe, countered irrelevant, non-countered irrelevant) repeated-measures ANOVAs on P300 amplitude and latency. Mean P300 amplitudes can be found in Figure 11, and mean P300 latencies are presented in Figure 12. The ANOVA for P300 amplitude yielded a significant main effect of stimulus type, $F(1.36,25.8) = 36.22$, p(g-g) < .001, $\eta^2 = .656$. Post-hoc t-tests revealed that across CM conditions, probe P300s (M = 13.26, SD = 4.84) were larger than countered irrelevants (M = 9.21, SD = 4.50), $p < .001$, and non-countered irrelevants (M = 6.77, SD = 3.40) $p < .001$. Countered irrelevants were also larger than non-countered irrelevants, $p < .001$. As predicted, the tests for a main effect of CM type (p = .46) and a CM type X stimulus interaction (p = .58) were both not significant. For P300 latency, the tests for a main effect of CM type, (p = .63) stimulus, (p(g-g) = .44), and CM type X stimulus interaction (p = .23) were all not significant.

Figure 11: Chapter 4 – Experiment 1 P300 Amplitudes

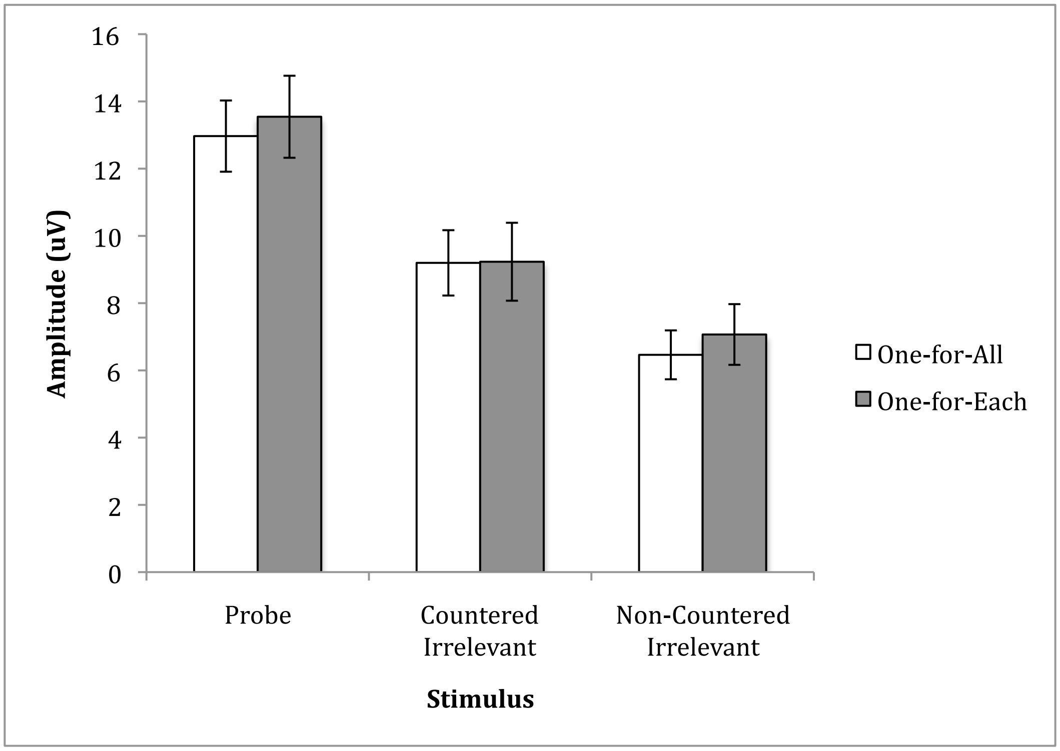

Figure 11: P300 amplitudes to probe, countered-irrelevant, and non-countered irrelevant stimuli in the *one-for-all* and *one-for-each* blocks. P300 amplitudes between the one-for-all and one-for-each methods did not differ.

Figure 12: Experiment 1 P300 Latencies

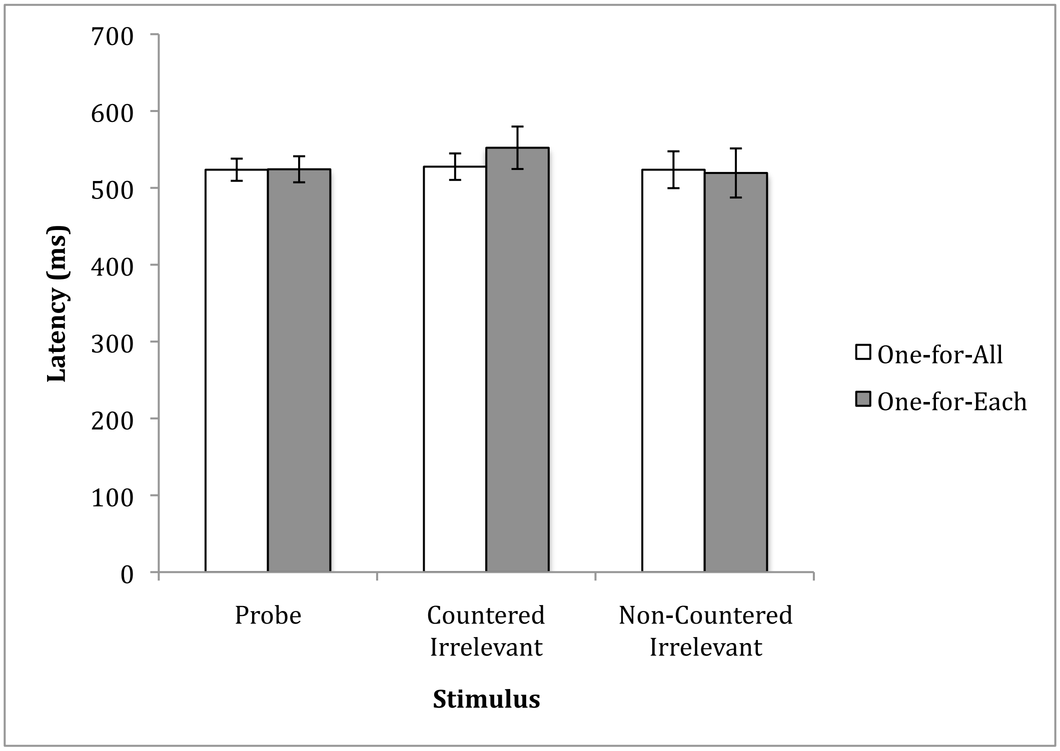

Figure 12: P300 latencies for probe, countered-irrelevant, and non-countered irrelevant stimuli in the *one-for-all* and *one-for-each* blocks. No differences were found.

Individual Bootstrapped Classifications

Correct classification rates can be found in Table 9. As in the previous studies, hit rates were compared using *N*-1 chi-squared tests (Campbell, 2007). Of the one-for-all blocks, 14/20 (70%) were classified as guilty along with 15/20 (75%) of the one-for-each blocks, $\chi^2 = 0.125$, $p = .72$. There was also no difference in the number of significant bootstrap iterations between the one-for-all ($M = 868.75$, $SD = 209.99$) and the one-for-each ($M = 888.50$, $SD = 204.20$) blocks, $t(19) = -.47$, $p = .643$. Additionally, while detection rates in the first block (16/20, 80%) were higher than in the second block (13/20, 65%), this difference was not significant, $\chi^2 = 1.129$, $p = .288$. However, the number of significant bootstrap iterations in the first block ($M = 920.75$, $SD = 147.03$) was higher than in the second block ($M = 836.50$, $SD = 246.26$), $t(19) = 2.25$, $p = .036$, so participants were more effective at using CMs in the second block than the first.

Table 9: Chapter 4 – Experiment 1 Correct Detection Rates

CM Method	Total		First Block		Second Block	
	Correct	Prop	Correct	Prop	Correct	Prop
One-for-All	15/20	0.75	8/10	0.80	7/10	0.70
One-for-Each	14/20	0.70	8/10	0.80	6/10	0.60
Overall	29/40	0.73	16/20	0.80	13/20	0.65

Note: Detection rates are based on a .9 bootstrap criterion

Table 10: Chapter 4 – Experiment 1 "I Saw It" Response RTs (ms)

CM Method	Probe		Countered Irrelevants		Non-countered Irrelevants	
	M	SD	M	SD	M	SD
One-for-All	553.25	235.46	565.7	230.53	534.00	196.01
One-for-Each	548.75	182.33	562.2	211.05	532.45	181.99

Reaction Times

Reaction times are presented in Table 10. We subjected the RT data to the same 2 (CM type) X 3 (stimulus) repeated-measures ANOVA. The results revealed a significant main effect of stimulus, $F(2,38) = 4.57$, $p = .017$, $\eta^2 = .194$. Post-hoc t-tests collapsing across the two CM types revealed that RTs to countered irrelevants ($M = 563.95$, $SD = 213.70$) were longer than to non-countered irrelevants ($M = 533.23$, $SD = 184.31$), $p = .005$. There were no differences between probe and countered irrelevant or probe and non-countered irrelevant RTs (both p's > .13). The tests for a main effect of CM type and a CM type X stimulus interaction were not significant (both p's > .88).

Experiment 1 Discussion

Our hypothesis that it is a recognition process during stimulus evaluation (rather than response selection) that evokes P300s to countered-irrelevant stimuli was supported by the results of *Experiment 1*. No differences in either P300 latency or amplitude were observed between the one-for-all and one-for-each CM methods. Further supporting the hypothesis is the fact that a reaction time difference between countered irrelevant and non-countered irrelevant stimuli was observed in absence of any latency effects. In contrast, P300 amplitudes to countered irrelevant stimuli were larger than those to non-countered irrelevants. This result replicates substantial prior evidence of the dissociation between RT and P300 latency (Magliero et al., 1984; McCarthy & Donchin, 1981). It is noted that while we considered the one-for-each method of CMs to be more difficult to execute (and this sentiment was expressed by many participants, but not recorded for all), there was no statistical difference in RTs between it and the one-for-all method. Reaction times have long been known to increase with higher task demand or response difficulty, a fact that has been used by a number of researchers to attempt to discriminate based on RTs between lying and truth telling persons (see Verschuere & De Houwer, 2011, for a review). While it is possible that the one-for-all and one-for-each methods may not differ much in difficulty, another potential explanation for a lack of an RT difference between the two methods was that we used the method of “lumping” CMs (where the CM is executed simultaneously with the “I Saw It” response) in which the instructions emphasized the need to respond quickly. In a previous study in our laboratory, these methods prevented CM users’ RTs from being distinguishable from those in an easier simple guilty condition (Sokolovsky et al., 2011). Overall, the P300 results from *Experiment 1* matched what was predicted based on the hypothesis that a simple recognition process is responsible for evoking P300s during CM use. In *Experiment 2*, we further examined this idea.

Experiment 2 Introduction

In the previous study, we showed that manipulating the difficulty of CMs had no effects on either P300 amplitude or latency. To further examine whether a simple recognition process is responsible for evoking P300s during CM use, we also wanted to manipulate the salience and meaningfulness of the mental CM response used. The amplitude of the P300 component is known to be sensitive to increased stimulus meaningfulness or salience (Johnson Jr., 1993), with more meaningful stimuli evoking larger P300s. In a CIT context, Rosenfeld et al. (2006) and Rosenfeld et al. (2007)

demonstrated that P300s to probe items in a P300 CIT were significantly larger when the probe was a well-rehearsed and meaningful stimulus (the participant's first name) versus an incidentally acquired detail (the experimenter's name). Given this, in the current experiment, the effectiveness of a mental CM (using the one-for-all method) where the assigned word was highly meaningful and salient (the participant's first name) was compared to a mental CM where the assigned word was relatively innocuous (the word "table"). Again, we hypothesized that the P300s evoked during CM use are due to recognizing the to-be-countered stimulus as meaningful, so there would be no difference in P300 amplitudes (and as a direct result, detection rates) or latencies between the two CM methods. As in *Experiment 1*, we expected the largest P300s to be evoked by probe stimuli, followed by countered irrelevant P300s, with non-countered irrelevant P300s the smallest. However, if the actual execution of the particular CM matters, one would expect to see larger P300s evoked to countered irrelevant stimuli in the block where the participant's name is used as a CM than in the "table" block, based on previous findings that P300 amplitude is sensitive to the meaningfulness and significance of a stimulus (Rosenfeld et al., 2006; Rosenfeld, Ellwanger, & Sweet, 1995; Rosenfeld et al., 2007).

Method

Participants

Experiment 2 consisted of twenty participants from an introductory psychology course who received course credit for completing the experiment. Four subjects were removed from analyses due to excessive eye blink artifacts (fewer than 20 accepted probe trials). The remaining 16 participants (8 male) included in the analyses ranged in age from 18-20 years old (*M* = 19.0, *SD* = 0.82). Each participant was randomly assigned to one of four conditions based upon which method of CM and which detail they would be tested on in the first block. Each of the four conditions had a total of four participants.

Procedure

Once again, we employed a partial Latin square design. Participants completed two blocks, one testing for knowledge of their hometown or their birthdate using two different methods of CMs (counterbalanced across participants). Overall, the methods between Experiment 1 and Experiment 2 were the same, except for specific mental CM methods used.

Countermeasures. Since there were no differences between the one-for-all and one-for-each methods of CM use in Experiment 1, we chose to use the one-for-all method for this study, but the word the participants silently said as their mental CM was manipulated. In one block, participants said their own first name silently in their head as the CM to each of the four countered-irrelevant (of eight total) stimuli. In the other block, they said the word “table.”

Results

P300 Amplitudes and Latencies

Grand average ERPs are presented in Figure 13. All analyses were completed using an alpha level of 0.05. Partial eta squared effect sizes are reported for significant effects. Greenhouse-Geisser correction was applied in ANOVAs when sphericity was violated.

Figure 13: Chapter 4 - Experiment 2 Grand Average ERPs

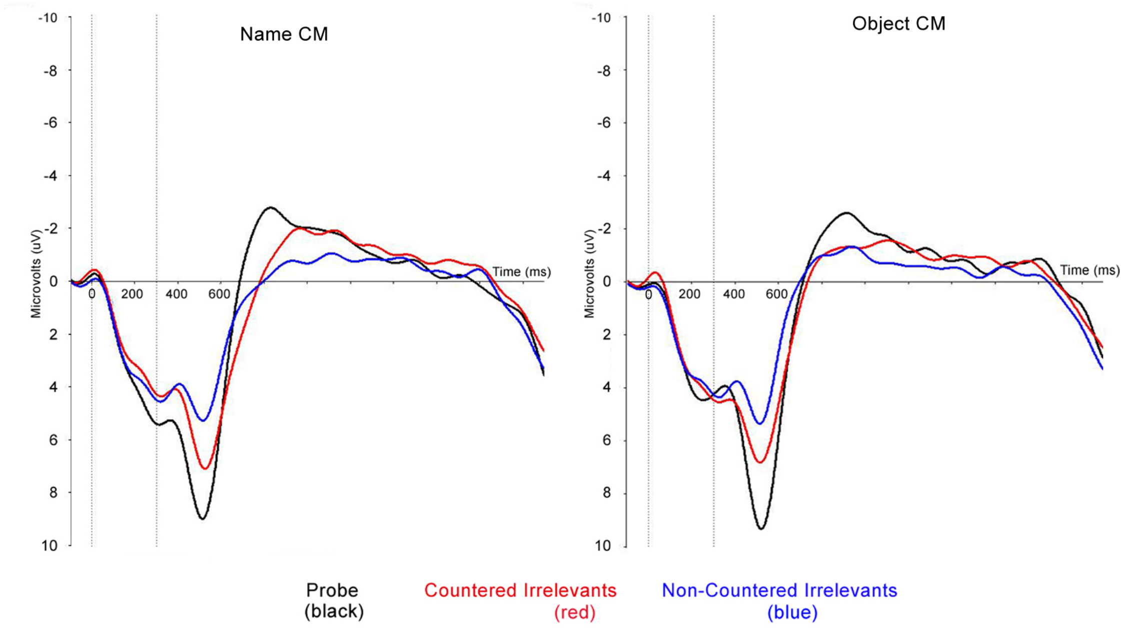

Figure 13: Grand average ERPs for the Name and Object CM blocks using the one-for-all CM method. The ERPs again clearly show that P300s to the countered irrelevants were larger than the P300s to the non-countered irrelevants. As predicted, there were no effects of CM method on any P300 measures.

As in Study 1, we conducted a 2 (block: first, second; or guilty knowledge detail: hometown, birthdate) X 3 (stimulus: probe, countered irrelevant, non-countered irrelevant) ANOVA to determine if there were any overall effects of block or guilty knowledge details or interactions with stimulus type on P300 amplitude. Once again, the tests for the main effect of block ($p = .26$) and the block X stimulus interaction ($p = .18$) were not significant. The same was true for the main effect of guilty knowledge detail ($p = .60$) and the guilty knowledge detail X stimulus interaction ($p = .46$). As such, all further analyses were conducted collapsing across blocks.

Again, we conducted two 2 (CM type: name CM, object CM) X 3 (stimulus: probe, countered irrelevant, non-countered irrelevant) ANOVAs on P300 amplitude (see Figure 14) and latency (see Figure 15). As in Study 1, there was no significant effect of CM type ($p = .84$) or a CM type X stimulus interaction ($p = .62$) for P300 amplitude. Once again, the only significant P300 amplitude effect was a main effect of stimulus type, $F(2,30) = 19.41$, $p < .001$, $\eta^2 = .56$. Post-hoc t-tests found significant differences between each stimulus type. Probe P300s ($M = 13.68$, $SD = 7.06$) were larger than those to countered irrelevants ($M = 9.77$, $SD = 5.88$), $p = .003$, and non-countered irrelevants ($M = 7.61$, $SD = 1.53$), $p < .001$. Countered irrelevant P300s were also significantly larger than non-countered irrelevant P300s, $p = .007$. The ANOVA on P300 latencies (see Figure 15) did not find any significant effects for CM type ($p = .24$), stimulus type (p(g-g) = .177), or a CM type X stimulus interaction (p(g-g) = .552).

Figure 14: Chapter 4 – Experiment 2 P300 Amplitudes

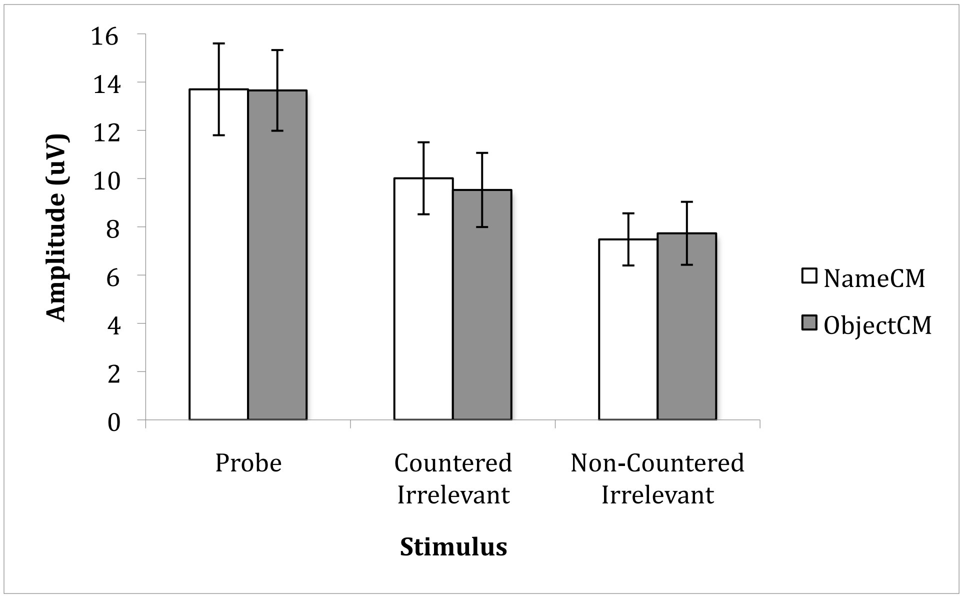

Figure 14: P300 amplitudes to probe, countered-irrelevant, and non-countered irrelevant stimuli in the Name and Object CM blocks. P300 amplitudes between the one-for-all and one-for-each methods did not differ. No differences were found between the two methods of CMs.

Figure 15: Chapter 4 – Experiment 2 P300 Latencies

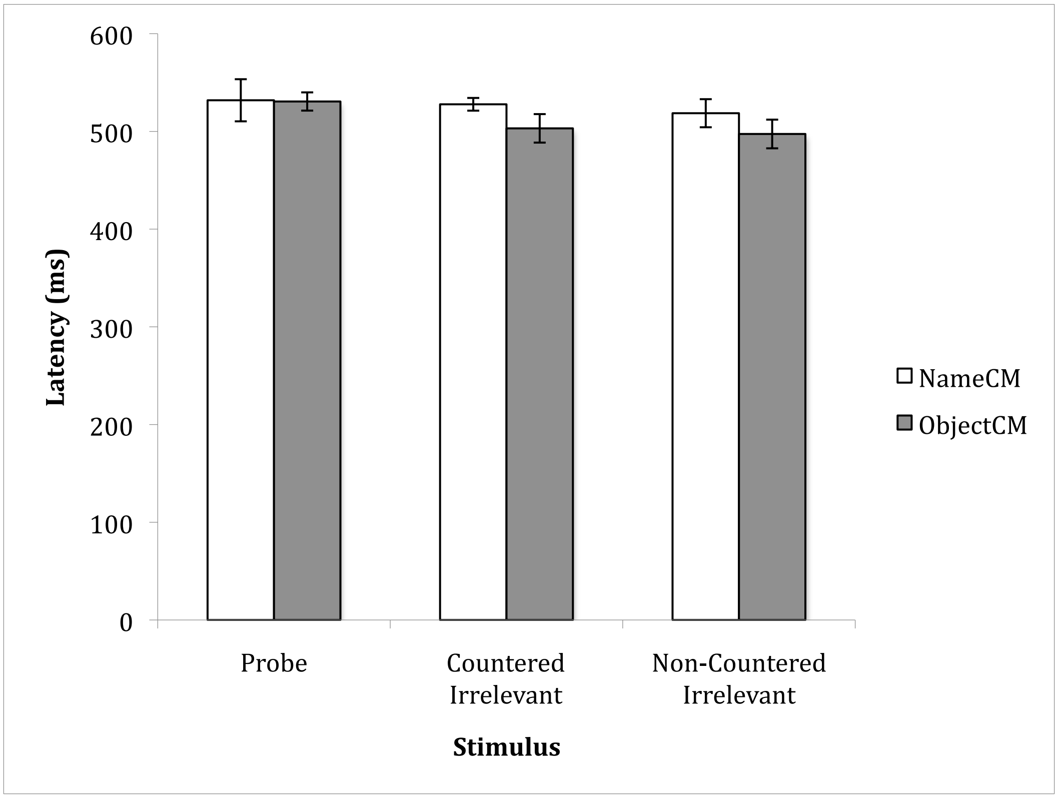

Figure 15: P300 latencies for probe, countered-irrelevant, and non-countered irrelevant stimuli in the one-for-all and one-for-each blocks. Again, no differences were found between the two CM methods or any stimulus types.

Individual Bootstrapped Classifications

Correct detection rates for the name CM and object CM blocks can be found in Table 11. As in Experiment 1, the detection rates between the two CM methods were not significantly different, $\chi^2 = .582$, $p = .446$. Again, we also found no difference in the number of significant bootstrap iterations between the name CM ($M = 857.88$, $SD = 227.44$) and the object CM ($M = 856.19$, $SD = 178.79$) blocks, $t(15) = .033$, $p = .97$. Similar to Experiment 1, we found a marginally significant effect of more significant bootstrap iterations in the first block ($M = 903.13$, $SD = 113.91$) than in the second block ($M = 810.94$, $SD = 257.27$), $t(15) = 2.03$, $p = .061$, however the number of correct classifications between the first and second blocks was equal (11/16).

Reaction Times

The RT data from one participant's object CM block was lost due to computer error, so the following analysis was conducted on the remaining 15 participants. Mean RTs for Experiment 2 are presented in Table 12. Again, we conducted a 2(CM type) X 3(stimulus) repeated-measures ANOVA on RTs. No significant effects were found (all p's > .24).

Table 11: Chapter 4 – Experiment 2 Correct Detection Rates

	Total		First Block		Second Block	
CM Type	**Correct**	**Prop**	**Correct**	**Prop**	**Correct**	**Prop**
Name CM	12/16	0.75	7/8	0.88	5/8	0.63
Object CM	10/16	0.63	4/8	0.50	6/10	0.60
Overall	22/32	0.69	11/16	0.69	11/16	0.69

Note: Detection rates are based on a .9 bootstrap criterion.

Table 12: Chapter 4 – Experiment 2 "I Saw It" Response RTs (ms)

	Probe		Countered Irrelevant		Non-countered Irrelevants	
CM Type	**M**	**SD**	**M**	**SD**	**M**	**SD**
Name CM	523.67	107.82	513.40	112.46	511.73	113.58
Object CM	517.27	133.75	549.80	149.70	519.80	146.41

Experiment 2 Discussion

While in *Experiment 1* we were interested in whether or not different CM methods (and their relative difficulty of response selection) would affect P300 amplitudes and latencies, in the current study, the main variable of interest was whether varying the significance and salience of a CM, while holding the specific method the same, would affect P300 amplitude alone. As predicted, no differences were found for either P300 amplitude between the two blocks using the participant's first name or the word "table" as the mental CM. P300 amplitudes to countered-irrelevant stimuli were found to be larger than to non-countered irrelevants, confirming that participants did execute the CMs. Along with this, there again were no P300 latency differences between the two CM blocks, though one would not expect to observe this effect (even if some process other than recognition was responsible for evoking the P300s to countered irrelevants) since the specific method of CMs (one-for-all) was the same in each block. On an individual level, while the mean number of significant iterations was lower in the second block than the first, the detection rates did not differ. Additionally, while the detection rates between the name and object CM blocks were slightly different, this result was not statistically significant. Collapsing across both, the overall 69% correct classification rate in *Experiment 2* was close to the 75% correct classification rate for the one-for-all block of *Experiment 1*.

Discussion

We hypothesized that the cognitive mechanism responsible for evoking P300s to countered irrelevant stimuli in a P300-CIT is simple recognition of the to-be-countered stimuli. Normally, without the application of CMs, these stimuli are truly irrelevant to the participant and have no significant meaning compared to other stimuli. However, they become significant when a participant assigns CMs to them, turning them in to covert targets, similar to standard overt target stimuli in a three-stimulus CIT protocol (Rosenfeld et al., 2004). While manipulating the difficulty and salience of CMs in the current experiment, we observed no differences in either P300 amplitude or latency to countered irrelevant stimuli, suggesting that some process other than those involved in the actual response selection and execution of the CM is responsible for the evoked P300 responses to these stimuli. Also supporting this theory, there were no differences in P300 latencies between countered irrelevant and probe stimuli, which do not require a unique response. If the actual execution of the CM were responsible for evoking P300, one would expect to see later P300 latencies to the countered irrelevant stimuli than to the probes, as the process of recognizing an irrelevant stimulus, identifying, and then executing its assigned CM would take longer than the simple recognition process involved with the probe stimulus.

The results fit well with what is known from extensive prior research into the major antecedents of P300. In his triarchic model of P300, Johnson Jr. (1986) argued that the primary psychological variables that affect P300 amplitude are subjective probability (rareness) and stimulus meaning (significance). During CM use, the countered irrelevant stimuli become meaningful and task relevant (which has been shown to increase P300 amplitude, Donchin & Coles, 1988). Previous research has shown that the P300s to a secondary oddball task are reduced when a concurrent and primary oddball task competes for perceptual resources, (Wickens et al., 1983). During CM use in a P300-CIT (be it the CTP or the three-stimulus method), the participant is essentially completing two tasks: an explicit task of identifying and recognizing the to-be-countered stimuli, and an involuntary implicit task of the recognition of the probe stimulus (in the three-stimulus method, the participant is also explicitly identifying the assigned target response).

Taken in this context, our finding of no differences in either countered-irrelevant or probe P300 amplitudes between the *one-for-all* and *one-for-each* methods may seem difficult to explain, as one would expect to see decreases in amplitude to the probe stimulus during the completion of a more difficult (*one-for-each*) competing task. However, this would only hold true if the perceptual task that is responsible for evoking P300s is different or changed between the two CM methods. Once again, we argue that it is simply a recognition process of each to-be-countered stimulus (and additionally, the implicit recognition task with the probe stimulus) that is important during CM use. Therefore, using a CM that is more difficult to execute does not affect P300 amplitude. In contrast, decreases in probe amplitude are observed when the number of countered stimuli is increased (Hu et al., 2012), which increases the difficulty of the explicit CM recognition task (as evidenced by increases in RT). In the current study, participants were tasked with recognizing equal numbers of to-be-countered stimuli in both CM blocks they completed. Decreases in probe P300 amplitudes due to having multiple tasks competing for perceptual resources may also explain why the CTP has been found to be more resistant to CMs than the three-stimulus CIT (Rosenfeld et al., 2008; Rosenfeld et al., 2004). In the CTP, the implicit probe recognition task only competes with the task of recognizing the to-be-countered stimuli. In contrast, in the three-stimulus CIT, the participant also must complete a third recognition task of identifying the target stimulus. This extra competing explicit oddball task may serve to further take perceptual resources away from the implicit recognition of the probe, reducing its evoked P300 response, thus making the CMs more effective.

We interpret the present results as suggesting that what makes CMs effective is not actually their execution, but the task relevance that is given to the irrelevant stimuli when they have a CM assigned to them. Once a participant assigns a CM to a stimulus, they actively search for it among the string of stimuli during the test. When the to-be-countered stimulus is presented, the participant recognizes it as meaingful or relevant, and this process evokes a P300 (just as probe stimuli evoke P300s in the absence of a task-relevant response). So, if a participant were to stop executing CMs to a stimulus, they should still recognize it as having been meaningful, and enhanced P300s should be observed in comparison to other non-countered irrelevant stimuli, but would likely be smaller than when executing CMs since they would no longer be involved in an active perceptual task. In a somewhat analogous situation, van Hooff, Brunia, and Allen (1996) presented participants with three types of words in a single block: recently learned target words, new, unstudied, words, and previously learned words that had served as targets on previous blocks. Results showed that in comparison to new, unstudied words, the words that had previously been designated as targets (and had required a unique response in the previous block) evoked larger P300s. These amplitudes were also found to be smaller than to the recently studied words that served as targets during the block. The previously studied words retained their significance even after their assigned target response had been removed. This suggests that while recognition of a meaningful stimulus is responsible for evoking P300s, the actual execution of an explicit response is important for maintaining or enhancing its significance and its evoked P300 response.

The current findings have some significant applied implications. We have shown that a simpler method of CM use (one-for-all) can be effective against the P300-CIT. As in Sokolovsky et al. (2011), we found that using the "lumping" method of CMs (executing the CM simultaneously with the "I Saw It" response) leads to no within-block group differences in RTs to countered- and non-countered stimuli, a metric that had previously been used to detect CM use (Hu et al., 2012; Rosenfeld et al., 2008). However, since we did not run a simple guilty group, we cannot say whether or not these RTs are elevated over a baseline level. (In a similar paradigm, Hu et al., 2012, found RTs of 498ms and 460ms for probe and average irrelevant stimuli, respectively.) In the current study, the use of CMs lowered detection rates in the CTP to around 70 percent, somewhat lower than most previous studies applying CMs in this protocol (Table 8), using both mock crime and self-referring information. However, these detection rates still compare favorably to the rates in three-stimulus P300-CIT (Mertens & Allen, 2008; Rosenfeld et al., 2004) and the ANS-CIT (Ben-Shakhar, 2011) when CMs are used. In the future, any researchers interested in examining the effects of CM use on a given P300-CIT protocol can teach participants to use a relatively simple method that is still fairly successful at defeating the test.

While the number of variations of CM use in the current study was limited, the results provide good evidence suggesting that a simple recognition process is responsible for evoking P300s to countered irrelevant stimuli during a P300-CIT. The detection rates from the current study and Hu et al. (2012) show that while the CTP remains resistant to CMs, it is not immune to them. Further research is needed to determine ways in which the CTP or any other protocol could be designed to more strongly resist these CM methods, perhaps through methods that focus attention on the implicit probe recognition task through processes such as those in (Rosenfeld, Hu, & Pederson, 2012). Overall, CM use remains an important area of study for all physiological CITs (ANS, fMRI, and ERP). Common sense suggests that understanding the cognitive and physiological mechanisms behind them can lead to the development of protocols that are more resistant to their use. Finding ways to focus participants or suspects' attention on the probe items, thus enhancing their significance, without inducing false-positives, could make for an accurate and CM resistant test.

CONCLUDING DISCUSSION

Three issues regarding the P300-CIT were examined in the presented studies. In the first, (Chapter 3 - The Impact of Prior Knowledge from Participant Instructions in a Mock Crime P300 Concealed Information Test), a prior finding from the ANS-CIT literature (e.g. Nahari & Ben-Shakhar, 2011) of inducing false-positives in innocent participants through exposure to crime-relevant information was extended to P300 CITs. We found that informed-innocent participants were indistinguishable from truly guilty participants in a P300-CIT even after only minimal exposure to probe items through participant instructions. This study also expanded on the previous research by comparing the effects of exposure to crime-relevant information on guilty participants. The results of this manipulation showed that informing participants or having them rehearse details later tested for in a mock crime biases the test toward enhanced sensitivity. While this is not always problematic, as aspects of many crimes may be planned in advance, when replicating scenarios where information would be acquired purely incidentally, it becomes important to avoid using such a procedure. The first finding demonstrates the importance of controlling the release of crime-relevant information in the field. If details that could be used in a P30-CIT are leaked to the general public, the accuracy of the test can be compromised. The second finding illustrates the importance of proper experimental controls for the ecological validity of P300-CIT studies, and calls into question the accuracy rates of some prior studies.

In the second study reported here (Chapter 4 - Effects of Presentation Modality on the Detection of Central and Peripheral Details in a P300 CIT), we attempted to replicate, extend, and combine two prior studies of P300-CITs (Ambach et al., 2010; Gamer & Berti, 2012). In contrast to Ambach et al. (2010), we observed evidence suggesting that pictorial stimuli may provide an advantage over verbal stimuli in P300-CITs, shown by significant modality X stimulus interactions with a baseline—peak P300 measure and trends using a peak—peak measure. There were also clear effects of central versus peripheral details, with central details evoking larger probe – irrelevant P300 amplitude differences than peripheral details. However, when controlling for rates of explicit memory, our results more closely resembled those in Gamer and Berti (2012), although, we still observed a trend for an enhanced difference in central details. Unexpectedly, when analyzing the two variables together, pictorial stimuli enhanced probe – irrelevant P300 differences, but only for the central item. Given these findings, even if the effect is small, it may be beneficial for law enforcement to use pictorial stimuli of actual crime-relevant items in P300-CITs. Additionally, the rates of explicit memory demonstrated that stimuli designated as “central” or “peripheral” in one study might not be directly comparable to those in another.

One threat to the use of these P300-CITs in the field, CMs, was examined in the third study (Chapter 5 – Cognitive Mechanism of Countermeasures in P300 Concealed Information Tests). In two experiments, we presented simple evidence that it is a recognition process that is responsible for evoking P300s during CM use in a P300-CIT. In *Experiment 1*, the difficulty and complexity of stimulus evaluation was manipulated by comparing P300s evoked by a single, repeated CM, to the concurrent use of four unique CMs. In the second experiment, the salience of a single, repeated, CM was manipulated. More complex stimulus evaluation processes, such as that manipulated in the first experiment, are known to increase P300 latency (Kutas et al., 1977; Magliero et al., 1984), however, no differences in latency were observed in the current study. Additionally, despite the fact that salience and meaningfulness are known to increase P300 amplitudes (Johnson Jr., 1986), no differences in amplitudes were found in *Experiment 2* between a highly meaningful CM (the participant's name) and another CM with low salience (the word "table"). Taken together these results are suggestive that another cognitive process, prior to ones involved in the actual execution of a CM, is responsible for evoking P300s to countered stimuli during CM use. We argue that simple recognition of the to-be-countered stimulus as meaningful is this mechanism. While the assigning and executing of CMs is important, as they are what transforms the to-be-countered irrelevant stimulus into a meaningful and task-relevant target, it is the initial recognition of the stimulus when it is presented that initiates the P300 response. Applied, this result means that a person seeking to defeat a P300-CIT need only employ a single, simple, repeated CM in an attempt to defeat the test, rather than a more elaborate and difficult strategy (e.g. *one-for-each*). For an examiner, it then becomes important to find ways to focus a participant or suspect's attention on the probe stimulus to help overcome the use of a CM.

The results from these three studies add to what we know about P300-CITs in different ways. The first study addressed implications of specific methods that are used in laboratory and their potential threats to its validity. The second examined two variables that can affect the test's accuracy, which are directly relevant for choosing what kind of items to test for in the field and how to present them. The third study helps us to better understand CMs and the threat they present to the efficacy of P300-CITs. Currently, no version of the P300-CIT has been sufficiently tested and vetted for field use. However, with the amount of research being conducted, and the number of researchers who currently support its use (Iacono, 2011; Iacono & Lykken, 1997), we may not be far from seeing a P300-CIT properly applied for the first time in a criminal investigation in the United States. The development and implementation of such a test would not only be useful for the identification and prosecution of guilty suspects, but also a valuable tool that could help prevent the false-convictions of innocent people.

REFERENCES

Abootalebi, V., Moradi, M. H., & Khalilzadeh, M. A. (2006). A comparison of methods for ERP assessment in a P300-based GKT. *International Journal of Psychophysiology, 62*(2), 309-320. doi: 10.1016/j.ijpsycho.2006.05.009

Abootalebi, V., Moradi, M. H., & Khalilzadeh, M. A. (2009). A new approach for EEG feature extraction in P300-based lie detection. *Computer Methods and Programs in Biomedicine, 94*(1), 48-57. doi: 10.1016/j.cmpb.2008.10.001

Ambach, W., Bursch, S., Stark, R., & Vaitl, D. (2010). A Concealed Information Test with multimodal measurement. *Int J Psychophysiol, 75*(3), 258-267. doi: 10.1016/j.ijpsycho.2009.12.007

Ben-Shakhar, G. (2002). A critical review of the control question test (CQT). In M. Kleiner (Ed.), *Handbook of Polygraph Testing*. London: Academic Press.

Ben-Shakhar, G. (2011). Countermeasures. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 200-214). New York, New York: Cambridge Univeristy Press.

Ben-Shakhar, G., & Elaad, E. (2002). The Guilty Knowledge Test (GKT) as an application of psychophysiology: Future prospects and obstacles. In M. Kleiner (Ed.), *Handbook of Polygraph Testing* (pp. 87-102). San Diego, California: Academic Press.

Ben-Shakhar, G., & Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. *Journal of Applied Psychology, 88*(1), 131-151. doi: 10.1037/0021-9010.88.1.131

Ben-Shakhar, G., Frost, R., Gati, I., & Kresh, Y. (1996). Is an apple a fruit? Semantic relatedness as reflected by psychophysiological responsivity. *Psychophysiology, 33*(6), 671-679. doi: 10.1111/j.1469-8986.1996.tb02363.x

Ben-Shakhar, G., & Gati, I. (1987). Common and distinctive features of verbal and pictorial stimuli as determinants of psychophysiological responsivity. *J Exp Psychol Gen, 116*(2), 91-105.

Bradley, M. T., & Warfield, J. F. (1984). Innocence, information, and the guilty knowledge test in the detection of deception. *Psychophysiology, 21*(6), 683-689. doi: 10.1111/j.1469-8986.1984.tb00257.x

Campbell, I. (2007). Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine, 26*(19), 3661-3675. doi: 10.1002/sim.2832

Carmel, D., Dayan, E., Naveh, A., Raveh, O., & Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: the external validity of mock crime studies. *Journal of Experimental Psychology: Applied, 9*(4), 261-269. doi: 10.1037/1076-898X.9.4.261

Combs, L. A., & Polich, J. (2006). P3a from auditory white noise stimuli. *Clinical neurophysiology, 117*(5), 1106-1112. doi: 10.1016/j.clinph.2006.01.023

Council, N. R. (2003). *The polygraph and lie detection*. Joseph Henry Press.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalogr Clin Neurophysiol, 39*(2), 131-143. doi: 10.1016/0013-4694(75)90003-6

Curran, T. (2004). Effects of attention and confidence on the hypothesized ERP correlates of recollection and familiarity. *Neuropsychologia, 42*(8), 1088-1106. doi: 10.1016/j.neuropsychologia.2003.12.011

Daubert v. Merrell Dow Pharmaceuticals, Inc, No. 92-102, 509 579 (Supreme Court 1993).

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*(3), 355 - 372. doi: 10.1017/S0140525X00058027

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biol Psychol, 14*(1), 1-52. doi: 10.1016/0301-0511(82)90016-3

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). The definition, identification, and reliability of measurment of the P300 component of the event-related brain potential. In P. K. Ackles, J. R. Jennings & M. G. H. Coles (Eds.), *Advances in Psychophysiology* (Vol. 2, pp. 1-78). Greenwich, CT: JAI Press.

. Facts on post-conviction DNA exonerations. (2012) Retrieved July 2, 2013, from http://www.innocenceproject.org/Content/DNA_Exonerations_Nationwide.php

Farwell, L. A., & Donchin, E. (1991). The truth will out: interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology, 28*(5), 531-547. doi: 10.1111/j.1469-8986.1991.tb01990.x

Fisher, R. A. (1922). On the interpretation of chi-squared from contingency tables, and the calcuation of P. *Journal of the Royal Statistical Society, 85*, 87-94.

Frye v. United States, 293 1013 (D.C. Cir 1923).

Gamer, M. (2010). Does the Guilty Actions Test allow for differentiating guilty participants from informed innocents? A re-examination. *International Journal of Psychophysiology, 76*(1), 19-24. doi: 10.1016/j.ijpsycho.2010.01.009

Gamer, M. (2011). Detecting concealed information using autonomic measures. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 27-45). New York, NY: Cambridge University Press.

Gamer, M., & Berti, S. (2010). Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials. *Psychophysiology, 47*(2), 355-364. doi: 10.1111/j.1469-8986.2009.00933.x

Gamer, M., & Berti, S. (2012). P300 amplitudes in the concealed information test are less affected by depth of processing than electrodermal responses. *Front Hum Neurosci, 6*, 308. doi: 10.3389/fnhum.2012.00308

Gamer, M., Gödert, H. W., Keth, A., Rill, H. G., & Vossel, G. (2008). Electrodermal and phasic heart rate responses in the Guilty Actions Test: Comparing guilty examinees to informed and uninformed innocents. *International Journal of Psychophysiology, 69*(1), 61-68. doi: 10.1016/j.ijpsycho.2008.03.001

Gamer, M., Kosiol, D., & Vossel, G. (2010). Strength of memory encoding affects physiological responses in the Guilty Actions Test. *Biological Psychology, 83*(2), 101-107. doi: 10.1016/j.biopsycho.2009.11.005

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

Gati, I., & Ben-Shakhar, G. (1990). Novelty and significance in orientation and habituation: a feature-matching approach. [Research Support, Non-U.S. Gov't]. *J Exp Psychol Gen, 119*(3), 251-263.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1974). New York: Wiley.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148*(3), 839-843.

Hu, X., Hegeman, D., Landry, E., & Rosenfeld, J. P. (2012). Increasing the number of irrelevant stimuli increases ability to detect countermeasures to the P300-based Complex Trial Protocol for concealed information detection. *Psychophysiology, 49*(1), 85-95. doi: 10.1111/j.1469-8986.2011.01286.x

Hu, X., Pornpattananangkul, N., & Rosenfeld, J. P. (2013). N200 and P300 as orthogonal and integrable indicators of distinct awareness and recognition processes in memory detection. *Psychophysiology*. doi: 10.1111/psyp.12018

Hu, X., & Rosenfeld, J. P. (2012). Combining the P300-complex trial-based concealed information test and the reaction time-based autobiographical implicit association test in concealed memory detection. *Psychophysiology, 49*(8), 1090-1100. doi: 10.1111/j.1469-8986.2012.01389.x

Iacono, W. G. (2011). Encouraging the use of the Guilty Knowledge Test (GKT): What the GKT has to offer law enforcement. In B. Verschuere, G. Ben-Shakhar & E. Meijer

(Eds.), *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 12-26). New York, New York: Cambridge Univeristy Press.

Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychophysiology, 82*(3), 426-433. doi: 10.1037/0021-9010.82.3.426

Johnson Jr., R. (1986). A triarchic model of P300 amplitude. *Psychophysiology, 23*, 367-384.

Johnson Jr., R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology, 30*(1), 90-97.

Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalization, and confabulation. *Psychological science, 7*(3), 125-128.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science, 197*(4305), 792-795. doi: 10.1126/science.887923

Labkovsky, E., & Rosenfeld, J. P. (2012a). The P300-based, complex trial protocol for concealed information detection resists any number of sequential countermeasures against up to five irrelevant stimuli. *Appl Psychophysiol Biofeedback, 37*(1), 1-10. doi: 10.1007/s10484-011-9171-0

Labkovsky, E., & Rosenfeld, J. P. (2012b). The P300-based, complex trial protocol for concealed information detection resists any number of sequential countermeasures against up to five irrelevant stimuli. *Applied Psychophysiology and Biofeedback, 37*(1), 1-10. doi: 10.1007/s10484-011-9171-0

Lykken, D. T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychophysiology, 43*(6), 385-388. doi: 10.1037/h0046060

Magliero, A., Bashore, T. R., Coles, M. G., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology, 21*(2), 171-186. doi: 10.1111/j.1469-8986.1984.tb00201.x

McCarthy, G., & Donchin, E. (1981). A metric for thought: a comparison of P300 latency and reaction time. [Research Support, U.S. Gov't, Non-P.H.S.]. *Science, 211*(4477), 77-80.

McNeil, B. J., & Hanley, J. A. (1984). Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. *Medical Decision Making, 4*(2), 137-150. doi: 10.1177/0272989X8400400203

Meijer, E. H., Smulders, F. T., Merckelbach, H. L., & Wolf, A. G. (2007). The P300 is sensitive to concealed face recognition. *International Journal of Psychophysiology, 66*(3), 231-237. doi: 10.1016/j.ijpsycho.2007.08.001

Meijer, E. H., Smulders, F. T., & Wolf, A. (2009). The contribution of mere recognition to the p300 effect in a concealed information test. *Appl Psychophysiol Biofeedback, 34*(3), 221-226. doi: 10.1007/s10484-009-9099-9

Meijer, E. H., Verschuere, B., & Ben-Shakhar, G. (2011). Practical guidelines for developing a CIT. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 293-302). New York, New York: Cambridge Univeriesty Press.

Meixner, J. B., Haynes, A., Winograd, M. R., Brown, J., & Rosenfeld, J. P. (2009). Assigned versus random, countermeasure-like responses in the P300 based complex trial protocol for detection of deception: task demand effects. *Applied Psychophysiology and Biofeedback, 34*(3), 209-220. doi: 10.1007/s10484-009-9091-4

Meixner, J. B., & Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology*. doi: 10.1111/j.1469-8986.2010.01050.x

Mertens, R., & Allen, J. J. (2008). The role of psychophysiology in forensic assessments: deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology, 45*(2), 286-298. doi: 10.1111/j.1469-8986.2007.00615.x

Nahari, G., & Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: the role of memory for crime details. *Psychophysiology, 48*(6), 733-744. doi: 10.1111/j.1469-8986.2010.01148.x

Nakayama, M. (2002). Practical use of the concealed information test for criminal investigation in Japan. In M. Kleiner (Ed.), *Handbook of polygraph testing*. London: Academic Press.

Osugi, A. (2011). Daily application of the Concealed Information Test: Japan. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test*. New York, NY: Cambridge Univeristy Press.

Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalogr Clin Neurophysiol, 67*(4), 360-371.

Peth, J., Vossel, G., & Gamer, M. (2012). Emotional arousal modulates the encoding of crime-related details and corresponding physiological responses in the Concealed Information Test. *Psychophysiology, 49*(3), 381-390. doi: 10.1111/j.1469-8986.2011.01313.x

Podlesny, J. A. (1993). Is the guilty knowledge polygraph technique applicable in criminal investigations? A review of FBI case records. *Crime Laboratory Digest, 20*(3), 57-61.

Polich, J. (1991). P300 in clinical applications: meaning, method, and measurement. *American Journal of EEG Technology, 31*(3), 201-231.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology, 118*(10), 2128-2148.

Raskin, D. C., & Honts, C. R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of polygraph testing*. London: Academic Press.

Rosenfeld, J. P. (2011). P300 in detecting concealed information. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 63-89). New York, NY: Cambridge University Press.

Rosenfeld, J. P., Biroschak, J. R., & Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *International Journal of Psychophysiology, 60*(3), 251-259. doi: 10.1016/j.ijpsycho.2005.06.002

Rosenfeld, J. P., Biroschak, J. R., Kleschen, M. J., & Smith, K. M. (2005). Subjective and objective probability effects on P300 amplitude revisited. *Psychophysiology, 42*(3), 356-359. doi: 10.1111/j.1469-8986.2005.00283.x

Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanov, S., & Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *International Journal of Neuroscience, 42*(1-2), 157-161. doi: 10.3109/00207458808985770

Rosenfeld, J. P., Ellwanger, J., & Sweet, J. (1995). Detecting simulated amnesia with event-related brain potentials. *Int J Psychophysiol, 19*(1), 1-11. doi: 10.1016/0167-8760(94)00057-L

Rosenfeld, J. P., Hu, X., Labkovsky, E., Meixner, J., & Winograd, M. R. (*in review*). Review of recent studies and issues regarding the P300-based Complex Trial Protocol for detection of concealed information. *International Journal of Psychophysiology*.

Rosenfeld, J. P., Hu, X., & Pederson, K. (2012). Deception Awareness Improves P300-based Deception Detection in Concealed Information Tests. *International Journal of Psychophysiology*. doi: 10.1016/j.ijpsycho.2012.06.007

Rosenfeld, J. P., & Labkovsky, E. (2010). New P300-based protocol to detect concealed information: resistance to mental countermeasures against only half the irrelevant stimuli and a possible ERP indicator of countermeasures. *Psychophysiology, 47*(6), 1002-1010. doi: 10.1111/j.1469-8986.2010.01024.x

Rosenfeld, J. P., Labkovsky, E., Winograd, M. R., Lui, M. A., Vandenboom, C., & Chedid, E. (2008). The Complex Trial Protocol (CTP): a new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. *Psychophysiology, 45*(6), 906-919. doi: 10.1111/j.1469-8986.2008.00708.x

Rosenfeld, J. P., Nasman, V. T., Whalen, R., Cantwell, B., & Mazzeri, L. (1987). Late vertex positivity in event-related potentials as a guilty knowledge indicator: a new method of life detection. *Int J Neurosci, 34*(1-2), 125-129.

Rosenfeld, J. P., Shue, E., & Singer, E. (2007). Single versus multiple probe blocks of P300-based concealed information tests for self-referring versus incidentally obtained information. *Biological Psychology, 74*(3), 396-404. doi: 10.1016/j.biopsycho.2006.10.002

Rosenfeld, J. P., & Skogsberg, K. R. (2006). P300-based Stroop study with low probability and target Stroop oddballs: the evidence still favors the response selection hypothesis. *International Journal of Psychophysiology, 60*(3), 240-250. doi: doi:10.1016/j.ijpsycho.2005.05.010

Rosenfeld, J. P., Soskins, M., Bosh, G., & Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology, 41*(2), 205-219. doi: 10.1111/j.1469-8986.2004.00158.x

Rosenfeld, J. P., Tang, M., Meixner, J., Winograd, M. R., & Labkovsky, E. (2009). The effects of asymmetric vs. symmetric probability of targets following probe and irrelevant stimuli in the complex trial protocol for detection of concealed information with P300. *Physiology and Behavior, 98*(1-2), 10-16. doi: 10.1016/j.physbeh.2009.03.030

Rushby, J. A., Barry, R. J., & Doherty, R. J. (2005). Separation of the components of the late positive complex in an ERP dishabituation paradigm. *Clin Neurophysiol, 116*(10), 2363-2380. doi: 10.1016/j.clinph.2005.06.008

Seymour, T. L., & Kerlin, J. R. (2008). Successful detection of verbal and visual concealed knowledge using an RT-based paradigm. *Applied Cognitive Psychology, 22*(4), 475-490. doi: 10.1002/acp.1375

Sokolov, E. N. (1990). The orienting response, and future directions of its development. *Integrative Physiological and Behavioral Science, 25*(3), 142-150. doi: 10.1007/BF02974268

Sokolovsky, A., Rothenberg, J., Labkovsky, E., Meixner, J., & Rosenfeld, J. P. (2011). A novel countermeasure against the reaction time index of countermeasure use in the P300-based complex trial protocol for detection of concealed information. *International Journal of Psychophysiology, 81*(1), 60-63. doi: 10.1016/j.ijpsycho.2011.03.008

Soskins, M., Rosenfeld, J. P., & Niendam, T. (2001). Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: complex vs. simple paradigms. *International Journal of Psychophysiology, 40*(2), 173-180. doi: 10.1016/S0167-8760(00)00154-9

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychol Bull, 118*(3), 315.

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science*. doi: 10.1126/science.959831

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol, 38*(4), 387-401. doi: 10.1016/0013-4694(75)90263-1

Stenberg, G., Radeborg, K., & Hedman, L. R. (1995). The picture superiority effect in a cross-modality recognition task. *Memory & Cognition, 23*(4), 425-441. doi: 10.3758/BF03197244 van Hooff, J., Brunia, C. H. M., & Allen, J. J. B. (1996). Event-related potentials as indirect measures of recognition memory. *International Journal of Psychophysiology, 21*(1), 15-31. doi: 10.1016/0167-8760(95)00043-7

Verschuere, B., Crombez, G., Degrootte, T., & Rosseel, Y. (2010). Detecting concealed information with reaction times: Validity and comparison with the polygraph. *Applied Cognitive Psychology, 24*(7), 991-1002. doi: 10.1002/acp.1601

Verschuere, B., & De Houwer, J. (2011). Detecting concealed information in less than a second: response latency-based measures. In B. Verschuere, G. Ben-Shakhar & E. Meijer (Eds.), *Memory Detection: Theory and Application of the Concealed Information Test* (pp. 46-62). New York, New York: Cambridge Univeristy Press.

Vrij, A. (2008). *Detecting lies and deceit : pitfalls and opportunities* (2nd ed.). Chichester ; Hoboken, NJ: John Wiley.

Wasserman, S., & Bockenholt, U. (1989). Bootstrapping: applications to psychophysiology. *Psychophysiology, 26*(2), 208-221. doi: 10.1111/j.1469-8986.1989.tb03159.x

Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. *Science, 221*(4615), 1080-1082. doi: 10.1126/science.6879207

Winograd, M. R., Labkovsky, E., Noriega, A., Lamano, J., & Rosenfeld, J. P. (2011). One (countermeasure, CM) for all (irrelevants) versus one for each (irrelevant) in the P300-based Complex Trial Protocol for detecting concealed information. *Psychophysiology, 48*(s1), s108. doi: 10.1111/j.1469-8986.2011.01259.x

Winograd, M. R., & Rosenfeld, J. P. (2011). Mock crime application of the Complex Trial Protocol (CTP) P300-based concealed information test. *Psychophysiology, 48*(2), 155-161. doi: 10.1111/j.1469-8986.2010.01054.x